ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ & ΑΕΙ ΠΕΙΡΑΙΑ Τ.Τ.
Τμήματα Ναυτιλίας και Επιχειρηματικών Υπηρεσιών & Μηχ. Αυτοματισμού ΤΕ

ΔΙΙΔΡΥΜΑΤΙΚΟ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΝΕΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΣΤΗ ΝΑΥΤΙΛΙΑ ΚΑΙ ΤΙΣ ΜΕΤΑΦΟΡΕΣ»

Ανίχνευση και Διάγνωση Βλαβών Δίχρονων Αργόστροφων Ναυτικών Κινητήρων με Χρήση Αλγορίθμων Μηχανικής Μάθησης

Fault Detection and Diagnosis of Two-Stroke Low Speed Marine Engine with Machine Learning Algorithms

Ονοματεπώνυμο Σπουδαστή: Γεώργιος Τσαγκανός

Ονοματεπώνυμο Υπεύθυνου Καθηγητή: Δρ Γρηγόρης Νικολάου

ΔΙΑΤΡΙΒΗ

Σεπτέμβριος 2017

Ανίχνευση και Διάγνωση Βλαβών Δίχρονων Αργόστροφων Ναυτικών Κινητήρων με Χρήση Αλγορίθμων Μηχανικής Μάθησης

Γεώργιος Τσαγκανός

Μεταπτυχιακή Διατριβή που υποβάλλεται στο καθηγητικό σώμα για την μερική εκπλήρωση των υποχρεώσεων απόκτησης του μεταπτυχιακού τίτλου του Διιδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες στη Ναυτιλία και τις Μεταφορές» του Τμήματος Ναυτιλίας και Επιχειρηματικών Υπηρεσιών του Πανεπιστημίου Αιγαίου και του Τμήματος Μηχανικών Αυτοματισμού Τ.Ε. του ΑΕΙ Πειραιά ΤΤ.

Δήλωση συγγραφέα διπλωματικής διατριβής

Ο κάτωθι υπογεγραμμένος Τσαγκανός Γεώργιος, του Ιωάννου, με αριθμό μητρώου …….. φοιτητής του. Διιδρυματικού Προγράμματος Μεταπτυχιακών Σπουδών «Νέες Τεχνολογίες στη Ναυτιλία και τις Μεταφορές» του Τμήματος Ναυτιλίας και Επιχειρηματικών Υπηρεσιών του Πανεπιστημίου Αιγαίου και του Τμήματος Μηχανικών Αυτοματισμού Τ.Ε. του ΑΕΙ Πειραιά ΤΤ, δηλώνω ότι: *«Είμαι συγγραφέας αυτής της μεταπτυχιακής διπλωματικής διατριβής και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην διατριβή. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η διατριβή προετοιμάστηκε από εμένα προσωπικά ειδικά για τη συγκεκριμένη μεταπτυχιακή διπλωματική διατριβή».*

Ο δηλών Ημερομηνία

ΠΕΡΙΛΗΨΗ

Η ανίχνευση βλαβών ναυτικού κινητήρα είναι εξαιρετικά σημαντικές στη μεταφορά των αγαθών, το θαλάσσιο περιβάλλον και στον ανταγωνισμό των ναυτιλιακών εταιριών γενικότερα. Η έγκαιρη ανίχνευση βλαβών, οδηγούν στη βελτίωση της αξιοπιστίας του κινητήρα, στις μειωμένες εμφανίσεις βλαβών του κινητήρα και στην μη διακοπή της λειτουργίας του. Η παρούσα διπλωματική εργασία περιγράφει και αξιολογεί την ανάπτυξη και εφαρμογή ευφυών διαγνωστικών μεθόδων, που βασίζονται στην χρήση αλγορίθμων μηχανικής μάθησης, επιτρέποντας την αποτελεσματική ανίχνευση και διάγνωση βλαβών ενός δίχρονου αργόστροφου ναυτικού κινητήρα diesel. Η έρευνα υλοποιήθηκε με το ελεύθερο εργαλείο εξόρυξης δεδομένων Weka το οποίο αναλύει τα δεδομένα των λειτουργικών παραμέτρων του κινητήρα που είναι εκτός ορίων, χρησιμοποιώντας την τεχνική ταξινόμησης για να προβλέψει με ακρίβεια την τάξη στόχο κάθε περίπτωσης.

Στόχος μας είναι η διερεύνηση επίδοσης διαφορετικών μεθόδων ταξινόμησης, συγκεκριμένα υλοποιήθηκαν και αξιολογήθηκαν επτά βασικοί αλγόριθμοι ταξινόμησης που είναι αντιπροσωπευτικοί των σημαντικότερων τεχνικών ταξινόμησης, επίσης, υλοποιήθηκαν και τρεις συνδυαστικές (ensemble) μέθοδοι με στόχο την βελτίωση της απόδοσης των βασικών αλγορίθμων ταξινόμησης.

Για την τελική επιλογή του αλγορίθμου πραγματοποιήθηκαν πειράματα σύγκρισης αξιολογούμενα από τις μετρικές αξιολόγησης, της ακρίβεια (Accuracy), και του F-Measure ο αρμονικός μέσος όρος της ορθότητας (Precision) και της ανάκλησης (Recall).

Τέλος, προτείνεται μια μέθοδος η οποία στηρίζεται στην κατασκευή ενός συνδυαστικού μοντέλου ταξινόμησης AdaBoost, ο οποίος βελτιώνει την απόδοση του βασικού ταξινομητή Simple Cart, αφού τα πειραματικά διαγνωστικά αποτελέσματα τον κατέταξαν με την υψηλότερη απόδοση στην ανίχνευση βλαβών ίση με 96,5%. Συνεπώς η προτεινόμενη μέθοδος είναι εφικτή για τη ανίχνευση και διάγνωση βλαβών δίχρονων αργόστροφων ναυτικών κινητήρων diesel.

Λέξεις-κλειδιά: Αλγόριθμοι Ταξινόμησης, Weka, Διασταυρωμένη επικύρωση, F-Measure, Accuracy, Πίνακας Σύγχυσης, Συνδυαστικοί μέθοδοι.

ABSTRACT

The detection of faults marine engine is extremely important in the transport of the goods, the marine environment and in the competition of shipping companies in general. The early detection of faults, leads to improvement of the reliability of the engine, reduced incidents of engine breakdowns and the non-interruption of the operation of the.

This thesis describes and evaluates the development and implementation of intelligent diagnostic methods based on the use of algorithms machine learning, allowing the effective detection and diagnosis of faults of a two-stroke slow speed marine diesel engine. The research was implemented with the free Weka data mining tool, which analyzes the data of the operating parameters of the engine that are out of bounds, using the technique of classification to predict with accuracy the class objective each case.

Our aim is to investigate the performance of different classification methods, in particular seven basic classification algorithms that are representative of the most important classification techniques were implemented and evaluated, also and three ensemble methods were developed, with the aim of improving the performance of the basic algorithms of classification.

For the final selection of the algorithm were performed experiments for comparison evaluated by the evaluation metrics, the Accuracy, and the F-measure is the harmonic average of the Precision and the Recall.

Finally, a method is proposed which is based on the construction of an ensemble classification model AdaBoost, which improves the performance of the basic classifier Simple Cart, after the experimental diagnostic results the ranked with the highest performance in the detection of faults equal to 96,5%. Consequently, the proposed method is feasible for the detection and diagnosis of faults of two-stroke slow speed marine diesel engines.

Keywords: Classification Algorithms,, Weka, Cross validation, F-Measure, Accuracy, Confusion Matrix, Ensemble methods.

Αφιερώνω τη διπλωματική εργασία

στους γονείς μου!

Πίνακας Περιεχομένων

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κατάλογος Διαγραμμάτων

1. ΕΙΣΑΓΩΓΗ

Για πολλά χρόνια, κύριος στόχος της ανάπτυξης των ναυτικών κινητήρων πρόωσης ήταν η αύξηση του βαθμού απόδοσης καθώς και της αξιοπιστίας τους κατά την λειτουργία τους.

Η ανάγκη μεγαλύτερων ελίκων, στα όρια κατασκευαστικών δυνατοτήτων, με χαμηλούς ρυθμούς περιστροφής για αύξηση της υδροδυναμικής αποδόσεως, οδήγησε στην επικράτηση των αργόστροφων δίχρονων μηχανών μεγάλης διαδρομής εμβόλου για την πρόωση εμπορικών πλοίων.

Οι δίχρονες μηχανές έχουν πολύ υψηλό βαθμό αποδόσεως και δυνατότητα να κάψουν χαμηλής ποιότητας βαριά καύσιμα.

Τα τελευταία χρόνια για περίπου στα νεότευκτα πλοία άνω των 2000dtw το 99% είχε κινητήρες προώσεως diesel, οι πλειοψηφία των οποίων ήταν δίχρονες αργόστροφες μηχανές (Κυρτάτος, 1999),(Pedersen & Engineer, n.d.).

Στην αξιολόγηση καλής λειτουργίας των κινητήρων πρόωσης πλοίου τα μεγαλύτερα προβλήματα είναι η συλλογή αξιόπιστων μετρήσεων λειτουργικών παραμέτρων του κινητήρα καθώς και η ύπαρξη στοιχείων αναφοράς για σύγκριση με τις μετρήσεις και εξαγωγή συμπερασμάτων.

Τα στοιχεία αναφοράς (reference data) που προ-υπάρχουν για την κύρια μηχανή είναι οι δοκιμές κατασκευαστή (shop tests) και οι δοκιμές παραλαβής πλοίου (sea trials).

Οι μετρούμενες τιμές κατά τη λειτουργία του πλοίου συγκρίνονται με τις τιμές αναφοράς του κατασκευαστή και προκύπτει ποσοστιαία διαφορά σε όλες τις μετρούμενες λειτουργικές παραμέτρους, αν η διαφορά σε οποιαδήποτε λειτουργική παράμετρο ξεπερνά τα όρια του κατασκευαστή τότε πρέπει να γίνει έλεγχος.

Ο έλεγχος, η εύρυθμη λειτουργία και η αξιοπιστία του κινητήρα είναι η πιο σημαντική απαίτηση που υπάρχει μέσα σε ένα εμπορικό πλοίο και απασχολεί τους μηχανικούς και τους ερευνητές είτε για την επιδιόρθωση είτε ακόμα και για την εξέλιξη τους.

Το γεγονός αυτό έχει δώσει κίνητρα στην επιστημονική κοινότητα για ανάπτυξη και εφαρμογή καινοτόμων τεχνολογιών με κύριο στόχο τη συνεχή και αξιόπιστη απόδοση των κινητήρων. Η αποτελεσματικότητα των νέων αυτών

τεχνολογιών και η επίδρασή τους στη συνολική λειτουργία των κινητήρων διερευνάται μέσω διεξαγωγής σύνθετων πειραμάτων.

Η συνεχώς εξελισσόμενη τεχνολογία και οι σύγχρονες απαιτήσεις επιβάλουν την εφαρμογή νέων βελτιωμένων μεθόδων ανίχνευσης και διάγνωσης των βλαβών τους (τεχνητή νοημοσύνη, μηχανική μάθηση, fuzzy logic), που προσφέρουν βέλτιστη αξιοποίηση του κινητήρα για πρόωση πλοίου σε όλες τις συνθήκες λειτουργίας καθώς και ασφάλεια σε ακραίες συνθήκες (Xiros & Kyrtatos, 2000) (Lan, Katagi, & Hashimoto, 1996).

Το διαγνωστικό σύστημα είναι υπεύθυνο για το έλεγχο των λειτουργικών παραμέτρων της μηχανής, για τον εντοπισμό διαφοροποιήσεων που επέρχονται από την κατάσταση του κινητήρα, που οφείλονται σε συγκεκριμένες βλάβες και επιτυγχάνει τόσο τη μείωση του ανθρώπινου δυναμικού που απαιτείται, μειώνοντας την πιθανότητα ανθρώπινου λάθους, όσο και την παροχή αυξημένης ποιότητας ασφάλειας, με την έγκαιρη αποφυγή καταστροφής του κινητήρα, προκαλώντας οικονομικές και περιβαλλοντολογικές ζημιές.

Αναζητούνται διαρκώς νέες τεχνικές και συστήματα που να καθιστούν τα διαγνωστικά συστήματα ανίχνευσης και διάγνωσης βλαβών κινητήρα πιο αξιόπιστα και γρήγορα.

1.1. Σκοπός της Διπλωματικής Εργασίας

Η παρούσα εργασία εντάσσεται στο χώρο της διαγνωστικής δίχρονων αργόστροφων ναυτικών κινητήρων diesel.

Απώτερος σκοπός της παρούσας εργασίας είναι, να συμβάλει στη σύγχρονη αυτή έρευνα μέσω της ανάπτυξης μεθόδων, που επιτρέπουν την εκτίμηση των λειτουργικών παραμέτρων του κινητήρα με χρήση αλγορίθμων μηχανικής μάθησης, για την ανάπτυξη ενός διαγνωστικού συστήματος, με στόχο να πραγματοποιεί αξιόπιστη εκτίμηση της λειτουργικής κατάστασης, την έγκαιρη ανίχνευση και διάγνωση βλαβών των δίχρονων αργόστροφων ναυτικών κινητήρων diesel, που εμφανίζονται κατά τη λειτουργία τους και οι οποίες θα μπορούσαν να έχουν καταστρεπτικές συνέπειες τόσο οικονομικές όσο και περιβαντολογικές και να βελτίωση και οργάνωση τη λειτουργίας τους και τη συντήρηση τους.

Τέλος, να προτείνει την καλύτερη μέθοδο αλγορίθμου μηχανικής μάθησης,

που θα επιτυγχάνει ένα υψηλό ποσοστό ακρίβειας στην ανίχνευση και διάγνωση των βλαβών.

Τα παραπάνω επιτυγχάνονται μέσω:

- της εξέτασης διαφόρων αλγορίθμων μηχανικής μάθησης και της σύγκρισης των λαμβανομένων από αυτούς αποτελεσμάτων (ποσοστό σωστών προβλέψεων, μερτικών απόδοσης και χρόνο κατασκευής μοντέλου)
- της εφαρμογής συνδυαστικών μεθόδων (ensemble methods) για την ενίσχυση και βελτίωση του αποτελέσματος.

1.2. Δομή της Διπλωματικής Εργασίας

Η παρούσα εργασία χωρίζεται σε 8 κεφάλαια.

Στο κεφάλαιο 2, παρατίθενται γενικά στοιχεία για τους κινητήρες Diesel και τη λειτουργία τους. Γίνεται ιδιαίτερη αναφορά στους δίχρονους αργόστροφους ναυτικούς κινητήρες μιας και οι συγκεκριμένες μηχανές συνιστούν τον πρωτεύοντα τομέα εφαρμογής διαγνωστικών μεθόδων.

Στο κεφάλαιο 3, γίνεται γενική αναφορά στις βλάβες στη διάγνωση βλαβών και στα συστήματα διάγνωσης. Ο ρόλος των λειτουργικών παραμέτρων στους κινητήρες και η αναγκαιότητα χρήσης τους, για την ανίχνευση και τη διάγνωση των βλαβών του κινητήρα. Αναδεικνύεται η σημασία που έχει για τη διαγνωστική διαδικασία, η πίεση καύσης. Στην συνέχεια παρατίθεται πίνακας που αναφέρει τις ομάδες των λειτουργικών παραμέτρων τα όρια των κατασκευαστικών αποκλίσεων του συγκεκριμένου κινητήρα και τις αντίστοιχες βλάβες που προσδιορίζουν.

Στο κεφάλαιο 4, αναφέρεται στην εξόρυξη δεδομένων, στον τομέα της μηχανικής μάθησης και στη τεχνική της ταξινόμησης όπου ανήκουν και οι αλγόριθμοι επιβλεπόμενης μάθησης που χρησιμοποιούνται στην παρούσα εργασία.

Στο κεφάλαιο 5, περιγράφεται το εργαλείο εξόρυξης δεδομένων WEKA, μέσω του οποίου γίνεται η χρήση των αλγορίθμων ταξινόμησης, και οι μετρικές απόδοσης για την αξιολόγηση των μοντέλων.

Στο κεφάλαιο 6, γίνεται αναφορά στις κατηγορίες αλγορίθμων ταξινόμησης του Weka.

Στο κεφάλαιο 7, παρουσιάζεται η μεθοδολογία έρευνας της εργασίας και υλοποιείται και παρουσιάζονται συγκριτικοί πίνακες και γραφήματα με τα

αποτελέσματα της έρευνας.

Συγκρίνονται και αξιολογούνται τα αποτελέσματα των μεθόδων ταξινόμησης και προτείνεται το μοντέλο, το οποίο πραγματοποιεί με βάση το σύνολο των δεδομένων την καλύτερη πρόβλεψη ταξινόμησης ανίχνευσης βλαβών.

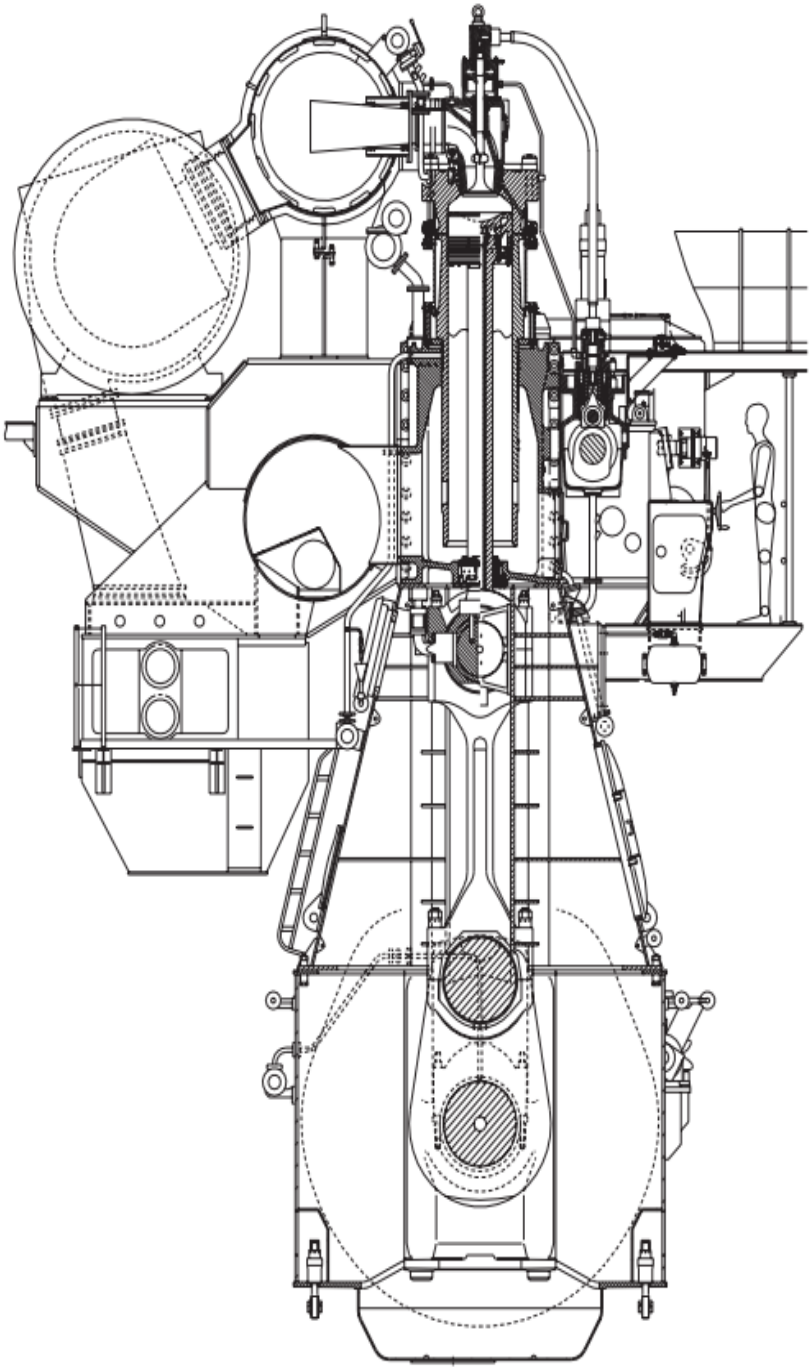
Τέλος, στο Κεφάλαιο 8, συνοψίζονται τα συμπεράσματα στα οποία κατέληξε η παρούσα έρευνα και αποτιμάται τελικώς η δυνατότητα χρήσης αλγοριθμικών μεθόδων ταξινόμησης μηχανικής μάθησης, για την επιτυχή ανίχνευση και διάγνωση βλαβών, στους δίχρονους αργόστροφους ναυτικούς κινητήρες diesel.

Στο τέλος του κεφαλαίου παρουσιάζονται οι δυσκολίες που προέκυψαν κατά την υλοποίηση της εργασίας, καθώς και τυχόν επεκτάσεις και βελτιώσεις στο σύστημα.

2. ΔΙΧΡΟΝΟΙ ΑΡΓΟΣΤΡΟΦΟΙ ΝΑΥΤΙΚΟΙ ΚΙΝΗΤΗΡΕΣ

Οι αργόστροφοι κινητήρες κυριαρχούν στο χώρο της πρόωσης των μεγάλων ποντοπόρων εμπορικών πλοίων (δεξαμενόπλοια, πλοία μεταφοράς εμπορευματοκιβωτίων, μεταφοράς χύδην φορτίου). Ο σημαντικότερος λόγος της επικράτησης τους είναι το γεγονός ότι ο βαθμός απόδοσής του σε υψηλές τιμές ισχύος μπορεί να ξεπεράσει το 50%. Είναι τυπικά δίχρονοι με μεγάλο λόγο διαδρομής - διαμέτρου εμβόλου και διάταξη βάκτρου με ζύγωμα - διωστήρα - στρόφαλο. Μπορούν να λειτουργήσουν αρκετά καλά με ένα μεγάλο εύρος ποιότητας καυσίμου, ακόμα και πολύ χαμηλής. Επιπλέον, είναι δυνατό να χρησιμοποιηθεί η απορριπτόμενη θερμότητα που βρίσκεται στα καυσαέρια και το νερό ψύξης, αυξάνοντας έτσι περαιτέρω το βαθμό απόδοσης. Τα παραπάνω οδηγούν σε χαμηλότερο κόστος χρήσης και μεγαλύτερη αξιοπιστία. Οι κινητήρες αυτοί συνδέονται άμεσα με την έλικα, χωρίς μειωτήρα χάρη στην χαμηλή ταχύτητα περιστροφής τους (Λαζάρου Χ. Κλιάνη, Ιωάννη Κ. Νικαλάου, 2003).

Στις μέρες μας μόνο τρεις κατασκευαστές αργόστροφων κινητήρων Diesel έχουν επιβιώσει (MAN B&W , Mitsubishi, Wartsila NSD). Οι παραπάνω κατασκευαστές προσφέρουν διαφορετικές οικογένειες δίχρονων αργόστροφων κινητήρων Diesel.



Σχήμα 2-1 Τομή δίχρονου αργόστροφου κινητήρα MAN B&W 7S60MC
Πηγή:**(Engine Selection Guide Two-stroke MC/MC-C Engines, 2000)**

3. ΔΙΑΓΝΩΣΗ ΒΛΑΒΩΝ ΔΙΧΡΟΝΩΝ ΑΡΓΟΣΤΡΟΦΩΝ ΝΑΥΤΙΚΩΝ ΚΙΝΗΤΗΡΩΝ DIESEL

Στις μέρες μας, οι δίχρονοι αργόστροφοι ναυτικοί κινητήρες diesel κυριαρχούν στον τομέα της ναυτιλίας ως μέσο πρόωσης πλοίων και χρησιμοποιούνται ευρύτατα.

Ο σημαντικός ρόλος που διαδραματίζουν οι κινητήρες diesel στην εύρυθμη λειτουργία του πλοίου, έχει σαν αποτέλεσμα τον πολύ ισχυρό αντίκτυπο μιας πιθανής βλάβης τους. Ο αντίκτυπος αυτός μπορεί να είναι απλά οικονομικές συνέπειες, περιβαλλοντολογικές συνέπειες ή, στην χειρότερη περίπτωση, κίνδυνος ανθρώπινων ζωών**.**

Όταν ένας κινητήρας λειτουργεί για μεγάλο χρονικό διάστημα είναι λογικό τα διάφορα εξαρτήματά του να εμφανίσουν σημάδια φθοράς. Αυτό μπορεί να οφείλεται τόσο σε τυχαίους παράγοντες όσο και στις συνθήκες λειτουργίας του. Οι βλάβες που μπορεί να εμφανίσει ένας κινητήρας δεν είναι δυνατό να προβλεφθούν και ειδικά οι φθορά των εξαρτημάτων του (Ζώτος, 2008).

Μια απρόβλεπτη αστοχία αυτών των μηχανών, μπορεί να επιφέρει ιδιαίτερα ανεπιθύμητα αποτελέσματα, θέτοντας σε κίνδυνο ανθρώπινες ζωές ή ακόμα και το περιβάλλον, όπως έχει διαπιστωθεί αρκετές φορές. Σε ηπιότερες μορφές αστοχίας, ο κινητήρας ενδέχεται να λειτουργεί με χαμηλό βαθμό απόδοσης ή ακόμα και να παύσει να λειτουργεί, προκαλώντας δαπανηρές καθυστερήσεις.

Ως βλάβη (fault) ενός κινητήρα ορίζεται η απομάκρυνση κάποιας λειτουργικής παραμέτρους του από ένα αποδεκτό όριο τιμών. Με τον τρόπο αυτό η βλάβη ορίζεται σαν μία ανωμαλία ή σύμπτωμα της διεργασίας που εμφανίζει αλλαγές στη λειτουργία του κινητήρα και επηρεάζει δυσμενώς την παρούσα ή/και την μελλοντική συμπεριφορά του. Η βλάβη έχει πραγματική έννοια με την ύπαρξη σύγκρισης μεταξύ δύο διαφορετικών καταστάσεων του συστήματος, η μία εκ των οποίων θεωρείται ως η αρχική χωρίς βλάβη κατάσταση (Σκουντριανός, 2005).
Στις αιτίες βλαβών των τμημάτων των κινητήρων μπορούν να συμπεριληφθούν.

1. Η διάβρωση.
2. Η ρύπανση.

3. Η αστοχία σε κόπωση.
4. Η αστοχία λόγο τάσεων μηχανικών.
5. Η αστοχία λόγο τάσεων θερμικών.
6. Η φθορά με συνέπεια την αλλαγή των διαστάσεων.
7. Η αποσύνθεση
8. Η μόνιμη παραμόρφωση της επιφάνειας τμημάτων της μηχανής.

Οι πιο πάνω αιτίες βλαβών υπάρχουν σε όλη τη διάρκεια ζωής του κινητήρα. Όταν αποκτήσουν σημαντικό μέγεθος και αρχίσουν να επηρεάζουν την λειτουργία του κινητήρα τότε απασχολούν τη διάγνωση που σκοπό έχει τον ποιοτικό και ποσοτικό εντοπισμό τους (Μαργαρώνης, 1986).

Για να αποτραπεί η πιθανότητα μιας απρόβλεπτης βλάβης ή το ενδεχόμενο λειτουργίας του κινητήρα σε χαμηλή απόδοση και σε συνδυασμό με τη φιλικότητα προς το περιβάλλον και την ασφάλεια των εργαζομένων, κρίνεται αναγκαία η παρακολούθηση τους κατά τη διάρκεια της λειτουργίας των κινητήρων diesel και η ανάγκη διάγνωσης των βλαβών τους, δίνοντας τη δυνατότητα στο μηχανικό να επαναφέρει τον δυσλειτουργικό κινητήρα στη φυσιολογική του συμπεριφορά.

Η πολυπλοκότητα των σημερινών κινητήρων, καθώς και η ανάγκη για ελάττωση του κόστους λειτουργίας τους, αλλά και οι απαιτήσεις απόδοσης και δυνατότητας παραγωγής έρχου από τους κινητήρες κάνουν το πρόβλημα της διάγνωσης οξύτερο και τις απαιτήσεις από τη διάγνωση μεγάλες έτσι επιβάλουν την εκμετάλλευση των δυνατοτήτων που παρέχουν τα διαγνωστικά συστήματα.

Οι διάφορες βλάβες ανάλογα με τη σοβαρότητα τους, μπορούν να οδηγήσουν σε σημαντικά προβλήματα λειτουργικότητας ή/και αστοχίας, ακόμη και σε καταρρεύσεις. Η αναγνώριση πιθανών βλαβών, τόσο όσον αφορά στην ύπαρξη αλλά και αναφορικά με τη σπουδαιότητα τους μπορεί να αποτρέψει τέτοιες αστοχίες και να προλάβει τις συνέπειές τους.

Ως διάγνωση ορίζεται η τέχνη ή η ενέργεια για τον εντοπισμό, προσδιορισμό ή/και απομόνωση της βλάβης ενός εξοπλισμού, βάσει της αξιολόγησης συγκεκριμένων σημάτων και συμπτωμάτων, καθώς και η απόφαση που λαμβάνεται (Τσανάκας, 2013).

Η διάγνωση βλαβών (Fault Detection and Diagnosis) είναι ένα σημαντικό πρόβλημα στην περιοχή της διαχείρισης διεργασιών. Μια διάγνωση βλαβών μπορεί

να πραγματοποιηθεί με την εν λειτουργία χρήση αισθητηρίων και εποπτικών συσκευών, για τη συλλογή της απαιτούμενης πληροφορίας σχετικά με την εκάστοτε κατάσταση ενός κινητήρα. Ο έγκαιρος εντοπισμός και διάγνωση βλαβών ενώ ο κινητήρας εξακολουθεί να λειτουργεί σε μία ελέγξιμη περιοχή μπορεί να συμβάλει στην αποφυγή της εξέλιξης ανώμαλων συμβάντων και να μειώσει την απώλεια αποδοτικότητας επιπλέον αποσκοπεί στον εντοπισμό, καταστάσεων που μπορεί να οδηγήσουν τον κινητήρα σε πρόωρη απώλεια, εκτεταμένη φθορά ή πλήρη αστοχία (Σκουντριανός, 2005).

Με την πολυπλοκότητα των νέων κινητήρων η διάγνωση αντιμετωπίζει το ερώτημα "Εάν υπάρχει πρόβλημα και αν υπάρχει πρόβλημα, να εντοπιστεί η βλάβη". Οπότε έτσι γίνεται ποιοτικός έλεγχος των κινητήρων (Μαργαρώνης, 1986).

3.1. Λειτουργικές Παράμετροι στην Διάγνωση

Διάγνωση με βάση τις λειτουργικές παραμέτρους είναι εκείνη η διαδικασία που χρησιμοποιεί λειτουργικά στοιχεία της θερμοδυναμικής κατάστασης του εργαζόμενου μέσου για τον εντοπισμό των βλαβών των κινητήρων.

Η διαδικασία διάγνωσης χρησιμοποιεί μετρήσεις πολλών θερμοδυναμικών μεγεθών και με μια κατάλληλη επεξεργασία καθορίζει τη βλάβη.

Η διάγνωση έχει τρείς φάσεις: (Μαργαρώνης, 1986)

1. Η μέτρηση των κατασκευαστικών και λειτουργικών παραμέτρων.
2. Η συγκέντρωση στοιχείων σύγκριση παλιότερης εμπειρίας για τον ίδιο κινητήρα.
3. Η επεξεργασία των στοιχείων αυτών και ο καθορισμός των προβλημάτων και των αιτίων που τις προκάλεσαν.

Ένα διαγνωστικό σύστημα πρέπει να είναι ικανό να εντοπίζει τις λειτουργικές παραμέτρους του συστήματος που έχουν υποστεί αλλοιώσεις και προκαλούν δυσλειτουργίες, και να αναγνωρίζει την ύπαρξη τέτοιων δυσλειτουργιών που επέρχονται στη κατάσταση των συνιστωσών του κινητήρα diesel, που μπορεί να οφείλονται σε συγκεκριμένες βλάβες ή σε γενικότερη φθορά του λόγω χρήσης.

Το βασικό στοιχείο της διαγνωστικής κινητήρων diesel είναι ότι όλα τα παραπάνω πρέπει να γίνουν με χρήση πληροφοριών από το σύστημα σε λειτουργία,

χωρίς να απαιτείται αποσυναρμολόγηση ή άλλου είδους άμεση πρόσβαση στο εσωτερικό του συστήματος, με εκμετάλλευση μόνο εξωτερικών πληροφοριών ή παρατηρήσεων.

Από τα μετρούμενα μεγέθη που σχετίζονται με τις θερμοδυναμικές επιδόσεις του κινητήρα (πιέσεις και θερμοκρασίες), γίνεται επιλογή εκείνων που παρουσιάζουν υψηλή ευαισθησία στην κατάσταση υγείας συγκεκριμένων συνιστωσών του συστήματος, των οποίων η παρακολούθηση μας ενδιαφέρει. Η επιλογή αυτή γίνεται καθορίζοντας αρχικά τις συνιστώσες ενδιαφέροντος και στη συνέχεια τις απαιτούμενες μετρήσεις για την παρακολούθησή τους.

Η διαγνωστική πληροφορία, μπορεί να αποκαλυφθεί, αν αυτά μετασχηματιστούν σε κατάλληλες διαγνωστικές παραμέτρους. Για παράδειγμα, οι ίδιες οι τιμές θερμοκρασιών και πιέσεων κατά μήκος ενός κινητήρα μπορούν να δώσουν πολύ μικρή πληροφορία σχετικά με την παρουσία κάποιας βλάβης και την αποτίμηση της συνολικής της κατάστασης, ενώ αντίθετα οι αποκλίσεις των μεγεθών αυτών από τις τιμές τους, για υγιή λειτουργία του κινητήρα, σχετίζονται άμεσα με τις διάφορες συνιστώσες της, παρέχοντας σημαντική διαγνωστική πληροφορία.

Η επιλογή των διαγνωστικών παραμέτρων βασίζεται αφ’ ενός στη φύση των βλαβών που εξετάζονται από το σύστημα και αφ’ ετέρου στην φυσική γνώση του κινητήρα.

Το τελικό στάδιο της διαγνωστικής διαδικασίας αφορά την εκτίμηση της κατάστασης της "υγείας" του κινητήρα και την εξαγωγή διαγνωστικών συμπερασμάτων, με κατάλληλη αξιολόγηση της διαθέσιμης πληροφορίας. Πρόκειται για τη λειτουργία του διαγνωστικού συστήματος που επιτελεί τη διάγνωση και πραγματοποιείται, συνήθως, με χρήση ενός ή περισσοτέρων αυτοματοποιημένων διαγνωστικών μεθόδων. Για την υλοποίηση των μεθόδων αυτών, απαραίτητη είναι η ύπαρξη κατάλληλου λογισμικού και αντίστοιχης βάσης γνώσης για το κινητήρα.

Η βάση γνώσης ενός κινητήρα αποτελείται από 'τιμές αναφοράς’ και από ‘υπογραφές βλαβών’. Οι ‘τιμές αναφοράς' (ή 'ονομαστικές τιμές ’), είναι οι τιμές των διαφόρων παραμέτρων που αντιστοιχούν στην υγιή κατάσταση του κινητήρα. Η γνώση των τιμών αυτών είναι απαραίτητη για όλες τις παραμέτρους που χρησιμοποιούνται στα πλαίσια της διάγνωσης γιατί αποτελούν τη βάση σύγκρισης με σκοπό την εκτίμηση της κατάστασης του κινητήρα. Η ‘υπογραφή βλάβης’ από

την άλλη πλευρά, είναι ένας γενικός όρος που αναφέρεται στις διαφοροποιήσεις από την υγιή κατάσταση που προκαλεί η παρουσία μιας βλάβης. Κάθε βλάβη προκαλεί αλλαγές στις διάφορες παραμέτρους κατά συγκεκριμένο τρόπο. Το σύνολο των διαφοροποιήσεων αυτών ονομάζεται 'υπογραφή της βλάβης'. Οι υπογραφές αυτές πρέπει να είναι διαθέσιμες προκειμένου να υπάρχει η δυνατότητα για αναγνώριση των βλαβών (Λαζάρου Χ. Κλιάνη, Ιωάννη Κ. Νικαλάου, 2003).

Η παραγωγή υπογραφών βλαβών προέρχεται από επεξεργασία μετρήσεων που λαμβάνονται από λειτουργούντες δίχρονους κινητήρες diesel, πριν και μετά την εμφάνιση κάποιας βλάβης.

Όταν στο κινητήρα προκύψει κάποια ανωμαλία, το διαγνωστικό σύστημα θα προτείνει ένα σύνολο από υποθέσεις ή σφάλματα (fault set) που αιτιολογεί την ανωμαλία αυτή. Η πληρότητα του διαγνωστικού συστήματος προϋποθέτει ότι το σύνολο των πραγματικών βλαβών είναι υποσύνολο του συνόλου που προτείνει το διαγνωστικό σύστημα. Η διακριτική ικανότητα του διαγνωστικού συστήματος προϋποθέτει ότι το σύνολο των πιθανών (ή προτεινόμενων) βλαβών θα πρέπει να είναι κατά το δυνατό μικρότερο.

Ακολούθως παρουσιάζεται ένα σύνολο επιθυμητών χαρακτηριστικών για ένα διαγνωστικό σύστημα (Σκουντριανός, 2005).

- Γρήγορος εντοπισμός και διάγνωση των βλαβών του συστήματος.
- Ικανότητα απομόνωσης (Isolability) που του επιτρέπει να ξεχωρίζει τις διάφορες βλάβες μεταξύ τους.
- Ευρωστία ως προς τις διάφορες μορφές θορύβου.
- Αναγνωρισιμότητα νέων βλαβών (Novelty identifiability) η ικανότητα να αποφασίζει ένα διαγνωστικό σύστημα εάν λειτουργεί ανώμαλα, εάν η αιτία είναι γνωστή βλάβη ή μία άγνωστη καινούργια βλάβη.

Ένα διαγνωστικό σύστημα πρέπει να έχει τη δυνατότητα εντοπισμού ενός μεγάλου αριθμού βλαβών. Επιπλέον, πρέπει να υπάρχει και η δυνατότητα για προσθήκη νέων αποκαλυπτόμενων βλαβών, οι οποίες δεν περιλαμβάνονταν αρχικά στο σύστημα (Λούκης Ε., 1993).

Κάθε προυπάρχουσα εμπειρία αλλά και κάθε νέο στοιχείο που εμφανίζεται στην γραμμή παραγωγής, θα πρέπει να αξιολογείται κατάλληλα από το αυτοματοποιημένο διαγνωστικό σύστημα.

Η επίδοση ενός διαγνωστικού συστήματος μπορεί να αξιολογηθεί με βάση τις ακόλουθες παραμέτρους: (Τσελέντη, 1998)

- Ακρίβεια (αριθμός σωστών απαντήσεων) Πρέπει να ανιχνεύει όσο το δυνατό μεγαλύτερο αριθμό πραγματικών βλαβών και ταυτόχρονα να κρατά σε χαμηλό επίπεδο τον αριθμό των εσφαλμένων συναγερμών (false alarms).
- Χρόνος διάγνωσης
- Αριθμός επιτυχών διαγνώσεων.
- Αριθμός λανθασμένων διαγνώσεων.

Ο διαχωρισμός αυτός είναι σημαντικός για την αξιολόγηση της διάγνωσης και την εκπαίδευση.

3.2. Μέτρηση Πίεση Κυλίνδρου Κινητήρα

Η παρακολούθηση της λειτουργίας του κινητήρα μεταφράζεται σε παρακολούθηση κάποιων συγκεκριμένων λειτουργικών παραμέτρων, η εξέλιξη των οποίων αποτελεί δείκτη της κατάστασης στην οποία βρίσκεται ο κινητήρας και επιτρέπει την έγκαιρη πρόβλεψη μιας πιθανής μελλοντικής βλάβης.

Οι περισσότερες δυσλειτουργίες των κινητήρων Diesel, είναι συνδεδεμένες με τη διαδικασία της καύσης εντός των κυλίνδρων. Στην πολύπλοκη αυτή διαδικασία λαμβάνει μέρος πλήθος υποσυστημάτων, όπως οι βαλβίδες ο εκκεντροφόρος άξονας, το σύστημα της έγχυσης του καυσίμου, οι αντλίες και άλλα. Η πίεση των αερίων στο εσωτερικό του κυλίνδρου συνιστά την πιο σημαντική πληροφορία για την εποπτεία όλων των μηχανισμών και των διαδικασιών που συμβαίνουν εντός του θαλάμου καύσης ενός κινητήρα Diesel. Η μέτρησή της αποτελεί το πρώτο σημαντικό ζητούμενο της πλειοψηφίας των μεθόδων παρακολούθησης της λειτουργίας του κινητήρα και διάγνωσης βλαβών. Η επεξεργασία των τιμών της πίεσης στον κύλινδρο είναι ιδιαίτερα σημαντική, καθώς μπορεί να δώσει στον μηχανικό χρήσιμες πληροφορίες σχετικά με την ισχύ της μηχανής, το ρυθμό έκλυσης θερμότητας, τη γωνία ανάφλεξης, τη διάρκεια της καύσης και την ποιότητα της συμπίεσης. Συνεπώς, αναδεικνύεται η πίεση εντός του κυλίνδρου και η εκτίμηση των παράγωγων μεγεθών που προκύπτουν από την επεξεργασία της (όπως η μέγιστη πίεση καύσης, η ενδεικνυόμενη ισχύς, η μέση ενδεικνυόμενη πίεση κλπ.), ως το κύριο μέγεθος επίβλεψης της λειτουργίας ενός

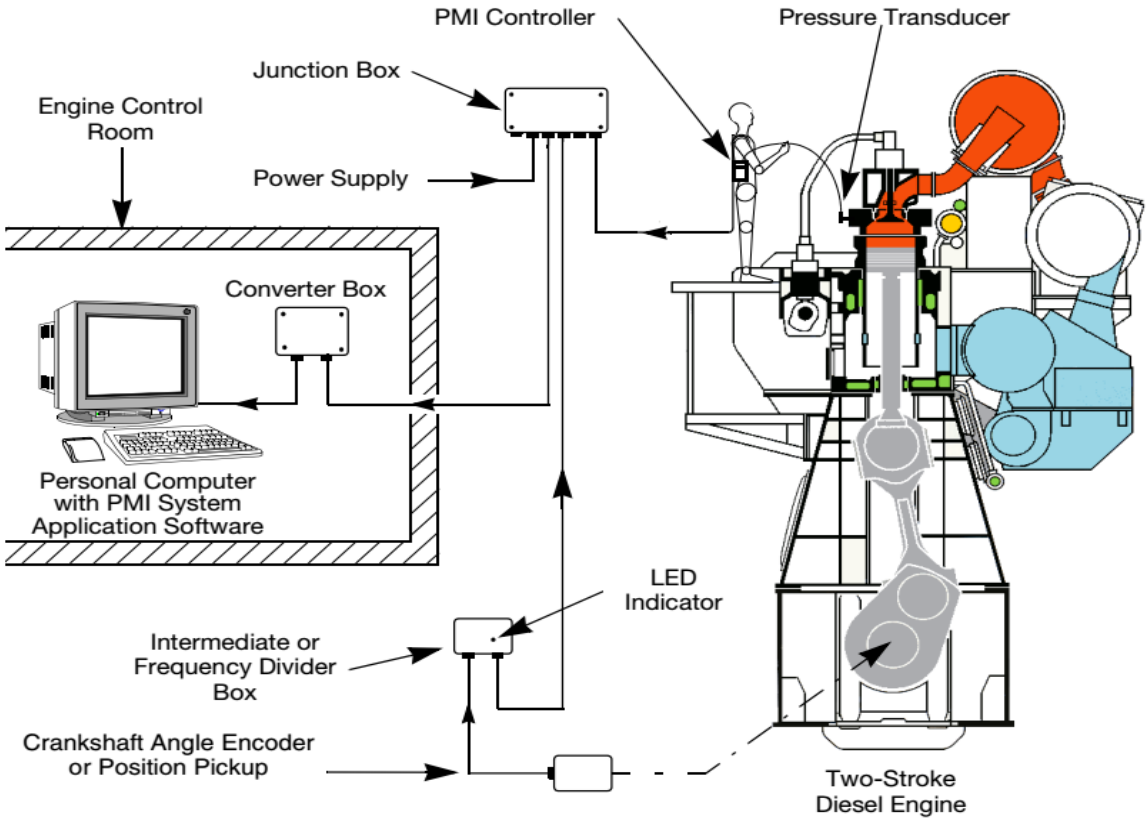
κινητήρα, στη μέτρηση του οποίου βασίζεται η διάγνωσης βλαβών. Το σύστημα MAN B&W Diesel's Off-line PMI (Pressure Management Indicator) είναι σχεδιασμένο να παρέχει στους μηχανικούς και στο τεχνικό προσωπικό μέσα στα πλοία ένα φορητό ηλεκτρονικό εργαλείο για την μέτρηση και ανάλυση των πιέσεων των κυλίνδρων σε MAN B&W δίχρονων κινητήρων diesel.

3.3. Μετρητικός Εξοπλισμός (PMI System)

Το σύστημα MAN–B&W Diesel's Off-line PMI (Pressure Management Indicator) είναι σχεδιασμένο να παρέχει στους μηχανικούς και στο τεχνικό προσωπικό μέσα στα πλοία ένα φορητό ηλεκτρονικό εργαλείο για την μέτρηση και ανάλυση των πιέσεων των κυλίνδρων σε MAN B&W δίχρονων κινητήρων diesel.

Το PMI System, χρησιμοποιείται για τη διαδικασία λήψης μετρήσεων πίεσης κυλίνδρου από κινητήρα Diesel, και παρέχει τη δυνατότητα της ειδικής επεξεργασίας των μετρήσεων που θα χρησιμοποιηθούν για τη διάγνωσή του.

Το σύστημα είναι εφοδιασμένο με το απαραίτητο λογισμικό και υλικό και είναι σχεδιασμένο να υπολογίζει αυτόματα, να εμφανίζει και να αποθηκεύει τα αποτελέσματα των μετρήσεων.



Σχήμα 3-1 MAN Off-line Cylinder Pressure Analyzer PMI System

Πηγή : http://marengine.com/ufiles/MAN-PMI_off.pdf

Τα κύρια προτερήματα χρήσης είναι:

- Εύκολο στη χρήση και χρειάζεται μόνο ένα άτομο να το χρησιμοποιεί.
- Ευαίσθητος και παράλληλα ισχυρός μετατροπέας πίεσης.
- Ευέλικτο σύστημα σκανδάλης. Λειτουργεί με μια ποικιλία από διαφορετικές εισροές (κωδικοποιητής γωνίας, εκπομπές φωτός και εισροές εγγύτητας) για ανίχνευση της γωνίας/περιστροφής του στροφαλοφόρου άξονα. Παράγει έναν μεγάλο αριθμό παλμών σε γωνίες/θέσεις του στροφαλοφόρου για βέλτιστη ακρίβεια.
- Γρήγορα και αξιόπιστα αποτελέσματα. Ένα σύνολο από μετρήσεις και αποτελέσματα παίρνει λιγότερο από δέκα λεπτά να παραχθεί.
- Άριστη ακρίβεια σε σύγκριση με τις μηχανικές συσκευές μέτρησης, συμπεριλαμβανόμενων άλλων ηλεκτρονικών συστημάτων για μετρήσεις σε κινητήρες diesel.
- Χρησιμοποιεί έναν απλό ηλεκτρονικό υπολογιστή και παράλληλα δεν απαιτεί λεπτομερή γνώση στους υπολογιστές.

Επιπλέον έχει δοκιμαστεί στο τομέα από μηχανικούς και τεχνικούς της MAN B&W και ενσωματώνει την τεχνογνωσία της και την εμπειρογνωμοσύνη σχετικά με τον σχεδιασμό των κινητήρων diesel και τον έλεγχο απόδοσης και μετρήσεων (“User’s Guide PMI System,” 2005).

3.4. Λειτουργικοί Παράμετροι Κινητήρα – Όρια Κατασκευαστικών Αποκλίσεων

Για την παρακολούθηση και τον έλεγχο των πιέσεων και των θερμοκρασιών των κυλίνδρων κατά τη λειτουργία του κινητήρα, δεν αρκεί η λήψη ενός δυναμοδεικτικού διαγράμματος για κάθε κύλινδρο, αλλά απαιτείται η λήψη διαδοχικών διαγραμμάτων, ώστε να εξαχθούν μέσες τιμές για κάθε κύλινδρο.

Από τη σύγκριση των διαγραμμάτων αυτών με τα πρότυπα διαγράμματα, αλλά και ελέγχοντας τις τυχόν διαφορές που παρουσιάζονται μεταξύ των κυλίνδρων, εξάγονται σημαντικά συμπεράσματα για τη λειτουργία του κάθε κυλίνδρου. Με τη λήψη των διαγραμμάτων ο μηχανικός γνωρίζει ανά πάσα στιγμή την παραγόμενη ισχύ από κάθε κύλινδρο, καθώς είναι εύκολο να υπολογισθεί η μέση πίεση

λειτουργίας. Από τη μορφή και τις διαφοροποιήσεις των διαγραμμάτων είναι δυνατόν να εξαχθούν συμπεράσματα για την κατάσταση των ελατηρίων του εμβόλου, των βαλβίδων ή των θυρίδων, των αντλιών πετρελαίου, των εγχυτήρων καυσίμου καθώς και για τη σωστή λειτουργία του συστήματος σαρώσεως. Διαπιστώνοντας διαφορές, μπορεί να γνωρίζει εάν αυτές είναι μέσα στα επιτρεπόμενα όρια που ορίζει ο κατασκευαστής.

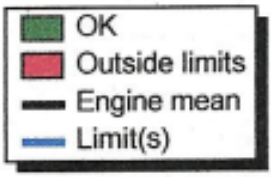
Το πεδίο που ερευνούμε είναι κυρίως η διαφοροποίηση των πιέσεων και θερμοκρασιών. Η διατήρησή τους παίζει βασικό ρόλο στην απόδοση του κινητήρα στα πλαίσια των κατασκευαστών. Γι' αυτό όταν γίνεται εντοπισμός σημαντικής αποκλίσεις παραμέτρων ελέγχονται οι αιτίες που τις προκαλούν.

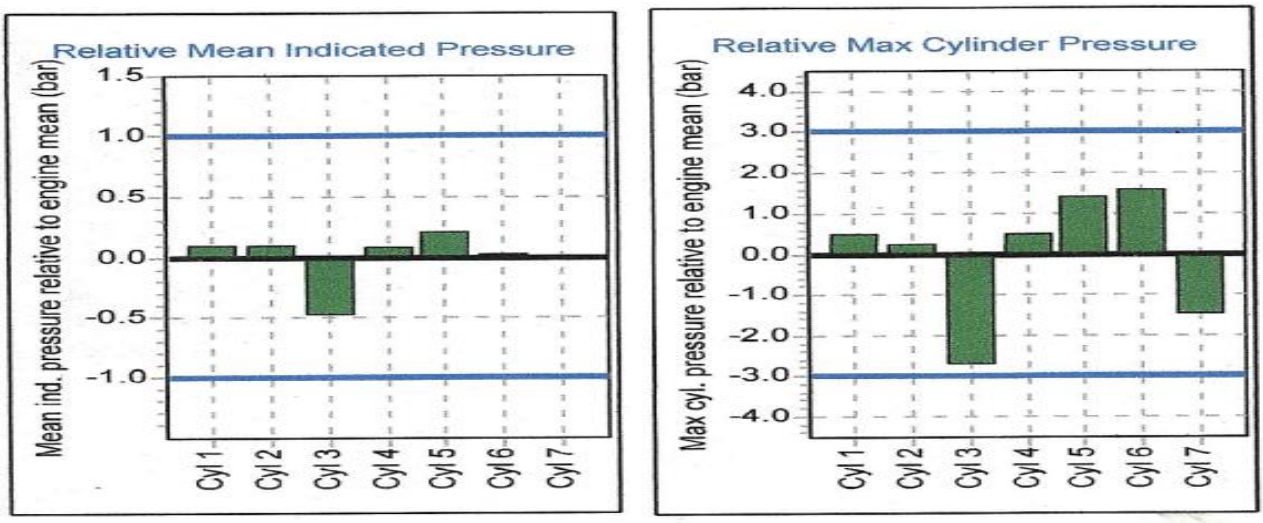
Η εκτιμώμενη απόκλιση για μια λειτουργική παράμετρο, υπολογίζεται από τη μέση τιμή των αποκλίσεων που προέκυψαν από τις μετρήσεις των κυλίνδρων, είναι ο αριθμητικός μέσος όρος των μετρήσεων πίεσης ενός συγκεκριμένου κυλίνδρου. Από τον υπολογισμό προκύπτουν οι εκτιμώμενες τιμές των λειτουργικών παραμέτρων όλων των περιπτώσεων.

Από το σχήμα 3-2 βλέπουμε τα κατασκευαστικά όρια σε δύο λειτουργικές παραμέτρους και τις αποκλίσεις των πιέσεων, μέσης ενδεικτικής πίεσης ±1 bar και μέγιστης πίεσης κυλίνδρου ±3 bar.

Cyl #	Power (kW)	Rpm (rpm)	Pmi (bar)	Pcomp (bar)	Pmax (bar)	Pmax pos (deg)	Pexp (bar)	Pmax-c (bar)	Ignition (deg)
1	1596	96.8	14.6	93.4	114.9	13.2	40.8	21.5	2.6
2	1588	96.3	14.6	94.0	114.6	12.9	41.7	20.6	2.3
3	1531	96.6	14.0	94.2	111.6	12.8	39.7	17.4	2.2
4	1584	96.1	14.6	93.5	114.9	14.6	40.6	21.4	2.7
5	1594	95.9	14.7	95.2	115.8	13.1	41.7	20.6	3.2
6	1585	96.6	14.5	94.2	116.0	13.7	41.2	21.8	2.4
7	1588	96.8	14.5	94.4	112.9	13.4	40.7	18.4	2.8
Mean	1581	96.4	14.5	94.1	114.4	13.4	40.9	20.2	2.6

Σχήμα 3-2 Πίνακας Condition Report επτακύλινδρης μηχανής



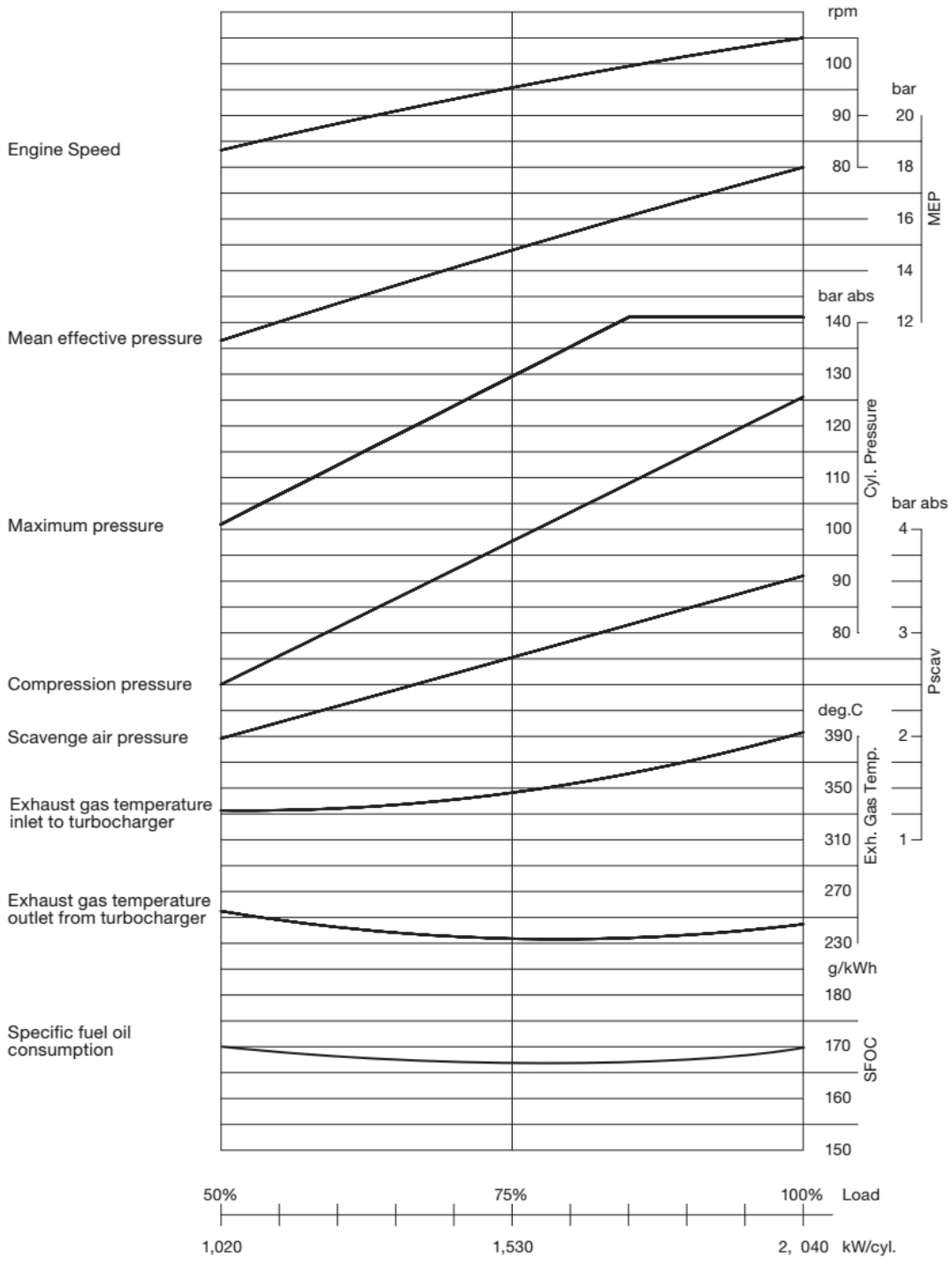


Σχήμα 3-3 Διαφορά μεταξύ της πραγματικής πίεσης και της μέσης μετρούμενης πίεσης (απόκλιση) για κάθε κύλινδρο της μηχανής

Ομάδα Χαρακτηριστικών	*Λειτουργικοί παράμετροι μηχανής ανά κύλινδρο*	*Όρια Κατασκευαστικών Αποκλίσεων*	*Κατηγορίες - Τάξεις*
Όλες	*Όλοι*	Αποκλίσεις Εντός Ορίων	*OK*
power_c1 έως power_c7	*Ισχύς (KW)*	Μεγαλύτερη από 2000 KW	*Power*
rpm_c1 έως rpm_c7	*Ταχύτητα Στροφές ανά λεπτό (rpm)*	Μικρότερη από μέση τιμή κυλίνδρων -5%	*RPM_low*
rpm_c1 έως rpm_c7	*Ταχύτητα Στροφές ανά λεπτό (rpm)*	Μικρότερη από μέση τιμή κυλίνδρων -10%	*RPM_very_low*
pmi_c1 έως pmi_c7	*Μέση ενδεικτική πίεση (bar)*	Μεγαλύτερη από μέση τιμή κυλίνδρων + 1 bar	*Pmi_high*
		Μικρότερη από μέση τιμή κυλίνδρων - 1 bar	*Pmi_low*
pmi_c1 έως pmi_c7 exhaust_gass_temp_c1 έως exhaust_gass_temp_c7	*Μέση ενδεικτική πίεση (bar) και Θερμοκρασία των καυσαερίων* ^{o}C	Μεγαλύτερη από μέση τιμή κυλίνδρων + 1 bar και Μεγαλύτερη από 410 ^{o}C	*Pmi_high Exhaust_gass_temp_high*
		Μικρότερη από μέση τιμή κυλίνδρων - 1 και Μικρότερη από 330 ^{o}C	*Pmi_low Exhaust_gass_temp_low*
pcomp_c1 έως pcomp_c7	*Πίεση συμπίεσης (bar)*	Μεγαλύτερη από μέση τιμή κυλίνδρων + 3 bar	*Pcomp_high*
		Μικρότερη από μέση τιμή κυλίνδρων - 3 bar	*Pcomp_low*
pmax_pos_c1 έως pmax_pos_c7	*Θέση εκδήλωσης μεγίστης πίεσης καύσης (deg)*	Μεγαλύτερη από 20 μοίρες	*Pmax_pos*
pmax_c_c1 έως pmax_c_c7	*Μεγίστη Πίεση καύσεως (bar)*	Μεγαλύτερη από μέση τιμή κυλίνδρων +2 bar	*Pmax_c_high*
		Μικρότερη από μέση τιμή κυλίνδρων -2 bar	*Pmax_c_low*
ignition_c1 έως ignition_c7	*Γωνία εγχύσεως καυσίμου (deg)*	Μεγαλύτερη από μέση τιμή κυλίνδρων +0,7 μοίρες	*Ignition_ angle _high*
		Μικρότερη από μέση τιμή κυλίνδρων -0,7 μοίρες	*Ignition_ angle _low*
exhaust_gass_temp_c1 έως exhaust_gass_temp_c7	*Θερμοκρασία των καυσαερίων* ^{o}C	Μεγαλύτερη από 410 ^{o}C	*Exhaust_gass_temp_high*
		Μικρότερη από 330 ^{o}C	*Exhaust_gass_temp_low*

Πίνακας 3-1 Ομάδες λειτουργικών παραμέτρων τα όρια των κατασκευαστικών αποκλίσεων της συγκεκριμένης μηχανής και οι αντίστοιχες κατηγορίες βλάβης

Performance Curves

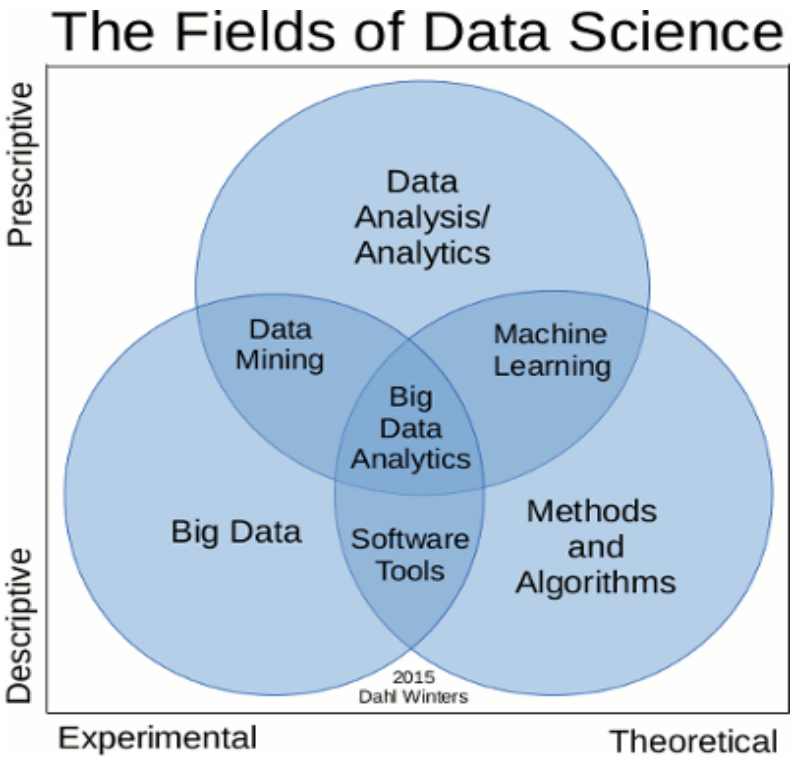


Σχήμα 3-4 Engine Condition report πλοίου, για κύρια δίχρονη αργόστροφη μηχανή τύπου: MAN B&W 7S60MC -C7 (7 Κυλίνδρων)

Πηγή: [("Project Guide Camshaft Controlled Two-stroke Engines," 2009)]

4. ΕΞΟΡΥΞΗ ΓΝΩΣΗΣ ΑΠΟ ΔΕΔΟΜΕΝΑ

Η αύξηση στην ποσότητα των διαθέσιμων στοιχείων άνοιξε την πόρτα σε ένα νέο πεδίο μελέτης που ονομάζεται Big Data η οποία αναφέρεται στον τεράστιο όγκο διαθέσιμων πληροφοριών που μπορεί να αξιοποιηθεί. Έτσι έκανε την εμφάνισή της η Επιστήμη των δεδομένων. Η Επιστήμη των δεδομένων είναι ένας αναδυόμενος τομέας που ασχολείται με την δυνατότητα πρόβλεψης από τα δεδομένα, την εξαγωγή γνώσης, την ερμηνεία των τεράστιων ποσοτήτων των αδόμητων δεδομένων και την γρήγορη και αποτελεσματική λήψη αποφάσεων.



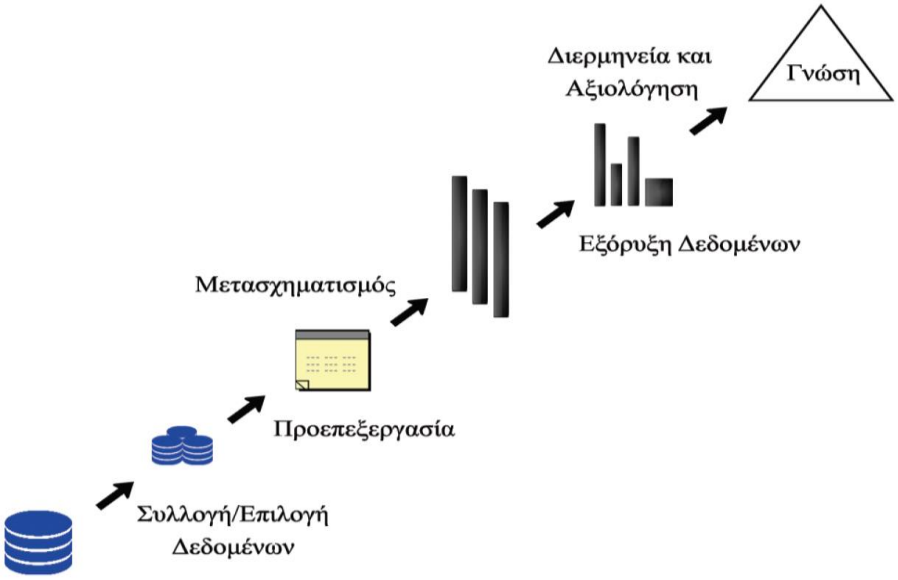
Σχήμα 4-1 Τα πεδία της Επιστήμης Δεδομένων περιλαμβάνουν τη Μηχανική Μάθηση και άλλες μεθόδους
Πηγή :http://scraping.pro/data-analytics-data-analysis-data-mining-data-science-machine-learning-big-data/

4.1. Ανακάλυψη Γνώσης (KDD-Knowledge Discovery in Databases)

Τις τελευταίες δεκαετίες υπήρξε μια εκρηκτική αύξηση στην ποσότητα των πληροφοριών και των δεδομένων που είναι αποθηκευμένα σε ηλεκτρονική μορφή. Σύμφωνα με τελευταίες εκτιμήσεις ο όγκος των πληροφοριών στον κόσμο διπλασιάζεται κάθε 20 μήνες. Οι υπολογιστές λαμβάνουν τα δεδομένα ως είσοδο σε διάφορες μορφές, τα οποία στη συνέχεια μπορούν να τα επεξεργαστούν. Τα δεδομένα είτε είναι αποθηκευμένα σε απλά αρχεία είτε σε υπολογιστικά φύλλα είτε σε πίνακες βάσης δεδομένων, ή σε κάποια άλλη μορφή αποθήκευσης (Radhi, Essa, & Bach, n.d. , 2014).

Τα δεδομένα εκτός του ότι μπορεί να είναι διαφορετικών μορφών μπορούν να προέρχονται και από διαφορετικές πηγές, με άσχετα χαρακτηριστικά μεταξύ τους και αρκετές φορές με αρκετές ελλείψεις.

Πριν από την εφαρμογή κάθε είδους εξόρυξης δεδομένων τα δεδομένα προετοιμάζονται κατάλληλα. Η εξόρυξη δεδομένων είναι επίσης γνωστή με πολλά ονόματα, συμπεριλαμβανομένης της εξόρυξης γνώσης, ανακάλυψη πληροφορίας, συλλογή πληροφοριών και την αρχαιολογία δεδομένων (Fayyad, 1996). Πολλοί ερευνητές και επαγγελματίες χρησιμοποιούν την εξόρυξη δεδομένων ως συνώνυμο για την ανακάλυψη της γνώσης, αλλά τα δεδομένα εξόρυξης είναι επίσης ένα μόνο βήμα της διαδικασίας ανακάλυψης γνώσης. Όλες οι τεχνικές που ακολουθούν μια αυτοματοποιημένη διαδικασία της Ανακάλυψης γνώσης (KDD), δηλαδή, καθαρισμός δεδομένων, ενσωμάτωση δεδομένων, επιλογή δεδομένων, μετατροπή δεδομένων, εξόρυξη δεδομένων και αναπαράσταση γνώσης (Wisaeng, 2013).



Σχήμα 4-2 Στάδια Ανακάλυψης της Γνώσεις KDD

Η διαδικασία της ανακάλυψης γνώσης KDD αποτελείται από μια επαναληπτική ακολουθία από διάφορα στάδια:

- Επιλογή δεδομένων (Selection): Επιλογή των κατάλληλων δεδομένων από τη βάση δεδομένων που σχετίζονται με την εργασία ανάλυσης.
- Προ-επεξεργασία Δεδομένων (Preprocessing): Οι υπάρχουσες βάσεις δεδομένων λόγω του τεράστιου μεγέθους και την πολυπλοκότητα είναι ιδιαίτερα ευαίσθητες σε θόρυβο, ελλιπή δεδομένα, και ασυνεπή δεδομένα. Μετά την επιλογή των δεδομένων πρέπει να μετατραπούν σε μια μορφή που είναι κατάλληλη για την τεχνική ταξινόμησης.
- Μετασχηματισμός δεδομένων (Transformation): Τα δεδομένα μετατρέπονται σε μια κατάλληλη μορφή εκτελώντας εξομάλυνση, γενίκευση, κανονικοποίηση και διακριτοποίησης ώστε να είναι έτοιμα να εκτελεστή η εξόρυξη δεδομένων.
- Εξόρυξης δεδομένων (Data Mining): Η εξόρυξη δεδομένων είναι μια διαδικασία όπου ευφυείς μέθοδοι αλγορίθμων εφαρμόζονται προκειμένου να εξαχθούν τα πρότυπα δεδομένων. Οι τεχνικές εξόρυξης δεδομένων ταξινόμησης, όπως τα δέντρα απόφασης, τα νευρωνικά δίκτυα, οι πλησιέστεροι γείτονες και οι κανόνες ταξινόμησης χρησιμοποιούνται για να εξαχθούν τα σχετικά πρότυπα δεδομένων με τις διάφορες ομάδες δεδομένων.

- Ερμηνεία των δεδομένων / Αξιολόγηση (Interpretation / Evaluation): Η ερμηνεία των προτύπων που εξορύσσονται περιλαμβάνουν την αξιολόγηση μοτίβων και την αναπαράσταση της γνώσης που αντιπροσωπεύονται από τεχνικές απεικόνισης που χρησιμοποιούνται για να βοηθήσει τους χρήστες να κατανοήσουν και να ερμηνεύσουν σωστά τα αποτελέσματα της εξόρυξης δεδομένων.

Η εξόρυξη δεδομένων αποτελεί τμήμα της διαδικασίας ανακάλυψης γνώσης από βάσεις δεδομένων (KDD-Knowledge Discovery in Databases).

4.2. Εξόρυξη Δεδομένων

Ο όρος εξόρυξη δεδομένων αναφέρεται στην εξόρυξη ή την ανακάλυψη νέων πληροφοριών με την μορφή κανόνων ή προτύπων από πηγές δεδομένων. Για να είναι πρακτικά χρήσιμες αυτές οι πληροφορίες πρέπει να εξαχθούν από μεγάλες βάσεις δεδομένων και αρχεία (Fig, 2000).

Το μέγεθος του όγκου από δεδομένα που χρησιμοποιούνται στις βάσεις δεδομένων, δεδομένης της συνεχούς αύξησης των δεδομένων έχει φτάσει μέχρι τα terabytes.

Από τα βασικότερα θέματα της εξόρυξη δεδομένων από τις βάσεις δεδομένων είναι η ανάκτηση, η ταξινόμηση, η ομαδοποίηση, η συσταδοποίηση και η απεικόνιση τους.

Τα στοιχεία της εξόρυξης δεδομένων περιλαμβάνουν την εξόρυξη, τη μετατροπή και τη φόρτωση των δεδομένων σε συστήματα αποθήκευσης δεδομένων, παρέχοντας πρόσβαση σε αναλυτές και ειδικούς της πληροφορικής, για την ανάλυση των δεδομένων με τα εργαλεία εξόρυξης δεδομένων, μετατρέποντάς τα σε μια κατανοητή δομή για περαιτέρω χρήση, όπου μπορούν να απεικονιστούν σε μια χρήσιμη μορφή, όπως ένα γράφημα ή ένας πίνακας.

Η εξόρυξη δεδομένων υποστηρίζεται από τρεις τεχνολογίες που είναι τώρα αρκετά ώριμες:

1. Συλλογή μαζικών δεδομένων
2. Ισχυρούς υπολογιστές με πολλούς επεξεργαστές
3. Αλγόριθμοι εξόρυξης δεδομένων.

Η εξόρυξη δεδομένων είναι μια ευρεία περιοχή που ενσωματώνει τεχνικές από διάφορους τομείς, συμπεριλαμβανομένης της μηχανικής μάθησης.

4.3. Μηχανική Μάθηση

Στην μηχανική μάθηση ο υπολογιστής περιέχει ενσωματωμένους αλγορίθμους εκμάθησης, ο στόχος είναι κάνει μέσω της αυτόματης εκμάθησης να μάθει να αποβλέπει στην εκπαίδευση ενός συστήματος, στην πραγματοποίηση ακριβών προβλέψεων, και στην έξυπνη λήψη αποφάσεων με μικρή ή καθόλου ανθρώπινη παρέμβαση ή βοήθεια. Η μάθηση γίνεται πάντα με βάση κάποιου είδους παρατηρήσεων στα δεδομένων, σημαντικό ρόλο παίζουν τα καταχωρημένα παραδείγματα, η εμπειρία και η εκπαίδευση του συστήματος (Schapire, 2013).

Η μηχανική μάθηση έχει χρησιμοποιηθεί και αξιοποιηθεί σε διάφορους τομείς όπως στην οπτική αναγνώριση χαρακτήρων, στην ανίχνευση προσώπου, στο φιλτράρισμα ανεπιθύμητων μηνυμάτων, στην επισήμανση θεμάτων: όπως κατηγοριοποίηση άρθρων ειδήσεων για την πολιτική, τον αθλητισμό, την ψυχαγωγία, κ.λπ., στην κατανόηση ομιλούμενης γλώσσας, στην ιατρική διάγνωση, στην κατηγοριοποίηση των πελατών στην ανίχνευση βλαβών, στην ανίχνευση της απάτης και την πρόβλεψη του καιρού.

4.3.1. Βήματα Μηχανικής Μάθησης

Υπάρχουν 5 βασικά βήματα που χρησιμοποιείται για να εκτελέσει μιας εργασίας στη μηχανική μάθηση:

1. Συλλογή δεδομένων: Πρόκειται για τα ανεπεξέργαστα δεδομένα προερχόμενα είτε από το Excel, την Access, αρχεία κειμένου κλπ, αυτό το βήμα αποτελεί το θεμέλιο της μελλοντικής μάθησης. Όσο καλύτερη είναι η ποικιλία, η πυκνότητα και ο όγκος των σχετικών δεδομένων, τόσο καλύτερες γίνονται οι προοπτικές μάθησης για το μηχάνημα.
2. Προετοιμασία των δεδομένων: Κάθε αναλυτική διαδικασία ευδοκιμεί σχετικά με την ποιότητα των δεδομένων που χρησιμοποιήθηκαν. Χρειάζεται να γίνει καθορισμός της ποιότητας των δεδομένων όπως τα ελλείποντα στοιχεία και την επεξεργασία των ακραίων τιμών.

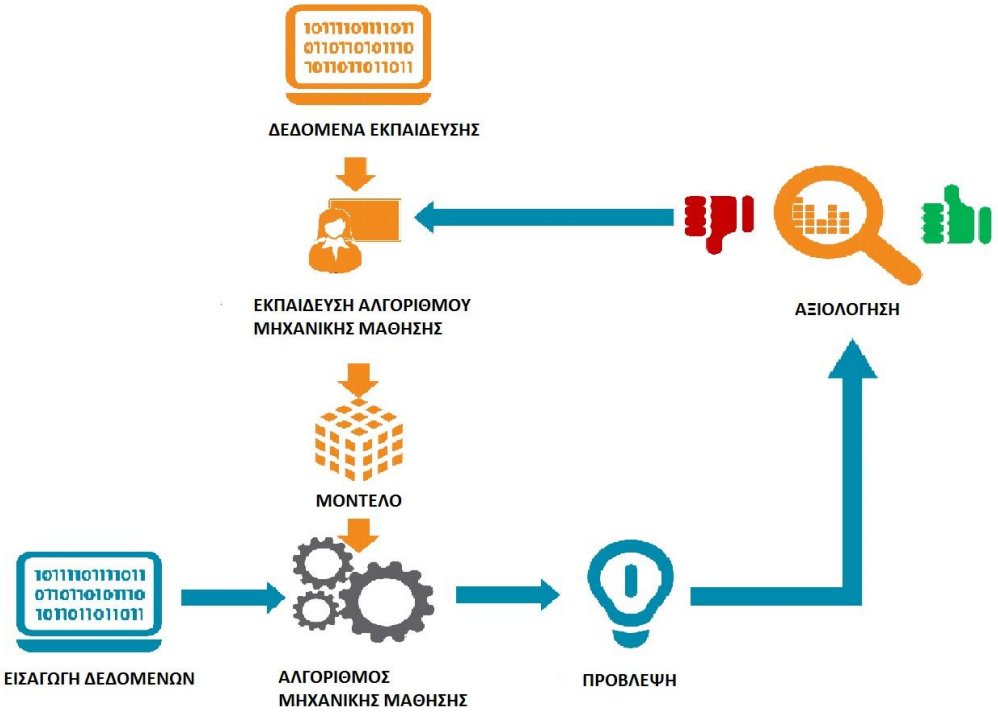
3. Εκπαίδευση ενός μοντέλου: Το βήμα αυτό ενέχει την επιλογή του κατάλληλου αλγορίθμου και αναπαράσταση των δεδομένων με τη μορφή του μοντέλου. Τα καθαρισμένα δεδομένα είναι χωρισμένα σε δύο μέρη, εκπαίδευσης και δοκιμής. Το πρώτο μέρος (δεδομένα εκπαίδευσης) χρησιμοποιείται για την ανάπτυξη του μοντέλου. Το δεύτερο μέρος (δεδομένα δοκιμών), χρησιμοποιείται ως αναφορά.
4. Αξιολόγηση του μοντέλου: Για τον έλεγχο της ακρίβειας, χρησιμοποιείται το δεύτερο μέρος των δεδομένων (δεδομένα δοκιμή). Αυτό το βήμα προσδιορίζει την ακρίβεια στην επιλογή του αλγορίθμου με βάση το αποτέλεσμα.
5. Βελτίωση της απόδοσης: Αυτό το βήμα περιλαμβάνει την επιλογή ενός διαφορετικού μοντέλου ή την εισαγωγή περισσότερων μεταβλητών για να αυξηθεί η αποτελεσματικότητα. Γι 'αυτό σημαντικό χρονικό διάστημα πρέπει να δαπανηθεί στη συλλογή και την προετοιμασία δεδομένων.

Σχήμα 4-3 Τα 5 βασικά βήματα στη μηχανική μάθηση
Πηγή:**https://medium.freecodecamp.org/every-single-machine-learning-course-on-the-internet-ranked-by-your-reviews-3c4a7b8026c0**

Το παρακάτω σχήμα 4-4 απεικονίζεται η διαδικασία κατασκευής ενός μοντέλου.

Ο αλγόριθμος μηχανικής μάθησης εκπαιδεύεται χρησιμοποιώντας ένα σύνολο από επισημανθέντα δεδομένα εκπαίδευσης για την κατασκευή ενός μοντέλου. Εισάγονται νέα δεδομένα εισόδου στον αλγόριθμο μηχανικής μάθησης και κάνουν μια πρόβλεψη με βάση το μοντέλο. Η πρόβλεψη αξιολογείται για την

ακρίβεια και αν η ακρίβεια είναι αποδεκτή, κατασκευάζεται το μοντέλο του αλγορίθμου μηχανικής μάθησης. Εάν η ακρίβεια δεν είναι αποδεκτή, ο αλγόριθμος μηχανικής μάθησης εκπαιδεύεται πάλι με ένα αυξημένο σύνολο δεδομένων εκπαίδευσης.

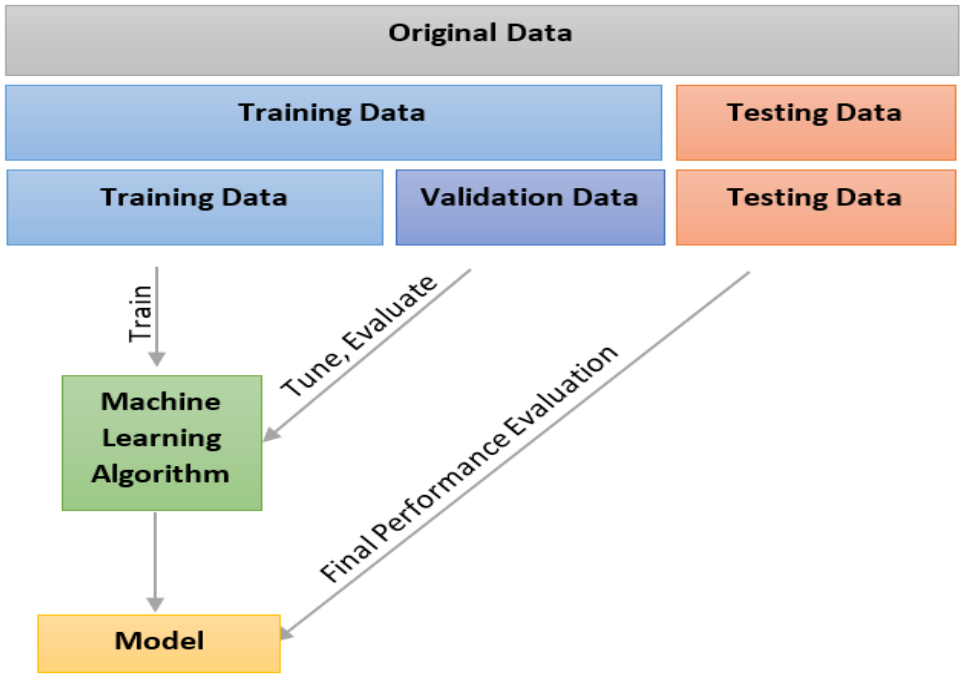


Σχήμα 4-4 Κατασκευή μοντέλου αλγορίθμου μηχανικής μάθησης
Πηγή : http://blogs.teradata.com/data-points/building-machine-learning-infrastructure-2/

Η εκπαίδευση και δοκιμή του μοντέλου όπως απεικονίζεται στο σχήμα 4-5.

Αρχικά, χωρίζονται τα δεδομένα σε ένα σύνολο δεδομένων εκπαίδευσης και δοκιμής. Στη συνέχεια, τα δεδομένα εκπαίδευσης χωρίζονται περαιτέρω σε σύνολα εκπαίδευσης και επικύρωσης.

Τα δεδομένα εκπαίδευσης χρησιμοποιούνται για την εκπαίδευση διαφορετικών μοντέλων. Στη συνέχεια, τα δεδομένα επικύρωσης χρησιμοποιούνται για τον υπολογισμό της απόδοσης του κάθε μοντέλου για να επιλεχθεί το καλύτερο. Τέλος, το μοντέλο χρησιμοποιεί τα σύνολα δοκιμών για την αξιολόγηση της απόδοσης.



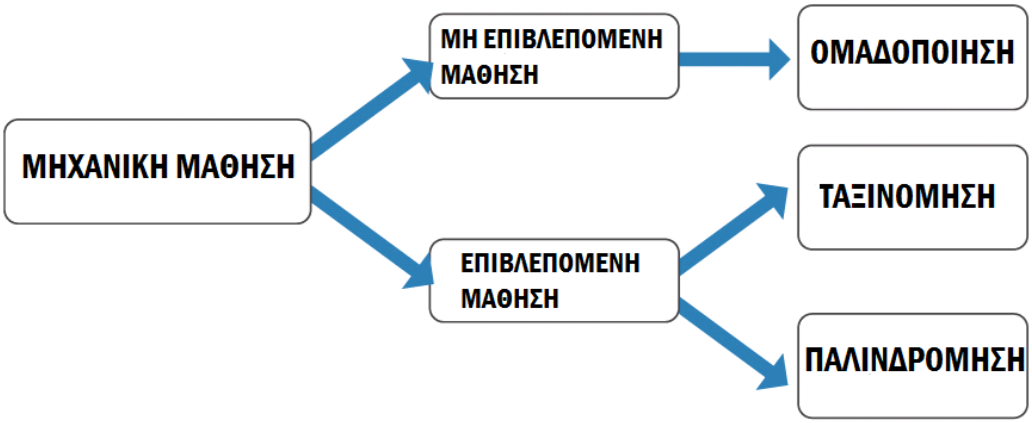
Σχήμα 4-5 Εκπαίδευση και δοκιμή μοντέλου με αλγορίθμους μηχανικής μάθησης
Πηγή :https://www.codeproject.com/Articles/1146582/Introduction-to-Machine-Learning

4.3.2. Κατηγορίες Μηχανικής Μάθησης

Η μάθηση χωρίζεται σε δυο βασικές κατηγορίες στην επιβλεπόμενη μάθηση (Supervised Learning) και στην μη επιβλεπόμενη μάθηση (Unsupervised Learning).

Στην επιβλεπόμενη μάθηση, είναι διαθέσιμα στον υπολογιστή τόσο τα δεδομένα εισόδου (input), όσο και τα αποτελέσματα, κατηγορίες στόχου (output). Οι αλγόριθμοι της επιβλεπόμενης μάθησης περιλαμβάνουν μια φάση που ονομάζεται εκπαίδευση (training) για να αναλύσουν τα δεδομένα και μια φάση δοκιμών (testing) για να εφαρμόσουν το μοντέλο πρόβλεψης σε καινούργια δεδομένα που πρωτοεισάγονται.

Στην μη επιβλεπόμενη μάθηση, είναι διαθέσιμες μόνο τα δεδομένα εισόδου (input) χωρίς τα αποτελέσματα (output). Οι αλγόριθμοι πρέπει να αποφασίσουν από μόνοι τους πώς θα κατατάξουν τα δεδομένα ανάλογα με την τοπολογία τους.



Σχήμα 4-6 Κατηγορίες Μηχανικής Μάθησης

4.3.3. Ταξινόμηση Δεδομένων

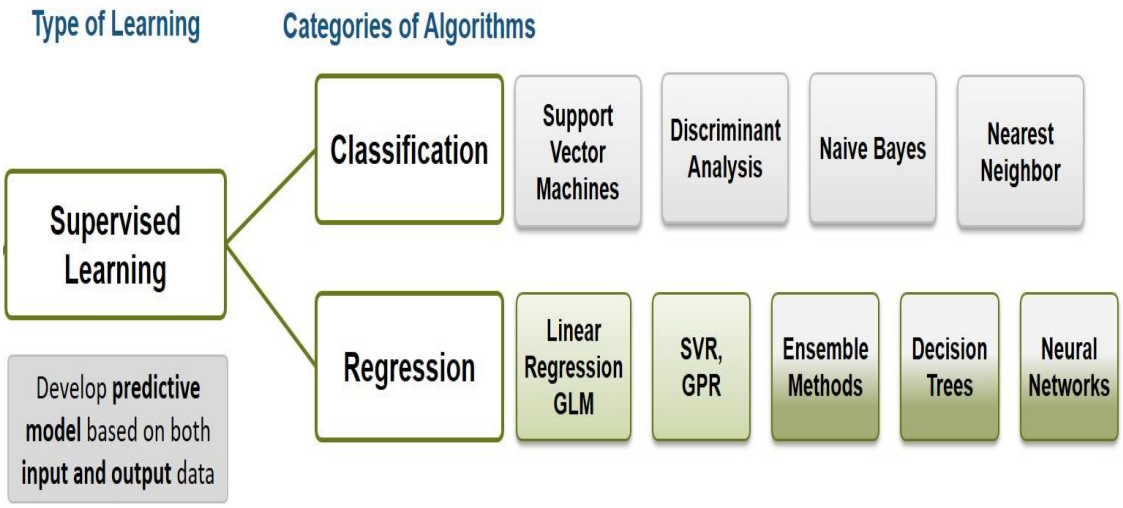
Ένα βασικό πρόβλημα στη διαχείριση δεδομένων είναι η ταξινόμησή τους.

Η ταξινόμηση είναι μια επιβλεπόμενη τεχνική μάθησης. Η ταξινόμησης των δεδομένων είναι ένα πρόβλημα που βρίσκει τη σωστή κατηγορία - τάξη (ή κατηγορίες -τάξεις) από τα δεδομένα όταν ένα σύνολο κατηγοριών - τάξεων και μια συλλογή από ένα σύνολο χαρακτηριστικών δεδομένων είναι δεδομένη. Ο αλγόριθμος ταξινόμησης απαιτεί να καθοριστούν οι κατηγορίες - τάξεις που πρέπει να ορίζονται με βάση τις τιμές των χαρακτηριστικών δεδομένων. Περιγράφει αυτές τις κατηγορίες - τάξεις, ανάλογα με τα χαρακτηριστικά των δεδομένων που ήδη είναι γνωστό σε ποιες κατηγορίες - τάξεις ανήκουν (Kotsiantis, 2007).

Η παρούσα εργασία επικεντρώνεται στην χρήση των τεχνικών ταξινόμησης. Οι πιο συχνά χρησιμοποιούμενες τεχνικές στην ταξινόμηση δεδομένων είναι:

- Τα Τεχνητά Νευρωνικά Δίκτυα: Μη-γραμμικά μοντέλα πρόβλεψης που μαθαίνουν μέσω της εκπαίδευσης και μοιάζουν με βιολογικά νευρωνικά δίκτυα στη δομή.
- Τα Δέντρα Απόφασης: Είναι δομές σε σχήμα δέντρου που αντιπροσωπεύουν τα σύνολα των αποφάσεων. Οι αποφάσεις αυτές δημιουργούν κανόνες για την ταξινόμηση ενός συνόλου δεδομένων.
- Η Μέθοδος του Πλησιέστερου Γείτονα: Μια τεχνική που ταξινομεί κάθε εγγραφή σε ένα σύνολο δεδομένων που βασίζεται σε ένα συνδυασμό από τάξεις. Μερικές φορές ονομάζεται τεχνική k-κοντινότερου γείτονα.

- Οι Κανόνες Επαγωγής: Η εξαγωγή των χρήσιμων κανόνων αν-τότε από δεδομένα βάσει στατιστικής σημαντικότητας.



Σχήμα 4-7 Τεχνικές στην ταξινόμηση δεδομένων
Πηγή : https://channels.theinnovationenterprise.com/articles/analytics-driven-embedded-systems-part-2-developing-analytics-and-prescriptive-controls

5. ΕΡΓΑΛΕΙΑ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΜΩΝ - ΤΕΧΝΙΚΕΣ

Τα εργαλεία εξόρυξης δεδομένων αποτελούν σημαντική βοήθεια στους χρήστες των οποίων το ενδιαφέρον επικεντρώνεται:

- στην ανάλυση δεδομένων
- στην σωστή ταξινόμηση δεδομένων
- στην ταχεία επεξεργασία για την κατανόηση πολύπλοκων δεδομένων.

Αυτό σημαίνει ότι είναι απαραίτητη η ύπαρξη του κατάλληλου λογισμικού εξόρυξης δεδομένων.

Τα εργαλεία εξόρυξης δεδομένων εξυπηρετούν στη διατήρηση και αξιοποίηση της χρήσιμης πληροφορίας των βάσεων δεδομένων, και απαντούν σε ερωτήσεις που προηγουμένως ήταν υπερβολικά χρονοβόρες για την επίλυση τους. Τα εργαλεία εξόρυξης δεδομένων εφαρμόζονται σε συστήματα παράλληλης επεξεργασίας υψηλής απόδοσης, έτσι μπορούν να αναλύουν τεράστιες βάσεις δεδομένων σε λίγα λεπτά. Ταχύτερη επεξεργασία σημαίνει ότι οι χρήστες μπορούν να πειραματιστούν αυτόματα με περισσότερα μοντέλα για την κατανόηση πολύπλοκων δεδομένων. Η υψηλή ταχύτητα καθιστά πρακτικά την δυνατότητα

στους χρήστες να αναλύουν τεράστιες ποσότητες δεδομένων. Οι μεγαλύτερες βάσεις δεδομένων, με τη σειρά τους, βελτίωσαν την απόδοση πρόβλεψης.

Βάσεις δεδομένων μπορούν να είναι μεγαλύτερες τόσο σε εύρος (στήλες) όσο και σε βάθος (σειρές):

- Περισσότερες στήλες - εύρος. Η υψηλή ταχύτητα εξόρυξη δεδομένων επιτρέπει στους χρήστες να εξερευνήσουν το πλήρες βάθος μιας βάσης δεδομένων, χωρίς να περιορίσουν των αριθμό των μεταβλητών.
- Περισσότερες σειρές - βάθος. Τα μεγαλύτερα δείγματα αποδίδουν μικρότερο σφάλμα της εκτίμησης, και επιτρέπουν στους χρήστες να εξάγουν περισσότερα συμπεράσματα.

Κάθε εργαλείο εξόρυξης δεδομένων έχει διαφορετικούς μεθόδους ανάλυσης και ερμηνείας των πληροφοριών από ομαδοποιημένα δεδομένα και έχει τα δικά του πλεονεκτήματα και μειονεκτήματα (Wimmer & Powell, 2015).

Σήμερα, πολλά είδη λογισμικού εξόρυξη δεδομένων είναι διαθέσιμα στο διαδίκτυο και χωρίζονται σε Εμπορικά και Ελεύθερα - Ανοικτού Κώδικα.

Παρακάτω είναι μερικά από τα πιο δημοφιλή εμπορικά λογισμικά εργαλεία εξόρυξης δεδομένων που διατίθενται στο εμπόριο.

Το SAS Enterprise Miner, είναι μια ολοκληρωμένη σουίτα η οποία παρέχει ένα φιλικό προς το χρήστη περιβάλλον GUI ("Data Mining Software, Model Development and Deployment, SAS Enterprise Miner | SAS," n.d.).[1]

Το IBM DB2 Intelligent Miner for Data είναι ένα ανεξάρτητο προϊόν που παρέχει λειτουργίες εξόρυξης δεδομένων για τη δημιουργία και την εφαρμογή

[1] https://www.sas.com/en_us/software/enterprise-miner.html

μοντέλων εξόρυξης δεδομένων. ("IBM Knowledgecenter - IBM DB2 Intelligent Miner for Data," n.d.)[2].

Το Oracle Data Mining είναι μια επιλογή από το σύστημα διαχείρισης βάσης δεδομένων της Oracle Corporation της Relational (RDBMS) Enterprise Edition (EE). Περιέχει αρκετούς αλγορίθμους εξόρυξης δεδομένων και ανάλυσης για την ταξινόμηση, την πρόβλεψη, την παλινδρόμηση και την ανίχνευση ανωμαλιών ("Oracle Data Mining," n.d.)[3].

Το KnowledgeSTUDIO μια ολοκληρωμένη σουίτα της εξόρυξης δεδομένων και πρόβλεψης εργαλείων μοντελοποίησης ("Advanced modeling and decision making software | KnowledgeSTUDIO," n.d.)[4].

Η TIMi Suite είναι μια πλήρης και ολοκληρωμένη σουίτα εργαλείων εξόρυξης δεδομένων ("Business-Intelligence: A categorization of the different B-I solutions," n.d.)[5].

Τα πιο κοινά εργαλεία εξόρυξης δεδομένων ανοικτού κώδικα που χρησιμοποιούν τεχνικές ταξινόμησης είναι:

Το RapidMiner είναι μια πλατφόρμα λογισμικού για την επιστήμη των δεδομένων που αναπτύχθηκε από την ίδια εταιρία που παρέχει ένα ολοκληρωμένο περιβάλλον για την προετοιμασία των δεδομένων , την εκμάθηση μηχανών , τη βαθιά εκμάθηση, την εξόρυξη κειμένου και την ανάλυση προγνωστικών. Το RapidMiner Studio Free Edition, περιορίζεται σε 1 λογικό επεξεργαστή και 10.000 σειρές δεδομένων, διατίθεται με την άδεια AGPL ("Data Science Platform | RapidMiner," n.d.)[6].

Το Apache Mahout είναι ένα έργο του Apache Software Foundation για την παραγωγή δωρεάν εφαρμογών κατανεμημένων ή άλλως κλιμακούμενων αλγορίθμων μηχανικής μάθησης που εστιάζονται κυρίως στους τομείς

[2] https://www.ibm.com/support/knowledgecenter/en/SSEPGG_9.5.0/com.ibm.im.overview.doc/c_ibm_db2_intelligent_miner_for_data.html

[3] http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html

[4] http://www.angoss.com/predictive-analytics-software/software/knowledgestudio/

[5] http://www.business-insight.com/html/products/products.html

[6] https://rapidminer.com/

φιλτραρίσματος , ομαδοποίησης και ταξινόμησης (“Apache Mahout: Scalable machine learning and data mining,” n.d.). http://mahout.apache.org/

Το Orange είναι μια ολοκληρωμένη σουίτα λογισμικού για μηχανική μάθηση και την εξόρυξη δεδομένων, που αναπτύχθηκε στο Εργαστήριο Βιοπληροφορικής, Τμήμα Υπολογιστών και Πληροφορικής, Πανεπιστήμιο της Λιουμπλιάνα, στη Σλοβενία, μαζί με την κοινότητα ανοικτού κώδικα (“Orange,” n.d.)[7].

Το DataMelt (ή, εν συντομία, DMelt) ένα περιβάλλον υπολογισμού και οπτικοποίησης αποτελεί ένα διαδραστικό πλαίσιο για επιστημονικούς υπολογισμούς, ανάλυση δεδομένων και οπτικοποίηση δεδομένων σχεδιασμένο για επιστήμονες, μηχανικούς και μαθητές. Το DataMelt είναι μια πολύ πλατφόρμα δεδομένων γραμμένο σε Java , έτσι τρέχει σε οποιοδήποτε λειτουργικό σύστημα όπου μπορεί να εγκατασταθεί η εικονική μηχανή Java (“DataMelt,” n.d.)[8].

Το KEEL (Εξαγωγή γνώσης βασισμένο στην εξελικτική μάθηση) είναι ένα εργαλείο λογισμικού Java ανοικτού κώδικα (GPLv3) που μπορεί να χρησιμοποιηθεί για μεγάλο αριθμό διαφορετικών εργασιών εντοπισμού δεδομένων γνώσης. Το KEEL παρέχει ένα απλό GUI βασισμένο στη ροή δεδομένων για το σχεδιασμό πειραμάτων με διαφορετικά σύνολα δεδομένων και αλγορίθμους υπολογιστικής νοημοσύνης (δίνοντας ιδιαίτερη προσοχή στους εξελικτικούς αλγόριθμους) προκειμένου να εκτιμηθεί η συμπεριφορά των αλγορίθμων (“KEEL: A software tool to assess evolutionary algorithms for Data Mining problems (regression, classification, clustering, pattern mining and so on),” n.d.)[9].

Το SPMF είναι μια βιβλιοθήκη εξόρυξης δεδομένων εξόρυξης δεδομένων γραμμένη σε Java , εξειδικευμένη στην εξόρυξη προτύπων. Διανέμεται υπό την άδεια GPL v3 (“SPMF: A Java Open-Source Data Mining Library,” n.d.)[10].

Το Rattle GUI είναι ένα πακέτο λογισμικού ελεύθερου και ανοιχτού κώδικα (GNU GPL v2) το οποίο παρέχει ένα γραφικό περιβάλλον χρήστη (GUI) για

[7] http://orange.biolab.si/license/

[8] http://jwork.org/dmelt/

[9] http://www.keel.es/

[10] http://www.philippe-fournier-viger.com/spmf/

την εξόρυξη δεδομένων χρησιμοποιώντας τη στατιστική γλώσσα προγραμματισμού R ("Rattle: A Graphical User Interface for Data Mining using R," n.d.)[11].

5.1. Εργαλείο Εξόρυξης Δεδομένων WEKA

Για την εξαγωγή των αποτελεσμάτων ως προς την ταξινόμηση των δεδομένων μας, θα χρησιμοποιήσουμε το δημοφιλές και ισχυρό εργαλείο WEKA.

5.1.1. Γενικά

Το Weka (Waikato Περιβάλλον για την Ανάλυση Γνώσης) είναι μια δημοφιλής σουίτα λογισμικού μηχανικής μάθησης γραμμένο σε Java , που αναπτύχθηκε και εξελίσσεται στο Πανεπιστήμιο του Waikato, Νέα Ζηλανδία και είναι διαθέσιμο δωρεάν στο παρακάτω link: http://www.cs.waikato.ac.nz/~ml/weka. Αποτελεί ελεύθερο λογισμικό κάτω από την άδεια GNU GPL. Το εργαλείο Weka συγκεντρώνει ένα ολοκληρωμένο σύνολο εργαλείων από αλγόριθμους μηχανικής μάθησης για εργασίες σχετικές με την εξόρυξη δεδομένων, περιέχει εργαλεία για την προεπεξεργασία των δεδομένα, την ταξινόμηση, την παλινδρόμηση, την ομαδοποίηση, τους κανόνες συσχέτισης, την οπτικοποίηση δεδομένων και για τη σύγκριση των αλγορίθμων μάθησης (Frank et al., 2005) (Bouckaert et al., 2013).

5.1.2. Πλεονεκτήματα WEKA

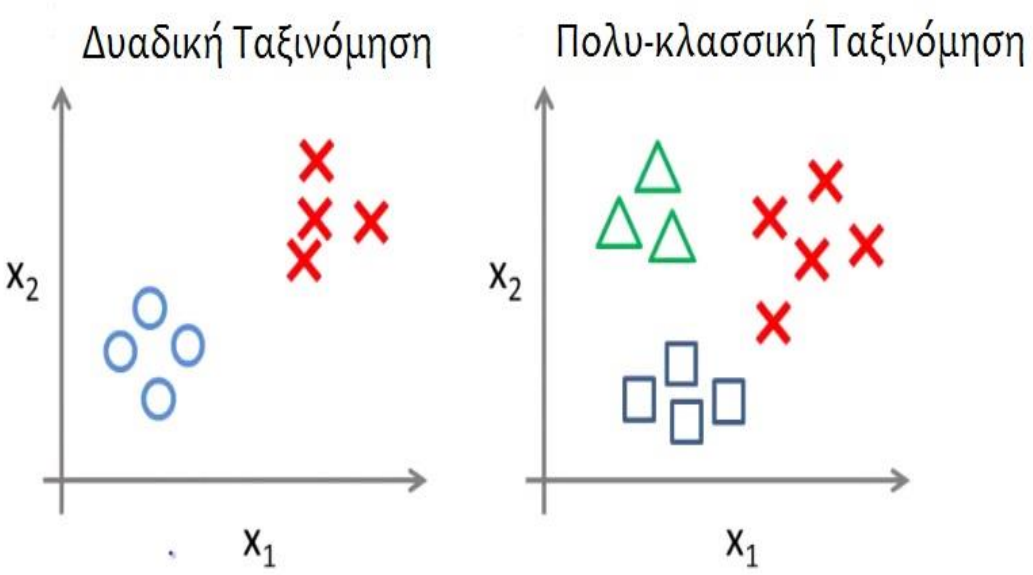
Τα πλεονεκτήματα του εργαλείου εξόρυξης δεδομένων Weka περιλαμβάνουν:(Jović, Brkić, & Bogunović, 2014)(Chauhan & Gautam, 2015)

- Την δωρεάν διαθεσιμότητα υπό την GNU General Public License και προσφέρει πολλά ισχυρά χαρακτηριστικά.
- Φορητότητα, δεδομένου ότι εφαρμόζεται πλήρως στη γλώσσα προγραμματισμού Java και έτσι τρέχει σε σχεδόν κάθε σύγχρονη υπολογιστική πλατφόρμα
- Ευκολία στη χρήση του λόγω γραφικών διεπαφών, το Weka παρέχει πρόσβαση σε SQL βάσεις δεδομένων χρησιμοποιώντας Java Database

[11] https://rattle.togaware.com/

Connectivity και μπορεί να επεξεργαστεί το αποτέλεσμα που επιστρέφεται από ένα ερώτημα βάσης δεδομένων.

- Ισχυρότερη σε τεχνικές μηχανικής μάθησης.
- Είναι επίσης κατάλληλο για την ανάπτυξη νέων συστημάτων μηχανικής μάθησης.
- Μπορεί να ενσωματωθεί σε άλλα πακέτα java.
- Δεν υπάρχει κανένας προγραμματισμός και η γλώσσα κωδικοποίησης που απαιτούνται.
- Μεταξύ αυτών, το εργαλείο WEKA έχει επιτύχει τις υψηλότερες βελτιώσεις στην απόδοση της ακρίβειας(“Weka 3 - Data Mining with Open Source Machine Learning Software in Java,” n.d.).
- Το WEKA μπορεί να χειριστεί τα πρόβλημα των πολλαπλών κλάσεων (multi-class) δεδομένων, το οποίο δεν συμβαίνει σε άλλα εργαλεία εξόρυξης δεδομένων.



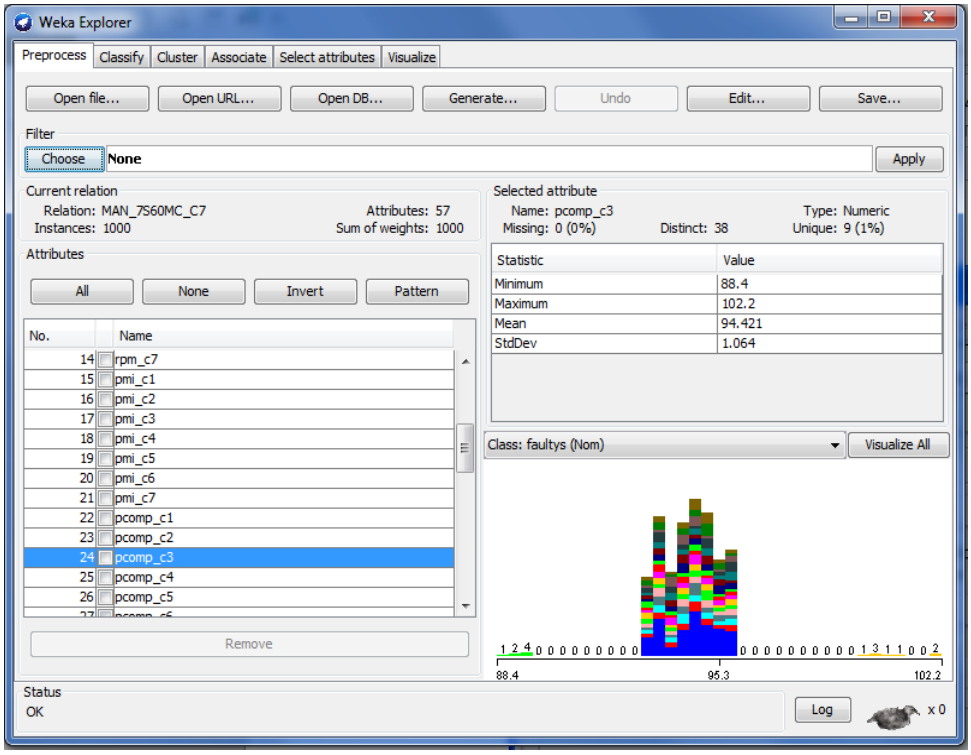
Σχήμα 5-1 Δυαδική και πολύ-κλασσική ταξινόμηση
Πηγή :http://www.geeksforgeeks.org/getting-started-with-classification/

5.1.3. Μενού Επιλογών του WEKA

Σχήμα 5-2 Μενού επιλογών του WEKA

Το περιβάλλον διεπαφών του Weka περιέχει τις ακόλουθες 5 επιλογές:

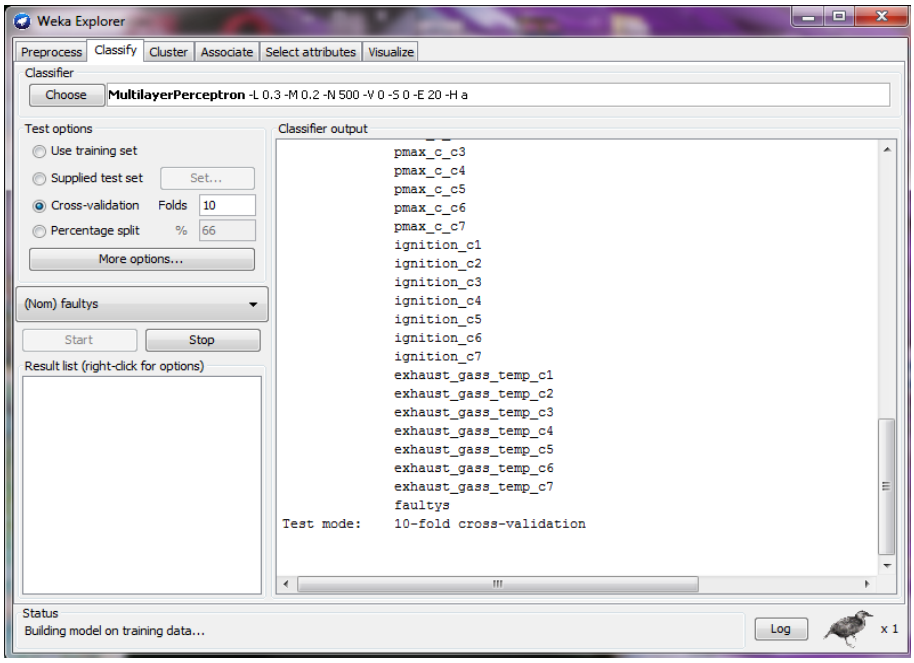
Στην επιλογή Explorer εκτελούνται όλες τις κύριες εργασίες Εξόρυξης Δεδομένων, όπως ταξινόμηση, παλινδρόμηση, ανάλυση συστάδων, ανακάλυψη κανόνων συσχέτισης, προεπεξεργασία των δεδομένων και οπτικοποίηση. Είναι η πιο δημοφιλής διεπαφή αλλά σε αυτό το περιβάλλον ο κύριος στόχος μας είναι η καρτέλα ταξινόμησης (Classify) (Aksenova, 2004).



Σχήμα 5-3 Καρτέλες του WEKA Explorer

5.1.4. Καρτέλα Ταξινόμησης (Classify)

Το WEKA προσφέρει μια μεγάλη ποικιλία εργαλείων για ταξινόμηση. Οι σχετικές εργασίες μπορούν να εκτελεστούν στην καρτέλα 'Classify'.



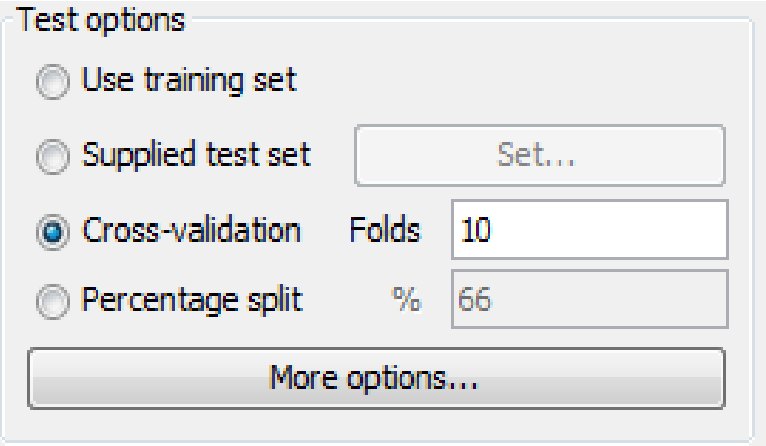
Σχήμα 5-4 Μέθοδοι ταξινόμησης του πεδίου Classifier

Η μέθοδος ταξινόμησης επιλέγεται μέσω του πεδίου 'Classifier'. Οι μέθοδοι χωρίζονται σε ομάδες ανάλογα την κατηγορία που ανήκουν και παρουσιάζονται σε μορφή δένδρου όπως παρουσιάζεται στο σχήμα 5-4.

5.1.5. Επιλογές Εκπαίδευσης και Δοκιμής Ταξινομητή

Τα αποτελέσματα της απόδοσης του μοντέλου του επιλεγμένου ταξινομητή μπορούν να ελεγχθούν με τέσσερις διαφορετικούς τρόπους ελέγχου απόδοσης ταξινομητή, με σκοπό την αξιολόγηση και εξαγωγή μοντέλου.

Για τον έλεγχο της απόδοσης των μοντέλων στην παρούσα εργασία επιλέχθηκε η τεχνική 10 folds cross-validation.



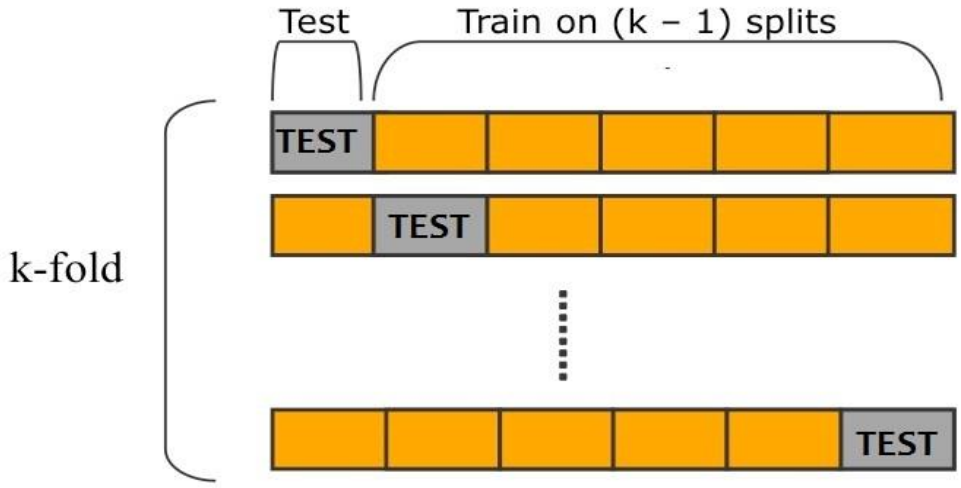
Σχήμα 5-5 Τρόποι ελέγχου απόδοσης ταξινομητή

Ο στόχος της διασταυρούμενης επικύρωσης είναι:

- να ξεπεραστεί το πρόβλημα της υπερεκπαίδευσης
- να γίνει γενικότερη η πρόβλεψη
- να μειώσει τη διακύμανση της εκτίμησης μεταξύ των δεδομένων.

Εκτεταμένα πειράματα έχουν δείξει ότι αυτή είναι η καλύτερη επιλογή για μια ακριβή εκτίμηση (Ross et al., 2009).

Με τη διαδικασία διασταυρωμένης επικύρωσης 'Cross-validation' τα στιγμιότυπα εισόδου διασπούνται τυχαία σε k ομάδες στιγμιότυπων ίσου μεγέθους (folds) και κάθε ομάδα αποτελεί και ένα training set και χρησιμοποιείται για εκπαίδευση, ενώ τα υπόλοιπα για δοκιμή test set, εκ περιτροπής έτσι πραγματοποιείται μια διαδικασία επαναλαμβανόμενης εκπαίδευσης για την πρόβλεψη. Το τελικό ποσοστό σφάλματος είναι ο μέσος όρος των k ποσοστών. Η μέθοδος ονομάζεται k-fold cross-validation. Ο χρήστης μπορεί να ορίσει το πλήθος των τμημάτων και η πιο διαδεδομένη τιμή για το k είναι το 10, χρησιμοποιώντας τον αριθμό των folds που εισάγονται στο ανάλογο πεδίο καθώς έχει αποδειχθεί ότι είναι το καλύτερο για μια ακριβέστερη εκτίμηση (Krčadinac, n.d.).



Σχήμα 5-6 Διαδικασία διασταυρωμένης επικύρωσης σε k ομάδες στιγμιότυπων
Πηγή :http://qingkaikong.blogspot.gr/2017/02/machine-learning-9-more-on-artificial.html

5.1.6. Αποτελέσματα Ταξινομητή (Classifier output)

Τα αποτελέσματα παρουσιάζονται στο παράθυρο εξόδου ταξινομητή (Classifier output), τα οποία περιλαμβάνουν τις ακόλουθες πληροφορίες:

1. Πληροφορίες εκτέλεσης (Run information): Εμφανίζονται τα στοιχεία που αφορούν τον αλγόριθμο μάθησης και τις επιλογές του, στη μέθοδο ελέγχου του αλγορίθμου, καθώς και το αρχείο εκπαίδευσης που χρησιμοποιήθηκε (αριθμός στιγμιότυπων, αριθμός χαρακτηριστικών κλπ.).
2. Μοντέλο ταξινομητή (Classifier model – full training set): Παρουσιάζει το μοντέλο ταξινομητή που δημιουργήθηκε από τα δεδομένα εκπαίδευσης.
3. Περίληψη (Summary): πρόκειται για μια λίστα στατιστικών στοιχείων για την ακρίβεια πρόβλεψης του αλγορίθμου, με βάση τη μέθοδο ελέγχου που επιλέχθηκε.
4. Λεπτομερής ακρίβεια για κάθε κατηγορία (Detailed Accuracy by Class): Η ακρίβεια του ταξινομητή για την πρόβλεψη της κάθε κατηγορίας.
5. Πίνακας σύγχυσης (Confusion Matrix): Πρόκειται για έναν πίνακα που δείχνει πόσα στυγμιότυπα αποδόθηκαν σε κάθε κατηγορία. Σε κάθε κελί του πίνακα υπάρχει ένας αριθμός στυγμιοτύπων που η πραγματική του κλάση είναι η γραμμή στην οποία βρίσκεται και η κατηγορία στην οποία αποδόθηκε από τον αλγόριθμο είναι η στήλη του.

5.1.7. Αξιολόγηση Απόδοσης Αλγορίθμων Ταξινόμησης

Αφού κατασκευαστεί ένα μοντέλο, αξιολογείτε για την ποιότητα του και την ακρίβεια της ταξινόμησης που πετυχαίνει.

Παρακάτω αναφέρονται τα κριτήρια εκείνα που μετρούν την απόδοση του συστήματος ταξινόμησης.

Γενικά, η απόδοση του συστήματος ταξινόμησης δίνεται μέσα από έναν πίνακα που ονομάζεται πίνακας σύγχυσης (confusion matrix).

5.1.7.1. Πίνακας Σύγχυσης

Το σχήμα 5-7 παρουσιάζεται πίνακας σύγχυσης (confusion matrix) πολλαπλών κατηγοριών, που αναπαριστά τις δυνατές εκβάσεις μιας προσπάθειας ταξινόμησης: True Positive (TP), True Negative (TN), False Positive (FP), False Negative (FN) και στη συνέχεια αξιολογείτε ο ταξινομητής υπολογίζοντας τις μετρικές απόδοσης μιας τάξης και στη συνέχεια υπολογίζεται ο μέσος όρος τους για να πάρει μια μέτρηση έναν μόνον αριθμό για την απόδοση όλων των τάξεων.

		PREDICTED				
		A	B	C	D	E
ACTUAL	A	TP_A	E_{AB}	E_{AC}	E_{AD}	E_{AE}
	B	E_{BA}	TP_B	E_{BC}	E_{BD}	E_{BE}
	C	E_{CA}	E_{CB}	TP_C	E_{CD}	E_{CE}
	D	E_{DA}	E_{DB}	E_{DC}	TP_D	E_{DE}
	E	E_{EA}	E_{EB}	E_{EC}	E_{ED}	TP_E

Σχήμα 5-7 Πίνακας σύγχυσης (confusion matrix) πολλαπλών κλάσεων

Οι πραγματικές τιμές αντιπροσωπεύονται από τις στήλες και οι προβλεπόμενες τιμές αντιπροσωπεύονται από σειρές.

Το άθροισμα κάθε γραμμής είναι το πλήθος των πραγματικών στιγμιότυπων κάθε τάξης. Οι αριθμοί κατά μήκος των κυρίων διαγωνίων κάθε γραμμής αποτελούν τις σωστές προβλέψεις του ταξινομητή για την αντίστοιχη τάξη True Positive (TP). Οι αριθμοί εκτός της διαγωνίου κάθε γραμμής είναι τα στιγμιότυπα της αντίστοιχης τάξεις τα οποία αντιπροσωπεύουν τα λάθη τη σύγχυση μεταξύ των διαφόρων

κατηγοριών και έχουν ταξινομηθεί στις υπόλοιπες τάξεις. Άρα κάθε στήλη μας δείχνει τα σωστά και λάθος ταξινομημένα στιγμιότυπα στην τάξη.

5.1.7.2. Μετρικές Απόδοσης

Οι Μετρικές απόδοσης είναι:

Η ορθότητα (accuracy) χρησιμοποιείται ως ένα στατιστικό μέτρο, του πόσο καλά μία δοκιμή σε μία ταξινόμηση προσδιορίζει σωστά ή αποκλείει μια κατάσταση, είναι δηλαδή το ποσοστό των σωστά ταξινομημένων προβλέψεων.

Σε περιπτώσεις εξόδων πολλαπλών κλάσεων (multi class) όπως στην παρούσα εργασία, μετρικές όπως η ορθότητα (accuracy) χάνουν την αξιοπιστία τους. Η ορθότητα (accuracy) από μόνη της είναι μερικές φορές παραπλανητική, καθώς μπορεί να έχει ένα μοντέλο με σχετικά 'υψηλή' ακρίβεια, αλλά το μοντέλο μπορεί να κάνει όλα τα είδη των λαθών για τις τάξεις που είναι πραγματικά κρίσιμα για το πρόβλημα. Τότε για την αξιολόγηση ενός συστήματος χρησιμοποιούνται οι μετρικές: Ακρίβεια (Precision), Ανάκληση (Recall) και ο σταθμισμένος μέσος όρος αυτών η συνάρτηση F (F-Measure) (Singh Sabharwal, n.d.).

5.1.7.2.1. Ακρίβεια (Precision)

Ως Ακρίβεια (Precision) ορίζεται ο λόγος του αριθμού των σωστά προσδιορισμένων τάξεων από το σύστημα προς τον αριθμό των συνολικών τάξεων που ανίχνευσε το σύστημα.

Η Ακρίβεια (Precision) υπολογίζεται ως: $\frac{TP}{TP+FP}$

Όπου TP και FP είναι οι αριθμοί των θετικών και ψευδώς θετικών προβλέψεων για την εξεταζόμενη τάξη.

- Ο συνολικός αριθμός των FP για μια κατηγορία είναι το άθροισμα των τιμών στην αντίστοιχη στήλη (εξαιρουμένων της κύριας διαγώνιας των TP)

5.1.7.2.2. Ανάκληση (Recall)

Ως Ανάκληση (Recall) ορίζεται ο λόγος του αριθμού των σωστά προσδιορισμένων τάξεων από το σύστημα προς τον αριθμό όλων των δεδομένων.

Η Ανάκληση (Recall), ονομάζεται επίσης ευαισθησία (Sensitivity), Αντιστοιχεί στο πραγματικό θετικό ποσοστό της τάξης

Υπολογίζεται ως: $\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$

Όπου TP και FN είναι οι αριθμοί των θετικών και ψευδώς αρνητικών προβλέψεων για την εξεταζόμενη τάξη.

- Ο συνολικός αριθμός των FN για μια κατηγορία είναι το άθροισμα των τιμών στην αντίστοιχη γραμμή (εξαιρουμένων της κύριας διαγώνιας των TP)
- Ο συνολικός αριθμός των δοκιμασμένων παραδειγμάτων της κάθε κατηγορίας θα είναι το άθροισμα της αντίστοιχης γραμμής (είναι το TP+FN για αυτή τη κλάση).

5.1.7.2.3. F-Measure

Στην πράξη οι δύο παραπάνω μετρικές δεν μπορούν να εκτιμηθούν χωριστά, καθώς παρέχουν μια αλληλοσυμπληρούμενη εικόνα της αποτελεσματικότητας ενός ταξινομητή. Δεδομένου ότι και τα δύο μέτρα είναι σημαντικά, συνήθως το μέτρο που τα συνδυάζει είναι η συνάρτηση F (F-Measure), αυτός είναι ο σταθμισμένος μέσος όρος της Ακρίβειας (Precision) και της Ανάκλησης (Recall), που ορίζεται ως εξής:

$$\text{F\_Measure} = \frac{2 * \text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}}$$

Ακριβώς επειδή το F-measure συνδυάζει και τα δύο μεγέθη αποκτά εκ των πραγμάτων μεγαλύτερη βαρύτητα στην αξιολόγηση του αλγορίθμου.

Για κάθε τάξη υπολογίζονται οι μετρικές και στη συνέχεια ο μέσος όρος με χρήση βαρών (Sokolova & Lapalme, 2009).

5.1.7.2.4. Ορθότητα (Accuracy)

Ως ορθότητα (accuracy) ορίζουμε την αναλογία όλων των προβλέψεων που ήταν σωστές (και τα δύο, πραγματικά θετικά (TP) και πραγματικά αρνητικά (TN)) στον πληθυσμό των προβλέψεων και δίνεται από τον ακόλουθο τύπο:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

Accuracy 100% σημαίνει ότι οι μετρούμενες τιμές είναι ακριβώς ίδιες με τις τιμές που δίνονται.

Η Μέση Ακρίβεια (Accuracy): υπολογίζεται ως το άθροισμα των σωστών ταξινομήσεων διαιρούμενο με το συνολικό αριθμό των ταξινομήσεων.

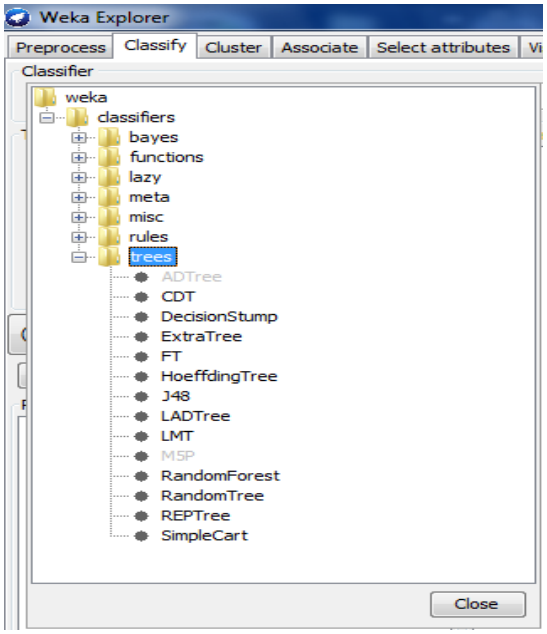
6. ΑΛΓΟΡΙΘΜΟΙ ΤΑΞΙΝΟΜΗΣΗΣ - ΚΑΤΗΓΟΡΙΕΣ

Οι διάφορες κατηγορίες αλγορίθμων ταξινόμησης που χρησιμοποιούνται ευρέως στην προγνωστική μοντελοποίηση είναι:

- Bayes: Αλγόριθμοι που βασίζονται στο Θεώρημα Bayes
- function: Αλγόριθμοι που βασίζονται στην εκτίμηση μιας λειτουργίας.
- lazy: Αλγόριθμοι που χρησιμοποιούν την επονομαζόμενη «τεμπέλικη μάθηση».
- meta: Αλγόριθμοι που χρησιμοποιούν ή συνδυάζουν πολλαπλούς αλγορίθμους.
- misc: Υλοποιήσεις που δεν χωρούν τακτοποιημένα σε άλλες κατηγορίες.
- rules: Αλγόριθμοι που χρησιμοποιούν κανόνες.
- trees: Αλγόριθμοι που χρησιμοποιούν δέντρα απόφασης.

Κάθε μια από τις παραπάνω κατηγορίες παρουσιάζει πλήθος από αλγόριθμους.

6.1. Περιγραφή Κατηγοριών Αλγορίθμων Ταξινόμησης



Σχήμα 6-1 Κατηγορίες αλγορίθμων ταξινόμησης στο Weka

Το είδος του προβλήματος καθορίζει με τι αλγόριθμους μπορούν να εργαστούν καθορίζεται από:

- το είδος – φύση του προβλήματος που επιλύεται
- τη μεταβλητή στόχο που θα προβλεφθεί.

Στην καρτέλα 'classifier' είναι επιλεγμένο κάτω από τις επιλογές της δοκιμής. Από προεπιλογή, το Weka επιλέγει το τελευταίο χαρακτηριστικό στο σύνολο δεδομένων. Εάν το χαρακτηριστικό είναι ονομαστικό, τότε το Weka υποθέτει ότι εργάζεται σε ένα πρόβλημα ταξινόμησης. Εάν το χαρακτηριστικό είναι αριθμητικό, το Weka υποθέτει ότι εργάζεται σε ένα πρόβλημα παλινδρόμησης. Το είδος του προβλήματος καθορίζει με τι αλγόριθμους μπορούμε να εργαστούμε.

Σχήμα 6-2 Επιλογή χαρακτηριστικού εξόδου προβλέψης στο Weka.

Όταν εργαζόμαστε πάνω σε ένα πρόβλημα μηχανικής μάθησης δεν μπορούμε να γνωρίζουμε ποιος αλγόριθμος είναι καλύτερος για το συγκεκριμένο πρόβλημα εκ των προτέρων. Κατά συνέπεια, η λύση είναι η δοκιμή και ανάλυση μιας σειράς από αλγόριθμους για το συγκεκριμένο πρόβλημά για να δούμε ποιος λειτουργεί καλύτερα.

Στο σημείο αυτό εξετάζονται οι πιο αντιπροσωπευτικοί αλγόριθμοι που διαθέτει το WEKA στην βιβλιοθήκη του για κάθε κατηγορία ταξινόμησης.

- Τον Naive Bayes ως εκπρόσωπο των Μπεϋζιανών αλγορίθμων
- Τον J48(C4.5) και Simple Cart ως εκπροσώπους των δέντρων αποφάσεων
- Τον LWL ως εκπρόσωπο των Τεμπέληδων αλγορίθμων
- Τον MultilayerPerceptron ως εκπρόσωπο των Νευρωνικών Δικτύων
- Τον SMO ως εκπρόσωπο των μηχανών διανυσμάτων υποστήριξης
- Τον MODLEN ως εκπρόσωπο των κανόνων ταξινόμησης
- Τον AdaBoost, MultiBoost και Decorate ως εκπροσώπους των μετα-αλγορίθμων

6.1.1. Bayes

Η κατηγορία αυτή αφορά σε στατιστικούς ταξινομητές που βασίζονται στις πιθανότητες εμφάνισης ενός αποτελέσματος.

Διάφοροι αλγόριθμοι της κατηγορίας Bayes έχουν αναπτυχθεί, όπως τα δίκτυα Bayesian και οι Bayes. Η διαφοροποίησή τους έγκειται στην επίδραση-εξάρτηση που ένα χαρακτηριστικό παίζει σε μια δεδομένη κλάση μέσα από κοινές ή μη κατανομές πιθανότητας. Επομένως τα χαρακτηριστικά μπορούν να λαμβάνονται είτε ως ανεξάρτητα μεταξύ τους είτε ως εξαρτημένα.

6.1.2.1. Αλγόριθμος Naive Bayes

Ο Naive Bayes χρησιμοποιεί μια απλή εφαρμογή του Θεωρήματος Bayes (εξ ου και αφελείς), όπου η εκ των προτέρων πιθανότητα για κάθε κατηγορία υπολογίζεται από τα δεδομένα εκπαίδευσης και υποθέτει ότι οι μεταβλητές είναι ανεξάρτητες μεταξύ τους. Η υπόθεση αυτή οδηγεί στον γρήγορο και εύκολο υπολογισμό των πιθανοτήτων. Ο Naive Bayes υπολογίζει την μεταγενέστερη πιθανότητα για κάθε κατηγορία και κάνει μια πρόβλεψη για την κατηγορία με την υψηλότερη πιθανότητα. Οι υπολογισμοί των πιθανοτήτων είναι εύκολο να γίνουν χωρίς την ανάγκη πολύπλοκων επαναληπτικών παραμέτρων εκτίμησης συστημάτων και μπορούν να εφαρμοστούν άμεσα σε τεράστια σύνολα δεδομένων. Είναι ισχυρός και μπορεί είναι να ερμηνευτεί εύκολα, αλλά απαιτεί μεγάλο σύνολο δεδομένων για την εκπαίδευσή του (Amancio et al., 2014).

6.1.3. Functions

Οι μέθοδοι στην κατηγορία Function περιλαμβάνουν τους ταξινομητές που μπορούν να γραφτούν ως μαθηματικές εξισώσεις με αρκετά φυσικό τρόπο. Η μηχανική μάθηση μπορεί να συνοψιστεί ως εκμάθηση μιας συνάρτησης (f) που χαρτογραφεί τις μεταβλητές εισόδου (X) στις μεταβλητές εξόδου (Y).

$Y = f(X)$

Ένας αλγόριθμος μαθαίνει αυτή τη λειτουργία χαρτογράφησης στόχου βασιζόμενος στα δεδομένα εκπαίδευσης.

Οι Function ταξινομητές χρησιμοποιούν την έννοια των νευρωνικών δικτύων και της παλινδρόμησης.

6.1.3.1. Αλγόριθμος Multilayer Perceptron

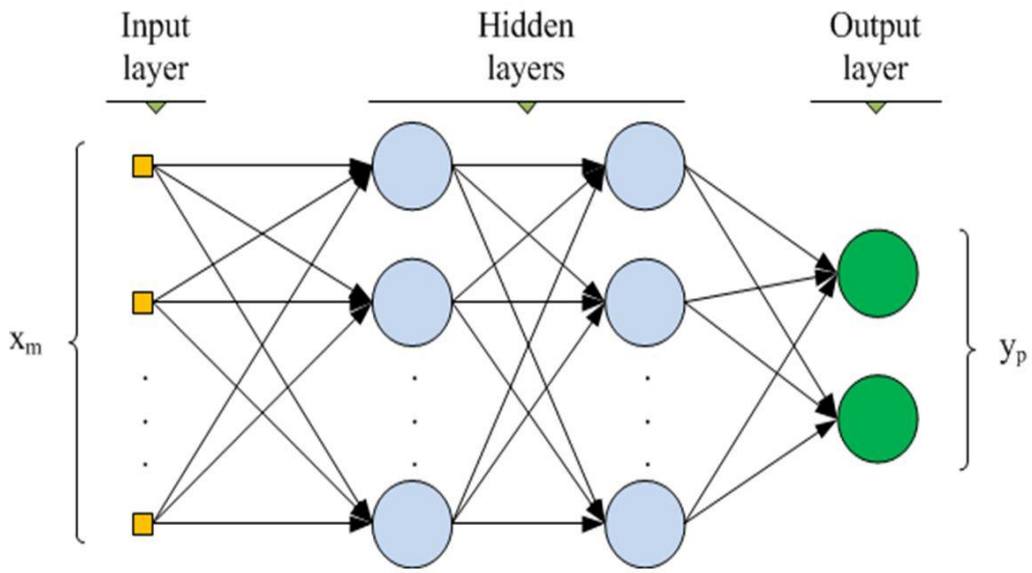
Είναι αλγόριθμος εμπνευσμένος από ένα μοντέλο προσομοίωσης βιολογικού νευρωνικού δικτύου του εγκεφάλου όπου μικρές μονάδες επεξεργασίας που ονομάζονται νευρώνες οργανώνονται σε στρώματα και εάν ρυθμιστούν σωστά είναι ικανά να προσεγγίσουν οποιαδήποτε λειτουργία. Στην εκπαίδευση προσπαθεί να προσεγγίσει την υποκείμενη συνάρτηση για καλύτερες διακρίσεις μεταξύ των τάξεων. Ο Multi-Layer Perceptron αλγόριθμος υποστηρίζει τόσο προβλήματα παλινδρόμησης όσο και ταξινόμησης (Kotsiantis, 2007).

Ένας Multilayer Perceptron είναι ένα μοντέλο τεχνητού νευρωνικού δικτύου που προωθεί τα σύνολα των δεδομένων εισόδου σε ένα σύνολο κατάλληλων εξόδων. Πρόκειται για μια τροποποίηση του τυπικού γραμμικού perceptron με το ότι χρησιμοποιεί ένα ή περισσότερα στρώματα νευρώνων με λειτουργίες μη γραμμικής ενεργοποίησης και είναι πιο ισχυρό από το perceptron, διότι μπορεί να διακρίνει δεδομένα που δεν είναι γραμμικά διαχωρίσιμα ή διαχωρίσιμα από ένα υπερεπίπεδο.

Ο Multilayer Perceptron έχει ιδιαίτερα χαρακτηριστικά. Το μοντέλο του κάθε νευρώνα στο δίκτυο περιλαμβάνει μια μη γραμμική λειτουργία ενεργοποίησης. Το δίκτυο περιέχει ένα ή περισσότερα στρώματα των κρυφών νευρώνων που δεν είναι μέρος της εισόδου ή εξόδου του δικτύου.

Αυτοί οι νευρώνες επιτρέπουν στο δίκτυο να μάθει πολύπλοκα καθήκοντα εξάγοντας προοδευτικά πιο σημαντικά χαρακτηριστικά από τα μοτίβα εισόδου.

Το δίκτυο παρουσιάζει υψηλό βαθμό συνδεσιμότητας που καθορίζεται από το δίκτυο. Μια αλλαγή στη συνδεσιμότητα του δικτύου απαιτεί μια αλλαγή στον πληθυσμό των συναπτικών συνδέσεων στα βάρη τους.



Σχήμα 6-3 Νευρωνικό δίκτυο Αλγορίθμου MultiLayer Perceptron.
Πηγή : http://www.cs.us.es/~fsancho/?e=135

Ο στόχος της διαδικασίας εκπαίδευσης είναι να καθορίσει το σύνολο των τιμών βάρους που θα προκαλέσουν την απόδοση του νευρωνικού δικτύου να ταιριάζει όσο το δυνατόν πιο κοντά στις πραγματικές τιμές στόχου.

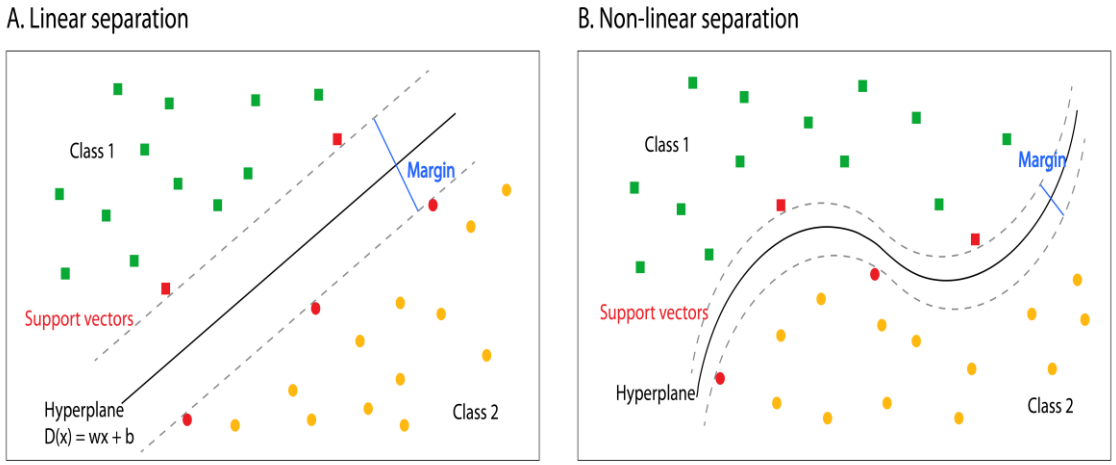
Ο Multilayer Perceptron για την εκπαίδευσή του απαιτεί πολύ χρόνο και μεγάλο όγκο δεδομένων, επίσης τα μοντέλα που κατασκευάζει δεν είναι ερμηνεύσιμα.

6.1.3.2. Αλγόριθμος SMO

Ο SMO εφαρμόζει τον αλγόριθμο ελάχιστης διαδοχικής βελτιστοποίησης (Sequential Minimal Optimization) του John C. Platt που χρησιμοποιείται στο εσωτερικό της εφαρμογής SVM για την εκπαίδευση ενός ταξινομητή support vector χρησιμοποιώντας πυρήνες πολυωνύμων, Gaussian ή RBF (Platt, 1998). Αυτή η εφαρμογή αντικαθιστά συνολικά όλες τις ελλείπουσες τιμές και μετατρέπει τα ονομαστικά χαρακτηριστικά σε δυαδικά. Οι συντελεστές στην έξοδο βασίζονται στα κανονικοποιημένα δεδομένα και όχι στα αρχικά δεδομένα. Η κανονικοποίηση μπορεί να απενεργοποιηθεί ή η είσοδος τυποποιείται σε μηδενική μέση τιμή και διακύμανση μονάδας. Αρχικά, το SVM αναπτύχθηκε για δυαδική ταξινόμηση και δεν είναι απλό να το επεκτείνουμε για πρόβλημα ταξινόμησης πολλαπλών κατηγοριών. Η βασική ιδέα να εφαρμοστεί η πολλαπλή ταξινόμηση στο SVM είναι

να αποσυντεθούν τα προβλήματα πολλαπλών τάξεων σε αρκετά προβλήματα κατηγορίας δύο που μπορούν να αντιμετωπιστούν άμεσα χρησιμοποιώντας διάφορα SVM (Madzarov, Gjorgjevikj, & Chorbev, 2009).

Ο στόχος είναι η εύρεση μιας γραμμής ώστε να διαχωριστούν οι κατηγορίες σε υπερεπίπεδα ώστε τα σημεία δεδομένων να βρίσκονται πλησιέστερα στην επιφάνεια απόφασης. Γεωμετρικά, το περιθώριο αντιστοιχεί στη μικρότερη απόσταση μεταξύ των πιο κοντινών σημείων δεδομένων σε ένα σημείο στο υπερεπίπεδο. Αυτό γίνεται χρησιμοποιώντας μια διαδικασία βελτιστοποίησης, οι περιπτώσεις ονομάζονται διανύσματα υποστήριξης, εξ ου και το όνομα της τεχνικής (Auria & Moro, 2008)(Chih-Wei Hsu, Chih-Chung Chang, 2008). Σε όλα σχεδόν τα προβλήματα, δεν μπορεί μια γραμμή να διαχωρίσει σωστά τις τάξεις, ως εκ τούτου, ένα περιθώριο προστίθεται γύρω από τη γραμμή για να χαλαρώσει τον περιορισμό, επιτρέποντας σε ορισμένες περιπτώσεις οι τάξεις να ταξινομούνται λανθασμένα, αλλά επιτρέποντας συνολικά ένα καλύτερο αποτέλεσμα. Τέλος, λίγα σύνολα δεδομένων μπορούν να διαχωριστούν με μόνο μια ευθεία γραμμή. Μερικές φορές μια γραμμή με καμπύλες ή ακόμα και πολυγωνικά σχήματα πρέπει να δοκιμάζονται. Ο SVM επιτυγχάνει προβάλλοντας τα δεδομένα σε ένα υψηλότερο χώρο διαστάσεων, προκειμένου να καθορίσει τις γραμμές και να κάνει προβλέψεις. Διαφορετικοί πυρήνες μπορούν να χρησιμοποιηθούν για τον έλεγχο της πρόβλεψης και την ευελιξία στο διαχωρισμό των κατηγοριών.



Σχήμα 6-4 Απεικόνιση του διαχωρισμού των δύο κλάσεων με τη χρήση SVM. (A) Γραμμικός και (B) μη γραμμικός
Πηγή:http://www.coxdocs.org/doku.php?id=perseus:user:activities:matrixprocessing:learning: classification parameteroptimization

6.1.4. Lazy

Οι ταξινομητές Lazy (Instance-based Learning) δεν χρησιμοποιούν τα στιγμιότυπα δεδομένων εκπαίδευσης για να κάνουν οποιαδήποτε γενίκευση, αλλά «απομνημονεύουν» το σύνολο δεδομένων εκπαίδευσης αντ 'αυτού. Διαφέρουν από τους πρόθυμους ταξινομητές όπως δέντρα αποφάσεων, νευρωνικά δίκτυα και τα δίκτυα Bayes που γενικεύουν τα δεδομένα εκπαίδευσης πριν λάβουν το ερωτήματα.

Οι Lazy αλγόριθμοι, είναι επίσης γνωστοί ως just-in-time μάθηση και μεταθέτουν την επεξεργασία των στιγμιότυπων μέχρι να λάβουν ρητή αίτηση παροχής πληροφοριών. Όταν συμβαίνει αυτό, η διαθέσιμη βάση δεδομένων αναζήτα εκείνα τα στιγμιότυπα που, σύμφωνα με κάποιο μέτρο της απόστασης, θεωρούνται πιο σχετικά για να απαντήσουν στο ερώτημα. Τόσο η απάντηση όσο και τα τυχόν ενδιάμεσα αποτελέσματα στη συνέχεια απορρίπτονται και για κάθε επόμενη αίτηση πληροφοριών θα κάνει η πλήρη διαδικασία να ξεκινήσει και πάλι.

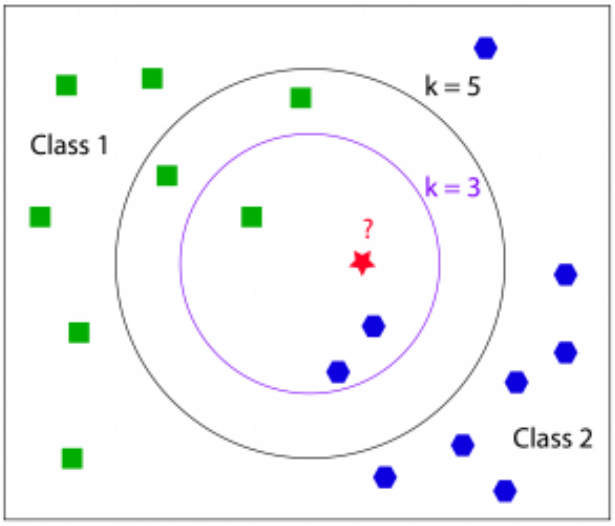
Οι ταξινομητές Lazy, είναι ιδιαίτερα κατάλληλοι όταν τα παραδείγματα δεν είναι όλα διαθέσιμα από την αρχή αλλά συλλέγονται on-line. Σε αυτή την περίπτωση ένα νέο στιγμιότυπο που παρατηρείται απαιτεί μόνο μια ενημέρωση στη βάση δεδομένων. Επίσης, χρειάζεται λιγότερο χρόνο στην εκπαίδευση, αλλά περισσότερο χρόνο στην πρόβλεψη. Γι’ αυτό απαιτούν μεγάλη ποσότητα μνήμης για την αποθήκευση των δεδομένων, και με το γεγονός ότι για κάθε αίτηση παροχής νέων πληροφοριών που παραλαμβάνουν, ξεκινάει η ταυτοποίηση ενός τοπικού μοντέλου από το μηδέν.

Προσφέρουν ελάχιστες εξηγήσεις ή την κατανόηση της οργάνωσης των δεδομένων. Ωστόσο, οι τεμπέληδες μαθητές, υποστηρίζουν την εκ βάθρων μάθηση. Είναι σε θέση να μοντελοποιήσουν περίπλοκους χώρους αποφάσεων που έχουν υπερπολυγωνικά σχήματα και μπορεί να μην είναι τόσο εύκολα περιγραφόμενοι από άλλους αλγορίθμους μάθησης.

Βασίζονται σε στιγμιότυπα που κατά την εκπαίδευση βρίσκουν την πλησιέστερα Ευκλείδεια απόσταση από το δοκιμαστικό στιγμιότυπο που προβλέπει την ίδια τάξη με αυτό το στιγμιότυπο εκπαίδευσης. Ο αριθμός των πλησιέστερων γειτόνων μπορεί να οριστεί στις παραμέτρους ή να καθοριστεί αυτόματα με τη χρήση cross-validation, με την επιφύλαξη ενός ανώτατου ορίου που δίνεται από την καθορισμένη τιμή (Witten, Frank, Hall, & Pal, n.d.).

Στο σχήμα 6-3 φαίνεται η εικόνα ταξινόμησης ενός νέου στοιχείου χρησιμοποιώντας αριθμό πλησιέστερων γειτόνων.

Χρησιμοποιώντας την πλειοψηφία των πλησιέστερων γειτόνων k, το καθορισμένο k μπορεί να αλλάξει την καθορισμένη κατηγορία του κόκκινου αστεριού. Εάν το k=3 (μοβ κύκλος) το άστρο αντιστοιχεί στην κλάση του μπλε πολυγώνου, επειδή οι τρεις πλησιέστεροι γείτονες περιλαμβάνουν δύο μπλε πολύγωνα και ένα πράσινο ορθογώνιο. Ενώ αν το k=5 (μαύρος κύκλος) το αστέρι έχει ανατεθεί στην πράσινη τάξη, επειδή οι πέντε πλησιέστεροι γείτονες περιλαμβάνουν περισσότερα πράσινα ορθογώνια παρά τα μπλε πολύγωνα.



Σχήμα 6-5 Εικόνα ταξινόμησης ενός νέου στοιχείου χρησιμοποιώντας αριθμό πλησιέστερων γειτόνων.
Πηγή : http://www.coxdocs.org/doku.php?id=perseus:user:activities:matrixprocessing:learning:classificationparameteroptimization

6.1.4.1. Αλγόριθμος LWL

Ο LWL (Locally weighted learning) είναι ένας τεμπέλης (lazy) αλγόριθμος για την τοπικά σταθμισμένη εκμάθηση και αναθέτει βάρη χρησιμοποιώντας μια μέθοδο βασισμένη σε στιγμές και δημιουργεί έναν ταξινομητή από τις σταθμισμένες περιπτώσεις (Englert, n.d.).

6.1.5. Trees

Μια προσέγγιση ‘διαίρει και βασίλευε’ στο πρόβλημα της μάθησης από ένα σύνολο ανεξάρτητων στιγμιότυπων οδηγεί σε ένα στυλ εκπροσώπησης που ονομάζεται δέντρο απόφασης.

Τα δέντρα απόφασης είναι ένας κλασσικός αλγόριθμος επιβλεπόμενης μάθησης, εύκολος να γίνει κατανοητός και εύκολος στη χρήση.

Ένα δέντρο απόφασης έχει μια δομή δέντρου, η οποία ξεκινά από τα χαρακτηριστικά ρίζας, και τελειώνει σε κόμβους φύλλων, χρησιμοποιείται κυρίως σε προβλήματα ταξινόμησης. Λειτουργεί και για τις κατηγορηματικές και τις συνεχείς μεταβλητές εισόδου και εξόδου (Rokach & Maimon, n.d.).

Ο στόχος είναι να επιτευχθεί η τέλεια ταξινόμηση με τον ελάχιστο αριθμό αποφάσεων, αν και αυτή δεν είναι πάντα δυνατή λόγω θορύβου ή ασυνεπειών στα δεδομένα. Οι κόμβοι σε ένα δέντρο απόφασης περιλαμβάνουν τη δοκιμή ενός συγκεκριμένου χαρακτηριστικού. Συνήθως, η δοκιμή συγκρίνει μια τιμή χαρακτηριστικού με μια σταθερά (Witten et al., n.d.).

Οι κόμβοι φύλλων δίνουν μια ταξινόμηση που ισχύει για όλες τις περιπτώσεις που φτάνουν στο φύλλο. Για την ταξινόμηση ενός άγνωστου στιγμιότυπου, δρομολογείται προς τα κάτω στο δέντρο σύμφωνα με τις τιμές των χαρακτηριστικών που ελέγχονται σε διαδοχικούς κόμβους και όταν φτάσει σε ένα φύλλο, η περίπτωση ταξινομείται σύμφωνα με την κατηγορία που αντιστοιχεί στο φύλλο. Εάν το χαρακτηριστικό που δοκιμάζεται σε έναν κόμβο είναι ονομαστικό, ο αριθμός των παιδιών είναι συνήθως ο αριθμός των δυνατών πιθανών τιμών του χαρακτηριστικού. Σε αυτήν την περίπτωση, επειδή υπάρχει ένας κλάδος για κάθε πιθανή τιμή, το ίδιο χαρακτηριστικό δεν θα επανεξεταστεί περαιτέρω κάτω από το δέντρο.

Η κατασκευή ενός δέντρου απόφασης μπορεί να εκφραστεί αναδρομικά. Αρχικά, επιλέγετε ένα χαρακτηριστικό που θα τοποθετηθεί στον ριζικό κόμβο και δημιουργείτε ένας κλάδος για κάθε πιθανή τιμή. Αυτό χωρίζει το στιγμιότυπο που ορίζεται σε υποσύνολα, ένα για κάθε τιμή του χαρακτηριστικού. Η διαδικασία μπορεί να επαναληφθεί επαναληπτικά για κάθε κλάδο και χρησιμοποιεί μόνο εκείνες τις περιπτώσεις που φτάνουν πραγματικά στον κλάδο. Αν οποιαδήποτε στιγμή σε όλες τις περιπτώσεις σε έναν κόμβο έχουν την ίδια ταξινόμηση, σταματάει να

αναπτύσσεται αυτό το τμήμα του δέντρου. Απομένει μόνο να καθοριστεί ποιο χαρακτηριστικό θα χωριστεί, δεδομένου ενός συνόλου στιγμιότυπων με διαφορετικές τάξεις.

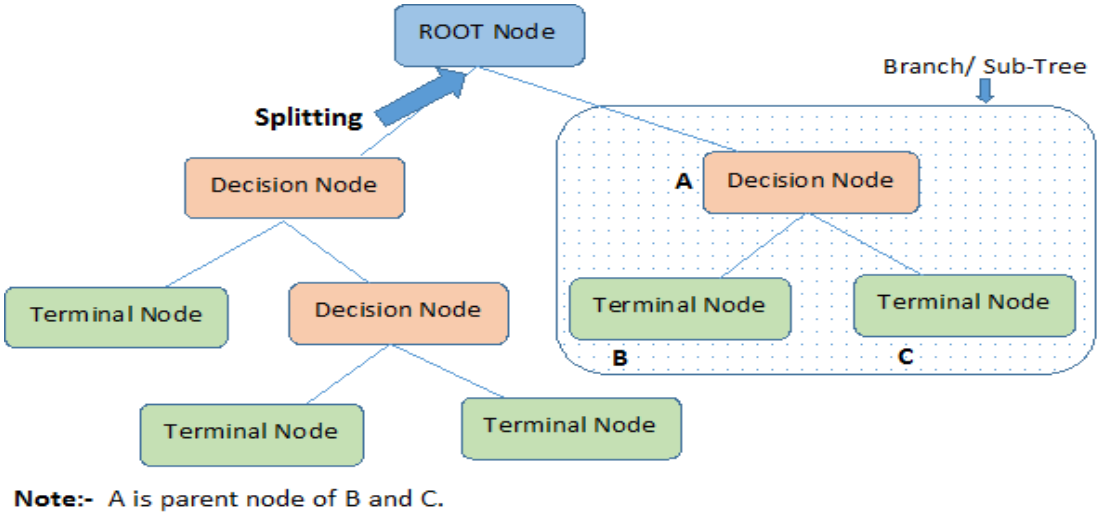
Τα δέντρα απόφασης χρησιμοποιούν πολλούς αλγόριθμους για να αποφασίσουν να χωρίσουν έναν κόμβο σε δύο ή περισσότερους υπο-κόμβους. Η δημιουργία των επιμέρους κόμβων αυξάνει την ομοιογένεια του προκύπτοντος υπο-κόμβου. Έτσι η καθαρότητα του κόμβου αυξάνεται σε σχέση με τη μεταβλητή-στόχο. Το δέντρο απόφασης χωρίζει τους κόμβους σε όλες τις διαθέσιμες μεταβλητές και στη συνέχεια επιλέγει τη διάσπαση που οδηγεί σε πιο ομοιογενή υπο-κόμβους.

Για καλύτερο χωρισμό χαρακτηριστικών που αναφέρεται στην καθαρότητα του κόμβου, κατά την οικοδόμηση του δέντρου είναι η απόφαση του κέρδους πληροφοριών (information gain). Το κέρδος πληροφορίας είναι η διαφορά μεταξύ της εντροπίας πριν και μετά από μια απόφαση. Η εντροπία (entropy) είναι ένα μέτρο της αβεβαιότητας που περιέχεται σε μια πληροφορία, και επιλέγει το χαρακτηριστικό που μειώνει την εντροπία. Ο τύπος για την εντροπία είναι ο ακόλουθος:

Εντροπία = -PP * log2 (ρP) -ρN * log2 (ρN)
PP = το ποσοστό θετικών παραδειγμάτων
PN = το ποσοστό αρνητικών παραδειγμάτων

6.1.5.1. Ανατομία του Δέντρου Απόφασης

Στο παρακάτω σχήμα 6-4 είναι μια οπτική απεικόνιση ενός δέντρου απόφασης.



Σχήμα 6-6 Απεικόνιση δέντρου απόφασης
Πηγή : https://www.analyticsvidhya.com/wp-content/uploads/2015/01/Decision_Tree_2.png

Οι τερματικοί κόμβοι (ή φύλλα) βρίσκονται στο κάτω μέρος του δέντρου απόφασης. Αυτό σημαίνει ότι τα δέντρα απόφασης συντάσσονται ανάποδα έτσι ώστε τα φύλλα να είναι στο κάτω μέρος και οι ρίζες του στην κορυφή.

Η βασική ορολογία που χρησιμοποιείται στα δέντρα απόφασης είναι:

1. Root Node (Κόμβος Ρίζα): Αντιπροσωπεύει σύνολο του πληθυσμού ή του δείγματος και αυτό μπορεί να διαχωριστεί περαιτέρω σε δύο ή περισσότερες ομοιογενεί σύνολα.
2. Splitting (Διάσπαση): Είναι μια διαδικασία της διαίρεση ενός κόμβου σε δύο ή περισσότερους υπο-κόμβους.
3. Decision Node (Απόφαση Κόμβος): Όταν ένας υπο-κόμβος χωρίζεται σε περαιτέρω υπο-κόμβους, τότε καλείται κόμβος απόφασης.
4. Leaf/Terminal Node (Φύλλο/Τερματικός Κόμβος): Ο κόμβος που δεν χωρίζεται ονομάζεται φύλλο ή τερματικός κόμβος.
5. Pruning (Κλάδεμα): Όταν αφαιρείτε ένας υπο-κόμβος ενός κόμβου απόφασης, η διαδικασία αυτή ονομάζεται κλάδεμα. Είναι η αντίστροφη διαδικασία της διάσπασης.
6. Branch/Sub-Tree (Κλαδί/Υπο-Δέντρο): Ένα επιμέρους τμήμα ολόκληρου του δέντρου ονομάζεται κλαδί ή υπο-δέντρο.

7. Parent and Child Node (Γονέας και παιδί Κόμβος): Ένας κόμβος, ο οποίος διαιρείται σε υπο-κόμβους ονομάζεται γονέας κόμβος του υπο-κόμβου, λαμβάνοντας υπόψη ότι ως υπο-κόμβοι είναι το παιδί του γονικού κόμβου.

Τα δέντρα απόφασης κατασκευάζονται με έναν αναδρομικό αλγόριθμο κατάτμησης που περιγράφεται εύκολα με τη χρήση του παρακάτω ψευδοκώδικα.

1. Για κάθε χαρακτηριστικό μεταβλητής εισόδου, αξιολογείται ο καλύτερος τρόπος για να χωριστούν τα δεδομένα σε δύο ή περισσότερες υποομάδες. Επιλέγετε η καλύτερη διάσπαση και χωρίζονται τα δεδομένα σε υποομάδες που ορίζονται από τη διάσπαση.
2. Επιλέγετε μια από τις υποομάδες και επαναλαμβάνετε το Βήμα 1, για κάθε υποομάδα.
3. Συνεχίζετε η διάσπαση έως ότου όλα τα χαρακτηριστικά μετά τη διάσπαση ανήκουν στην ίδια μεταβλητή στόχο.

6.1.5.2. Κλάδεμα του δέντρου απόφασης

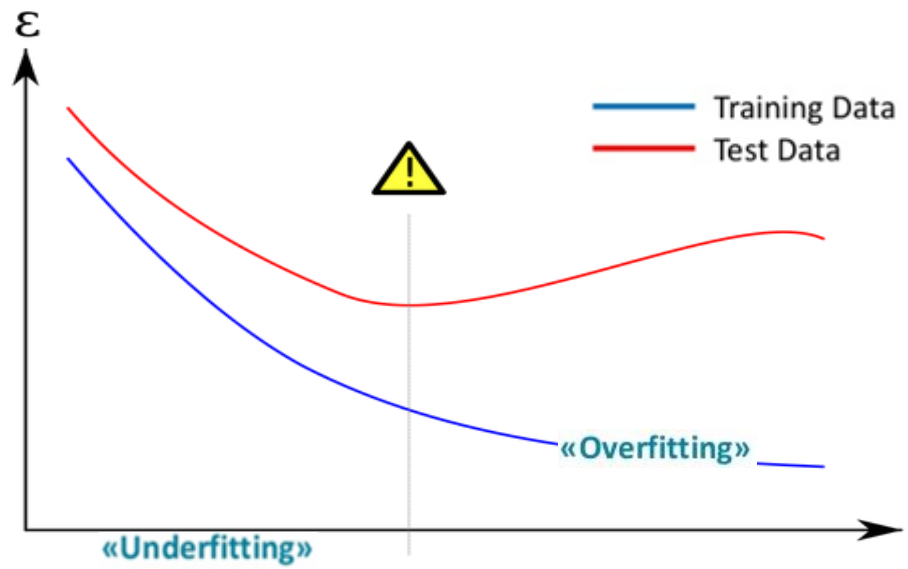
Τα πλήρως αναπτυγμένα δέντρα αποφάσεων συχνά περιέχουν περιττή δομή που οδηγεί σε υπερεκπαίδευση (overfitting) δηλαδή φαίνεται εσφαλμένα να έχουν μικρότερο σφάλμα κατά την εκπαίδευση των δεδομένων, γι’ αυτό πρέπει να απλοποιηθούν μέσω κλαδέματος προτού αναπτυχθούν. Η ταχύτερη και απλούστερη μέθοδος κλαδέματος είναι να δουλευτεί μέσα από κάθε κόμβο φύλλων στο δέντρο και να αξιολογείτε το αποτέλεσμα της αφαίρεσής του. Οι κόμβοι φύλλων αφαιρούνται μόνο εάν το αποτέλεσμα οδηγήσει σε μείωση της συνάρτησης του συνολικού κόστους σε ολόκληρο το σύνολο δοκιμών. Σταματάει να καταργεί τους κόμβους όταν δεν μπορούν να γίνουν περαιτέρω βελτιώσεις. Για το κλάδεμα του δέντρου εφαρμόζεται μια στρατηγική postpruning (μερικές φορές αποκαλούμενη προς τα πίσω κλάδεμα backward pruning) ή και σπάνια την prepruning (ή προς τα εμπρός κλάδεμα forward pruning) (Witten et al., n.d.).

Η εκτίμηση σφάλματος είναι η τυπική τεχνική επαλήθευσης. Κρατούνται κάποια από τα δεδομένα που είχαν δοθεί αρχικά και χρησιμοποιούνται ως ανεξάρτητο δοκιμαστικό σύνολο για να εκτιμηθεί το σφάλμα σε κάθε κόμβο. Αυτό ονομάζεται κλάδεμα με μειωμένο σφάλμα reduced-error pruning. Μειονεκτεί στο ότι το πραγματικό δέντρο βασίζεται σε λιγότερα δεδομένα.

Μπορούν να χρησιμοποιηθούν πιο εξελιγμένες μέθοδοι κλαδέματος, όπως η κλάδευση πολυπλοκότητας κόστους (που ονομάζεται επίσης και ασθενέστερη κλάδευση), όπου μια παράμετρος μάθησης χρησιμοποιείται για να υπολογίσει αν οι κόμβοι μπορούν να αφαιρεθούν με βάση το μέγεθος του υπο-δέντρου.

Στο σχήμα 6-5 εμφανίζεται το φαινόμενο της υπερεκπαίδευσης.

Ο κάθετος άξονας αντιπροσωπεύει το σφάλμα μιας υπόθεσης. Ο οριζόντιος άξονας αντιπροσωπεύει την πολυπλοκότητα της υπόθεσης. Η μπλε καμπύλη αντιπροσωπεύει το σφάλμα της εξόδου ενός αλγορίθμου εκμάθησης μηχανής στα δεδομένα εκπαίδευσης και η κόκκινη καμπύλη αντιπροσωπεύει τη γενίκευση αυτής της υπόθεσης στον πραγματικό κόσμο. Το φαινόμενο υπερεκπαίδευσης είναι σημάδι στη μέση του γραφήματος, πριν από το οποίο το σφάλμα της εκπαίδευσης και το σφάλμα γενίκευσης πέφτουν κάτω, αλλά μετά το οποίο το σφάλμα της εκπαίδευσης συνεχίζει να μειώνεται ενώ το σφάλμα γενίκευσης αυξάνεται.



Σχήμα 6-7 Το φαινόμενο υπερεκπαίδευσης
Πηγή : https://jeremykun.com/2015/09/21/the-boosting-margin-or-why-boosting-doesnt-overfit/

6.1.5.3. Δέντρα Ταξινόμησης και Παλινδρόμησης

Τα δέντρα ταξινόμησης και παλινδρόμησης ή το CART για συντομία είναι ένας όρος που εισήγαγε ο Leo Breiman για να αναφερθεί στους αλγόριθμους δέντρων απόφασης που μπορούν να χρησιμοποιηθούν για προβλήματα μοντελοποίησης ταξινόμησης ή παλινδρόμησης (Breiman, 1984).

Τα δέντρα λειτουργούν σχεδόν όμοια μεταξύ τους, υπάρχουν όμως βασικές διαφορές και ομοιότητες μεταξύ των δέντρων ταξινόμησης και παλινδρόμησης:

1. Τα δέντρα παλινδρόμησης χρησιμοποιούνται όταν οι εξαρτημένες μεταβλητές είναι συνεχής.
2. Τα δέντρα ταξινόμησης χρησιμοποιούνται όταν οι εξαρτημένες μεταβλητές είναι κατηγορηματικές.
3. Στην περίπτωση των δέντρων παλινδρόμησης, η τιμή που λαμβάνεται από τερματικούς κόμβους στα δεδομένα εκπαίδευσης είναι η μέση απόκριση της παρατήρησης που υπάγονται στην εν λόγω περιοχή. Έτσι, αν μια νέα παρατήρηση των δεδομένων εμπίπτει στην εν λόγω περιοχή, θα κάνει πρόβλεψη με μέση τιμή.
4. Στην περίπτωση των δέντρων ταξινόμησης, η τιμή (κατηγορία) που λαμβάνεται από τερματικό κόμβο στα δεδομένα εκπαίδευσης είναι η λειτουργία των παρατηρήσεων που υπάγονται στην εν λόγω περιοχή. Έτσι, αν μια νέα παρατήρηση των δεδομένων εμπίπτει στην εν λόγω περιοχή, θα κάνει πρόβλεψη της με την τιμή λειτουργίας.

6.1.5.4. Αλγόριθμος J48

Ο J48 είναι μια open source Java υλοποίηση του αλγορίθμου C4.5 (“J” για Java και 48 για C4.8) στο εργαλείο εξόρυξης δεδομένων Weka. Το C4.5 είναι ένα πρόγραμμα που δημιουργεί ένα δέντρο απόφασης που βασίζεται σε ένα σύνολο επισημασμένων δεδομένων εισόδου. Αυτός ο αλγόριθμος αναπτύχθηκε από τον Ross Quinlan (Quinlan, 1993). Τα δέντρα απόφασης που παράγονται από το C4.5 μπορούν να χρησιμοποιηθούν για ταξινόμηση και γι 'αυτό το λόγο το C4.5 αναφέρεται συχνά ως στατιστικός ταξινομητής "C4.5 (J48)".

Ο J48 χρησιμοποιεί την εντροπία πληροφοριών για την κατασκευή δέντρων αποφάσεων από το σύνολο των δεδομένων εκπαίδευσης. Κάθε κόμβος του δέντρου αντιπροσωπεύει την πιο αποτελεσματική διάσπαση των δειγμάτων που προσδιορίζεται από το υψηλότερο κανονικοποιημένο κέρδος πληροφοριών. Το Weka επιτρέπει επίσης τη οπτική αναπαράσταση του δέντρου απόφασης για τον αλγόριθμο J48.

6.1.5.5. Αλγόριθμος Simple Cart

Το Simple Cart (δέντρο ταξινόμησης και παλινδρόμησης) αναπτύχθηκε από τον Leo Breiman στις αρχές του 1980 (Moore, 1987). Το Simple Cart χρησιμοποιεί δείγμα μάθησης το οποίο είναι ένα σύνολο στοιχείων με προκαθορισμένες τάξεις για όλες τις παρατηρήσεις για την οικοδόμηση του δέντρου απόφασης. Το Simple Cart (δέντρο ταξινόμησης και παλινδρόμησης) είναι μια τεχνική ταξινόμηση και δημιουργεί δυαδικό δέντρο απόφασης. Δεδομένου ότι η παραγωγή είναι δυαδικό δένδρο, που παράγει μόνο δύο παιδιά. Η μέθοδος Εντροπίας χρησιμοποιείται για να επιλεγεί ο καλύτερος διαχωρισμός χαρακτηριστικού. Το Simple Cart χειρίζεται δεδομένα που λείπουν αγνοώντας αυτήν την εγγραφή (Timofeev & Härdle, 2004).

Το Simple Cart είναι μια τεχνική μάθησης, η οποία δίνει τα αποτελέσματα ως δέντρο παλινδρόμησης ή ταξινόμησης, ανάλογα με το εάν είναι κατηγορηματικό ή αριθμητικό το σύνολο δεδομένων. Αυτή η μεθοδολογία είναι ίσως η πιο γνωστή και πιο ευρέως χρησιμοποιούμενη. Χρησιμοποιεί επικύρωση ή μια μεγάλη ανεξάρτητη δοκιμή σε δείγμα δεδομένων για να επιλεγεί το καλύτερο δέντρο από την ακολουθία των δέντρων που θεωρούνται κατά τη διαδικασία κλαδέματος. Ο βασικός αλγόριθμος Simple Cart είναι ένας άπληστος αλγόριθμος που επιλέγει τα καλύτερα διαχωρίσιμα χαρακτηριστικά σε κάθε στάδιο της διαδικασίας. Κατά την εφαρμογή του Simple Cart, το σύνολο δεδομένων χωρίζεται σε τα δύο υποομάδες που είναι ο πιο διαφορετικές σε σχέση με το αποτέλεσμα. Η διαδικασία αυτή συνεχίζεται για κάθε υποομάδα έως ότου επιτυγχάνεται κάποιο μέγεθος ελάχιστης υποομάδας.

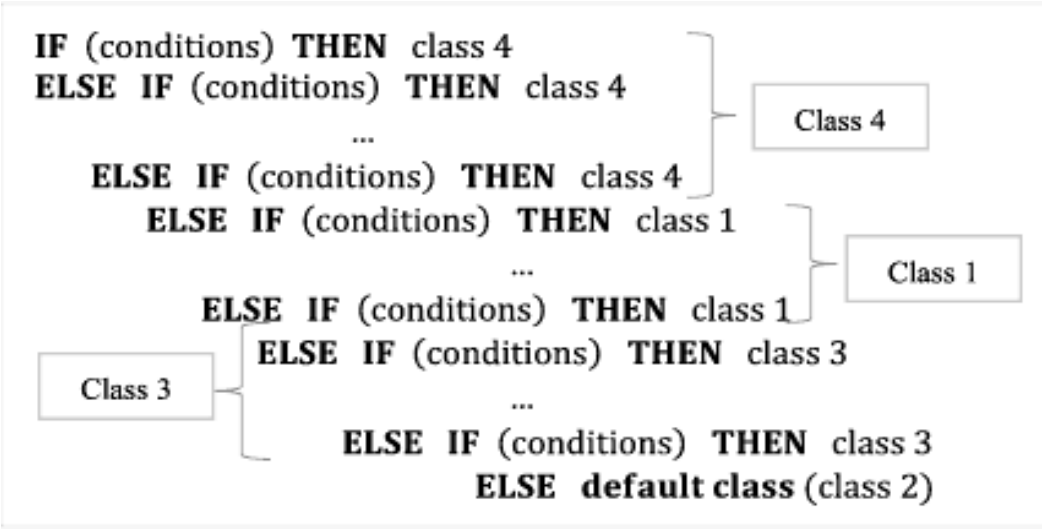
Το Simple Cart είναι μια τροποποίηση του C4.5. Στον αλγόριθμο αυτό, η απόφαση που εκπροσωπείται στο δέντρο μπορεί να περιλαμβάνει συνδυασμούς των χαρακτηριστικών, έτσι ώστε οι παραδοχές να μην περιορίζονται πλέον σε υπερ-ορθογώνια διαμερίσματα. Ενώ στον C4.5 χρησιμοποιείται η μετρική αναλογία κέρδους στις δοκιμές εκπαίδευσης, στον Simple Cart χρησιμοποιείται ο δείκτης ποικιλότητας Gini. Μετά το σχηματισμό του δέντρο για ταξινόμηση, εφαρμόζεται η στρατηγική κλαδέματος ελάχιστο κόστος-πολυπλοκότητας. Στο κλάδεμα εφαρμόζεται ο διαχωρισμός με την ελάχιστη συνεισφορά στη συνολική απόδοση για τα δεδομένα της εκπαίδευσης.

6.1.6. Rules

Οι κανόνες είναι μια δημοφιλής εναλλακτική λύση των δέντρων απόφασης. Η προϋπόθεση ενός κανόνα είναι μια σειρά δοκιμών όπως ακριβώς και οι δοκιμές στους κόμβους των δέντρων αποφάσεων, το αποτέλεσμα δίνει την τάξη που ισχύει σε περιπτώσεις που καλύπτονται από αυτόν τον κανόνα

Ένας κανόνας δημιουργείται για κάθε φύλλο. Στην αρχή δημιουργίας του κανόνα περιλαμβάνει μια συνθήκη για κάθε κόμβο στο μονοπάτι από τη ρίζα προς το φύλλο, και το επακόλουθο του κανόνα είναι η τάξη που έχει εκχωρηθεί από το φύλλο. Αυτή η διαδικασία παράγει κανόνες που είναι σαφείς, δεδομένου ότι η σειρά με την οποία εκτελούνται είναι άνευ σημασίας.

Το εκτιμώμενο ποσοστό σφάλματος παρέχει ακριβώς το απαραίτητο μηχανισμό για το κλάδεμα των κανόνων. Λαμβάνοντας υπόψη έναν συγκεκριμένο κανόνα, ανιχνεύονται ποια από τα παραδείγματα εκπαίδευσης καλύπτονται τώρα από τον κανόνα και καταργείται με την διαγραφή του, υπολογίζεται μια απαισιόδοξη εκτίμηση του ποσοστού σφάλματος του νέου κανόνα και συγκρίνεται με την απαισιόδοξη εκτίμηση του αρχικού κανόνα. Εάν ο νέος κανόνας είναι καλύτερος, διαγράφετε ο αρχικός κανόνας και συνεχίζει, ψάχνοντας για άλλες συνθήκες που πρέπει να διαγραφούν. Ολοκληρώνονται οι κανόνες όταν δεν υπάρχουν υπόλοιποι όροι που θα το βελτιώσουν εάν αφαιρεθούν. Αφού όλοι οι κανόνες έχουν κλαδευτεί με αυτόν τον τρόπο, τότε ελέγχονται αν υπάρχουν διπλότυπα για να τα αφαιρεθούν από το σύνολο κανόνων (Witten et al., n.d.).



Σχήμα 6-8 Λίστα κανόνων απόφασης σε ένα σύνολο δεδομένων με 4 τάξεις.
Πηγή : https://www.researchgate.net/publication/293016132_RipMC_RIPPER_for_Multiclass_Classification

6.1.6.1. Αλγόριθμος MODLEM

Ο αλγόριθμος MODLEM δημιουργεί ένα ελάχιστο σύνολο κανόνων. Αυτοί οι κανόνες μπορούν να υιοθετηθούν στην μηχανική μάθηση ως ταξινομητές. Είναι ένας αλγόριθμος διαδοχικής κάλυψης, ο οποίος δεν απαιτείται προκαταρκτική διακριτοποίηση αριθμητικών χαρακτηριστικών και χειρίζεται αυτά τα χαρακτηριστικά κατά την επαγωγή του κανόνα, όταν δημιουργούνται οι στοιχειώδεις προϋποθέσεις ενός κανόνα. Κατά συνέπεια, οι συνθήκες αριθμητικού χαρακτηριστικού είναι ακριβέστερες και περιγράφουν στενά την τάξη (Grzymala-Busse & Stefanowski, n.d.).

Το μοντέλο MODLEM χρησιμοποιεί επίσης τη θεωρία rough set για των χειρισμό των ασυνεπή παραδειγμάτων και υπολογίζει μια ενιαία τοπική κάλυψη για κάθε προσέγγιση της έννοιας.

Οι κανόνες που δημιουργούνται στον αλγόριθμο MODLEM είναι πιο εύκολο ερμηνεύσιμοι ακόμα και από τα δέντρα απόφασης.

6.1.7. Meta algorithms

Οι μετα-αλγόριθμοι (meta algorithms) παίρνουν τους ταξινομητές ως αδύναμους μαθητές και τους μετατρέπουν σε πιο ισχυρούς μαθητές. Λειτουργούν τόσο για ταξινόμηση όσο και για παλινδρόμηση, ανάλογα με τον βασικό μαθητή.

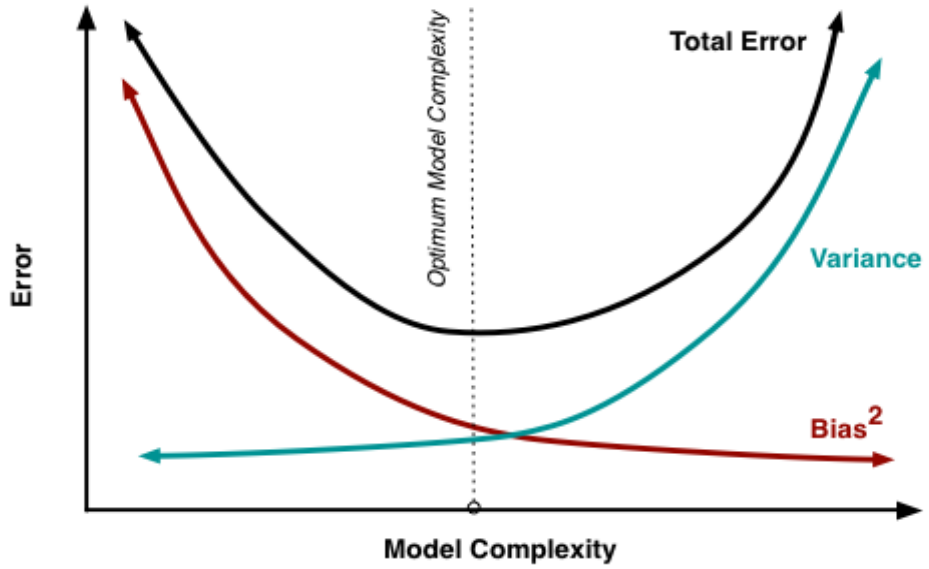
6.1.7.1. Συνδυαστικοί Μέθοδοι (Ensemble methods)

Συνήθως τα μοντέλα πάσχουν από τη μάστιγα του συστηματικού σφάλματος και του σφάλματος απόκλισης. Το συστηματικό σφάλμα (Bias) σημαίνει, «πόσο κατά μέσο όρο είναι οι προβλεπόμενες τιμές διαφορετικές από τις πραγματικές τιμές». Το σφάλμα απόκλισης (Variance) σημαίνει, «πόσο διαφορετικές είναι οι προβλέψεις του μοντέλου στο ίδιο σημείο, αν τα διάφορα δείγματα λαμβάνονται από τον ίδιο πληθυσμό».

Η αύξηση τις πολυπλοκότητας του μοντέλου, θα μειώσει το σφάλμα πρόβλεψης λόγω του χαμηλότερου συστηματικού σφάλματος στο μοντέλο. Όμως θα συνεχίσει να κάνει το μοντέλο πιο σύνθετο, και θα καταλήξει σε υπερεκπαίδευση (overffiting) του μοντέλου και το μοντέλο θα αρχίσει να υποφέρει από υψηλό σφάλμα απόκλισης.

Ένα καλό μοντέλο θα πρέπει να διαχειριστεί και να διατηρήσει μια ισορροπία μεταξύ αυτών των δύο τύπων σφαλμάτων, η λύση είναι η ensemble μάθηση.

Οι συνδυαστικοί μέθοδοι (Ensemble Methods) περιλαμβάνουν ομάδες από προβλεπτικά μοντέλα για να επιτευχθεί μια καλύτερη ακρίβεια και σταθερότητα στο μοντέλο. Οι συνδυαστικοί μέθοδοι είναι γνωστοί για την μετάδοση μέγιστης ώθησης στα μοντέλα. Για την κατασκευή τους, οι αλγόριθμοι μάθησης εκτελούνται ανεξάρτητα και μόνο οι προβλέψεις τους συνδυάζονται παράλληλα ή σειριακά.



Σχήμα 6-9 Συστηματικό σφάλμα (Bias) και σφάλμα απόκλισης (variance)
Πηγή :https://www.researchgate.net/post/How_does_model_complexity_impact_the_bias-variance_tradeoff

Όλοι οι ταξινομητές Ensemble στην πραγματικότητα είναι μετα-ταξινομητές που δέχονται οποιονδήποτε βασικό ταξινομητή ως παράμετρο.

Οι συνδυαστικοί μέθοδοι συνδυάζουν διάφορους ασθενείς επιβλεπόμενους αλγόριθμους μάθησης. Ένας συνδυασμός των πολύ διαφορετικών μοντέλων παράγουν συνήθως καλύτερα αποτελέσματα. Με το συνδυασμό των διαφόρων μεθόδων χειριζόμαστε σε ορισμένα μοντέλα το συστηματικό σφάλμα, το σφάλμα απόκλισης καθώς και τη μείωση υπερεκπαίδευσης (overfitting) (Rokach, 2010).

Υπάρχουν δύο σημαντικά πλεονεκτήματα των μοντέλων Ensemble:

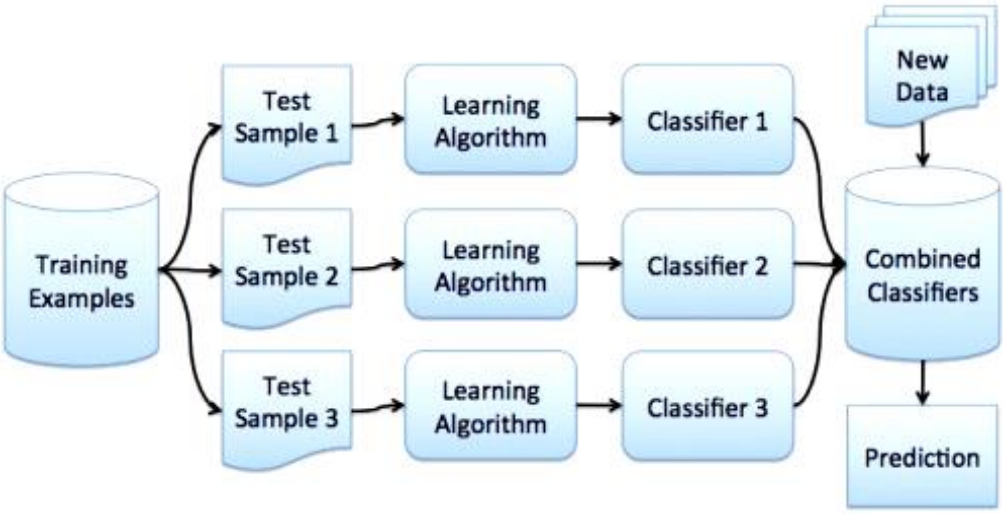
- Η καλύτερη πρόβλεψη και
- Η δημιουργία πιο σταθερού μοντέλου

Μερικοί από τους πιο συνηθισμένους συνδυαστικούς μεθόδους ensemble είναι: ο Boosting, ο Bagging και ο Stacking.

Και οι τρεις λεγόμενοι ‘μετα-αλγόριθμοι’ (meta-algorithms): συνδυάζουν διάφορες τεχνικές μηχανικής μάθησης σε ένα μοντέλο πρόβλεψης προκειμένου να μειωθεί το variance σφάλμα απόκλισης (bagging), και το bias συστηματικό σφάλμα (boosting) ή να βελτιωθεί η δύναμη πρόβλεψης (stacking ψευδώνυμο ensemble).

6.1.7.1.1. Μέθοδος Boosting

Ο όρος 'Boosting' αναφέρεται σε μια οικογένεια αλγορίθμων που μετατρέπει ασθενείς μαθητές σε ισχυρούς μαθητές. Η μέθοδος Boosting είναι μια επαναληπτική τεχνική που προσαρμόζουν το βάρος μιας παρατήρησης με βάση την τελευταία κατάταξη. Εάν μια παρατήρηση που ταξινομείται λανθασμένα, προσπαθεί να αυξήσει το βάρος αυτής της παρατήρησης και αντιστρόφως. Ξεκινά με ένα βασικό ταξινομητή με το σύνολο δεδομένων εκπαίδευσης και απόδοση ίσου βάρους σε κάθε παρατήρηση. Κατόπιν ένας δεύτερος ταξινομητής δημιουργείται πίσω από αυτόν για να επικεντρωθεί στις περιπτώσεις δεδομένων εκπαίδευσης που στον πρώτο ταξινομητή πήραμε λάθος, τότε δίνει μεγαλύτερο βάρος στην λάθος ταξινομημένη παρατήρηση. Μια επαναληπτική διαδικασία συνεχίζεται προσθέτοντας ταξινομητές έως ότου επιτευχθεί ένα όριο στον αριθμό των μοντέλων ή την ακρίβεια. Συγκεκριμένα είναι μια προσέγγιση δύο σταδίων, όπου στο πρώτο στάδιο χρησιμοποιεί υποσύνολα των αρχικών δεδομένων για να παράγει μια σειρά από μέτρια εκτέλεση μοντέλων και στη συνέχεια 'ενισχύει' την απόδοσή τους με το συνδυασμό των ταξινομητών χρησιμοποιώντας μια συγκεκριμένη συνάρτηση κόστους (πλειοψηφία). Η δημιουργία υποσυνόλου δεν είναι τυχαία και εξαρτάται από την απόδοση των προηγούμενων μοντέλων, κάθε νέα υποσύνολα περιέχουν τα στοιχεία που είχαν ταξινομηθεί εσφαλμένα από τα προηγούμενα μοντέλα. Αυτή η τεχνική είναι πολύ ευαίσθητη στο θόρυβο και είναι αποτελεσματική μόνο με τη χρήση αδύναμων ταξινομητών (Witten et al., n.d.). Υπάρχουν αρκετές παραλλαγές των Boosting μεθόδων όπως ο AdaBoost, ο MultiBoost, ο LogitBoost και ο καθένας έχει το δικό του κανόνα ενημέρωσης στα βάρη του, προκειμένου να αποφευχθούν κάποια συγκεκριμένα προβλήματα (θόρυβος, ανισορροπία τάξης κ.α.).



Σχήμα 6-10 Επαναληπτική διαδικασία πρόσθεσης ταξινομητών με την μέθοδο Boosting
Πηγή : http://www.kdnuggets.com/2016/11/data-science-basics-intro-ensemble-learners.html

6.1.7.1.2. Αλγόριθμος AdaBoostM1

Ο AdaBoost είναι ένας emsemble αλγόριθμος μηχανικής μάθησης για προβλήματα ταξινόμησης. Είναι μέρος μιας ομάδας emsemble μεθόδων που ονομάζεται boosting, προσθέτει νέα μοντέλα μηχανικής μάθησης σε μια σειρά, όπου τα επόμενα μοντέλα προσπαθούν να διορθώσουν τα σφάλματα πρόβλεψης που γίνονται από προηγούμενα μοντέλα. Ο AdaBoost ήταν η πρώτη επιτυχής εφαρμογή αυτού του τύπου μοντέλου ο οποίος προτάθηκε για πρώτη φορά από τους Freund & Schapire (Freund & Schapire, 1996).

Ο Adaboost σχεδιάστηκε για να χρησιμοποιεί τα μοντέλα δέντρων απόφασης, το καθένα με ένα μοναδικό σημείο απόφασης. Το πρώτο μοντέλο κατασκευάζεται ως κανονικό. Κάθε δείγμα στο σύνολο δεδομένων εκπαίδευσης σταθμίζεται και ανανεώνει τα βάρη με βάση την συνολική ακρίβεια του μοντέλου αν ένα παράδειγμα ταξινομήθηκε σωστά ή όχι. Τα μετέπειτα μοντέλα έχουν εκπαιδευτεί και προστίθενται μέχρι να επιτευχθεί μια ελάχιστη ακρίβεια ή εάν είναι δυνατόν να υπάρχουν περαιτέρω βελτιώσεις. Κάθε μοντέλο είναι σταθμισμένο με βάση την εμπειρία του και τα βάρη που χρησιμοποιούνται για να συνδυάζονται οι προβλέψεις από όλα τα μοντέλα με νέα δεδομένα (Bauer, Kohavi, Chan, Stolfo, & Wolpert, 1999).

Κατά τη διάρκεια κάθε βήματος μάθησης γίνεται:

- Αύξηση των βαρών των δειγμάτων που δεν έχουν εκπαιδευτή σωστά από τον αδύναμο μαθητή.
- Μείωση των βαρών των δειγμάτων που έχουν εκπαιδευτή σωστά από τον αδύναμο μαθητή.

Η τελική ταξινόμηση βασίζεται σε σταθμισμένη ψήφο αδύναμων ταξινομητών που παράγονται σε επαναλήψεις.

6.1.7.1.3. Αλγόριθμος MultiBoostAB

Ο MultiBoost επεκτείνει την προσέγγιση του AdaBoost με την τεχνική του wagging, η οποία είναι μια παραλλαγή του bagging από όπου τα βάρη της εκπαίδευσης που δημιουργούνται κατά τη διάρκεια της ενίσχυσης boosting χρησιμοποιούνται στην επιλογή των δειγμάτων bootstrap. Ο MultiBoostAB αποτελεί επέκταση της εξαιρετικά επιτυχημένης τεχνικής AdaBoost για τη διαμόρφωση επιτροπών αποφάσεων. Είναι σε θέση να αξιοποιήσει τόσο το συστηματικό σφάλμα του AdaBoost όσο και τη μείωση του σφάλματος απόκλισης με την ανώτερη μείωση της απόκλισης. Χρησιμοποιώντας δέντρο απόφασης ως βασικό αλγόριθμο εκμάθησης, ο Multiboost αποδεικνύει ότι παράγει επιτροπές αποφάσεων με χαμηλότερο σφάλμα από ό, τι είτε το AdaBoost, είτε του Bagging (Kundu, 2014) (Benbouzid, Busa-Fekete, Casagrande, Collin, & Com, 2012).

6.1.7.1.4. Αλγόριθμος DECORATE

Ο DECORATE (Diverse Ensemble Creation by Oppositional Relabeling of Artificial Training Examples) δημιουργεί σύνολα διαφόρων ταξινομητών χρησιμοποιώντας ειδικά κατασκευασμένα τεχνητά παραδείγματα εκπαίδευσης. Η τεχνική αυτή βελτιώνει σταθερά τον βασικό ταξινομητή.

Ξεπερνά τον Boosting σε μικρά σύνολα εκπαίδευσης και τους αντιπάλους του σε μεγαλύτερα. Μία παράμετρος είναι ο αριθμός των τεχνητών παραδειγμάτων για να χρησιμοποιηθεί ως ένα ποσοστό των δεδομένων εκπαίδευσης.

Ένας άλλος είναι ο επιθυμητός αριθμός ταξινομητών στο σύνολο, αν και η εκτέλεση μπορεί να τερματιστεί πρόωρα επειδή ο αριθμός των επαναλήψεων μπορεί επίσης να περιοριστεί. Μεγαλύτερα σύνολα συνήθως παράγουν πιο ακριβή μοντέλα

αλλά έχουν μεγαλύτερο χρόνο εκπαίδευσης και πολυπλοκότητα μοντέλων (Prem Melville, 2003).

6.2. Βιβλιογραφική Ανασκόπηση

Πρόσφατα, το θέμα της παρακολούθησης της κατάστασης των μηχανών και της διάγνωσης βλαβών ως μέρος του συστήματος συντήρησης έγινε παγκόσμιο λόγω των πιθανών πλεονεκτημάτων που θα προκύψουν από το μειωμένο κόστος συντήρησης, τη βελτιωμένη παραγωγικότητα και την αυξημένη διαθεσιμότητα μηχανών.

Αυτή η ενότητα συνοψίζει και επανεξετάζει διάφορα επιστημονικά άρθρα στα πεδίο της μηχανικής μάθησης που εφαρμόζονται στον τομέα της διάγνωσης βλαβών κινητήρων.

Τα τελευταία χρόνια έχουν πραγματοποιηθεί αρκετές εργασίες στη διάγνωση βλαβών κινητήρων, με χρήση αλγορίθμων μηχανικής μάθησης από πολλούς ερευνητές. Παρακάτω παρουσιάζουμε μια σύντομη περίληψη, βάση της βιβλιογραφίας.

Οι παρακάτω ερευνητές παρουσιάζουν τρόπους επίλυσης της διάγνωση βλαβών κινητήρα, χρησιμοποιώντας μηχανή διανυσματικής υποστήριξης (SVM).

Οι (S.-F. Yuan & Chu, 2006), παρουσιάζουν μια νέα ταξική ταξινόμηση SVM με τον αλγόριθμο 'one to other' για την επίλυση των προβλημάτων αναγνώρισης πολλών κατηγοριών. Η αποτελεσματικότητα της μεθόδου επαληθεύεται από την εφαρμογή στη διάγνωση βλαβών σε ρότορα τούρμπο αντλίας.

Οι (S. Yuan & Chu, 2007), χρησιμοποιούν τον αλγόριθμος τεχνητής ανοσοποίησης (artificial immunisation algorithm) για τη βελτιστοποίηση των παραμέτρων στο SVM. Η διάγνωση βλάβης στον ρότορα τούρμπο αντλίας δείχνει ότι το SVM βελτιστοποιημένο από το αλγόριθμο τεχνητής ανοσοποίησης (AIA) μπορεί να δώσει μεγαλύτερη ακρίβεια αναγνώρισης από το κανονικό SVM.

Οι (Widodo & Yang, 2011), παρουσιάζουν μια έρευνα σχετικά με την παρακολούθηση της κατάστασης μηχανής και τη διάγνωση βλαβών χρησιμοποιώντας μηχανή διανυσματικής υποστήριξης (SVM), παρέχοντας υψηλή ακρίβεια στην ταξινόμηση για την παρακολούθηση και τη διάγνωση της κατάστασης μηχανής, έχοντας εξαιρετική απόδοση στην γενίκευση .

Οι (Zhan, Shi, Shwe, & Wang, 2007), προτείνουν τη διάγνωση βλαβών του κύριου κυλινδρικού καλύμματος ναυτικού κινητήρα, που βασίζεται στο σήμα κραδασμών από τον κινητήρα, με τη χρήση της μηχανής διανυσματικής υποστήριξης (SVM), η μέθοδος τους είναι ικανή όχι μόνο να εντοπίζει βλάβες αλλά και να ταξινομεί με ακρίβεια διαφορετικούς τύπους βλαβών.

Η εργασία που ανέπτυξαν οι (Widodo & Yang, 2011), αφορά την ανάπτυξη έξυπνου προγνωστικού συστήματος μηχανών με χρήση μηχανής διανυσματικής υποστήριξης (SVM) για τον υπολογισμό της πιθανότητας επιβίωσης του χρόνου αποτυχίας των εξαρτημάτων της μηχανής. Τα αποτέλεσμα δείχνουν ότι η προτεινόμενη μέθοδος είναι πολλά υποσχόμενη να είναι ένα σύστημα προγνωστικών μηχανών με βάση την πιθανότητα.

Ο (Xu, 2012), προτείνει ένα νέο έξυπνο σύστημα για την on-line ανάλυση των σημάτων δόνησης με χρήση της τοπικής γραμμικής ενσωμάτωσης (Hessian-based locally linear embedding) και support vector machine (SVM) για την ανίχνευση βλαβών του κινητήρα diesel και την πρόληψη της δυσλειτουργίας της μηχανής. Τα αποτελέσματα της διάγνωσης αποδεικνύουν ότι η προτεινόμενη μέθοδος είναι πολύ αποτελεσματική για τη διάγνωση βλαβών των κινητήρων diesel. Οι βλάβες μπορούν να εντοπιστούν on-line με υψηλό ποσοστό 90%.

Οι (Gao & Hou, 2016), εξετάζουν την έρευνα και την ανάπτυξη της διάγνωσης και παρακολούθησης βλαβών με βάση τη μηχανή διανυσματικής υποστήριξης (SVM), η SVM παρουσιάζει το πλεονέκτημα της στην απόδοση γενίκευσης και σε περίπτωση μικρού δείγματος.

Οι παρακάτω ερευνητές ανάπτυξαν διάφορες διαγνωστικές μεθόδους με χρήση των τεχνητών νευρωνικών δικτύων (ANN), για τη διάγνωση βλαβών κινητήρα.

Οι (Sharkey, Chandroth, & Sharkey, 2000), δημιούργησαν ένα σύστημα διάγνωσης βλαβών πολλαπλών δικτύων που έχει σχεδιαστεί για να παρέχει έγκαιρη προειδοποίηση για βλάβες που σχετίζονται με την καύση σε κινητήρα diesel. Χρησιμοποιήθηκαν τεχνητά νευρωνικά δίκτυα για να αναγνωρίσουν τις βλάβες. Τα ατομικά εκπαιδευμένα δίκτυα, μερικά από τα οποία εκπαιδεύτηκαν σε υποτμήματα, συνδυάστηκαν για να σχηματίσουν ένα σύστημα πολλαπλών δικτύων. Το σύστημα πολλαπλών δικτύων φαίνεται να είναι αποτελεσματικό σε σύγκριση με την απόδοση

των δικτύων από τα οποία συνδυάστηκαν.

Οι (Xiros & Kyrtatos, 2000), χρησιμοποίησαν νευρωνικά δίκτυα για την εφαρμογή της συνεχούς πρόβλεψης ζήτησης ροπής έλικας, η οποία μπορεί να αξιοποιηθεί προληπτικά από τον έλεγχο του κινητήρα. Ο νευρωνικός προβλεπτικός παράγοντας επικυρώθηκε μέσω προσομοίωσης, χρησιμοποιώντας δεδομένα που καταγράφηκαν κατά τη διάρκεια προγραμματισμένης λειτουργίας ενός μεγάλου πλοίου μεταφοράς εμπορευματοκιβωτίων καθώς και από δοκιμές σε λεκάνη θαλάμου.

Ο (Ρωμέσης, 2005), ανάπτυξε διαγνωστικές μεθόδους με χρήση πιθανολογικών νευρωνικών δικτύων (PNN) είναι μια δομή τεχνητού νευρωνικού δικτύου που επιτρέπει την ταξινόμηση των υπογραφών με βάση το νόμο του Bayes και των δικτύων πιθανοτήτων που ανήκουν στη περιοχή των Στοχαστικών Έμπειρων Συστημάτων και επιτρέπουν την εκτίμηση της λειτουργικής κατάστασης αεριοστροβίλων και τον εντοπισμό βλαβών συνιστωσών τους με χρήση αεροθερμοδυναμικών μετρήσεων. Σε όλες τις εξεταζόμενες περιπτώσεις, οι διαγνωστικές μέθοδοι εντόπισαν επιτυχώς τις υφιστάμενες βλάβες, πιστοποιώντας την δυνατότητα τους να αντιμετωπίζουν αποτελεσματικά το πρόβλημα της διάγνωσης βλαβών αεριοστροβίλων.

Οι (Moosavian, Ahmadi, Tabatabaeefar, & Khazaee, 2013), εξετάζουν ένα νέο σχήμα για τη διάγνωση σφαλμάτων κύριων ρουλεμάν κινητήρων εσωτερικής καύσης, με βάση την τεχνική φασματικής πυκνότητας (power spectral density) και δύο ταξινομητές, δηλαδή του K-πλησιέστερου γείτονα (KNN) και του τεχνητού νευρωνικού δικτύου (ANN). Από τα αποτελέσματα προκύπτει ότι η απόδοση του ANN είναι καλύτερη από του KNN και μπορεί να διαχωρίσει με αξιόπιστο τρόπο τις διαφορετικές συνθήκες βλάβης στα κύρια ρουλεμάν του κινητήρα εσωτερικής καύσης.

Οι (Ayubi Rad & Yazdanpanah, 2015), προτείνουν την μέθοδο ενός τοπικού επιβλεπόμενου MultiLayer Perceptron ταξινομητή ενσωματωμένο με μοντέλα ανεξάρτητης ανάλυσης συνιστωσών, για την ανίχνευση και διάγνωση βλαβών βιομηχανικών συστημάτων. Τα αποτελέσματα από τα πειράματα έδειξαν την υπεροχή της προτεινόμενης μεθόδου σε σύγκριση με άλλες γνωστές δημοσιευμένες εργασίες.

Οι (Li et al., 2012), προτείνουν σε αυτή την εργασία μια νέα μέθοδος

σύντηξης πληροφοριών για την παρακολούθηση της κατάστασης και στη διάγνωση βλαβών των ναυτικών κινητήρων diesel. Χρησιμοποιήθηκε ένας ταξινομητής Fuzzy Νευρωνικού Δικτύου (FNN) για τον εντοπισμό των βλαβών του κινητήρα. Τα πραγματικά δεδομένα δόνησης που μετρήθηκαν σε ένα πλοίο χρησιμοποιώντας αισθητήρες τεσσάρων καναλιών χρησιμοποιήθηκαν για την αξιολόγηση της προτεινόμενης μεθόδου. Τα πειραματικά διαγνωστικά αποτελέσματα καταδεικνύουν ποσοστό ανίχνευσης βλάβης 90,5%.

Οι (Sahin, Yavuz, Arnavut, & Uluyol, 2007), παρουσιάζουν ένα σύστημα διάγνωσης βλαβών για κινητήρες αεροπλάνων, που χρησιμοποιούν Bayesian δίκτυα και κατανεμημένη βελτιστοποίηση σμήνους σωματιδίων (particle swarm optimization), που μπορεί να εντοπίζει επιτυχώς τα σφάλματα στα δεδομένα δοκιμών.

Οι παρακάτω ερευνητές χρησιμοποίησαν συστήματα ασαφών κανόνων, για τη διάγνωση βλαβών κινητήρα.

Η εργασία των (Twiddle & Jones, 2002), παρουσιάζει μια τεχνική για τη διάγνωση βλαβών σε μια τάξη του κινητήρα diesel, τα οποία επηρεάζουν αρνητικά την απόδοση της καύσης του. Έχει αναπτυχθεί ένα σύστημα ασαφών κανόνων (fuzzy rule) για τη διάγνωση βλαβών με βάση τις εκτιμήσεις φορτίου του κινητήρα. Η διάγνωση γίνεται συνδυάζοντας τα στοιχεία δύο χωριστών εκτιμήσεων του φορτίου κινητήρα. Η δοκιμή του διαγνωστικού συστήματος είχε ως αποτέλεσμα ποσοστό επιτυχίας ταξινόμησης μεγαλύτερο από 90%.

Οι (Vong, Wong, & Wong, 2014), προτείνουν ένα νέο πλαίσιο ταυτόχρονης διάγνωσης βλαβών, για να βελτιωθεί η χρονοβόρα διαδικασία διάγνωσης βλαβών του κινητήρα, ενσωματώνοντας την ασαφή ταξινόμηση πολλαπλών ετικετών. Αυτό το πλαίσιο ονομάζεται ασαφής και πιθανολογική διάγνωση ταυτόχρονης βλάβης (fuzzy and probabilistic simultaneous-fault diagnosis).

Χρησιμοποιήθηκαν επίσης και συνδυαστικοί μέθοδοι (ensemble methods) που συνδυάζουν τους απλούς αλγορίθμους βελτιώνοντας τελικά την απόδοση των διαγνωστικών τους συστημάτων.

Στην μελέτη τους οι (Sharkey et al., 2000), εστιάζουν στην ανίχνευση αρχικών βλαβών σε κινητήρα εσωτερικής καύσης χρησιμοποιώντας έναν ελάχιστο αριθμό αισθητηριακών πληροφοριών. Δύο σύνολα τεχνητών νευρωνικών δικτύων (ANN) εκπαιδεύονται χωριστά, χρησιμοποιώντας χαρακτηριστικά από τα δεδομένα

πίεσης και κραδασμών. Σε μια ξεχωριστή μελέτη, τα δεδομένα πίεσης και κραδασμών, συγχωνεύονται μαζί στο επίπεδο σήματος και στη συνέχεια χρησιμοποιούνται για να εκπαιδεύσουν ένα άλλο σύνολο των ANN που φαίνεται να παρουσιάζουν καλύτερη αξιοπιστία από ότι και τα δύο συστήματα. Στην τελική μελέτη, τα αποτελέσματα των τριών συστημάτων, συνδυάζονται μαζί σε ένα πλειοψηφικό σύστημα ψηφοφορίας, ensemble μέθοδος, αναγνωρίζοντας με επιτυχία 2854 από τις 3000 περιπτώσεις δοκιμών με επίπεδο εμπιστοσύνης 90%.

Οι (Hu Jinhai,Xie Shousheng,Cai Kailong,He Xiuran, 2007), παρουσιάζουν μια νέα προσέγγιση της διάγνωσης βλαβών που ονομάζεται Diverse AdaBoost-SVM, η οποία χρησιμοποιεί την SVM ως αδύναμο μαθητή για τον AdaBoost. Οι πρακτικές εφαρμογές σε ελαττωματικά δείγματα UCI Repository και αεροκινητήρες δείχνουν ότι η προτεινόμενη μέθοδος έχει καλύτερη απόδοση γενίκευσης και είναι πιο κατάλληλη για την ταξινόμηση των ελαττωματικών δειγμάτων που είναι διάσπαρτα σε μεγάλο βαθμό και επίσης πιο κατάλληλα για την ταξινόμηση των μη ισορροπημένων ελαττωματικών δειγμάτων.

Οι (Amozegar & Amozegar, 2015) σε αυτή τους τη διατριβή προτείνουν μια νέα προσέγγιση για την ανίχνευση και απομόνωση βλαβών με τη χρήση ensemble νευρωνικών δικτύων. Οι μέθοδοι ensemble συνδυάζουν διάφορες προβλέψεις μοντέλου για να μειώσουν το σφάλμα μοντελοποίησης και να αυξήσουν την ακρίβεια της πρόβλεψης. Για τους σκοπούς της παρακολούθησης της υγείας των κινητήρων αεριωθούμενων αεροπλάνων, το μοντέλο του τζετ κινητήρα εκπροσωπείται χρησιμοποιώντας τρεις διαφορετικούς αυτοδύναμους ή μεμονωμένους αλγορίθμους μάθησης νευρωνικών δικτύων. Συγκεκριμένα, ένα MultiLayer Perceptron (MLP), ένα a Radial-Basis Function (RBF) νευρωνικό δίκτυο και ένα Support Vector Machine (SVM) εκπαιδεύονται ώστε να μοντελοποιήσουν μεμονωμένα τη δυναμική των κινητήρων τζετ. Στη συνέχεια, χρησιμοποιούνται οι τρεις τεχνικές βασισμένες στη συνδυαστική (ensemble) μέθοδο που αντιπροσωπεύουν τη δυναμική του κινητήρα. Συμπεραίνεται ότι τα συνδυαστικά (ensemble) μοντέλα βελτιώνουν την ακρίβεια μοντελοποίησης σε σύγκριση με αυτόνομες λύσεις.

Οι (Wong, Zhong, Yang, & Vong, 2016) προτείνουν ένα νέο σύστημα για τη διάγνωση μεμονωμένων και ταυτόχρονων βλαβών για τους κινητήρες της αυτοκινητοβιομηχανίας που συνδυάζεται με τη χρήση μιας νέας μεθόδου

πιθανολογικού συνόλου (probabilistic ensemble method), η οποία μπορεί να βελτιώσει τη συνολική διαγνωστική ακρίβεια και να αυξήσει τον αριθμό ανιχνεύσιμων βλαβών.

Παρατηρούμε ότι οι περισσότερες εργασίες διάγνωσης βλαβών στην μηχανική μάθηση έγιναν με χρήση αλγορίθμων ταξινόμησης SVM και νευρωνικών δικτύων.

Επίσης διάφοροι συνδυαστικοί μέθοδοι (ensemble methods) έχουν προταθεί τα τελευταία χρόνια. Αυτοί οι μέθοδοι έχουν αποδειχθεί ότι βελτιώνουν σημαντικά την ακρίβεια ταξινόμησης.

Από την βιβλιογραφική ανασκόπηση δεν εντοπίσαμε προηγούμενες έρευνες στην ανίχνευση και διάγνωση βλαβών δίχρονων αργόστροφων ναυτικών κινητήρων.

Οι ναυτιλιακές εταιρείες κάνουν χρήση λογισμικού προγράμματος διάγνωσης βλαβών από τις αντίστοιχες κατασκευαστικές εταιρείες ναυτικών κινητήρων με υψηλό οικονομικό κόστος.

Η επιλογή και οριοθέτηση της έρευνας, βασίστηκε πρωταρχικά:

1. Στην ανάπτυξη μεθόδων εξόρυξης δεδομένων και στην εξέλιξη των δυνατοτήτων των αλγορίθμων μηχανικής μάθησης στην επίλυση προβλημάτων.
2. Στις διαπιστώσεις μιας σειράς ερευνών στις διαγνώσεις βλαβών κινητήρων, συγκλίνουν στο συμπέρασμα ότι η επίλυση της διάγνωσης βλαβών στους κινητήρες είναι από τα κυριότερα προβλήματα που αντιμετωπίζουν οι μηχανικοί πολύ περισσότερο μάλιστα σε έναν από τους μεγαλύτερους κινητήρες του κόσμου δηλαδή στους δίχρονους αργόστροφους ναυτικούς κινητήρες diesel.

7. ΜΕΘΟΔΟΛΟΓΙΑ ΕΡΕΥΝΑΣ

Τα βήματα της μεθοδολογίας έρευνας στην παρούσα εργασία ήταν τα εξής:

1. Καθορισμός δείγματος – training set

Καταγράφηκαν 1000 εγγραφές – σενάρια βλάβης μηχανής MAN B&W 7S60MC.

Αξιοποιώντας τον Προσομοιωτή της Σχολής Μηχανικών της ΑΕΝ Ασπροπύργου, δημιουργήθηκαν διάφορα σενάρια στα οποία η κύρια μηχανή του πλοίου, τύπου MAN B&W 7S60MC, παρουσίασε συγκεκριμένου τύπου βλάβες ανάλογες με αυτές τις 17 που ορίσαμε ότι θέλουμε να εξετάσουμε. Διαφορετικά σενάρια οδήγησαν σε βλάβες συγκεκριμένου τύπου έτσι ώστε για κάθε μια βλάβη να έχουμε περισσότερα από ένα διαφορετικά σενάρια που την προκάλεσαν.

Παρουσίαση Προσομοιωτή Σχολής στο Παράρτημα.

2. Συλλογή και επεξεργασία των δεδομένων.

Για κάθε βλάβη καταγράψαμε τις τιμές των προς εξέταση παραμέτρων της μηχανής, δημιουργώντας με τον τρόπο αυτό ένα πρότυπο αρχείο βλαβών MAN_7S60MC.arff. Το αρχείο αυτό οργανώθηκε σε βάση δεδομένων προκειμένου στη συνέχεια να αποτελέσει το αρχείο εισαγωγής στο Weka, εκμάθησης για το σύστημά μας.

Συνολικά καταγράφτηκαν 1200 εγγραφές, οι οποίες περιλάμβαναν όλες τις βλάβες και μάλιστα με παρόμοια σχεδόν συχνότητα εμφάνισής τους. Από τις εγγραφές αυτές προέκυψαν 1000 έγκυρες. Μέσω λογισμικού rules based, που αναπτύχθηκε ειδικά για την εγκυρότητα (validation) των δεδομένων, σύμφωνα με τα στοιχεία λειτουργίας της μηχανής όπως αυτά ορίζονται από τον κατασκευαστή της, διαμορφώθηκε και δημιουργήθηκε η πρότυπη βάση δεδομένων.

Το αρχείο δεδομένων της ερευνάς μας MAN_7S60MC.arff αποτελείται από 1000 στιγμιότυπα, που το καθένα έχει 56 χαρακτηριστικά και μία κατηγορία-τάξη.

Α.Α.	Χαρακτηριστικό	Τύπος	Είδος Τιμών
1	power_c1	αριθμητικό	KW
2	power_c2	αριθμητικό	KW
3	power_c3	αριθμητικό	KW
4	power_c4	αριθμητικό	KW
5	power_c5	αριθμητικό	KW
6	power_c6	αριθμητικό	KW
7	power_c7	αριθμητικό	KW
8	rpm_c1	αριθμητικό	Rpm
9	rpm_c2	αριθμητικό	Rpm
10	rpm_c3	αριθμητικό	Rpm
11	rpm_c4	αριθμητικό	Rpm
12	rpm_c5	αριθμητικό	Rpm
13	rpm_c6	αριθμητικό	Rpm
14	rpm_c7	αριθμητικό	Rpm
15	pmi_c1	αριθμητικό	Bar
16	pmi_c2	αριθμητικό	Bar
17	pmi_c3	αριθμητικό	Bar
18	pmi_c4	αριθμητικό	Bar
19	pmi_c5	αριθμητικό	Bar
20	pmi_c6	αριθμητικό	Bar
21	pmi_c7	αριθμητικό	Bar
22	pcomp_c1	αριθμητικό	Bar
23	pcomp_c2	αριθμητικό	Bar
24	pcomp_c3	αριθμητικό	Bar
25	pcomp_c4	αριθμητικό	Bar
26	pcomp_c5	αριθμητικό	Bar
27	pcomp_c6	αριθμητικό	Bar
28	pcomp_c7	αριθμητικό	Bar
29	pmax_pos_c1	αριθμητικό	Bar
30	pmax_pos_c2	αριθμητικό	Bar
31	pmax_pos_c3	αριθμητικό	Bar
32	pmax_pos_c4	αριθμητικό	Bar
33	pmax_pos_c5	αριθμητικό	Bar
34	pmax_pos_c6	αριθμητικό	Bar
35	pmax_pos_c7	αριθμητικό	Bar
36	pmax_c_c1	αριθμητικό	Bar
37	pmax_c_c2	αριθμητικό	Bar

38	pmax_c_c3	αριθμητικό	Bar
39	pmax_c_c4	αριθμητικό	Bar
40	pmax_c_c5	αριθμητικό	Bar
41	pmax_c_c6	αριθμητικό	Bar
42	pmax_c_c7	αριθμητικό	Bar
43	ignition_c1	αριθμητικό	Degree
44	ignition_c2	αριθμητικό	Degree
45	ignition_c3	αριθμητικό	Degree
46	ignition_c4	αριθμητικό	Degree
47	ignition_c5	αριθμητικό	Degree
48	ignition_c6	αριθμητικό	Degree
49	ignition_c7	αριθμητικό	Degree
50	exhaust_gass_temp_c1	αριθμητικό	^{o}C
51	exhaust_gass_temp_c2	αριθμητικό	^{o}C
52	exhaust_gass_temp_c3	αριθμητικό	^{o}C
53	exhaust_gass_temp_c4	αριθμητικό	^{o}C
54	exhaust_gass_temp_c5	αριθμητικό	^{o}C
55	exhaust_gass_temp_c6	αριθμητικό	^{o}C
56	exhaust_gass_temp_c7	αριθμητικό	^{o}C
57	faultys	κατηγορικό	0-16

Πίνακας 7-1 Χαρακτηριστικά αρχείου δεδομένων.

3. Εκτενή μελέτη της υπάρχουσας βιβλιογραφίας για τον καθορισμό των προς μελέτη αλγορίθμων μηχανικής μάθησης που θα εξεταστούν στη συνέχεια – Αλγόριθμοι ταξινόμησης.

Βιβλιογραφική έρευνα κατέδειξε τους ευρέως χρησιμοποιούμενους αλγορίθμους ταξινόμησης οι οποίοι δίνουν ακριβή και αξιόπιστα αποτελέσματα σε πολυκλασικά (multi-class) προβλήματα. Η απόδοση των αντίστοιχων αλγορίθμων μηχανικής μάθησης κρίνεται ιδιαίτερα ικανοποιητική σε αντίστοιχα προβλήματα διαφόρων πεδίων εφαρμογής.

4. Μελέτη των προς εξέταση χαρακτηριστικών των αλγορίθμων μηχανικής μάθησης που επιλέχθηκαν.
5. Παραμετροποίηση και εκτέλεση των αλγορίθμων.
6. Διαμόρφωση σχετικού μοντέλου.

7. Συγκριτική μελέτη των αποτελεσμάτων των αλγορίθμων με σκοπό την ανεύρεση του καταλληλότερου ανά περίπτωση αλγορίθμου και των προς ρύθμιση παραμέτρων αυτών.
8. Βελτιστοποίηση των αλγορίθμων μέσω των συνδυαστικών μεθόδων με σκοπό την αύξηση της ακρίβειάς τους.
9. Διαμόρφωση τελικού μοντέλου αντιστοίχησης βλαβών.
10. Κάθε αλγόριθμος εκπαιδεύτηκε χρησιμοποιώντας 1000 στιγμιότυπα. Η επαλήθευση του μοντέλου για την εκτίμηση της απόδοσης της μεθοδολογίας έγινε εφαρμόζοντας την διαδικασία της διασταυρωμένης επικύρωσης 10 τμημάτων (10-fold cross-validation). Δηλαδή, από το σύνολο των 1000 στιγμιότυπων χρησιμοποιήθηκαν 100 στιγμιότυπα για έλεγχο και τα υπόλοιπα για εκπαίδευση.

7.1. Υλοποίηση Έρευνας

Σε αυτή την ενότητα περιγράφονται οι αλγόριθμοι ταξινόμησης που χρησιμοποιούνται στην έρευνα και υλοποιούνται με το εργαλείο Weka.

Μετά την επιλογή ενός αλγορίθμου μηχανικής μάθησης, είναι αναγκαίο να πραγματοποιηθεί η ρύθμιση των παραμέτρων του κατάλληλα για την εκπαίδευση και κατασκευή μοντέλου με την βέλτιστη απόδοσή του. Το Weka επιλέγει έξυπνα λογικές προεπιλογές για κάθε αλγόριθμο μηχανικής μάθησης που σημαίνει ότι μετά την επιλογή ενός αλγορίθμου χρησιμοποιεί προκαθορισμένες παραμέτρους χωρίς να γνωρίζουμε πολλά γι 'αυτούς. Ένας αλγόριθμος πρέπει να δοκιμαστεί συστηματικά σε μια σειρά από τυπικές διαμορφώσεις παραμέτρων του, οι δοκιμές αυτές πρέπει να επαναληφθούν τόσες φορές όσες είναι και οι συνδυασμοί των παραμέτρων που έχει κάθε αλγόριθμος.

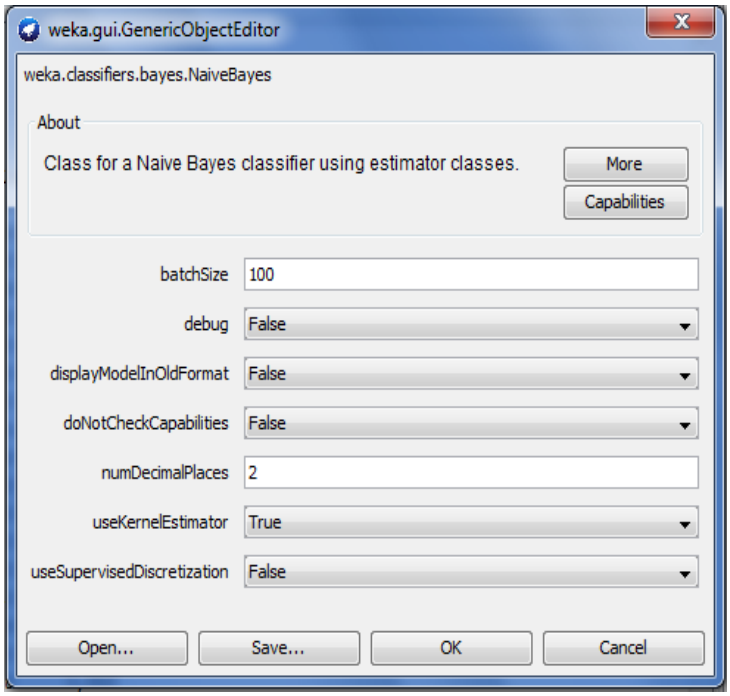
Μετά την κατασκευή του μοντέλου σε κάθε αλγόριθμο ταξινόμησης, το Weka εξάγει τα αποτελέσματα στον πίνακα σύγχυσης (Confusion Matrix) στον οποίο εμφανίζονται αναλυτικά οι προβλέψεις ανά κατηγορία βλάβης.

Στους παρακάτω πίνακες σύγχυσης που εμφανίζονται στα αποτελέσματα κάθε αλγορίθμου, φαίνονται αναλυτικά το πλήθος των σωστών και εσφαλμένων προβλέψεων του αλγορίθμου ανά κατηγορία βλάβης, στην κύρια διαγώνιο του πίνακα εμφανίζεται το πλήθος των σωστών προβλέψεων στην ανίχνευση βλαβών για

κάθε κατηγορία – τάξη βλάβης, καθώς και το πλήθος των λανθασμένων προβλέψεων βλαβών που βρίσκονται εκτός της κύριας διαγωνίου. Το συνολικό άθροισμα στην στήλη κάθε κατηγορίας – τάξης βλάβης (εκτός της κύριας διαγωνίου), είναι οι λάθος θετικές προβλέψεις βλαβών (FP), δηλαδή φαίνεται να υπάρχει ένδειξη για βλάβη στην συγκεκριμένη κατηγορία – τάξη βλάβης ενώ στην πραγματικότητα δεν υπάρχει. Ενώ το συνολικό άθροισμα στην γραμμή κάθε κατηγορίας – τάξης βλάβης (εκτός της κύριας διαγωνίου), είναι οι λάθος αρνητικές προβλέψεις βλαβών (FN), δηλαδή φαίνεται ότι δεν υπάρχει ένδειξη για βλάβη στην συγκεκριμένη κατηγορία – τάξη βλάβης ενώ στην πραγματικότητα υπάρχει.

Τέλος από τους πίνακες σύγχυσης υπολογίζονται οι μετρικές απόδοσης F-Measure, Accuracy (ποσοστό σωστών ταξινομημένων στιγμιότυπων) και ο χρόνος που απαιτείται για την κατασκευή του μοντέλου, που αξιολογούν την απόδοση κάθε μοντέλου και εμφανίζονται συγκεντρωτικά σε πίνακες και διαγράμματα.

Ρύθμισης παραμέτρων για τον Αλγόριθμο Naive Bayes



Σχήμα 7-1 Διαμόρφωση παραμέτρων του αλγορίθμου Naive Bayes

Δεδομένου ότι τα προς εξέταση δεδομένα μας είναι αριθμητικά, οι παράμετροι UseKernelEstimator και useSupervisedDiscretization ορίζονται σε TRUE και FALSE αντίστοιχα. Η παράμετρος UseKernelEstimator χρησιμοποιεί ένα

πυρήνα εκτιμητή για τα αριθμητικά χαρακτηριστικά και όχι μια κανονική κατανομή. Ο αριθμός των δεκαδικών ψηφίων που χρησιμοποιούνται ως έξοδος των αριθμών στο μοντέλο είναι σταθερός και ίσος με δύο και ορίζεται με την παράμετρο numDecimalPlaces. Δεδομένου ότι στην περίπτωσή μας χρησιμοποιούμε λίγες κλάσεις και πολλά χαρακτηριστικά, η παράμετρος displayModelInOldFormat ορίζεται ως FALSE.

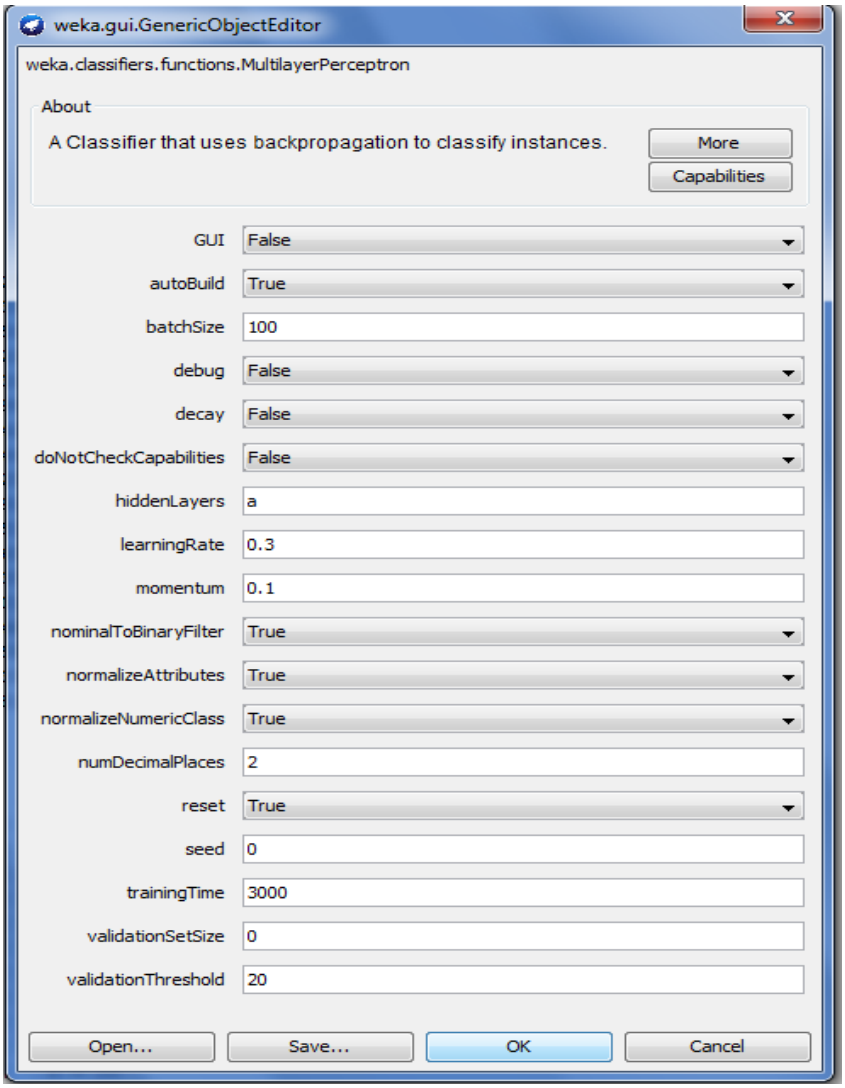
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
223	2	0	0	0	0	0	0	0	0	1	1	0	0	0	4	0	a=0	OK
0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	15	c=2	RPM_low
0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	0	d=3	Pmi_low
0	0	0	1	48	0	0	0	0	1	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
0	0	0	21	3	0	0	0	11	0	0	0	0	0	0	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	1	31	0	0	0	1	2	0	0	0	0	0	0	0	j=9	Pmi_high & Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	51	1	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	47	2	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
14	0	0	0	0	0	0	0	0	0	3	0	0	0	24	8	0	o=14	Exhaust_gass_temp _low
17	0	0	0	0	1	0	0	0	0	1	1	1	0	6	22	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-2 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου Naive Bayes.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
Naive Bayes	0.841	86.1 %	0.05

Πίνακας 7-3 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου Naive Bayes.

Ρύθμισης παραμέτρων για τον Αλγόριθμο Multilayer Perceptron.



Σχήμα 7-2 Διαμόρφωση παραμέτρων του αλγορίθμου Multilayer Perceptron

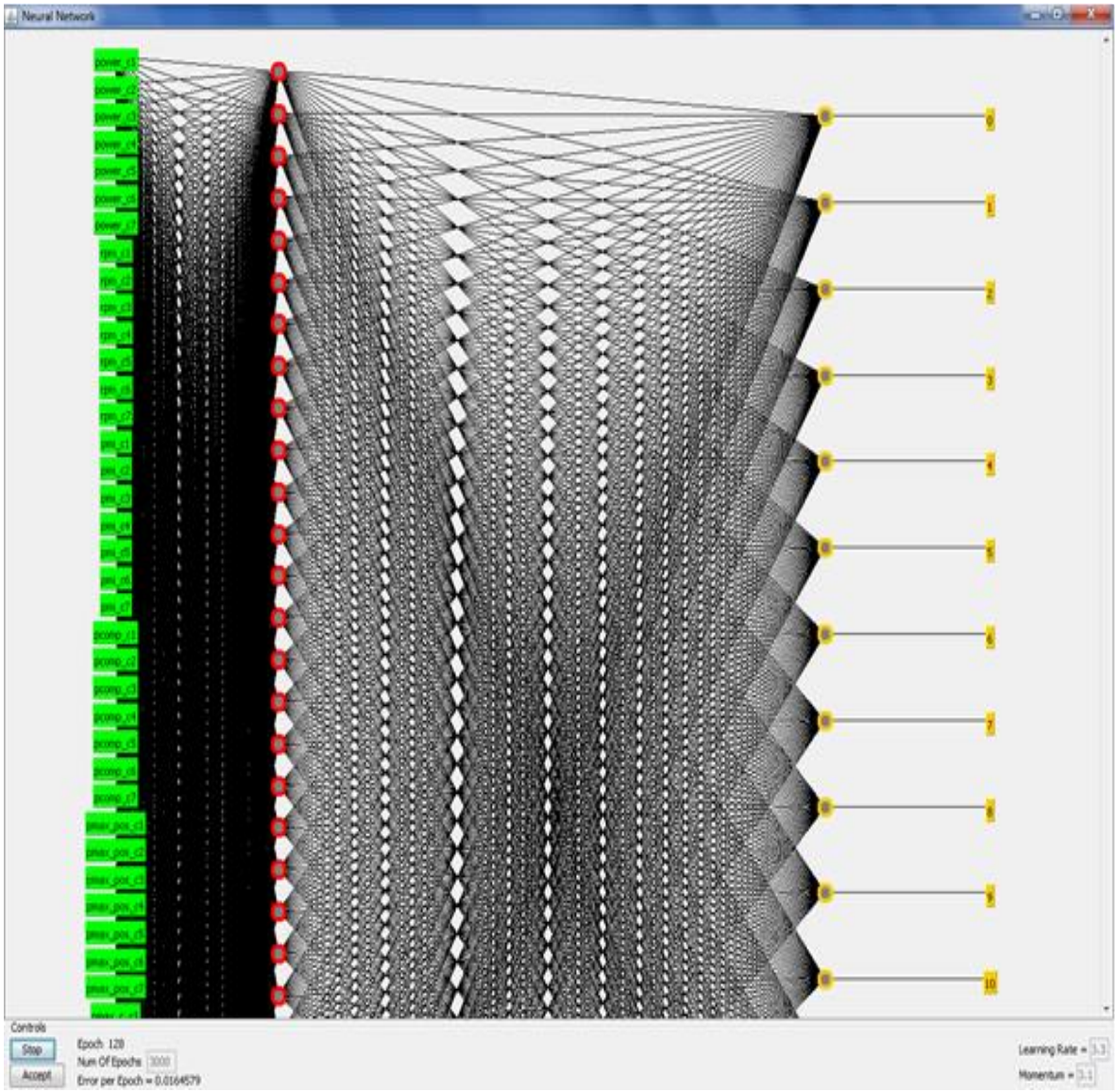
Μπορούν να οριστούν πολλοί παράμετροι από το παράθυρο παραμέτρων η σωστή επιλογή τους είναι αρκετά δύσκολη.

Με την επιλογή του GUI ενεργοποιείται ένα γραφικό περιβάλλον GUI για το σχεδιασμό της δομής του δικτύου. Η παράμετρος autoBuild θα σχεδιάσει αυτόματα το δίκτυο για να το εκπαιδεύσει στο σύνολο των δεδομένων, που προστίθενται και συνδέονται στρώματα κρυφών δικτύων. Η παράμετρος hiddenLayers ορίζει τα κρυφά επίπεδα που υπάρχουν και τον αριθμό των κόμβων που περιέχει ο καθένας. Ο αριθμός των κρυφών δικτύων στην παράμετρο hiddenLayers, είναι ορισμένη αυτόματα στο «a» από προεπιλογή (Amancio et al., 2014).

Το «a» υπολογίζεται από τον παρακάτω τύπο:

a = (αριθμός χαρακτηριστικών + αριθμός κλάσεων)/ 2.

Ορίζοντας το ρυθμό μάθησης learning rate, ρυθμίζεται η διαδικασία της μάθησης καθορίζοντας πόσο να ενημερώνεται το μοντέλο σε κάθε εποχή, οι εποχές είναι οι συνάψεις μεταξύ των νευρώνων. Οι συνήθεις τιμές είναι μικρές, η τιμή που επιλέχθηκε είναι 0,3. Η διαδικασία της μάθησης ρυθμίζεται περαιτέρω με τη παράμετρο momentum να έχει επιλεχθεί στην τιμή 0,2 για να συνεχίσει την ενημέρωση των βαρών και να ορίσει τον ρυθμό ταχύτητας, ακόμα και όταν χρειάζεται να γίνουν αλλαγές, Η παράμετρος reset επαναφέρει αυτόματα το δίκτυο με χαμηλότερο ρυθμό εκμάθησης και ξεκινά ξανά την εκπαίδευση, εάν αποκλίνει από την σωστή απάντηση. Η παράμετρος trainingTime ορίζει τον αριθμό των εποχών εκπαίδευσης δηλαδή τις συνάψεις που γίνονται μεταξύ των νευρώνων. Η εκπαίδευση συνεχίζεται έως ότου η απόδοση στο σύνολο επικύρωσης αρχίσει να επιδεινώνεται συνεχώς ή μέχρι να επιτευχθεί ο καθορισμένος αριθμός εποχών. Η παράμετρος NominalToBinaryFilter προεπεξεργάζεται τα δεδομένα με τη χρήση φίλτρου με σκοπό να βελτιώσει την απόδοση εάν υπάρχουν ονομαστικά χαρακτηριστικά στα δεδομένα. Η παράμετρος NormalizeAttributes εξομαλύνει τα χαρακτηριστικά με σκοπό να βελτιωθεί η εκτέλεση του δικτύου. Με την παράμετρο NormalizeNumericClass εξομαλύνεται μια κλάση εάν είναι αριθμητική με σκοπό πάλι τη βελτίωση της απόδοσης του δικτύου.



Σχήμα 7-3 Αρχιτεκτονική δικτύου MultiLayer Perceptron

Όταν το δίκτυο αρχίζει να εκπαιδεύεται, μια ένδειξη λειτουργίας της εποχής και το σφάλμα εκείνης της εποχής, εμφανίζεται στο κάτω αριστερό μέρος του πίνακα στο σχήμα 9-3. Το σφάλμα βασίζεται σε ένα δίκτυο που αλλάζει καθώς υπολογίζεται η τιμή. Το δίκτυο σταματά όταν φτάσει ο καθορισμένος αριθμός εποχών, οπότε είτε αποδέχεται το αποτέλεσμα είτε αυξάνετε ο επιθυμητός αριθμός των εποχών.

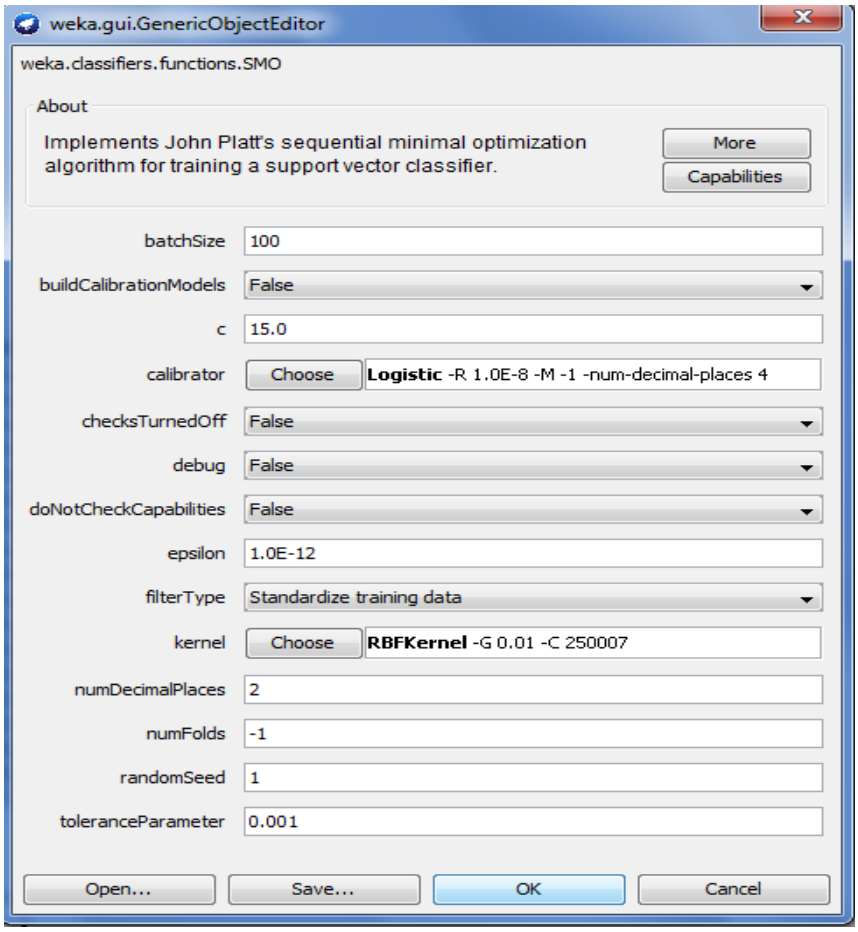
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
157	1	5	3	2	4	6	2	4	1	11	7	6	3	7	10	2	**a=0**	**OK**
1	45	0	0	1	0	0	0	0	0	0	0	1	0	0	0	1	**b=1**	**Power**
18	0	11	3	1	2	1	2	0	1	1	1	1	1	1	1	5	**c=2**	**RPM_low**
5	1	0	27	0	1	1	0	7	0	0	3	1	2	1	1	0	**d=3**	**Pmi_low**
8	0	1	0	23	0	1	0	0	11	1	2	1	0	1	1	0	**e=4**	**Pmi_high**
9	0	0	0	0	36	0	1	1	0	1	0	0	0	0	2	0	**f=5**	**Pcomp_ low**
6	0	0	0	2	0	35	0	0	0	1	2	1	2	2	0	0	**g=6**	**Pcomp_ high**
1	0	0	0	0	0	0	43	0	1	2	1	0	0	0	1	0	**h=7**	**Pmax_pos**
8	0	2	11	0	2	3	0	7	0	0	1	0	0	1	0	0	**i=8**	**Pmi_low & Exhaust_gass_temp _low**
3	0	0	0	17	0	1	1	0	10	2	0	0	1	0	0	0	**j=9**	**Pmi_high & Exhaust_gass_temp _high**
17	0	1	3	1	2	0	1	2	2	13	1	0	1	5	3	0	**k=10**	**Pmax_c_low**
17	0	3	0	0	2	4	0	0	0	1	12	3	0	2	3	1	**l=11**	**Pmax_c_high**
14	0	0	0	2	0	1	0	0	0	1	0	25	0	3	2	1	**m=12**	**Pmax_c_low Ignition_ angle _low**
8	0	0	1	0	1	0	0	0	0	2	0	0	35	2	2	1	**n=13**	**Ignition_ angle _high**
23	0	2	0	0	1	0	1	2	1	3	2	0	0	12	2	0	**o=14**	**Exhaust_gass_temp _low**
24	0	4	1	1	0	1	1	0	1	0	3	3	1	1	8	0	**p=15**	**Exhaust_gass_temp _high**
1	0	2	0	0	0	1	0	0	1	1	0	0	0	0	0	45	**q=16**	**RPM_very_low**

Πίνακας 7-4 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου MultiLayer Perceptron.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
MultiLayerPerceptron	0.530	54.4 %	333.53

Πίνακας 7-5 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου MultiLayer Perceptron

Ρύθμισης παραμέτρων για τον Αλγόριθμο SMO



Σχήμα 7-4 Διαμόρφωση παραμέτρων του αλγορίθμου SMO

Η παράμετρος C, ονομάζεται παράμετρος πολυπλοκότητας και ελέγχει πόσο ευέλικτη μπορεί να είναι η διαδικασία για τη σχεδίαση γραμμής για να διαχωριστούν οι τάξεις. Η τιμή 0 δεν επιτρέπει καμία παραβίαση του περιθωρίου, η τιμή που επιλέχθηκε είναι C=15.

Μια βασική παράμετρος του SMO είναι ο τύπος του πυρήνα kernel που θα χρησιμοποιηθεί. Ο απλούστερος πυρήνας είναι ο Linear Kernel που διαχωρίζει τα δεδομένα με μια ευθεία γραμμή ή υπερεπίπεδο. Η προεπιλογή στο Weka είναι ο Polynomial Kernel που θα διαχωρίζει τις τάξεις χρησιμοποιώντας καμπύλη μια κυρτή γραμμή ή κυματοειδής γραμμή, όσο μεγαλύτερο είναι η πολυώνυμο, τόσο πιο κυματοειδής γραμμή (εκθετική τιμή) (Amancio et al., 2014).

Ο πυρήνας kernel που επιλέχθηκε είναι ο RBFKernel ένας δημοφιλής και ισχυρός πυρήνας που είναι ικανός να μάθει κλειστά πολύγωνα και πολύπλοκα

σχήματα για το διαχωρισμό των τάξεων και έχει αποδειχθεί ότι στις περισσότερες περιπτώσεις δίνει καλύτερα αποτελέσματα από τον γραμμικό πυρήνα.

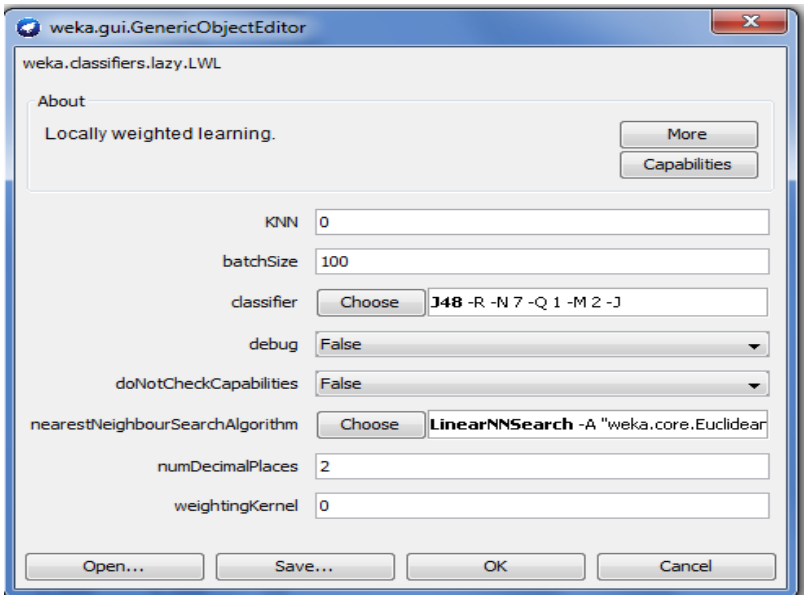
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
201	0	0	0	0	0	0	0	1	0	1	2	3	0	10	13	0	a=0	OK
0	48	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
1	0	38	0	1	0	0	0	0	0	0	2	2	0	1	0	5	c=2	RPM_low
0	0	0	42	0	0	0	0	8	0	0	0	0	0	0	0	0	d=3	Pmi_low
2	0	0	0	28	0	0	0	0	14	4	1	0	0	1	0	0	e=4	Pmi_high
0	0	1	0	0	46	0	0	1	0	1	0	0	0	0	1	0	f=5	Pcomp_ low
0	0	0	0	1	0	46	0	0	0	0	1	1	0	0	2	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
5	0	0	6	0	1	0	0	19	0	2	0	1	0	1	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
2	0	1	0	17	0	0	0	0	14	0	0	1	0	0	0	0	j=9	Pmi_high & Exhaust_gass_temp _high
15	0	0	0	1	0	0	0	2	0	25	1	2	0	2	4	0	k=10	Pmax_c_low
6	0	0	0	1	1	0	0	0	0	2	32	1	0	0	5	0	l=11	Pmax_c_high
7	0	0	0	0	0	0	0	0	0	0	0	40	0	1	1	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	1	0	0	0	0	0	0	0	51	0	0	0	n=13	Ignition_ angle _high
31	0	0	0	0	0	0	0	0	0	1	2	0	0	12	3	0	o=14	Exhaust_gass_temp _low
34	0	0	0	0	0	0	0	0	0	0	1	0	0	0	14	0	p=15	Exhaust_gass_temp _high
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	50	q=16	RPM_very_low

Πίνακας 7-6 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου SMO.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
SMO	0.747	75,5%	1,34

Πίνακας 7-7 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου SMO

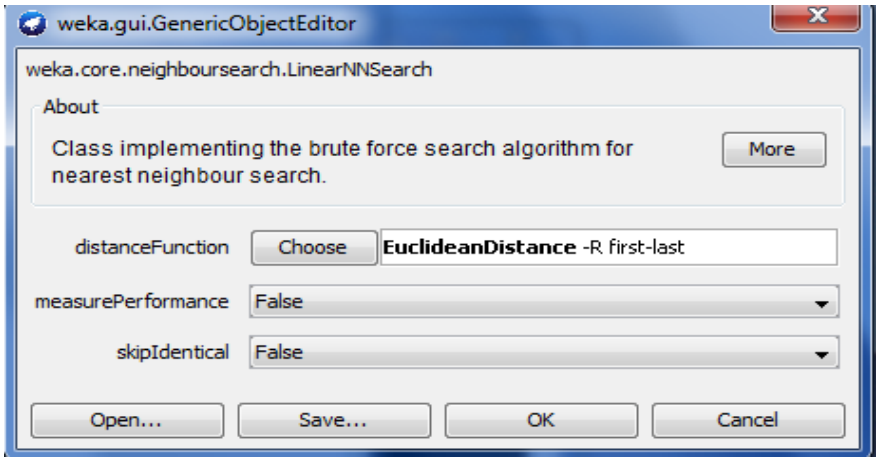
Ρύθμισης παραμέτρων για τον Αλγόριθμο LWL



Σχήμα 7-5 Διαμόρφωση παραμέτρων του αλγορίθμου LWL

Ορίζεται ο αριθμός των γειτόνων KNN = 0 (σημαίνει πως χρησιμοποιούνται όλοι οι γείτονες), ο οποίος καθορίζει το εύρος ζώνης του πυρήνα και το σχήμα του πυρήνα για να χρησιμοποιηθεί στη στάθμιση των linear, inverse ή Gaussian.

Το σχήμα του πυρήνα στάθμισης παράμετρος weightingKernels ρυθμίστηκε στην τιμή 0 (Linear). Για χρήση ταξινομητή επιλέχθηκε το δέντρο απόφασης J48. Μια άλλη σημαντική παράμετρος είναι το μέτρο απόστασης που χρησιμοποιείται. Αυτό έχει ρυθμιστεί στο nearestNeighbourSearchAlgorithm που ελέγχει τον τρόπο με τον οποίο τα δεδομένα εκπαίδευσης αποθηκεύονται και αναζητούνται.



Σχήμα 7-6 Παράθυρο διαλόγου nearestNeighbourSearchAlgorithm

Η προεπιλογή είναι LinearNNSearch. Κάνοντας κλικ στο όνομα αυτού του αλγορίθμου αναζήτησης θα παρέχει ένα άλλο παράθυρο ρυθμίσεων, όπου μπορείτε να επιλέξετε μια παράμετρο distanceFunction ως συνάρτηση απόστασης που θα χρησιμοποιηθεί για την εύρεση γειτόνων. Από προεπιλογή, η EuclideanDistance χρησιμοποιείται για τον υπολογισμό της απόστασης μεταξύ των περιπτώσεων, το οποίο είναι καλό για αριθμητικών δεδομένων με την ίδια κλίμακα.

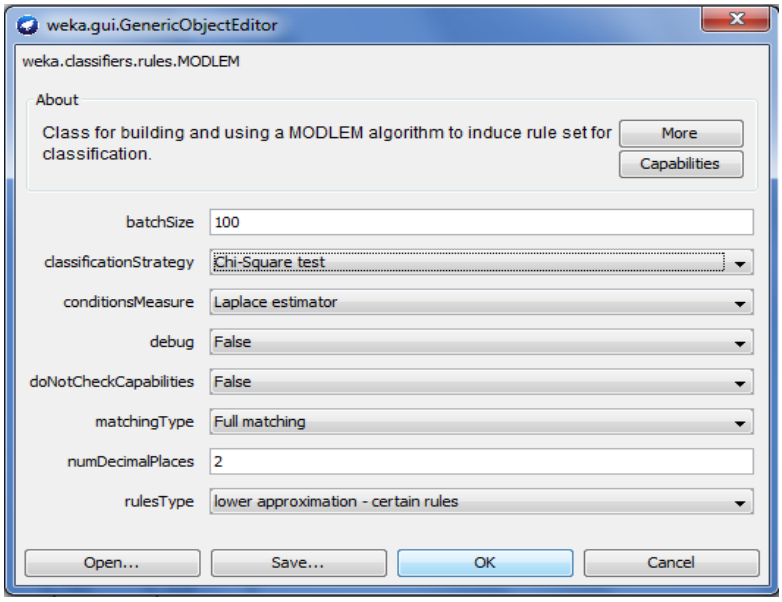
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
226	2	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	a=0	OK
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
2	0	45	0	0	0	0	0	0	0	0	0	0	0	0	0	3	c=2	RPM_low
0	0	0	49	0	0	0	0	1	0	0	0	0	0	0	0	0	d=3	Pmi_low
1	0	0	0	45	0	0	0	0	4	0	0	0	0	0	0	0	e=4	Pmi_high
7	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
1	0	0	11	0	0	0	0	8	0	0	0	0	0	15	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	10	0	0	0	0	9	0	0	0	0	0	16	0	j=9	Pmi_high & Exhaust_gass_temp _high
6	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
7	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
6	0	0	0	0	0	0	0	1	0	0	0	0	0	42	0	0	o=14	Exhaust_gass_temp _low
0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	47	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-8 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου LWL.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
LWL	0.890	90,1%	0

Πίνακας 7-9 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου LWL

Ρύθμισης παραμέτρων για τον Αλγόριθμο MODLEM



Σχήμα 7-7 Διαμόρφωση παραμέτρων του αλγορίθμου MODLEM

Στην παράμετρο ClassificationStrategy επιλέχθηκε ο αλγόριθμος διαχωρισμού Chi-Square.

Ο Chi-Square είναι ένας αλγόριθμος για να βρεθεί η στατιστική σημασία μεταξύ των διαφορών μεταξύ υπο-κόμβων και γονικού κόμβου. Μετριέται με το άθροισμα των τετραγώνων των τυποποιημένων διαφορών μεταξύ των παρατηρούμενων και των αναμενόμενων συχνοτήτων της μεταβλητής στόχου.

Η ανάλυση Chi-squared είναι χρήσιμη για τον προσδιορισμό του επιπέδου στατιστικής σημασίας των κανόνων συσχέτισης. Τα αποτελέσματά διευκολύνουν το κλάδεμα των κανόνων που λαμβάνονται με τη χρήση τυποποιημένων τεχνικών εξόρυξης κανόνων σύνδεσης, επιτρέπουν τον εντοπισμό στατιστικά σημαντικών κανόνων που μπορεί να έχουν αγνοηθεί από τον αλγόριθμο εξόρυξης και παρέχουν μια αναλυτική.

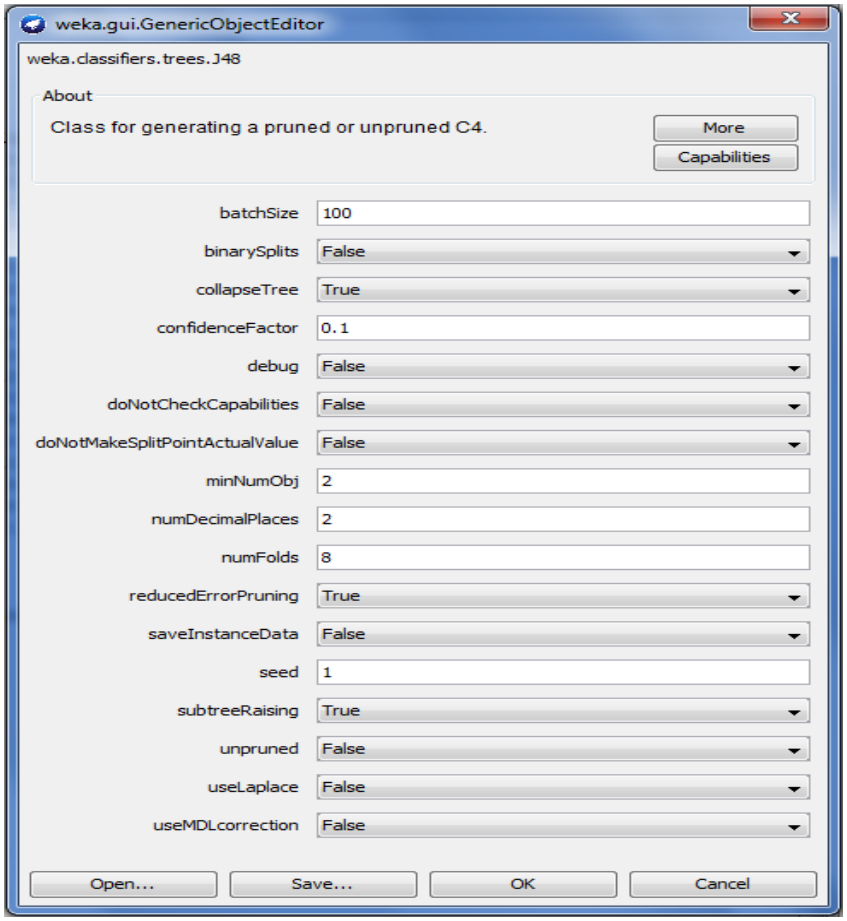
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
215	2	1	0	0	0	0	0	0	10	0	0	0	0	0	3	0	a=0	OK
2	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
2	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c=2	RPM_low
3	0	0	40	0	0	0	0	7	0	0	0	0	0	0	0	0	d=3	Pmi_low
2	0	0	0	37	0	0	0	0	11	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
1	0	0	4	0	0	0	0	18	0	0	0	0	0	12	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	8	0	0	0	0	18	0	0	0	0	0	9	0	j=9	Pmi_high & Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
4	0	0	0	0	0	0	0	5	0	0	0	0	0	40	0	0	o=14	Exhaust_gass_temp _low
5	0	0	0	0	0	0	0	0	5	0	0	0	0	0	39	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-10 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου MODLEM.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
MODLEM	0.905	90,4%	0,56

Πίνακας 7-11 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου MODLEM

Ρύθμισης παραμέτρων για τον Αλγόριθμο J48



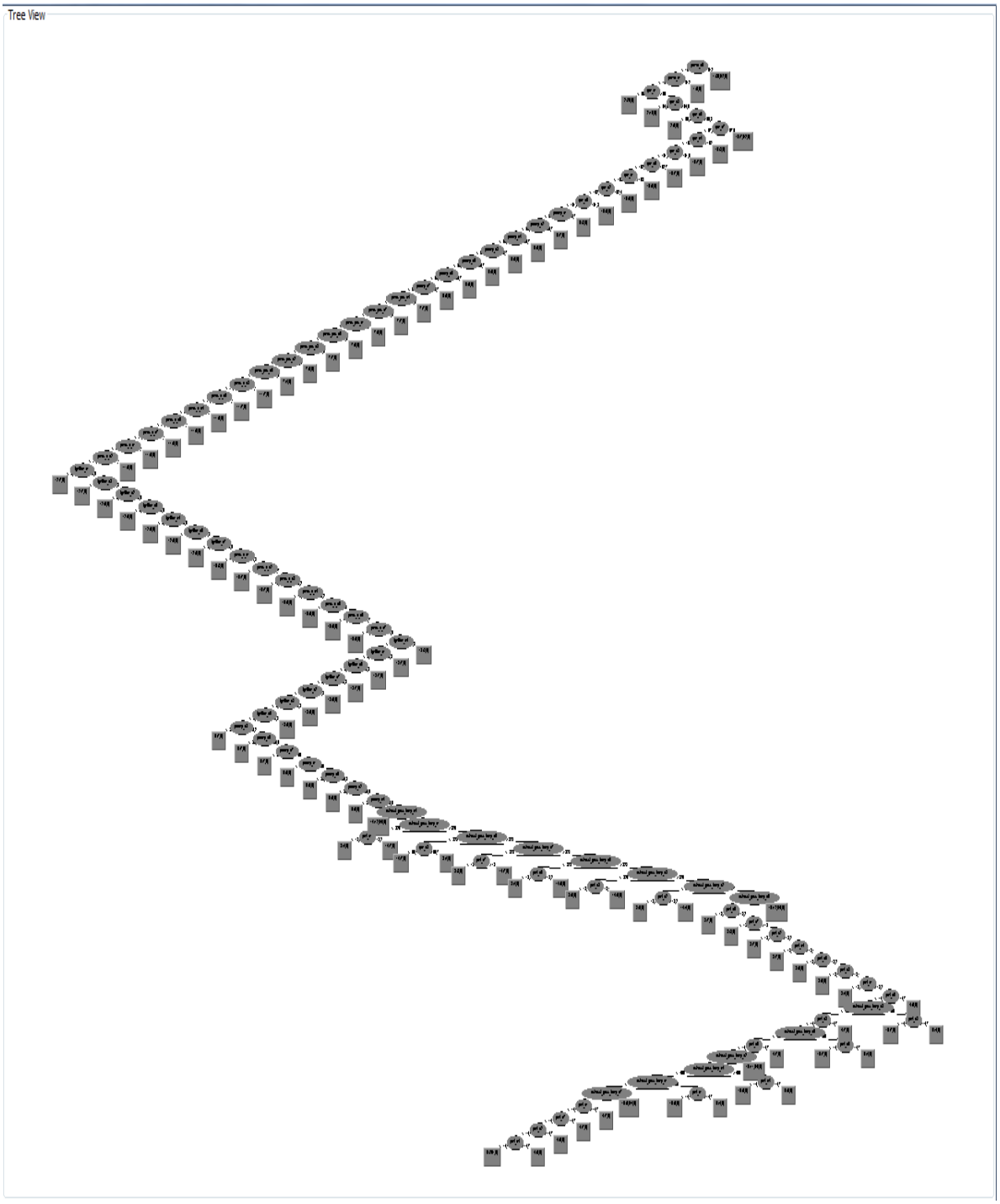
Σχήμα 7-8 Διαμόρφωση παραμέτρων του αλγορίθμου J48

Το όριο εμπιστοσύνης ConfidenceFactor που χρησιμοποιείται για κλάδεμα, η προεπιλεγμένη τιμή είναι 0.25 μικρύνουμε την τιμή αυτή, για να προκαλέσουμε ένα πιο δραστικό κλάδεμα, ρυθμίστηκε στην τιμή 0.1.

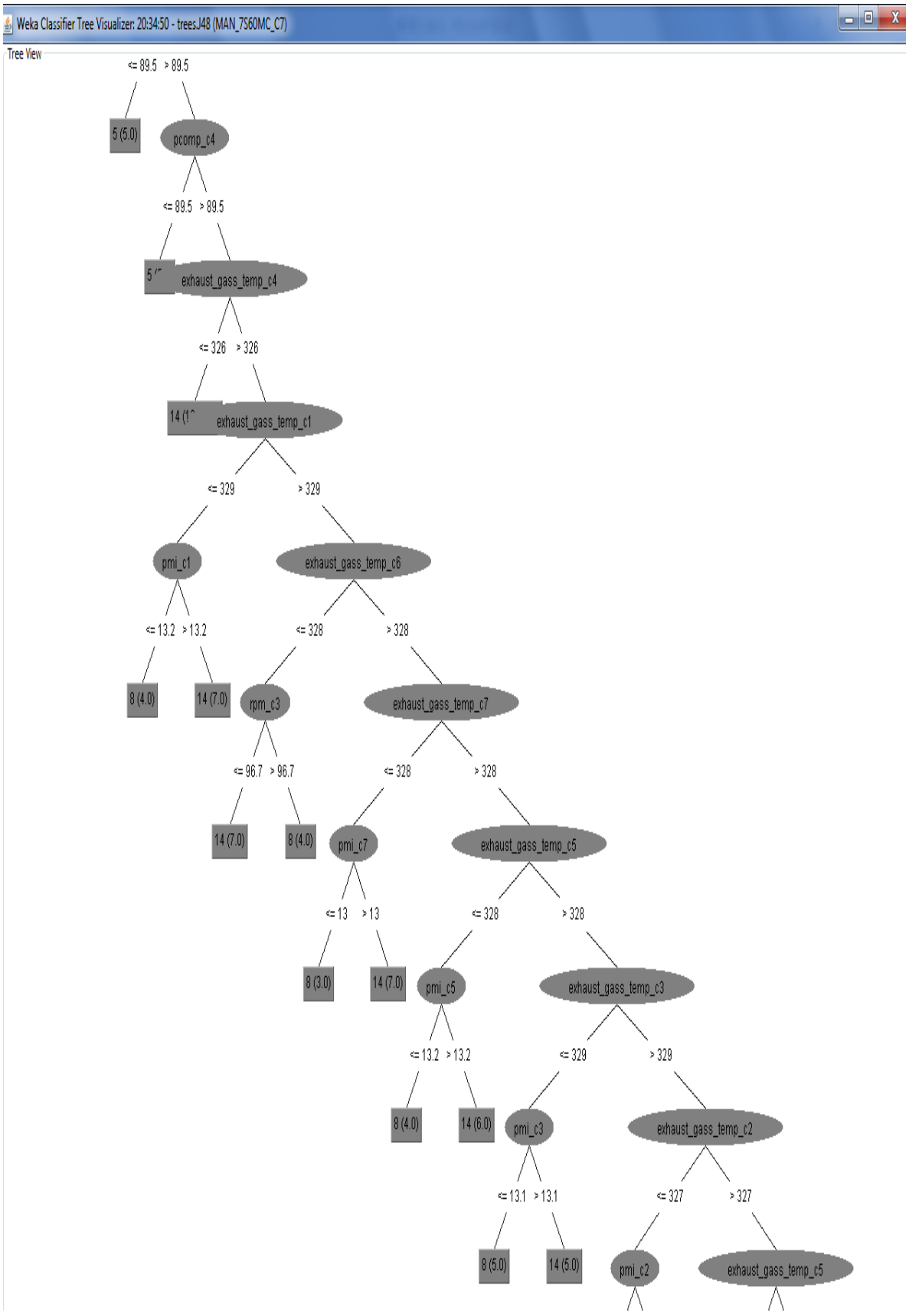
Ο ελάχιστος αριθμός επιτρεπόμενων περιπτώσεων σε ένα φύλλο κατά την κατασκευή του δέντρου από τα δεδομένα εκπαίδευσης, καθορίζετε με την παράμετρο MinNumObj ορίστηκε στην τιμή 2. Επιλέχθηκε η μέθοδος Reduced Error Pruning κλάδεμα μικρής κλίμακας, για την αποφυγή της περιττής δομής που οδηγεί σε υπερεκπαίδευση (overfitting). Η παράμετρος numFolds ο οποίος ορίστηκε στην τιμή 8 καθορίζει το μέγεθος του συνολικού κλαδέματος, δηλαδή τα δεδομένα κατανέμονται ισομερώς σε αυτόν τον αριθμό των τμημάτων που έχει οριστεί και το τελευταίο τμήμα χρησιμοποιείται για κλάδεμα pruning. Τα δεδομένα διαιρούνται εξίσου σε αυτόν τον αριθμό των τμημάτων και το τελευταίο χρησιμοποιείται για κλάδεμα. Με τη τιμή TRUE στην παράμετρο subtree raising καταργείται η αύξηση

των υπο-δέντρων, αποδίδοντας έναν πιο αποδοτικό αλγόριθμο, έτσι αναγκάζει τον αλγόριθμο να χρησιμοποιήσει το μη επεξεργασμένο δέντρο κατά το κλάδεμα αντί του κλαδεμένο δέντρου (post-pruning εκ των υστέρων κλάδεμα).

Η απεικόνιση του δέντρου απόφασης που δημιουργήθηκε για το σύνολο των δεδομένων εκπαίδευσης, είναι:



Σχήμα 7-9 Οπτικοποίηση του δέντρου απόφασης J48



Σχήμα 7-10 Μέρος της απεικόνισης του δέντρου απόφασης J48

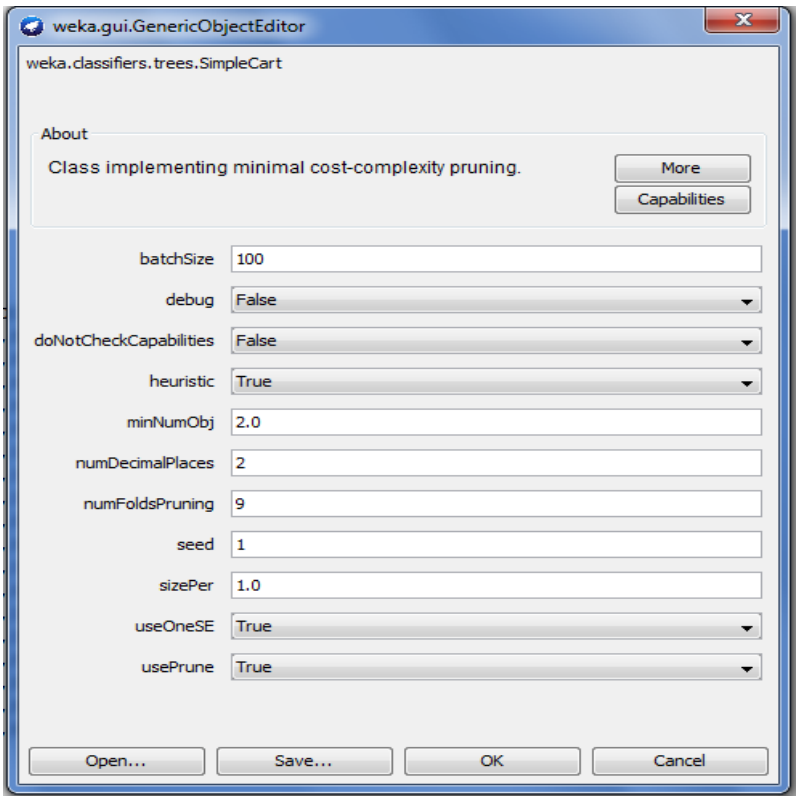
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
227	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	a=0	OK
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
1	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	3	c=2	RPM_low
0	0	0	49	0	0	0	0	1	0	0	0	0	0	0	0	0	d=3	Pmi_low
2	0	0	0	40	0	0	0	0	8	0	0	0	0	0	0	0	e=4	Pmi_high
7	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
1	0	0	9	0	0	0	0	10	0	0	0	0	0	15	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	22	0	0	0	0	5	0	0	0	0	0	8	0	j=9	Pmi_high & Exhaust_gass_temp _high
6	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
7	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
2	0	0	0	0	0	0	0	10	0	0	0	0	0	37	0	0	o=14	Exhaust_gass_temp _low
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	48	0	p=15	Exhaust_gass_temp _high
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	50	q=16	RPM_very_low

Πίνακας 7-12 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου J48.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
J48	0.881	89,1	0,58

Πίνακας 7-13 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου J48

Ρύθμισης παραμέτρων για τον Αλγόριθμο Simple Cart



Σχήμα 7-11 Διαμόρφωση παραμέτρων του αλγορίθμου Simple Cart

Στην παράμετρο NumFoldsPruning = 9 ο αριθμός επιλέχθηκε για τις επαναλήψεις της εσωτερικής μεθόδου διασταυρωμένης επικύρωσης, για τον έλεγχο των δοκιμών των δεδομένων. Η παράμετρος UseOneSe ορίστηκε TRUE και χρησιμοποιείται ο κανόνας «1SE», έτσι ώστε να δημιουργηθεί μια απόφαση κλαδέματος. Η παράμετρος UsePrune ορίστηκε TRUE χρησιμοποιεί το ελάχιστο κόστος-πολυπλοκότητα κλαδέματος, για την αποφυγή της περιττής δομής που οδηγεί σε υπερεκπαίδευση (overfitting).

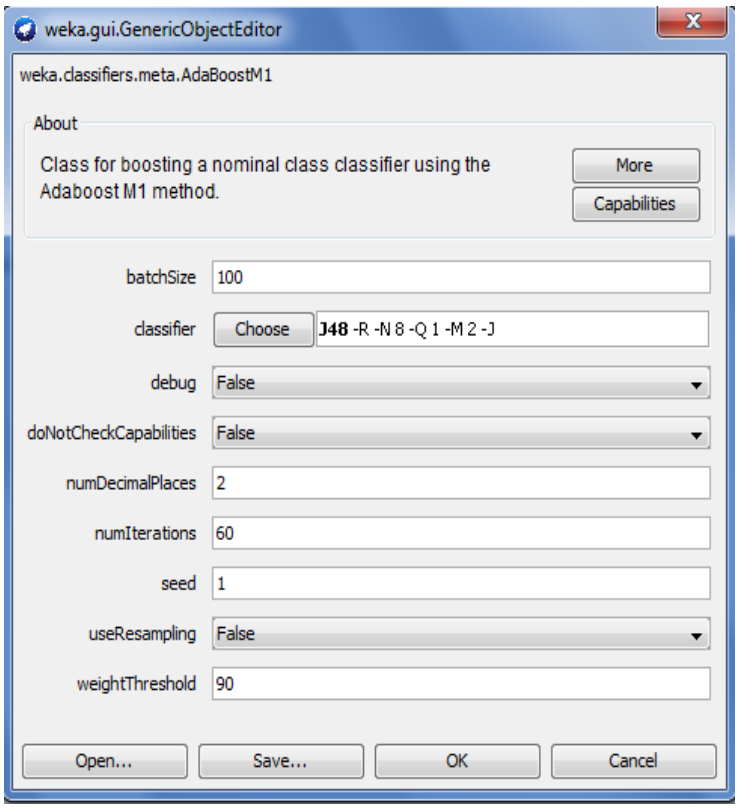
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
227	2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	a=0	OK
2	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
1	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	1	c=2	RPM_low
0	0	0	45	0	0	0	0	5	0	0	0	0	0	0	0	0	d=3	Pmi_low
0	0	0	0	44	0	0	0	0	6	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
2	0	0	6	0	0	0	0	24	0	0	0	0	0	3	0	0	i=8	Pmi_low & Exhaust_gass_temp_ low
0	0	0	0	3	0	0	0	0	30	0	0	0	0	0	2	0	j=9	Pmi_high & Exhaust_gass_temp_ high
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	47	0	0	1	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m=12	Pmax_c_low Ignition_ angle low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
1	0	0	0	0	0	0	0	1	0	0	0	0	0	47	0	0	o=14	Exhaust_gass_temp _low
2	0	0	0	0	0	0	0	0	0	0	0	0	0	2	45	0	p=15	Exhaust_gass_temp _high
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	48	q=16	RPM_very_low

Πίνακας 7-14 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου Simple Cart.

Αλγόριθμος	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
Simple Cart	0.955	95,5%	5,73

Πίνακας 7-15 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου Simple Cart

Ρύθμισης παραμέτρων για τον Αλγόριθμο AdaBoostM1



Σχήμα 7-12 Διαμόρφωση παραμέτρων του αλγορίθμου AdaBoostM1

Ο αδύναμος μαθητής εντός του μοντέλου AdaBoostM1 καθορίζεται από την παράμετρο classifier.

Τα μοντέλα του αλγορίθμου AdaBoostM1 υλοποιήθηκαν με βασικούς ταξινομητές τα δέντρα απόφασης, J48 και Simple Cart, που ως 'αδύναμοι μαθητές' απέδωσαν την πιο αξιοσημείωτη βελτίωση στην απόδοσή τους. Μια βασική παράμετρος εκτός από το αδύναμο μαθητή είναι ο αριθμός των μοντέλων που θα κατασκευαστούν και θα προστεθούν στη σειρά. Αυτό μπορεί να καθοριστεί με την παράμετρο numIterations. Αυξάνουμε την τιμή μέχρι να δούμε την καλύτερη βελτίωση στο μοντέλο όπου παρατηρήθηκε η βέλτιστη τιμή 60.

Η πρώτη επιλογή του πειράματος στο WEKA με το μοντέλο AdaBoostM1 υλοποιήθηκε με βασικό ταξινομητή το δέντρο απόφασης J48 ως αδύναμος μαθητής.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
228	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	a=0	OK
2	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
0	0	47	0	0	0	0	0	0	0	0	0	0	0	0	0	3	c=2	RPM_low
0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	0	d=3	Pmi_low
0	0	0	0	49	0	0	0	0	1	0	0	0	0	0	0	0	e=4	Pmi_high
7	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
1	0	0	6	0	0	0	0	22	0	0	0	0	0	6	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	1	0	0	0	0	34	0	0	0	0	0	0	0	j=9	Pmi_high & Exhaust_gass_temp _high
6	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
7	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
4	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	o=14	Exhaust_gass_temp _low
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-16 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου AdaBoostM1 με βασικό ταξινομητή τον J48.

Αλγόριθμος με Βασικό Ταξινομητή	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
AdaBoostM1 J48	0.950	95,1%	24,41

Πίνακας 7-17 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου AdaBoostM1 με Βασικό Ταξινομητή τον J48

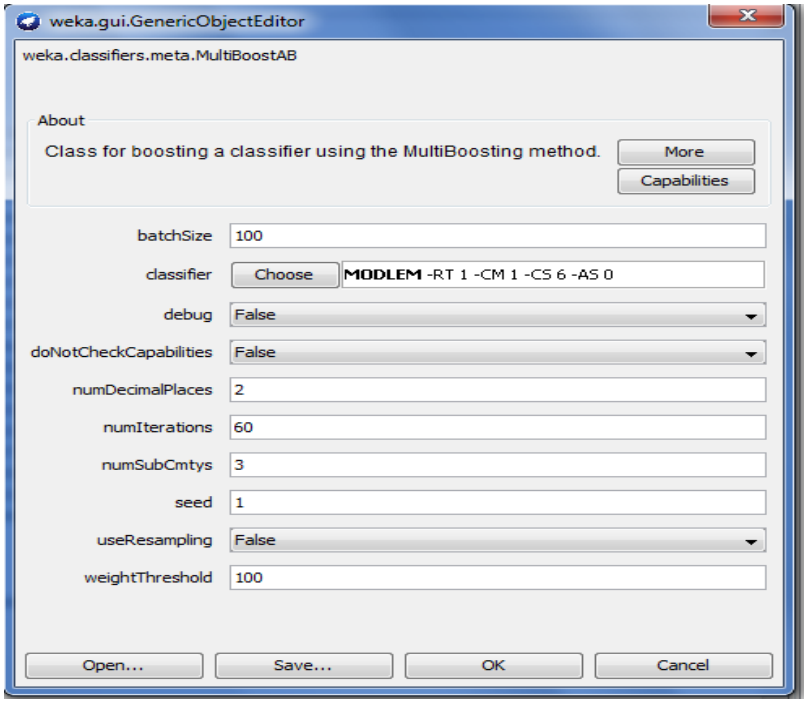
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες - Τάξεις	
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a=0	OK
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
1	0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c=2	RPM_low
1	0	0	48	0	0	0	0	1	0	0	0	0	0	0	0	0	d=3	Pmi_low
2	0	0	0	46	0	0	0	0	2	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
2	0	0	5	0	0	0	0	27	0	0	0	0	0	1	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	6	0	0	0	0	26	0	0	0	0	0	3	0	j=9	Pmi_high & Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
4	0	0	0	0	0	0	0	1	0	0	0	0	0	44	0	0	o=14	Exhaust_gass_temp _low
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-18 Πίνακας 1.1 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου AdaBoostM1 με βασικό ταξινομητή τον Simple Cart.

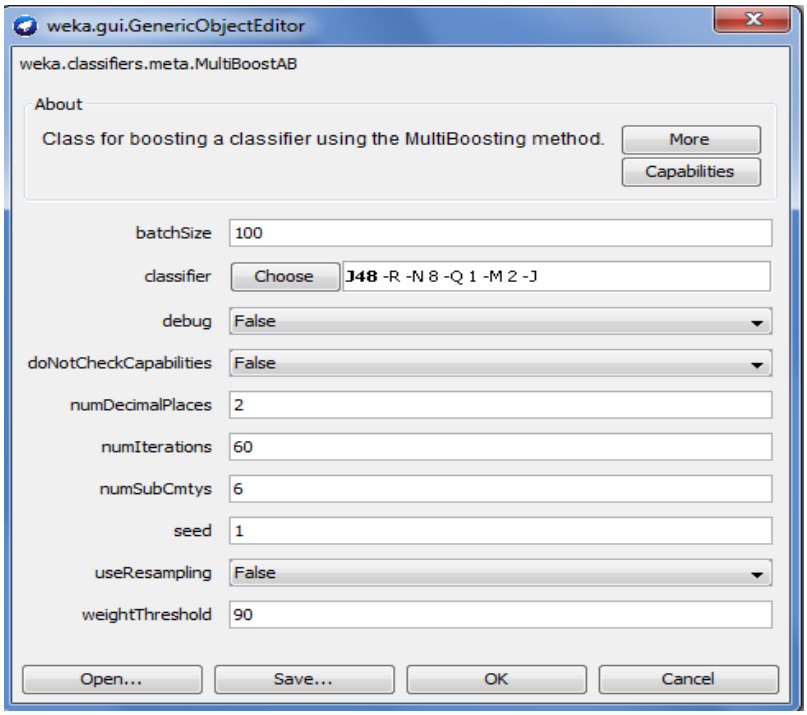
Αλγόριθμος με Βασικό Ταξινομητή	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
AdaBoostM1 **Simple Cart**	0.965	96,6%	238,99

Πίνακας 7-19 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου AdaBoostM1 με Βασικό Ταξινομητή τον Simple Cart

Ρύθμισης παραμέτρων για τον Αλγόριθμο MultiBoostAB



Σχήμα 7-13 Διαμόρφωση παραμέτρων του αλγορίθμου MultiBoostAB με βασικό αλγόριθμο τον MODLEM



Σχήμα 7-14 Διαμόρφωση παραμέτρων του αλγορίθμου MultiBoostAB με βασικό αλγόριθμο τον J48

Τα δέντρα απόφασης, J48, Simple Cart και δημιουργίας κανόνων MODLEM απέδωσαν την πιο αξιοσημείωτη βελτίωση στην απόδοσή τους, εντός του μοντέλου MultiBoost ως αδύναμοι μαθητές που καθορίζεται από την παράμετρο classifier.

Ο αριθμός των επαναλήψεων του μοντέλου που θα πραγματοποιηθούν καθορίστηκε στην παράμετρο NumIterations στην τιμή 60. Η παράμετρος WeightTreshold καθορίζει το κατώφλι βάρους για κλάδεμα η τιμή καθορίστηκε στην τιμή 100 για τους βασικούς αλγορίθμους MODLEM και Simple Cart. Με βασικό αλγόριθμο τον J48, η παράμετρος WeightTreshold μειώθηκε στην τιμή 90 για την επιτάχυνση της διαδικασίας εκμάθησης. Στην παράμετρο NumSubCmtys ορίζετε (κατά προσέγγιση) ο αριθμός των «subcommittees» για τη λήψη αποφάσεων. Στο πείραμα χρησιμοποιήθηκε τη τιμή 3.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες	
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a=0	OK
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c=2	RPM_low
0	0	0	47	0	0	0	0	3	0	0	0	0	0	0	0	0	d=3	Pmi_low
2	0	0	0	43	0	0	0	0	5	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
0	0	0	3	0	0	0	0	17	0	0	0	0	0	15	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	18	0	0	0	0	8	0	0	0	0	0	9	0	j=9	Pmi_high & Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
6	0	0	0	0	0	0	0	1	0	0	0	0	0	42	0	0	o=14	Exhaust_gass_temp _low
3	0	0	0	0	0	0	0	0	7	0	0	0	0	0	39	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-20 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου MultiBoostAB με βασικό ταξινομητή τον MODLEM.

Αλγόριθμος με Βασικό Ταξινομητή	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
MultiBoostAB MODLEM	0.919	92,5%	13,73

Πίνακας 7-21 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου MultiBoostAB με Βασικό Ταξινομητή τον MODLEM

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες	
228	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	a=0	OK
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
1	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	1	c=2	RPM_low
0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	0	d=3	Pmi_low
0	0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	e=4	Pmi_high
7	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
1	0	0	5	0	0	0	0	23	0	0	0	0	0	6	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	0	0	0	0	0	33	0	0	0	0	0	2	0	j=9	Pmi_high & Exhaust_gass_temp _high
6	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
7	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
4	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	0	o=14	Exhaust_gass_temp _low
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-22 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου MultiBoostAB με βασικό ταξινομητή τον J48.

Αλγόριθμος με Βασικό Ταξινομητή	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
MultiBoostAB J48	0.951	95,2%	16,3

Πίνακας 7-23 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου MultiBoostAB με Βασικό Ταξινομητή τον J48

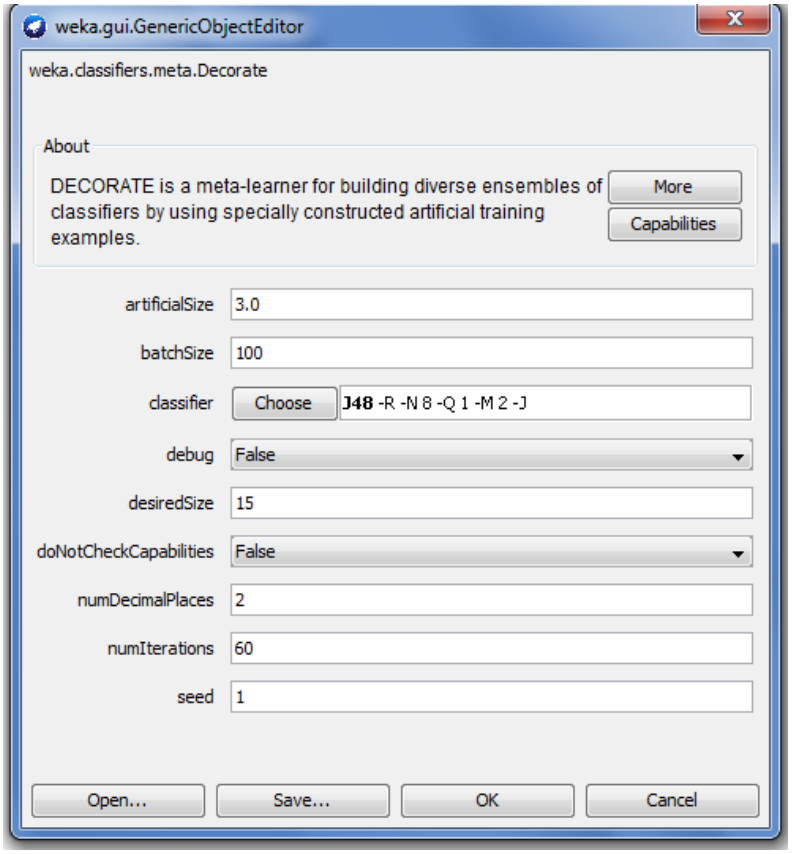
a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες	
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a=0	OK
0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
2	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c=2	RPM_low
1	0	0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	d=3	Pmi_low
2	0	0	0	46	0	0	0	0	2	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
2	0	0	7	0	0	0	0	22	0	0	0	0	0	4	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	4	0	0	0	0	29	0	0	0	0	0	2	0	j=9	Pmi_high & Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
6	0	0	0	0	0	0	0	0	0	0	0	0	0	43	0	0	o=14	Exhaust_gass_temp _low
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	p=15	Exhaust_gass_temp _high
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	q=16	RPM_very_low

Πίνακας 7-24 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου MultiBoostAB με βασικό ταξινομητή τον Simple Cart.

Αλγόριθμος με Βασικό Ταξινομητή	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
MultiBoostAB Simple Cart	0.959	96%	224,86

Πίνακας 7-25 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου MultiBoostAB με Βασικό Ταξινομητή τον Simple Cart

Ρύθμισης παραμέτρων για τον Αλγόριθμο Decorate με βασικό αλγόριθμο J48.



Σχήμα 7-15 Διαμόρφωση παραμέτρων του αλγορίθμου Decorate με βασικό ταξινομητή τον J48

Η επιλογή στην παράμετρο classifier του πειράματος με το μοντέλο Decorate, υλοποιήθηκε με βασικό ταξινομητή, το δέντρο απόφασης J48. Η παράμετρος artificialSize καθορίζει τον αριθμό των τεχνητών παραδειγμάτων που χρησιμοποιούνται κατά τη διάρκεια της εκπαίδευσης. Καθορίζεται ως ποσοστό των δεδομένων εκπαίδευσης. Οι υψηλότερες τιμές μπορούν να αυξήσουν την ποικιλομορφία του ensemble, η τιμή έχει οριστεί στην τιμή 3. Η παράμετρος desiredSize καθορίζει τον επιθυμητό αριθμό ταξινομητών στο σύνολο Decorate εδώ έχει οριστεί στην τιμή 15. Ο Decorate μπορεί να τερματιστεί πριν επιτευχθεί αυτό το μέγεθος (ανάλογα με την τιμή του αριθμού επαναλήψεων numIterations). Τα μεγαλύτερα μεγέθη του ensemble συνήθως οδηγούν σε πιο ακριβή μοντέλα, αλλά αυξάνουν το χρόνο εκπαίδευσης και την πολυπλοκότητα του μοντέλου. Η παράμετρος numIterations, καθορίζει το μέγιστο επιτρεπτό αριθμό επαναλήψεων Decorate που θα εκτελεστούν, εδώ ορίστηκε στην τιμή 60. Αυτή η παράμετρος πρέπει να είναι μεγαλύτερη από το desiredSize. Κάθε επανάληψη δημιουργεί έναν

ταξινομητή, αλλά δεν το προσθέτει αναγκαστικά στην μέθοδο ensemble. Ο Decorate σταματά όταν επιτευχθεί το επιθυμητό συνολικό μέγεθος.

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	Προβλεπόμενες ← Κατηγορίες	
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a=0	OK
2	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b=1	Power
1	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	1	c=2	RPM_low
0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	d=3	Pmi_low
0	0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	e=4	Pmi_high
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f=5	Pcomp_ low
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g=6	Pcomp_ high
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h=7	Pmax_pos
2	0	0	5	0	0	0	0	19	0	0	0	0	0	9	0	0	i=8	Pmi_low & Exhaust_gass_temp _low
0	0	0	0	7	0	0	0	0	23	0	0	0	0	0	5	0	j=9	Pmi_high & Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k=10	Pmax_c_low
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l=11	Pmax_c_high
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m=12	Pmax_c_low Ignition_ angle _low
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n=13	Ignition_ angle _high
3	0	0	0	0	0	0	0	1	0	0	0	0	0	45	0	0	o=14	Exhaust_gass_temp _low
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	p=15	Exhaust_gass_temp _high
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q=16	RPM_very_low

Πίνακας 7-26 Πίνακας σύγχυσης (Confusion Matrix) του αλγορίθμου Decorate με βασικό ταξινομητή τον J48.

Αλγόριθμος με Βασικό Ταξινομητή	*F-Measure*	*Accuracy*	*Χρόνος (seconds)*
Decorate J48	0.950	95,2%	84,94

Πίνακας 7-27 Μετρικές Απόδοσης και ο χρόνος ταξινόμησης του αλγορίθμου Decorate με Βασικό Ταξινομητή τον J48

7.2. Ανάλυση Αποτελεσμάτων

Σ' αυτό το κεφάλαιο έγινε η κατασκευή μοντέλων των αλγορίθμων ταξινόμησης, με το εργαλείο Weka, που αναπτύχθηκαν στα πλαίσια της παρούσας εργασίας.

Σε αυτή την ενότητα παρουσιάζονται και αναλύονται τα αποτελέσματα των μεθόδων αυτών.

Στην παρούσα εργασία χρησιμοποιήθηκαν για τα πειράματα, επτά βασικοί αλγόριθμοι από όλες τις κατηγορίες ταξινόμησης, ο NaiveBayes των μπευζιανών δικτύων, ο Multilayer Perceptron των νευρωνικών δικτύων, ο SMO των μηχανών διανυσματικής υποστήριξης, ο LWL των τεμπέλικων (lazy) αλγορίθμων, ο MODLEM των αλγορίθμων δημιουργίας κανόνων, και οι J48, και Simple Cart των δέντρων απόφασης. Επίσης σε αυτήν την εργασία έγινε προσπάθεια να βελτιωθεί η απόδοση των βασικών αλγορίθμων με τους συνδυαστικούς μεθόδους, AdaBoost, MultiBost και Decorate. Όλοι οι αλγόριθμοι υλοποιήθηκαν με τη χρήση του ελεύθερου εργαλείου εξόρυξης δεδομένων Weka, για την ανάλυση της ακρίβειας και της απόδοσης τους.

Για την αξιολόγηση της αποτελεσματικότητας κάθε αλγορίθμου χρησιμοποιήθηκαν οι μετρικές απόδοσης, F-Measure, Accuracy (ποσοστό ακρίβειας ταξινόμησης στη βάση των σωστά ταξινομημένων περιπτώσεων) και ο χρόνος που απαιτείται για την κατασκευή του μοντέλου. Οι αλγόριθμοι στη συνέχεια συγκρίθηκαν και αξιολογήθηκαν με τις παραπάνω μετρικές απόδοσης. Επίσης σύγκριση γίνεται και μεταξύ των συνδυαστικών μεθόδων και των βασικών αλγορίθμων που χρησιμοποιήθηκαν, στόχος είναι με βάση την σύγκριση αυτή να εξεταστούν κατά πόσο η χρήση των συνδυαστικών μεθόδων βοήθησαν στην βελτίωση της απόδοσης των βασικών αλγορίθμων στην διάγνωση βλαβών του κινητήρα.

Συγκρίνοντας τους βασικούς αλγόριθμους με βάση τη μετρική απόδοσης F-Measure, τα ποσοστά των σωστών ταξινομημένων στιγμιότυπων του στατιστικού μέτρου Accuracy, του πίνακα 7-28, βλέπουμε ότι ο πιο αποδοτικός και ακριβείς στις προβλέψεις του είναι ο αλγόριθμος Simple Cart με μικρό σχετικά χρόνο κατασκευής του μοντέλου του. Την χειρότερη απόδοση από τους αλγορίθμους που επιλέχθηκαν στην έρευνα, παρατηρούμε στον MultiLayer Perceptron, πέρα από την μικρή

απόδοση, ο χρόνος κατασκευής του μοντέλου του ήταν κατά πολύ μεγαλύτερος από όλους τους υπόλοιπους αλγόριθμους. Επίσης, και ο αλγόριθμος SMO δεν παρουσίασε ικανοποιητική απόδοση και ακρίβεια πρόβλεψης.

Οι αλγόριθμοι Naïve Bayes, LWL, MODLEM και J48 παρουσίασαν σχεδόν την ίδια απόδοση και ποσοστό ακρίβεια και με μικρό χρόνο κατασκευής των μοντέλων τους.

Συγκρίνοντας τους συνδυαστικούς μεθόδους με βάση τη μετρική απόδοσης F-Measure και το ποσοστό ακρίβειας στις προβλέψεις Accuracy του πίνακα 7-28, παρατηρούμε ότι τις πιο ακριβείς προβλέψεις βλαβών και με υψηλότερο μέτρο απόδοσης έχει ο AdaBoost με βασικό ταξινομητή τον Simple Cart, βελτιώνοντας την απόδοσή του απλού Simple Cart κατά 1,1%. Η βελτίωση όμως επιφέρει πολύ μεγαλύτερο χρόνο κατασκευής του μοντέλου του.

Αλγόριθμοι Ταξινόμησης	F-Measure	Accuracy %	Χρόνος Κατασκευής Μοντέλου (sec)
Naïve Bayes	0.841	86,1	0,05
Multilayer Perceptron	0.530	54,4	333,53
SMO	0.747	75,5	1,34
LWL	0.890	90,1	0
MODLEM	0.905	90,4	0,56
J48	0.881	89,1	0,58
Simple Cart	0.955	95,5	5,73
AdaBoost J48	0.950	95,1	24,41
AdaBoost SimpleCart	0.965	96,6	238,99
MultiBoost MODLEM	0.919	92,5	13,73
MultiBoost J48	0.951	95,2	16,3
MultiBoost SimpleCart	0.959	96	224,86
Decorate J48	0.950	95,2	84,94

Πίνακας 7-28 Κατάταξη των αλγορίθμων σύμφωνα με τα αποτελέσματα του ποσοστού σωστών ταξινομημένων στιγμιότυπων και του χρόνου που απαιτείται για την κατασκευή του μοντέλου.

Δεύτερος πιο αποτελεσματικός σε ποσοστό ακρίβειας στις προβλέψεις και μέτρο απόδοσης F-Measure είναι ο MultiBoost με βασικό ταξινομητή τον Simple

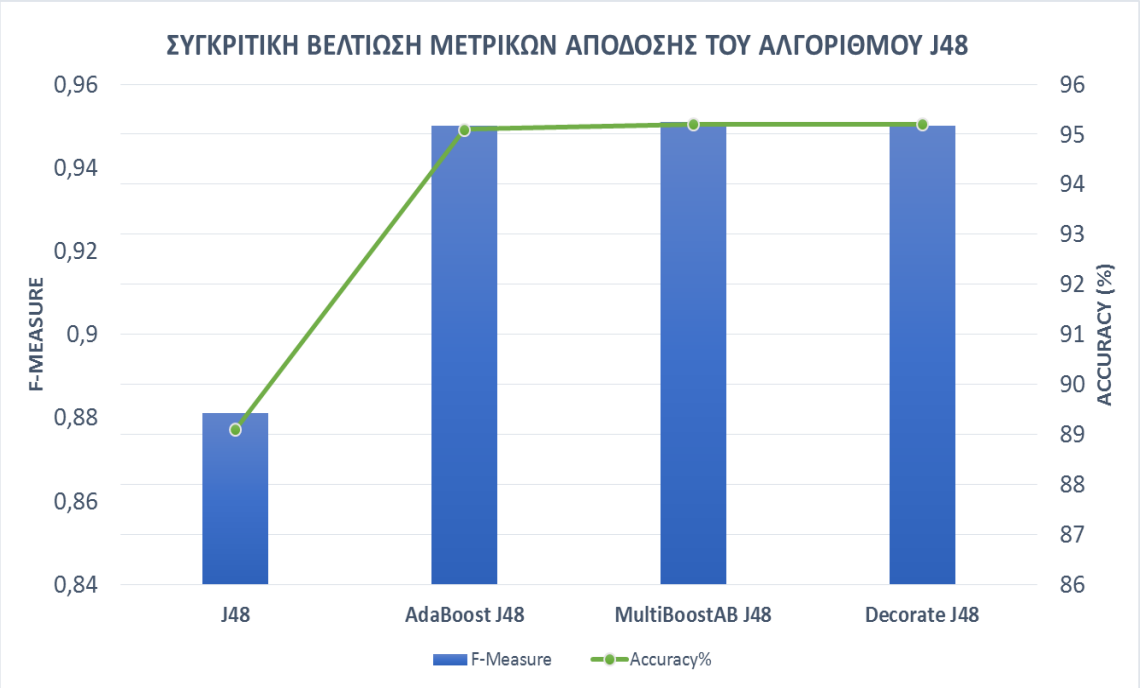
Cart, βελτιώνοντας την απόδοσή του απλού Simple Cart κατά 0,5% με επίσης μεγάλο χρόνο κατασκευής του μοντέλου του.

Ο βασικός ταξινομητής J48 ενισχύθηκε και απέδωσε ικανοποιητικά με τους συνδυαστικούς μεθόδους AdaBoost, MultiBoost και Decorate αυξάνοντας σημαντικά την απόδοσή του F-Measure και το ποσοστό ακρίβειας στις προβλέψεις Accuracy κατά 7%.

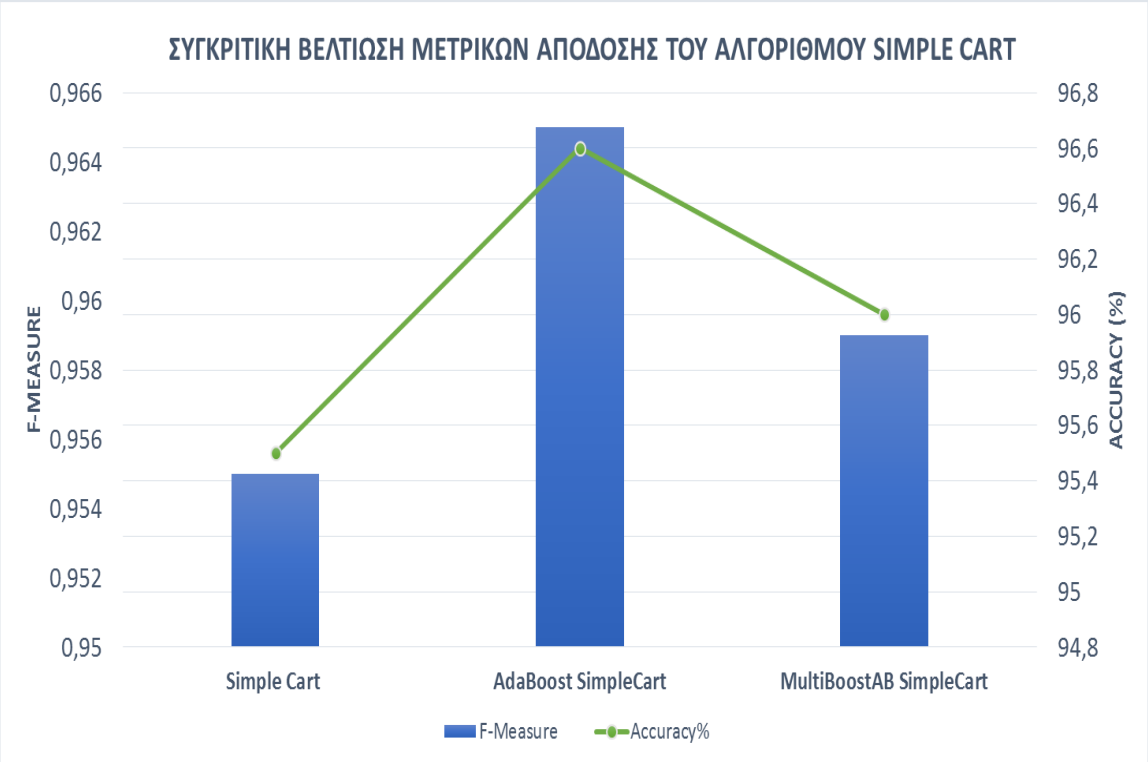
Ο λιγότερο πιο αποτελεσματικός είναι ο MultiBoost με βασικό ταξινομητή τον MODLEM, επιτυγχάνοντας μικρή βελτίωση.

Τα συγκριτικά αποτελέσματα βελτίωσης απόδοσης των βασικών αλγορίθμων ταξινόμησης MODLEM, J48 και Simple Cart με τις συνδυαστικές μεθόδους AdaBoost, MultiBoost και Decorate, απεικονίζονται γραφικά στα διαγράμματα 7-1, 7-2 και 7-3, τα γραφήματα εμφανίζουν δύο ξεχωριστούς άξονες, στον πρώτο άξονα εμφανίζεται με μπλε στήλες η μετρική απόδοσης F-Measure, στον δεύτερο άξονα εμφανίζεται με κόκκινη γραμμή το ποσοστό των σωστών ταξινομημένων στυγμιοτύπων (Accuracy).

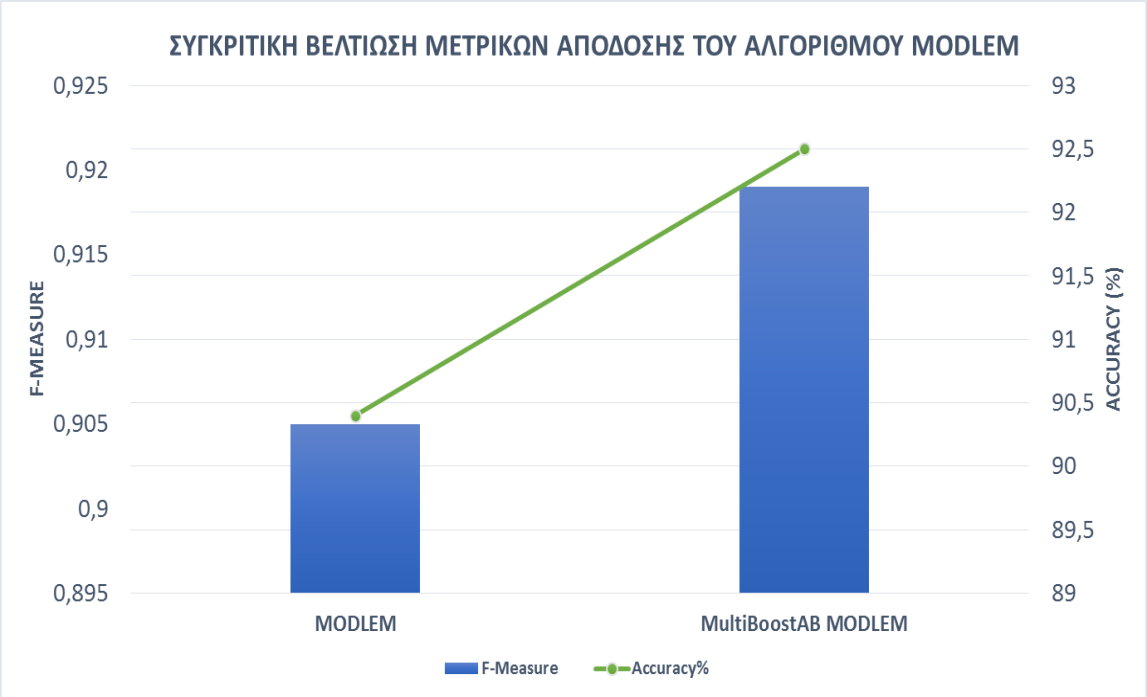
Τέλος, από τους πίνακες σύγχυσης όλων των αλγορίθμων παρατηρούμε τις τιμές των κελιών (εκτός της κύριας διαγωνίου) στις κατηγορίες 8 (Pmi_low & Exhaust_gass_temp _low) και 9 (Pmi_high & Exhaust_gass_temp _high), το ποσοστό των λάθος προβλέψεων είναι μεγαλύτερο από τις υπόλοιπες κατηγορίες-τάξεις. Οι βλάβες που ανήκουν σε αυτές τις κατηγορίες είναι περισσότερο συνδυαστικές και πολύπλοκες από τις βλάβες των άλλων κατηγοριών-τάξεων.



Διάγραμμα 7-1 Συγκριτική βελτίωση των μετρικών απόδοσης του βασικού αλγορίθμου J48 με τις συνδυαστικές μεθόδους AdaBoost, MultiBoost και Decorate



Διάγραμμα 7-2 Συγκριτική βελτίωση των μετρικών απόδοσης του βασικού αλγορίθμου Simple Cart με τις συνδυαστικές μεθόδους AdaBoost και MultiBoost



Διάγραμμα 7-3 Συγκριτική βελτίωση των μετρικών απόδοσης του βασικού αλγορίθμου MODLEM με την συνδυαστική μέθοδο MultiBoost

	NaiveBayes	Multilayer Perceptron	SMO	LWL	MODLEM	J48	Simple Cart	AdaBoost J48	AdaBoost SimpleCart	MultiBoostAB MODLEM	MultiBoostAB J48	MultiBoostAB SimpleCart	Decorate J48
Precision	0,864	0,529	0,752	0,899	0,907	0,881	0,955	0,939	0,966	0,919	0,941	0,962	0,953
Recall	0,861	0,544	0,755	0,901	0,904	0,891	0,955	0,937	0,966	0,925	0,939	0,96	0,952
F-Measure	0,841	0,53	0,747	0,89	0,905	0,881	0,955	0,935	0,965	0,919	0,938	0,959	0,95

Διάγραμμα 7-4 Σύγκριση των μετρικών απόδοσης (Precision, Recall, F-Measure) ανά αλγόριθμο

8. ΣΥΜΠΕΡΑΣΜΑΤΑ – ΠΡΟΤΑΣΕΙΣ

8.1. Συμπεράσματα

Στην συγκεκριμένη εργασία πραγματοποιήθηκε έρευνα στην κατασκευή διαγνωστικών μεθόδων, αναπτύσσοντας διάφορα μοντέλα ταξινόμησης με αλγορίθμους μηχανικής μάθησης, επιτρέποντας την εκτίμηση της λειτουργικής κατάστασης του δίχρονου αργόστροφου ναυτικού κινητήρα diesel για την ανίχνευση και διάγνωση βλαβών.

Οι διαγνωστικές μέθοδοι που παρουσιάζονται στην παρούσα εργασία φάνηκε ότι μπορούν να αποτελέσουν αξιόπιστα διαγνωστικά εργαλεία.

Τα κυριότερα συμπεράσματα που προέκυψαν από την έρευνα, μπορούν να συνοψιστούν στα εξής.

Από την παρουσίαση των αποτελεσμάτων στους βασικούς αλγόριθμους, υπερτέρησε ο Simple Cart, ο οποίος ανήκει στα δέντρα παλινδρόμησης, τα οποία χρησιμοποιούνται όταν οι εξαρτημένες τιμές όπως των δεδομένων της παρούσας εργασίας είναι συνεχείς, ως εκ τούτου πλεονέκτησε σε σχέση με τον J48 ο οποίος ανήκει στα δέντρα ταξινόμησης και είναι πιο κατάλληλος όταν οι εξαρτημένες τιμές είναι κατηγορηματικές.

Το μοντέλο που κατασκεύασε ο MultiLayer Perceptron είχε κακή απόδοση και αυτό οφείλεται, στο μικρό σχετικά όγκο δεδομένων που είχε για την εκπαίδευσή του, επίσης και ο SMO δεν παρουσίασε ικανοποιητική απόδοση, δεδομένου ότι οι κατηγορίες ταξινόμησης ήταν 17 πολύ παραπάνω από δύο, αυτό διαφοροποιείται από τα αποτελέσματα των προτάσεων στην βιβλιογραφική ανασκόπηση αφού σημαντικό ρόλο στα αποτελέσματα έχει ο τύπος και το εύρος των δεδομένων.

Στην συνέχεια με την παρουσίαση αποτελεσμάτων των συνδυαστικών μεθόδων, διαπιστώθηκε, όπως αποδεικνύεται από τα αποτελέσματα των προτεινόμενων μεθόδων στην βιβλιογραφική ανασκόπηση, οι συνδυαστικοί μέθοδοι, εκμεταλλευόμενη τα πλεονεκτήματα της μεθόδου ‘Boosting’, πέτυχαν βελτίωση της απόδοσης πρόβλεψης των βασικών αλγορίθμων, ανιχνεύοντας πολύ πιο αποτελεσματικά τις κατηγορίες βλαβών του κινητήρα, από ότι πετυχαίνει ο κάθε ένας αλγόριθμος ξεχωριστά, εξάγοντας ακριβέστερα και πιο αξιόπιστα διαγνωστικά συμπεράσματα.

Έτσι, ο προτεινόμενος αλγόριθμος που από τα αποτελέσματα προέκυψε η ξεκάθαρη υπεροχή του στην ανίχνευση και διάγνωση βλαβών, στηρίχθηκε στην χρήση της συνδυαστικής μεθόδου AdaBoost, με βασικό ταξινομητή το δέντρο απόφασης Simple Cart, που υπερτέρησε συγκριτικά με τους άλλους αλγόριθμους δίνοντας την καλύτερη απόδοση πρόβλεψης F-Measure 0.965, αλλά με κόστος την αύξηση του χρόνου που απαιτήθηκε για την κατασκευή του μοντέλου του.

Ειδικότερα, οι εξετασθέντες αλγόριθμοι δίνουν ως έναν βαθμό αξιόπιστα αποτελέσματα, ενώ η πειραματική τους εκτέλεση επιβεβαιώνει σε αρκετά σημεία το εξετασθέν θεωρητικό υπόβαθρο.

Κρίνοντας από τα αποτελέσματα, συμπεραίνουμε πως η επιτυχής ανίχνευση και διάγνωση βλαβών του κινητήρα με χρήση του προτεινόμενου μοντέλου είναι εφικτή.

Τα αποτέλεσμα αυτής της εργασίας υποδεικνύουν πόσο χρήσιμες είναι οι τεχνικές των αλγορίθμων μηχανικής μάθησης και τις δυνατότητες που παρέχει το εργαλείο εξόρυξης δεδομένων Weka στην ανάλυση των λειτουργικών παραμέτρων για την εφαρμογή του διαγνωστικού συστήματος ανίχνευσης και διάγνωσης βλαβών στους δίχρονους αργόστροφους ναυτικούς κινητήρες diesel.

8.2. Περιορισμοί

Η συλλογή μετρήσεων από λειτουργούντες ναυτικούς κινητήρες πριν και μετά την εμφάνιση κάποιας βλάβης ή με τεχνητά προκαλούμενης βλάβης είναι πρακτικά αδύνατο να συλλεχθούν όλες οι απαραίτητες μετρήσεις για την δημιουργία βάσης υπογραφών βλαβών σε εύλογο χρονικό διάστημα και με τον φόβο η εμφάνιση κάποιας βλάβης να οδηγήσει στην καταστροφή του κινητήρα.

Επιπλέον, μετά από επισκέψεις σε πολλές ναυτιλιακές εταιρείες, δεν κατέστη δυνατόν να βρεθεί βάση με πραγματικά δεδομένα, κυρίως από τις λειτουργικές παραμέτρους κύριας μηχανής.

Για την αντιμετώπιση αυτών των προβλημάτων, η εξαγωγή των μετρούμενων μεγεθών από τις λειτουργικές παραμέτρους, για την δημιουργία βάσης υπογραφής βλαβών, χρησιμοποιήθηκε προσομοίωση της λειτουργίας ενός μοντέλου ναυτικού

κινητήρα, από τον προσομοιωτή μηχανοστασίου πλοίου, της Ακαδημίας Εμπορικού Ναυτικού Ασπροπύργου.

8.3. Μελλοντική Έρευνα

Στο σημείο αυτό, έχοντας πια μια συνολική εικόνα της λειτουργίας των μεθόδων, παρουσιάζονται κάποιες πιθανές μελλοντικές εργασίες όπως η συλλογή και καταχώρηση περισσότερων δεδομένων στο σύνολο εκπαίδευσης των αλγορίθμων μηχανικής μάθησης για καλύτερες προβλέψεις. Συλλογή πραγματικών δεδομένων από τις ναυτιλιακές εταιρείες και από διάφορους τύπους μηχανής.

Τα αποτελέσματα που προκύπτουν από την παρούσα εργασία είναι ιδιαίτερα ικανοποιητικά και ενθαρρυντικά προς την κατεύθυνση της περαιτέρω αξιοποίησης των αλγορίθμων μηχανικής μάθησης για την αυτόματη διάγνωση και ανίχνευση βλαβών, με στόχο την έγκυρη και έγκαιρη πρόβλεψη ενός σοβαρού ναυτικού ατυχήματος. Θα μπορούσε επίσης να επιχειρηθεί η εξόρυξη περισσότερων χαρακτηριστικών που αφορούν τους ναυτικούς κινητήρες.

Είναι επιθυμητή η χρήση υβριδικών συστημάτων, δηλαδή περισσοτέρων της μίας διαγνωστικής μεθόδου και με τον κατάλληλο συνδυασμό τους να επωφελούνται από τα πλεονεκτήματα της κάθε μεθόδου, προσδίδοντας έτσι αυξημένες ικανότητες ανίχνευσης βλαβών και αξιοπιστίας, αφού είναι γνωστό ότι συχνά κάθε μέθοδος έχει στην πραγματικότητα συγκεκριμένες αδυναμίες εφαρμογής.

9. ΒΙΒΛΙΟΓΡΑΦΙΑ

Ζώτος, Ι. (2008). *Αναγνώριση Βλαβών Λειτουργίας Περιστρεφόμενων Εξαρτημάτων Μηχανών Μετάδοσης Κίνησης. Διδακτορική Διατριβή*. Εθνικό Μετσόβιο Πολυτεχνείο.

Κυρτάτος, Ν. (1999). Σημαντικά Θέματα Έρευνας και Εξέλιξης στους Ναυτικούς Κινητήρες Diesel, (Ifo 380), 1–12.

Λαζάρου Χ. Κλιάνη, Ιωάννη Κ. Νικαλάου, Ι. Α. Σ. (2003). *Μηχανές Εσωτερικής Καύσεως Τόμος Δεύτερος*.

Λούκης Ε. (1993). *Συμβολή στην Διάγνωση Βλαβών Αεριοστρόβιλων με Χρήση Μεθόδων Ανάλυσης Μετρήσεων Ταχείας Απόκρισης. Διδακτορική Διατριβή*. Εθνικό Μετσόβιο Πολυτεχνείο.

Μαργαρώνης, Ι. Ε. (1986). *Συστήματα Διάγνωσης Λειτουργίας σε Ναυτικές Μηχανές Diesel. Διδακτορική Διατριβή*. Εθνικό Μετσόβιο Πολυτεχνείο.

Ρωμέσης, Χ. Ν. (2005). *Χρήση Στοχαστικών Μεθόδων στην Ανάπτυξη Συστημάτων Παρακολούθησης Λειτουργίας και Διάγνωση Βλαβών Αεριοστρόβιλων. Διδακτορική Διατριβή*. Εθνικό Μετσόβιο Πολυτεχνείο.

Σκουντριανός, Η. (2005). *Μοντελοποίηση και Διάγνωση-Αναγνώριση Βλαβών μη Γραμμικών Δυναμικών Συστημάτων με Νευρωνικά Δίκτυα Τοπικών Μοντέλων. Διδακτορική Διατριβή*. Εθνικό Μετσόβιο Πολυτεχνείο.

Τσανάκας, Ι. Α. (2013). *Προηγμένες Τεχνολογίες Διάγνωσης-Πρόγνωσης Βλαβών σε Μηχανολογικές Κατασκευές με Χρήση Εποπτικών Μεθόδων: Περίπτωση Υπέρυθρης Θερμογραφίας*. Δημοκρίτειο Πανεπιστήμιο Κρήτης.

Advanced modeling and decision making software | KnowledgeSTUDIO. (n.d.). Retrieved August 18, 2017, from http://www.angoss.com/predictive-analytics-software/software/knowledgestudio/

Aksenova, S. S. (2004). Machine Learning with WEKA WEKA Explorer Tutorial for WEKA Version 3.4.3. Retrieved from http://csed.sggs.ac.in/csed/sites/default/files/WEKA Explorer Tutorial.pdf

Amancio, D. R., Comin, C. H., Casanova, D., Travieso, G., Bruno, O. M., Rodrigues, F. A., & Da Fontoura Costa, L. (2014). A systematic comparison of supervised classifiers. *PLoS ONE*, *9*(4), 1–13. https://doi.org/10.1371/journal.pone.0094137

Amozegar, M., & Amozegar, M. (2015). Aircraft Jet Engine Health Monitoring Through System Identification Using Ensemble Neural Networks. Retrieved from http://spectrum.library.concordia.ca/980252/

Apache Mahout: Scalable machine learning and data mining. (n.d.). Retrieved August 18, 2017, from http://mahout.apache.org/

Auria, L., & Moro, R. A. (2008). Support Vector Machines (SVM) as a Technique for Solvency Analysis. Retrieved from www.diw.de

Ayubi Rad, M. A., & Yazdanpanah, M. J. (2015). Designing supervised local neural network classifiers based on EM clustering for fault diagnosis of Tennessee Eastman process. *Chemometrics and Intelligent Laboratory Systems*, *146*, 149–

157. https://doi.org/10.1016/j.chemolab.2015.05.013
Bauer, E., Kohavi, R., Chan, P., Stolfo, S., & Wolpert, D. (1999). An Empirical Comparison of Voting Classification Algorithms: Bagging, Boosting, and Variants. *Machine Learning*, *36*, 105–139. Retrieved from http://download.springer.com/static/pdf/369/art%253A10.1023%252FA%253A1007515423169.pdf?originUrl=http%3A%2F%2Flink.springer.com%2Farticle%2F10.1023%2FA%3A1007515423169&token2=exp=1491847257~acl=%2Fstatic%2Fpdf%2F369%2Fart%25253A10.1023%25252FA%25253A1007
Benbouzid, D., Busa-Fekete, R., Casagrande, N., Collin, F.-D., & Com, B. K. (2012). MULTIBOOST: A Multi-purpose Boosting Package Balázs Kégl. *Journal of Machine Learning Research*, *13*, 549–553. Retrieved from file:///C:/Users/AΘHNA/Downloads/multiboost benbouzid12a.pdf
Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., & Scuse, D. (2013). WEKA Manual for Version 3-7-8, 1–327. Retrieved from papers3://publication/uuid/24E005A2-AA1B-4614-BAF5-4D92C4F37413
Breiman, L. (1984). *Classification and regression trees*. Wadsworth International Group. Retrieved from http://cds.cern.ch/record/2253780
Business-Intelligence: A categorization of the different B-I solutions. (n.d.). Retrieved August 18, 2017, from http://business-insight.com/html/intelligence/bi_intro.html
Chauhan, N., & Gautam, N. (2015). Parametric Comparison of Data Mining Tools, *v*, 291–298.
Chih-Wei Hsu, Chih-Chung Chang, and C.-J. L. (2008). A Practical Guide to Support Vector Classification. *BJU International*, *101*(1), 1396–400. https://doi.org/10.1177/02632760022050997
Data Mining Software, Model Development and Deployment, SAS Enterprise Miner | SAS. (n.d.). Retrieved August 18, 2017, from https://www.sas.com/en_us/software/enterprise-miner.html
Data Science Platform | RapidMiner. (n.d.). Retrieved August 18, 2017, from https://rapidminer.com/
DataMelt. (n.d.). Retrieved August 18, 2017, from http://jwork.org/dmelt/
Engine Selection Guide Two-stroke MC/MC-C Engines. (2000) (5th ed.). Retrieved from https://www.fsb.unizg.hr/ship-design/esg.pdf
Englert, P. (n.d.). Locally Weighted Learning. Retrieved from http://www.ias.informatik.tu-darmstadt.de/uploads/Teaching/AutonomousLearningSystems/Englert_ALS_2012.pdf
Fayyad, U. M. (1996). Data Mining and Knowledge Discovery in Databases: Applications in Astronomy and Planetary Science. Retrieved from http://www.research.microsoft.com/research/dtg
Fig, Y. (2000). Data mining and warehousing concepts 1.1, 1–7. Retrieved from http://www.asee.org/documents/zones/zone1/2014/Student/PDFs/19.pdf
Frank, E., Hall, M., Holmes, G., Kirkby, R., Pfahringer, B., Witten, I. H., & Trigg, L. (2005). Weka. In *Data Mining and Knowledge Discovery Handbook* (pp. 1305–1314). New York: Springer-Verlag. https://doi.org/10.1007/0-387-25465-X_62
Freund, Y., & Schapire, R. E. (1996). Experiments with a New Boosting Algorithm. Retrieved from http://www.research.att.com/orgs/ssr/people/
Gao, X., & Hou, J. (2016). An improved SVM integrated GS-PCA fault diagnosis

approach of Tennessee Eastman process. *Neurocomputing*, *174*, 906–911. https://doi.org/10.1016/j.neucom.2015.10.018

Grzymala-Busse, J. W., & Stefanowski, J. (n.d.). Three Discretization Methods for Rule Induction. Retrieved from http://sci2s.ugr.es/keel/pdf/specific/articulo/IJIS01.pdf

Hu Jinhai,Xie Shousheng,Cai Kailong,He Xiuran, P. J. (2007). Classification Method of DivClassification Method of Diverse AdaBoost-SVM and Its Application to Fault Diagnosis of Aeroengine. Retrieved from http://en.cnki.com.cn/Article_en/CJFDTOTAL-HKXB200705013.htm

IBM Knowledgecenter - IBM DB2 Intelligent Miner for Data. (n.d.). Retrieved August 12, 2017, from https://www.ibm.com/support/knowledgecenter/en/SSEPGG_9.5.0/com.ibm.im.overview.doc/c_ibm_db2_intelligent_miner_for_data.html

Jović, A., Brkić, K., & Bogunović, N. (2014). An overview of free software tools for general data mining. *2014 37th International Convention on Information and Communication Technology, Electronics and Microelectronics, MIPRO 2014 - Proceedings*, 1112–1117. https://doi.org/10.1109/MIPRO.2014.6859735

KEEL: A software tool to assess evolutionary algorithms for Data Mining problems (regression, classification, clustering, pattern mining and so on). (n.d.). Retrieved August 18, 2017, from http://sci2s.ugr.es/keel/category.php?cat=clas

Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. *Informatica*, *31*, 249–268. https://doi.org/10.1115/1.1559160

Krčadinac, U. (n.d.). Training and Testing. Retrieved from http://krcadinac.com

Kundu, M. K. (2014). *Advanced computing, networking and informatics*. Retrieved from https://books.google.gr/books?id=hTwqBAAAQBAJ&pg=PA579&lpg=PA579&dq=multiboostab&source=bl&ots=DvD2Q7_d2E&sig=VwHpmy10Un2hoqXrSQ0qaV0wu6Q&hl=el&sa=X&ved=0ahUKEwibgbX0r8_VAhWF5xoKHb_dC1A4ChDoAQhEMAk#v=onepage&q=multiboostab&f=false

Lan, W.-C., Katagi, T., & Hashimoto, T. (1996). Quasi Steady State Simulation of Diesel Engine Transient Performance and Design of Mechatronic Governor*. Retrieved from http://www.jime.jp/e/publication/bulletin/english/pdf/mv24n011996p01.pdf

Li, Z., Yan, X., Guo, Z., Zhang, Y., Yuan, C., & Peng, Z. (2012). Condition Monitoring and Fault Diagnosis for Marine Diesel Engines using Information Fusion Techniques. *Electronics and Electrical Engineering*, *123*(7), 109–112. https://doi.org/10.5755/j01.eee.123.7.2387

Madzarov, G., Gjorgjevikj, D., & Chorbev, I. (2009). A Multi-class SVM Classifier Utilizing Binary Decision Tree. *Informatica*, *33*, 233–241. Retrieved from https://pdfs.semanticscholar.org/7754/5697e84da29842832edfc369b6bc46757d02.pdf

Moore, D. H. (1987). Classification and regression trees, by Leo Breiman, Jerome H. Friedman, Richard A. Olshen, and Charles J. Stone. *Cytometry*, *8*(5), 534–535. https://doi.org/10.1002/cyto.990080516

Moosavian, A., Ahmadi, H., Tabatabaeefar, A., & Khazaee, M. (2013). Comparison of two classifiers; K-nearest neighbor and artificial neural network, for fault diagnosis on a main engine journal-bearing. *Shock and Vibration*, *20*(2), 263–272. https://doi.org/10.3233/SAV-2012-00742

Oracle Data Mining. (n.d.). Retrieved August 18, 2017, from http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html
Orange. (n.d.). Retrieved August 18, 2017, from https://orange.biolab.si/license/
Pedersen, B. F., & Engineer, S. M. (n.d.). Diesel Engine Combustion Analysis for Condition Monitoring and Energy Conservation.
Platt, J. (1998). *Sequential Minimal Optimization: A Fast Algorithm for Training Support Vector Machines*. Retrieved from https://www.microsoft.com/en-us/research/publication/sequential-minimal-optimization-a-fast-algorithm-for-training-support-vector-machines/
Prem Melville. (2003). Creating Diverse Ensemble Classifiers. Retrieved from http://www.cs.utexas.edu/~ml/papers/decorate-proposal-03.pdf
Project Guide Camshaft Controlled Two-stroke Engines. (2009). In *MAN B&W S60MC-C8* (1st ed.).
Quinlan, J. R. (John R. (1993). *C4.5 : programs for machine learning*. Morgan Kaufmann Publishers. Retrieved from https://books.google.gr/books?hl=el&lr=&id=b3ujBQAAQBAJ&oi=fnd&pg=PP1&dq=Quinlan+JR+(1993)+C4.5:+Programs+for+Machine+Learning.+Morgan+Kaufmann&ots=sPapQOBsz5&sig=59fxlRf9D5UfBxL6hM0FDn92ZVM&redir_esc=y#v=onepage&q=Quinlan JR (1993) C4.5%3A Progra
Rattle: A Graphical User Interface for Data Mining using R. (n.d.). Retrieved August 18, 2017, from https://rattle.togaware.com/
Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, *33*(1–2), 1–39. https://doi.org/10.1007/s10462-009-9124-7
Rokach, L., & Maimon, O. (n.d.). DECISION TREES. Retrieved from http://www.ise.bgu.ac.il/faculty/liorr/hbchap9.pdf
Ross, K. A., Jensen, C. S., Snodgrass, R., Dyreson, C. E., Jensen, C. S., Snodgrass, R., … Chen, L. (2009). Cross-Validation. In *Encyclopedia of Database Systems* (pp. 532–538). Boston, MA: Springer US. https://doi.org/10.1007/978-0-387-39940-9_565
Sahin, F., Yavuz, M. Ç., Arnavut, Z., & Uluyol, Ö. (2007). Fault diagnosis for airplane engines using Bayesian networks and distributed particle swarm optimization. *Parallel Computing*, *33*(2), 124–143. https://doi.org/10.1016/j.parco.2006.11.005
Schapire, R. (2013). Theoretical Machine Learning, 1–7.
Sharkey, A. J. C., Chandroth, G. O., & Sharkey, N. E. (2000). A Multi-Net System for the Fault Diagnosis of a Diesel Engine. *Neural Computing & Applications*, *9*(2), 152–160. https://doi.org/10.1007/s005210070026
Singh Sabharwal, J. (n.d.). Multi-Label Text Classification. Retrieved from http://jasneet.me/assets/files/multi_label_svm.pdf
Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*, 427–437. https://doi.org/10.1016/j.ipm.2009.03.002
SPMF: A Java Open-Source Data Mining Library. (n.d.). Retrieved August 18, 2017, from http://www.philippe-fournier-viger.com/spmf/
Timofeev, R., & Härdle, W. (2004). Classification and Regression Trees (CART) Theory and Applications. Retrieved from http://edoc.hu-berlin.de/master/timofeev-roman-2004-12-20/PDF/timofeev.pdf

Twiddle, J. A., & Jones, N. B. (2002). A high-level technique for diesel engine combustion system condition monitoring and fault diagnosis. *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, *216*(2), 125–134. https://doi.org/10.1243/0959651021541499

User's Guide PMI System. (2005). In *MAN B&W Diesel A/S* (2.3, p. 82). Retrieved from http://marengine.com/ufiles/MAN-PMI_off.pdf

Vong, C. M., Wong, P. K., & Wong, K. I. (2014). Simultaneous-fault detection based on qualitative symptom descriptions for automotive engine diagnosis. *Applied Soft Computing*, *22*, 238–248. https://doi.org/10.1016/j.asoc.2014.05.014

Weka 3 - Data Mining with Open Source Machine Learning Software in Java. (n.d.). Retrieved August 18, 2017, from http://www.cs.waikato.ac.nz/ml/weka/

Widodo, A., & Yang, B.-S. (2011). Machine health prognostics using survival probability and support vector machine. *Expert Systems with Applications*, *38*(7), 8430–8437. https://doi.org/10.1016/j.eswa.2011.01.038

Wimmer, H., & Powell, L. M. (2015). A Comparison of Open Source Tools for Data Science, 1–9. Retrieved from http://iscap.info

Wisaeng, K. (2013). An Empirical Comparison of Data Mining Techniques in Medical Databases. *International Journal of Computer Applications*, *77*(7), 23–27. https://doi.org/10.5120/13408-1061

Witten, I. H. (Ian H. ., Frank, E., Hall, M. A. (Mark A., & Pal, C. J. (n.d.). *Data mining : practical machine learning tools and techniques*. Retrieved from https://books.google.gr/books?hl=el&lr=&id=1SylCgAAQBAJ&oi=fnd&pg=PP1&dq=functions+algorithm+machine+learning+weka&ots=8HIKpdnvud&sig=E4bi22FTvvhGS_4R-cagYHzWc4s&redir_esc=y#v=onepage&q=functions algorithm machine learning weka&f=false

Wong, P. K., Zhong, J., Yang, Z., & Vong, C. M. (2016). Sparse Bayesian extreme learning committee machine for engine simultaneous fault diagnosis. *Neurocomputing*, *174*, 331–343. https://doi.org/10.1016/j.neucom.2015.02.097

Xiros, N. I., & Kyrtatos, N. P. (2000). A neural predictor of propeller load demand for improved control of diesel ship propulsion. In *Proceedings of the 2000 IEEE International Symposium on Intelligent Control. Held jointly with the 8th IEEE Mediterranean Conference on Control and Automation (Cat. No.00CH37147)* (pp. 321–326). IEEE. https://doi.org/10.1109/ISIC.2000.882944

Xu, H. F. (2012). New Intelligent Condition Monitoring and Fault Diagnosis System for Diesel Engines Using Embedded System. *Applied Mechanics and Materials*, *235*, 408–412. https://doi.org/10.4028/www.scientific.net/AMM.235.408

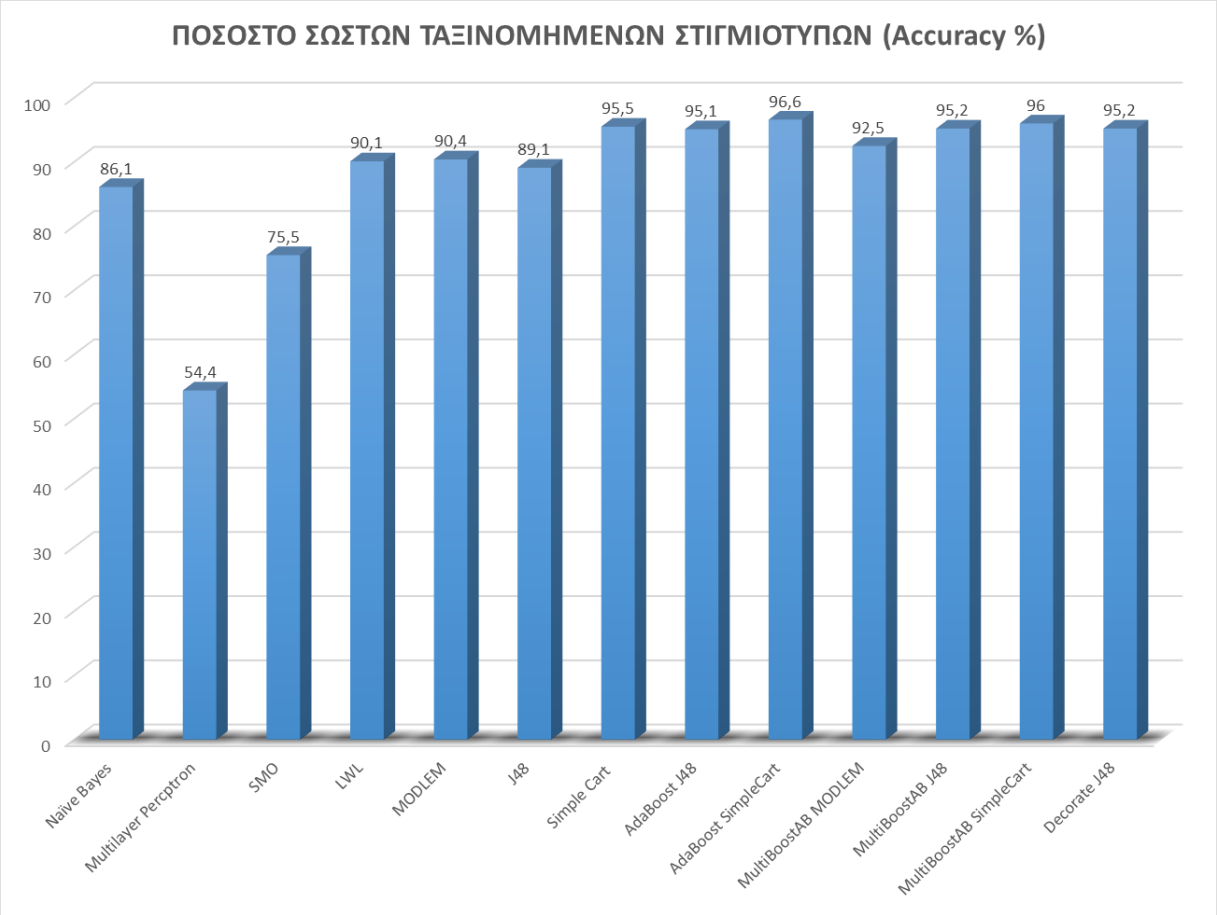
Yuan, S.-F., & Chu, F.-L. (2006). Support vector machines-based fault diagnosis for turbo-pump rotor. *Mechanical Systems and Signal Processing*, *20*(4), 939–952. https://doi.org/10.1016/j.ymssp.2005.09.006

Yuan, S., & Chu, F. (2007). Fault diagnosis based on support vector machines with parameter optimisation by artificial immunisation algorithm. *Mechanical Systems and Signal Processing*, *21*(3), 1318–1330. https://doi.org/10.1016/j.ymssp.2006.06.006

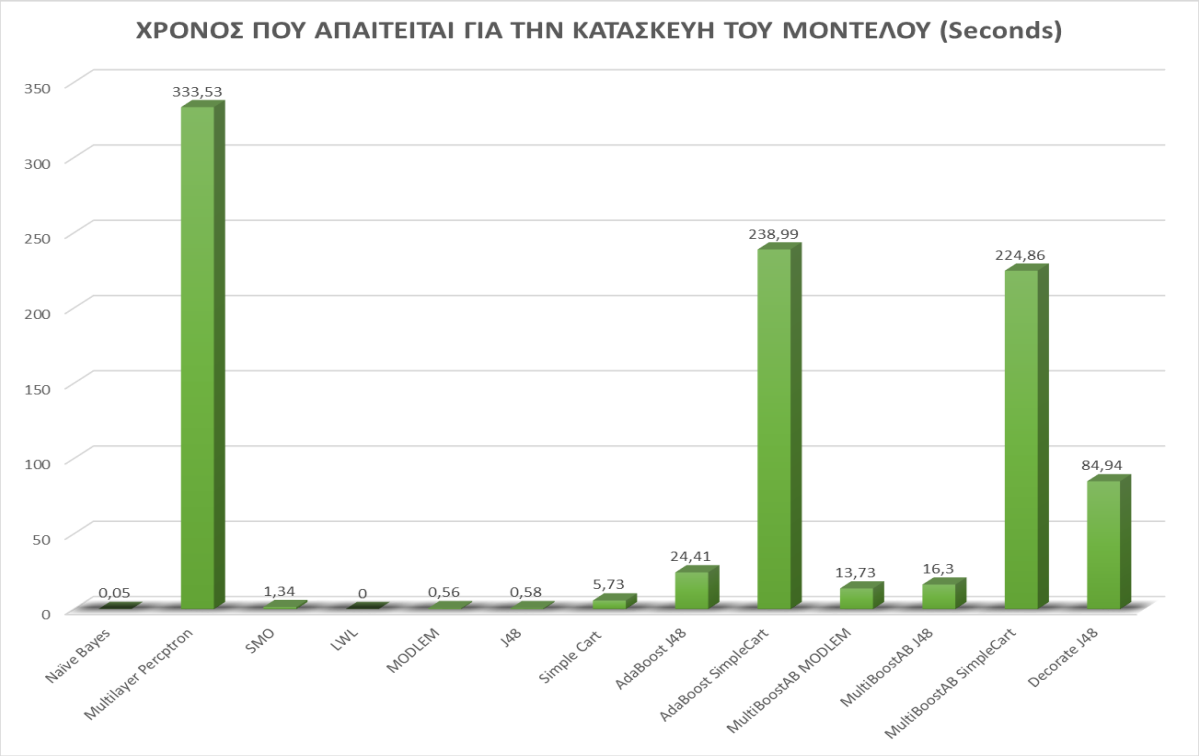
Zhan, Y.-L., Shi, Z.-B., Shwe, T., & Wang, X.-Z. (2007). Fault Diagnosis of Marine Main Engine Cylinder Cover Based on Vibration Signal. In *2007 International Conference on Machine Learning and Cybernetics* (pp. 1126–1130). IEEE. https://doi.org/10.1109/ICMLC.2007.4370313

ΠΑΡΑΡΤΗΜΑΤΑ

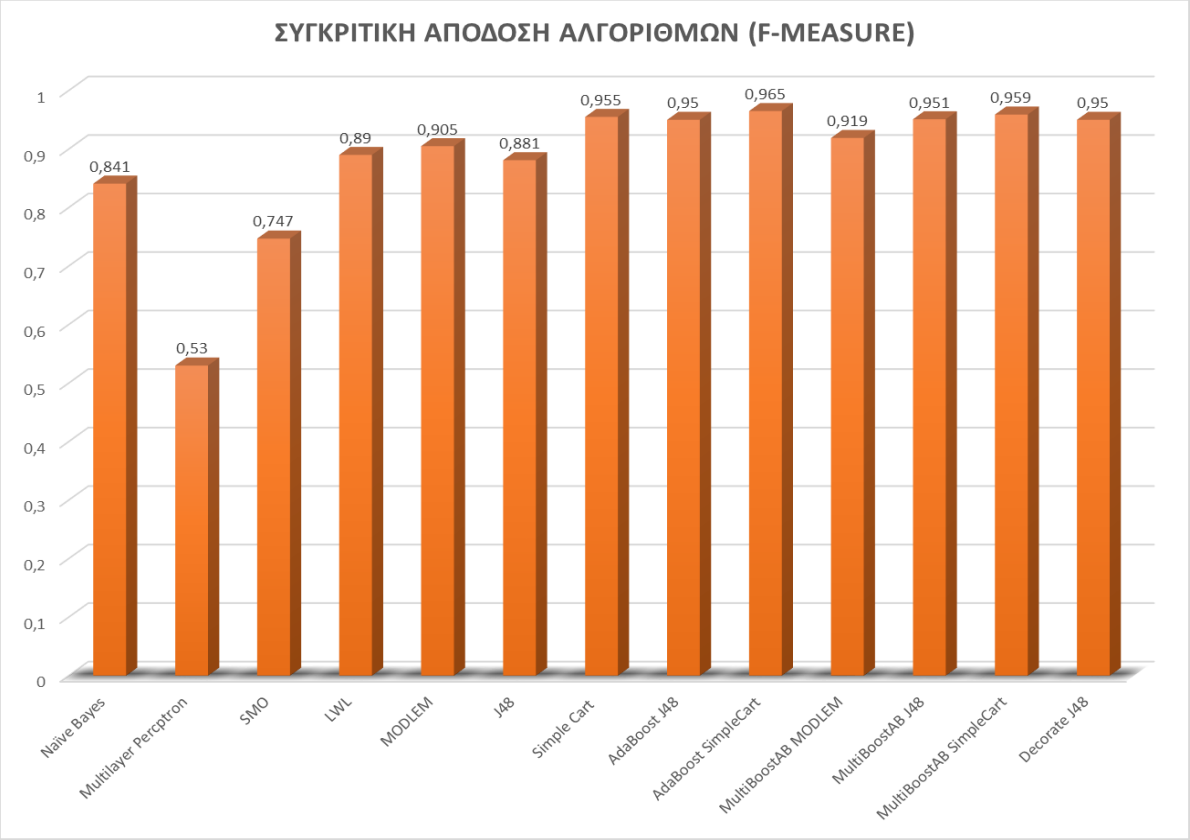
ΠΑΡΑΡΤΗΜΑ 1



Διάγραμμα 1 Ποσοστό σωστών ταξινομημένων στιγμιότυπων ανά αλγόριθμο (accuracy%).



Διάγραμμα 2 Χρόνος που απαιτείται για την κατασκευή των μοντέλων.



Διάγραμμα 3 Σύγκριση της μετρικής απόδοσης F-Measure ανά αλγόριθμο.

ΠΑΡΑΡΤΗΜΑ 2

Scheme: **weka.classifiers.bayes.NaiveBayes -K**
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
power_c1
power_c2
power_c3
power_c4
power_c5
power_c6
power_c7
rpm_c1
rpm_c2
rpm_c3
rpm_c4
rpm_c5
rpm_c6
rpm_c7
pmi_c1
pmi_c2
pmi_c3
pmi_c4
pmi_c5
pmi_c6
pmi_c7
pcomp_c1
pcomp_c2
pcomp_c3
pcomp_c4
pcomp_c5
pcomp_c6
pcomp_c7
pmax_pos_c1
pmax_pos_c2
pmax_pos_c3
pmax_pos_c4
pmax_pos_c5
pmax_pos_c6
pmax_pos_c7
pmax_c_c1
pmax_c_c2
pmax_c_c3
pmax_c_c4
pmax_c_c5
pmax_c_c6
pmax_c_c7
ignition_c1
ignition_c2
ignition_c3
ignition_c4
ignition_c5
ignition_c6
ignition_c7
exhaust_gass_temp_c1
exhaust_gass_temp_c2
exhaust_gass_temp_c3
exhaust_gass_temp_c4
exhaust_gass_temp_c5
exhaust_gass_temp_c6
exhaust_gass_temp_c7

faultys
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Naive Bayes Classifier

Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	861	86.1 %
Incorrectly Classified Instances	139	13.9 %
Kappa statistic	0.8463	
Mean absolute error	0.0252	
Root mean squared error	0.1149	
Relative absolute error	23.5475 %	
Root relative squared error	49.6855 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,965	0,040	0,878	0,965	0,920	0,896	0,984	0,931	0
	1,000	0,002	0,961	1,000	0,980	0,979	0,999	0,980	1
	0,700	0,000	1,000	0,700	0,824	0,830	1,000	0,996	2
	0,960	0,024	0,676	0,960	0,793	0,794	0,996	0,929	3
	0,960	0,036	0,585	0,960	0,727	0,734	0,980	0,577	4
	1,000	0,001	0,980	1,000	0,990	0,990	1,000	1,000	5
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
	0,314	0,003	0,786	0,314	0,449	0,487	0,984	0,659	8
	0,057	0,001	0,667	0,057	0,105	0,189	0,949	0,368	9
	0,981	0,005	0,911	0,981	0,944	0,942	1,000	0,998	10
	1,000	0,003	0,941	1,000	0,970	0,969	1,000	1,000	11
	0,959	0,001	0,979	0,959	0,969	0,968	1,000	0,998	12
	1,000	0,002	0,963	1,000	0,981	0,980	1,000	1,000	13
	0,490	0,006	0,800	0,490	0,608	0,612	0,963	0,689	14
	0,449	0,013	0,647	0,449	0,530	0,520	0,946	0,587	15
	1,000	0,016	0,773	1,000	0,872	0,872	1,000	0,997	16
Weighted Avg.	0,861	0,015	0,864	0,861	0,841	0,840	0,988	0,888	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
223	2	0	0	0	0	0	0	0	0	1	1	0	0	0	4	0	a = 0
0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
0	0	35	0	0	0	0	0	0	0	0	0	0	0	0	0	15	c = 2
0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	0	d = 3
0	0	0	1	48	0	0	0	0	1	0	0	0	0	0	0	0	e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h = 7
0	0	0	21	3	0	0	0	11	0	0	0	0	0	0	0	0	i = 8
0	0	0	1	31	0	0	0	1	2	0	0	0	0	0	0	0	j = 9
0	0	0	0	0	0	0	0	0	0	51	1	0	0	0	0	0	k = 10
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l = 11
0	0	0	0	0	0	0	0	0	0	0	0	47	2	0	0	0	m = 12
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n = 13
14	0	0	0	0	0	0	0	0	0	3	0	0	0	24	8	0	o = 14
17	0	0	0	0	1	0	0	0	0	1	1	1	0	6	22	0	p = 15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q = 16

Scheme: weka.classifiers.rules.MODLEM -RT 1 -CM 1 -CS 6 -AS 0
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Rule 1. (pcomp_c4 < 93.15)&(power_c4 < 1601.5)&(power_c3 >= 1588.5)&(rpm_c3 < 96.95)&(pmi_c5 >= 14.05)&(pmax_pos_c4 >= 13.25) => (faultys = 0) (14/14, 6.06%)
Rule 2. (pmax_pos_c7 < 12.95)&(pmax_pos_c2 >= 13.65)&(exhaust_gass_temp_c2 < 371.5)&(power_c6 < 1609.5)&(rpm_c5 >= 95.55)&(power_c2 < 1610.5)&(power_c5 >= 1586.5)&(pcomp_c1 >= 93.15)&(pcomp_c2 < 97.5) => (faultys = 0) (18/18, 7.79%)
Rule 3. (pmax_pos_c5 < 12.95)&(rpm_c2 < 96.25)&(pmax_c_c1 < 19.85)&(power_c5 < 1604.5)&(power_c1 < 1610.5)&(rpm_c7 < 96.95) => (faultys = 0) (16/16, 6.93%)
Rule 4. (pmax_pos_c2 >= 14.35)&(pmax_c_c1 >= 19.95)&(pcomp_c4 < 94.45)&(pmi_c2 >= 14.15)&(power_c1 < 1608.5)&(exhaust_gass_temp_c6 < 407.5)&(power_c2 < 1610.5)&(power_c5 >= 1589.5) => (faultys = 0) (16/16, 6.93%)
Rule 5. (pmax_c_c6 < 17.35)&(power_c1 < 1599.5)&(ignition_c3 < 2.65)&(ignition_c5 >= 2.25)&(rpm_c1 < 97.05)&(pcomp_c2 >= 93.15) => (faultys = 0) (11/11, 4.76%)
Rule 6. (pmax_c_c3 >= 20.85)&(pmax_c_c4 >= 19.05)&(pmi_c7 >= 14.15)&(pcomp_c1 >= 93.85)&(pmax_c_c5 >= 17.45)&(pmi_c3 < 15.3) => (faultys = 0) (13/13, 5.63%)
Rule 7. (pcomp_c1 < 93.25)&(pcomp_c6 >= 95.45)&(pmax_c_c7 >= 18.05)&(pmi_c7 < 15.15) => (faultys = 0) (8/8, 3.46%)
Rule 8. (rpm_c4 >= 96.85)&(rpm_c5 >= 96.05)&(rpm_c3 < 96.45)&(exhaust_gass_temp_c3 < 368.5)&(power_c1 >= 1593.5)&(pmax_pos_c1 >= 13.05)&(rpm_c7 < 96.85) => (faultys = 0) (14/14, 6.06%)
Rule 9. (pmax_pos_c2 < 12.85)&(pmax_c_c6 < 18.05)&(power_c3 >= 1595.5)&(pmi_c5 >= 14.35) => (faultys = 0) (6/6, 2.6%)
Rule 10. (pmax_c_c2 < 17.75)&(power_c4 >= 1607.5)&(power_c7 >= 1599.5)&(pmi_c1 >= 14.15)&(power_c1 < 1659) => (faultys = 0) (8/8, 3.46%)
Rule 11. (pmax_c_c5 < 17.35)&(pmax_c_c6 >= 18.95)&(rpm_c5 >= 96.45)&(power_c6 >= 1587.5)&(rpm_c3 >= 95.65) => (faultys = 0) (8/8, 3.46%)
Rule 12. (pmax_c_c4 < 17.75)&(pmax_pos_c2 >= 14.45)&(power_c1 >= 1596.5)&(rpm_c3 < 96.75) => (faultys = 0) (8/8, 3.46%)
Rule 13. (pcomp_c5 < 93.15)&(pmi_c1 >= 14.45)&(power_c5 < 1602.5)&(power_c2 < 1608.5) => (faultys = 0) (6/6, 2.6%)
Rule 14. (pcomp_c7 in [95.65, 97.35])&(rpm_c1 < 96.15)&(ignition_c6 >= 2.55)&(power_c3 < 1608.5) => (faultys = 0) (6/6, 2.6%)
Rule 15. (pcomp_c1 >= 95.65)&(pmax_c_c5 >= 19.45)&(ignition_c4 < 2.65)&(power_c1 >= 1586.5) => (faultys = 0) (5/5, 2.16%)
Rule 16. (power_c1 >= 1610.5)&(ignition_c2 < 2.45)&(ignition_c1 < 2.65)&(power_c7 < 1608.5)&(power_c3 >= 1586.5) => (faultys = 0) (8/8, 3.46%)
Rule 17. (pmax_c_c7 >= 21.05)&(exhaust_gass_temp_c5 >= 375.5)&(power_c6 >= 1596.5)&(pmax_pos_c5 >= 13.55)&(power_c7 < 1606.5) => (faultys = 0) (5/5, 2.16%)
Rule 18. (pmax_c_c1 >= 21.05)&(rpm_c6 >= 96.85)&(power_c7 >= 1588.5) => (faultys = 0) (3/3, 1.3%)
Rule 19. (power_c4 < 1587.5)&(pcomp_c1 < 93.65)&(pmax_c_c3 < 19.45)&(power_c1 < 1608.5) => (faultys = 0) (6/6, 2.6%)
Rule 20. (pmax_c_c4 < 17.35)&(ignition_c4 >= 3.05)&(exhaust_gass_temp_c5 < 371.5) => (faultys = 0) (4/4, 1.73%)
Rule 21. (pmax_pos_c3 >= 14.55)&(pmax_pos_c4 >= 13.85)&(power_c2 >= 1598.5)&(rpm_c2 < 96.55) => (faultys = 0) (6/6, 2.6%)
Rule 22. (pcomp_c7 < 93.15)&(rpm_c5 < 95.65)&(power_c4 < 1599.5) => (faultys = 0) (3/3, 1.3%)
Rule 23. (pcomp_c5 < 93.95)&(exhaust_gass_temp_c3 >= 400.5)&(pmi_c6 >= 14.35)&(power_c2 >= 1590.5)&(pcomp_c4 < 94.95)&(rpm_c2 >= 95.55) => (faultys = 0) (7/7, 3.03%)
Rule 24. (power_c6 < 1586.5)&(power_c5 < 1589.5)&(power_c3 < 1609.5) => (faultys = 0) (3/3, 1.3%)
Rule 25. (power_c2 >= 1608.5)&(pmax_c_c2 < 17.75)&(ignition_c1 < 2.55)&(power_c3 >= 1592.5)&(power_c5 < 1604.5) => (faultys = 0) (5/5, 2.16%)

Rule 26. (exhaust_gass_temp_c2 >= 399.5)&(pcomp_c7 >= 95.25)&(ignition_c3 >= 2.95)&(rpm_c1 >= 96.45) => (faultys = 0) (4/4, 1.73%)
Rule 27. (pcomp_c1 >= 95.45)&(power_c1 < 1590.5)&(pmax_pos_c6 < 13.65)&(power_c5 < 1604.5) => (faultys = 0) (5/5, 2.16%)
Rule 28. (pmax_c_c7 < 17.35)&(ignition_c2 >= 3.15)&(pmi_c7 < 14.25) => (faultys = 0) (5/5, 2.16%)
Rule 29. (exhaust_gass_temp_c2 >= 407.5)&(power_c1 >= 1606.5)&(rpm_c1 < 95.85) => (faultys = 0) (3/3, 1.3%)
Rule 30. (exhaust_gass_temp_c5 >= 408.5)&(pmi_c5 < 14.05) => (faultys = 0) (1/1, 0.43%)
Rule 31. (ignition_c5 < 2.15)&(pmax_pos_c2 < 13.05)&(power_c1 < 1595.5) => (faultys = 0) (2/2, 0.87%)
Rule 32. (power_c3 < 1588.5)&(pmi_c3 < 14.05)&(power_c5 < 1599.5)&(power_c1 >= 1593.5) => (faultys = 0) (3/3, 1.3%)
Rule 33. (pmax_c_c4 >= 21.05)&(exhaust_gass_temp_c2 < 333.5) => (faultys = 0) (2/2, 0.87%)
Rule 34. (exhaust_gass_temp_c6 >= 408.5)&(pmi_c3 >= 14.65) => (faultys = 0) (1/1, 0.43%)
Rule 35. (pcomp_c4 >= 95.65)&(rpm_c3 >= 96.75)&(pcomp_c2 >= 94.95) => (faultys = 0) (2/2, 0.87%)
Rule 36. (pmax_c_c2 < 17.35)&(power_c1 >= 1608.5)&(pmi_c6 >= 14.55) => (faultys = 0) (2/2, 0.87%)
Rule 37. (pmax_c_c4 < 17.25)&(pmi_c3 >= 14.65)&(power_c1 < 1595.5) => (faultys = 0) (1/1, 0.43%)
Rule 38. (pcomp_c3 < 93.25)&(pmax_c_c7 < 18.05)&(pmax_c_c5 >= 19.25)&(power_c1 >= 1588.5) => (faultys = 0) (5/5, 2.16%)
Rule 39. (power_c2 < 1586.5)&(pmax_pos_c5 >= 14.55)&(power_c3 < 1588.5) => (faultys = 0) (1/1, 0.43%)
Rule 40. (power_c3 < 1586.5)&(pmax_c_c1 >= 20.65) => (faultys = 0) (3/3, 1.3%)
Rule 41. (exhaust_gass_temp_c3 >= 404.5)&(rpm_c6 >= 97.05)&(power_c1 >= 1586.5) => (faultys = 0) (2/2, 0.87%)
Rule 42. (exhaust_gass_temp_c6 < 333.5)&(pcomp_c2 < 93.35)&(power_c2 >= 1592.5) => (faultys = 0) (3/3, 1.3%)
Rule 43. (ignition_c5 < 2.15)&(pmax_c_c5 < 17.45)&(power_c4 < 1601.5) => (faultys = 0) (1/1, 0.43%)
Rule 44. (pcomp_c6 >= 95.55)&(exhaust_gass_temp_c6 >= 401.5)&(power_c4 >= 1599.5) => (faultys = 0) (1/1, 0.43%)
Rule 45. (power_c6 >= 1622)&(pmax_c_c2 >= 17.35) => (faultys = 1) (44/45, 89.8%)
Rule 46. (power_c1 >= 1659) => (faultys = 1) (35/35, 71.43%)
Rule 47. (rpm_c1 < 95.25) => (faultys = 2) (29/29, 58%)
Rule 48. (rpm_c3 < 95.2) => (faultys = 2) (25/25, 50%)
Rule 49. (rpm_c6 < 95.4) => (faultys = 2) (22/22, 44%)
Rule 50. (rpm_c7 >= 97.15)&(rpm_c1 >= 96.95) => (faultys = 2) (9/9, 18%)
Rule 51. (pmi_c7 < 12.55)&(pcomp_c7 >= 93.25) => (faultys = 3) (8/8, 16%)
Rule 52. (pmi_c2 < 12.6) => (faultys = 3) (7/7, 14%)
Rule 53. (pmi_c3 < 12.5) => (faultys = 3) (7/7, 14%)
Rule 54. (pmi_c4 < 12.55) => (faultys = 3) (7/7, 14%)
Rule 55. (pmi_c5 < 12.6) => (faultys = 3) (7/7, 14%)
Rule 56. (pmi_c1 < 12.55)&(power_c7 >= 1591.5) => (faultys = 3) (7/7, 14%)
Rule 57. (pmi_c6 < 12.75)&(rpm_c1 < 96.45) => (faultys = 3) (7/7, 14%)
Rule 58. (pmi_c7 >= 15.85)&(exhaust_gass_temp_c7 < 407.5) => (faultys = 4) (8/8, 16%)
Rule 59. (pmi_c1 >= 15.95)&(exhaust_gass_temp_c1 < 406.5) => (faultys = 4) (7/7, 14%)
Rule 60. (pmi_c4 >= 15.75)&(exhaust_gass_temp_c4 < 394.5) => (faultys = 4) (7/7, 14%)
Rule 61. (pmi_c2 >= 15.25)&(exhaust_gass_temp_c2 < 409.5) => (faultys = 4) (7/7, 14%)
Rule 62. (pmi_c3 >= 15.3)&(exhaust_gass_temp_c3 < 402.5) => (faultys = 4) (7/7, 14%)
Rule 63. (pmi_c5 >= 15.25)&(exhaust_gass_temp_c5 < 402.5) => (faultys = 4) (7/7, 14%)
Rule 64. (pmi_c6 >= 15.25)&(exhaust_gass_temp_c6 < 408.5) => (faultys = 4) (7/7, 14%)
Rule 65. (pcomp_c7 < 91.55) => (faultys = 5) (8/8, 16%)
Rule 66. (pcomp_c1 < 91.55) => (faultys = 5) (7/7, 14%)
Rule 67. (pcomp_c2 < 91.3) => (faultys = 5) (7/7, 14%)
Rule 68. (pcomp_c3 < 91.15) => (faultys = 5) (7/7, 14%)
Rule 69. (pcomp_c4 < 91.3) => (faultys = 5) (7/7, 14%)
Rule 70. (pcomp_c5 < 91.45) => (faultys = 5) (7/7, 14%)
Rule 71. (pcomp_c6 < 91.3) => (faultys = 5) (7/7, 14%)
Rule 72. (pcomp_c1 >= 97.7) => (faultys = 6) (8/8, 15.69%)
Rule 73. (pcomp_c2 >= 97.5) => (faultys = 6) (8/8, 15.69%)
Rule 74. (pcomp_c3 >= 97.65) => (faultys = 6) (8/8, 15.69%)
Rule 75. (pcomp_c4 >= 97.75) => (faultys = 6) (7/7, 13.73%)
Rule 76. (pcomp_c5 >= 97.4) => (faultys = 6) (7/7, 13.73%)
Rule 77. (pcomp_c7 >= 97.35) => (faultys = 6) (7/7, 13.73%)
Rule 78. (pcomp_c6 >= 97.4) => (faultys = 6) (6/6, 11.76%)
Rule 79. (pmax_pos_c1 >= 17.45) => (faultys = 7) (7/7, 14.29%)
Rule 80. (pmax_pos_c2 >= 17.45) => (faultys = 7) (7/7, 14.29%)
Rule 81. (pmax_pos_c3 >= 17.4) => (faultys = 7) (7/7, 14.29%)

Rule 82. (pmax_pos_c4 >= 18.05) => (faultys = 7) (7/7, 14.29%)
Rule 83. (pmax_pos_c5 >= 17.35) => (faultys = 7) (7/7, 14.29%)
Rule 84. (pmax_pos_c6 >= 17.6) => (faultys = 7) (7/7, 14.29%)
Rule 85. (pmax_pos_c7 >= 17.65) => (faultys = 7) (7/7, 14.29%)
Rule 86. (exhaust_gass_temp_c7 < 327)&(rpm_c1 < 96.45) => (faultys = 8) (5/5, 14.29%)
Rule 87. (exhaust_gass_temp_c1 < 330.5)&(pmi_c1 < 14.15)&(power_c5 >= 1586.5) => (faultys = 8) (7/7, 20%)
Rule 88. (pmi_c2 in [12.6, 13.6]) => (faultys = 8) (5/5, 14.29%)
Rule 89. (pmi_c3 in [12.5, 13.55]) => (faultys = 8) (5/5, 14.29%)
Rule 90. (pmi_c5 in [12.6, 13.6]) => (faultys = 8) (5/5, 14.29%)
Rule 91. (exhaust_gass_temp_c4 < 325.5)&(pmi_c4 < 13.55) => (faultys = 8) (5/5, 14.29%)
Rule 92. (pmi_c6 < 13.6)&(exhaust_gass_temp_c6 < 329) => (faultys = 8) (5/5, 14.29%)
Rule 93. (pmi_c2 >= 15.85)&(exhaust_gass_temp_c2 >= 409.5) => (faultys = 9) (5/5, 14.29%)
Rule 94. (exhaust_gass_temp_c6 >= 415.5)&(pmi_c6 >= 14.65) => (faultys = 9) (5/5, 14.29%)
Rule 95. (pmi_c3 >= 15.95)&(exhaust_gass_temp_c3 >= 402.5) => (faultys = 9) (5/5, 14.29%)
Rule 96. (exhaust_gass_temp_c7 >= 412.5)&(pmi_c7 >= 15.15) => (faultys = 9) (5/5, 14.29%)
Rule 97. (pmi_c1 >= 15.2)&(exhaust_gass_temp_c1 >= 406.5) => (faultys = 9) (5/5, 14.29%)
Rule 98. (pmi_c4 >= 15.2)&(exhaust_gass_temp_c4 >= 394.5) => (faultys = 9) (5/5, 14.29%)
Rule 99. (pmi_c5 >= 15.25)&(exhaust_gass_temp_c5 >= 402.5) => (faultys = 9) (5/5, 14.29%)
Rule 100. (pmax_c_c1 < 16) => (faultys = 10) (8/8, 15.38%)
Rule 101. (pmax_c_c2 < 15.95) => (faultys = 10) (8/8, 15.38%)
Rule 102. (pmax_c_c3 < 16.2) => (faultys = 10) (8/8, 15.38%)
Rule 103. (pmax_c_c4 < 16.2) => (faultys = 10) (8/8, 15.38%)
Rule 104. (pmax_c_c5 < 16.1) => (faultys = 10) (7/7, 13.46%)
Rule 105. (pmax_c_c6 < 15.95) => (faultys = 10) (7/7, 13.46%)
Rule 106. (pmax_c_c7 < 16.05) => (faultys = 10) (6/6, 11.54%)
Rule 107. (pmax_c_c2 >= 22.6) => (faultys = 11) (7/7, 14.58%)
Rule 108. (pmax_c_c3 >= 22.6) => (faultys = 11) (7/7, 14.58%)
Rule 109. (pmax_c_c4 >= 22.6) => (faultys = 11) (7/7, 14.58%)
Rule 110. (pmax_c_c5 >= 22.5) => (faultys = 11) (7/7, 14.58%)
Rule 111. (pmax_c_c6 >= 22.6) => (faultys = 11) (7/7, 14.58%)
Rule 112. (pmax_c_c7 >= 22.6) => (faultys = 11) (7/7, 14.58%)
Rule 113. (pmax_c_c1 >= 22.5) => (faultys = 11) (6/6, 12.5%)
Rule 114. (ignition_c1 < 1.5) => (faultys = 12) (7/7, 14.29%)
Rule 115. (ignition_c2 < 1.5) => (faultys = 12) (7/7, 14.29%)
Rule 116. (ignition_c3 < 1.45) => (faultys = 12) (7/7, 14.29%)
Rule 117. (ignition_c4 < 1.5) => (faultys = 12) (7/7, 14.29%)
Rule 118. (ignition_c5 < 1.45) => (faultys = 12) (7/7, 14.29%)
Rule 119. (ignition_c6 < 1.5) => (faultys = 12) (7/7, 14.29%)
Rule 120. (ignition_c7 < 1.5) => (faultys = 12) (7/7, 14.29%)
Rule 121. (ignition_c2 >= 4.7) => (faultys = 13) (8/8, 15.38%)
Rule 122. (ignition_c3 >= 4.8) => (faultys = 13) (8/8, 15.38%)
Rule 123. (ignition_c4 >= 4.65) => (faultys = 13) (8/8, 15.38%)
Rule 124. (ignition_c1 >= 4.7) => (faultys = 13) (7/7, 13.46%)
Rule 125. (ignition_c5 >= 5) => (faultys = 13) (7/7, 13.46%)
Rule 126. (ignition_c6 >= 4.8) => (faultys = 13) (7/7, 13.46%)
Rule 127. (ignition_c7 >= 4.65) => (faultys = 13) (7/7, 13.46%)
Rule 128. (exhaust_gass_temp_c2 < 321.5)&(pmi_c2 >= 13.6) => (faultys = 14) (7/7, 14.29%)
Rule 129. (exhaust_gass_temp_c3 < 324.5)&(pmi_c3 >= 13.55) => (faultys = 14) (7/7, 14.29%)
Rule 130. (exhaust_gass_temp_c5 < 324.5)&(pmi_c5 >= 13.6) => (faultys = 14) (7/7, 14.29%)
Rule 131. (exhaust_gass_temp_c6 < 326.5)&(rpm_c3 < 96.75) => (faultys = 14) (7/7, 14.29%)
Rule 132. (exhaust_gass_temp_c1 < 322.5)&(power_c3 >= 1586.5) => (faultys = 14) (6/6, 12.24%)
Rule 133. (exhaust_gass_temp_c4 < 328)&(pmi_c4 >= 13.55) => (faultys = 14) (7/7, 14.29%)
Rule 134. (exhaust_gass_temp_c7 < 329)&(pmi_c7 >= 13.5) => (faultys = 14) (7/7, 14.29%)
Rule 135. (exhaust_gass_temp_c1 < 328.5)&(power_c1 < 1586.5) => (faultys = 14) (1/1, 2.04%)
Rule 136. (exhaust_gass_temp_c2 >= 417.5)&(pmi_c2 < 15.25) => (faultys = 15) (7/7, 14.29%)
Rule 137. (exhaust_gass_temp_c1 >= 414)&(pmi_c1 < 15.2) => (faultys = 15) (7/7, 14.29%)
Rule 138. (exhaust_gass_temp_c4 >= 412.5)&(pmi_c4 < 14.65) => (faultys = 15) (7/7, 14.29%)
Rule 139. (exhaust_gass_temp_c3 >= 411)&(pmi_c3 < 14.65) => (faultys = 15) (7/7, 14.29%)
Rule 140. (exhaust_gass_temp_c5 >= 410)&(pmi_c5 < 15.25) => (faultys = 15) (7/7, 14.29%)
Rule 141. (exhaust_gass_temp_c6 >= 411)&(pmi_c6 < 15.25) => (faultys = 15) (7/7, 14.29%)
Rule 142. (exhaust_gass_temp_c7 >= 410)&(pmi_c7 < 15.15) => (faultys = 15) (7/7, 14.29%)
Rule 143. (rpm_c4 >= 105.35) => (faultys = 16) (9/9, 17.65%)
Rule 144. (rpm_c2 >= 105.5) => (faultys = 16) (8/8, 15.69%)

Rule 145. (rpm_c3 >= 106.5) => (faultys = 16) (8/8, 15.69%)
Rule 146. (rpm_c5 >= 106.3) => (faultys = 16) (8/8, 15.69%)
Rule 147. (rpm_c6 >= 106) => (faultys = 16) (7/7, 13.73%)
Rule 148. (rpm_c1 >= 106.4) => (faultys = 16) (6/6, 11.76%)
Rule 149. (rpm_c7 >= 106.5) => (faultys = 16) (5/5, 9.8%)

Number of Rules: 149

Time taken to build model: 0.56 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	904	90.4 %
Incorrectly Classified Instances	96	9.6 %
Kappa statistic	0.8944	
Mean absolute error	0.0113	
Root mean squared error	0.1063	
Relative absolute error	10.5508 %	
Root relative squared error	45.9507 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,931	0,025	0,919	0,931	0,925	0,902	0,953	0,871	0
	0,959	0,002	0,959	0,959	0,959	0,957	0,979	0,922	1
	0,960	0,001	0,980	0,960	0,970	0,968	0,979	0,942	2
	0,800	0,004	0,909	0,800	0,851	0,846	0,898	0,737	3
	0,740	0,008	0,822	0,740	0,779	0,769	0,866	0,621	4
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	5
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
	0,514	0,012	0,600	0,514	0,554	0,541	0,751	0,326	8
	0,514	0,027	0,409	0,514	0,456	0,437	0,744	0,227	9
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	10
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	12
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
	0,816	0,013	0,769	0,816	0,792	0,781	0,902	0,637	14
	0,796	0,013	0,765	0,796	0,780	0,769	0,892	0,619	15
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	16
Weighted Avg.	0,904	0,009	0,907	0,904	0,905	0,897	0,947	0,844	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
215	2	1	0	0	0	0	0	0	10	0	0	0	0	0	3	0	a = 0
2	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
2	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c = 2
3	0	0	40	0	0	0	0	7	0	0	0	0	0	0	0	0	d = 3
2	0	0	0	37	0	0	0	0	11	0	0	0	0	0	0	0	e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h = 7
1	0	0	4	0	0	0	0	18	0	0	0	0	0	12	0	0	i = 8
0	0	0	0	8	0	0	0	0	18	0	0	0	0	0	9	0	j = 9
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k = 10
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l = 11
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m = 12
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n = 13
4	0	0	0	0	0	0	0	5	0	0	0	0	0	40	0	0	o = 14
5	0	0	0	0	0	0	0	0	5	0	0	0	0	0	39	0	p = 15

0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 51 | q = 16

Scheme:weka.classifiers.lazy.LWL -U 0 -K -1 -A "weka.core.neighboursearch.LinearNNSearch -A \"weka.core.EuclideanDistance -R first-last\"" -W weka.classifiers.trees.J48 -- -R -N 7 -Q 1 -M 2 -J
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Locally weighted learning
==========
Using classifier: weka.classifiers.trees.J48
Using linear weighting kernels
Using all neighbours

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**901**	**90.1 %**
Incorrectly Classified Instances	**99**	**9.9 %**
Kappa statistic	0.8905	
Mean absolute error	0.012	
Root mean squared error	0.0953	
Relative absolute error	11.2286 %	
Root relative squared error	41.1928 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,978	0,040	0,879	0,978	0,926	0,905	0,971	0,882	0
0,980	0,002	0,960	0,980	0,970	0,968	0,987	0,903	1
0,900	0,000	1,000	0,900	0,947	0,946	0,960	0,924	2
0,980	0,012	0,817	0,980	0,891	0,889	0,986	0,867	3
0,900	0,011	0,818	0,900	0,857	0,850	0,958	0,880	4
0,860	0,000	1,000	0,860	0,925	0,924	0,930	0,867	5
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
1,000	0,001	0,980	1,000	0,990	0,989	0,999	0,980	7
0,229	0,003	0,727	0,229	0,348	0,397	0,875	0,519	8
0,257	0,005	0,643	0,257	0,367	0,394	0,971	0,550	9
0,885	0,000	1,000	0,885	0,939	0,938	0,942	0,891	10
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
0,857	0,000	1,000	0,857	0,923	0,922	0,928	0,864	12
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
0,857	0,018	0,712	0,857	0,778	0,769	0,933	0,784	14
0,959	0,017	0,746	0,959	0,839	0,837	0,986	0,928	15
1,000	0,003	0,944	1,000	0,971	0,970	0,999	0,960	16
Weighted Avg.	**0,901**	**0,013**	**0,899**	**0,901**	**0,890**	**0,886**	**0,968**	**0,883**

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
226	2	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	a = 0
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
2	0	45	0	0	0	0	0	0	0	0	0	0	0	0	0	3	c = 2
0	0	0	49	0	0	0	0	1	0	0	0	0	0	0	0	0	d = 3
1	0	0	0	45	0	0	0	0	4	0	0	0	0	0	0	0	e = 4
7	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6

```
 0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0 |  h = 7
 1  0  0 11  0  0  0  0  8  0  0  0  0  0 15  0  0 |  i = 8
 0  0  0  0 10  0  0  0  0  9  0  0  0  0  0 16  0 |  j = 9
 6  0  0  0  0  0  0  0  0  0 46  0  0  0  0  0  0 |  k = 10
 0  0  0  0  0  0  0  0  0  0  0 48  0  0  0  0  0 |  l = 11
 7  0  0  0  0  0  0  0  0  0  0  0 42  0  0  0  0 |  m = 12
 0  0  0  0  0  0  0  0  0  0  0  0  0 52  0  0  0 |  n = 13
 6  0  0  0  0  0  0  0  1  0  0  0  0  0 42  0  0 |  o = 14
 0  0  0  0  0  0  0  0  0  1  0  0  0  0  1 47  0 |  p = 15
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 51 |  q = 16
```

Scheme: weka.classifiers.functions.SMO -C 15.0 -L 0.001 -P 1.0E-12 -N 1 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.RBFKernel -G 0.01 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

SMO

Kernel used:
RBF kernel: K(x,y) = e^-(0.01* <x-y,x-y>^2)

Classifier for classes: 0, 1

BinarySMO

Number of support vectors: 58

Number of kernel evaluations: 4515 (87.114% cached)

Time taken to build model: 1.34 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**755**	**75.5 %**
Incorrectly Classified Instances	245	24.5 %
Kappa statistic	0.7265	
Mean absolute error	0.1042	
Root mean squared error	0.2239	
Relative absolute error	97.3706 %	
Root relative squared error	96.8117 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	MCC	ROC Area	PRC Area	**Class**
0,870	0,134	0,661	0,870	0,751	0,675	0,930	0,692	0
0,980	0,000	1,000	0,980	0,990	0,989	0,988	0,977	1
0,760	0,004	0,905	0,760	0,826	0,821	0,955	0,736	2
0,840	0,006	0,875	0,840	0,857	0,850	0,995	0,851	3
0,560	0,022	0,571	0,560	0,566	0,543	0,965	0,507	4
0,920	0,003	0,939	0,920	0,929	0,926	0,996	0,929	5
0,902	0,000	1,000	0,902	0,948	0,947	0,997	0,960	6
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
0,543	0,012	0,613	0,543	0,576	0,562	0,960	0,458	8
0,400	0,015	0,500	0,400	0,444	0,429	0,958	0,403	9

	0,481	0,012	0,694	0,481	0,568	0,559	0,872	0,426	10
	0,667	0,011	0,762	0,667	0,711	0,699	0,944	0,597	11
	0,816	0,012	0,784	0,816	0,800	0,790	0,980	0,727	12
	0,981	0,000	1,000	0,981	0,990	0,990	0,997	0,986	13
	0,245	0,017	0,429	0,245	0,312	0,298	0,874	0,230	14
	0,286	0,030	0,326	0,286	0,304	0,272	0,876	0,208	15
	0,980	0,005	0,909	0,980	0,943	0,941	0,997	0,902	16
Weighted Avg.	**0,755**	**0,038**	**0,752**	**0,755**	**0,747**	**0,722**	**0,953**	**0,692**	

=== Confusion Matrix ===

```
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q  <-- classified as
201  0  0  0  0  0  0  0  1  0  1  2  3  0 10 13  0 |  a = 0
  0 48  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |  b = 1
  1  0 38  0  1  0  0  0  0  0  0  2  2  0  1  0  5 |  c = 2
  0  0  0 42  0  0  0  0  8  0  0  0  0  0  0  0  0 |  d = 3
  2  0  0  0 28  0  0  0  0 14  4  1  0  0  1  0  0 |  e = 4
  0  0  1  0  0 46  0  0  1  0  1  0  0  0  0  1  0 |  f = 5
  0  0  0  0  1  0 46  0  0  0  0  1  1  0  0  2  0 |  g = 6
  0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0 |  h = 7
  5  0  0  6  0  1  0  0 19  0  2  0  1  0  1  0  0 |  i = 8
  2  0  1  0 17  0  0  0  0 14  0  0  1  0  0  0  0 |  j = 9
 15  0  0  0  1  0  0  0  2  0 25  1  2  0  2  4  0 |  k = 10
  6  0  0  0  1  1  0  0  0  0  2 32  1  0  0  5  0 |  l = 11
  7  0  0  0  0  0  0  0  0  0  0  0 40  0  1  1  0 |  m = 12
  0  0  0  0  0  1  0  0  0  0  0  0  0 51  0  0  0 |  n = 13
 31  0  0  0  0  0  0  0  0  0  1  2  0  0 12  3  0 |  o = 14
 34  0  0  0  0  0  0  0  0  0  0  0  1  0  0  0 14 |  p = 15
  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0 50 |  q = 16
```

Scheme: weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.1 -N 3000 -V 0 -S 0 -E 20 -H a
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

Sigmoid Node 0
Inputs Weights
Threshold 10.197017305961296
Node 17 -1.7164704368065153
Node 18 2.334523379789572
Node 19 -12.834711682808514
Node 20 -0.10302284210054075
Node 21 -9.036521541537509
Node 22 -8.427135782681015
Node 23 5.378700266667454
Node 24 -9.306555578878024
Node 25 -9.47203621858671
Node 26 5.9790513845585185
Node 27 0.658033954464148
Node 28 -7.1650621400157295
Node 29 0.1894051816132461
Node 30 -5.213735936507506

Time taken to build model: 333.53 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances **544** **54.4 %**
Incorrectly Classified Instances **456** **45.6 %**
Kappa statistic 0.4893
Mean absolute error 0.0583
Root mean squared error 0.212
Relative absolute error 54.5004 %
Root relative squared error 91.649 %
Total Number of Instances 1000

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,680	0,212	0,491	0,680	0,570	0,423	0,807	0,501	0
	0,918	0,002	0,957	0,918	0,938	0,935	0,978	0,952	1
	0,220	0,021	0,355	0,220	0,272	0,250	0,520	0,155	2
	0,540	0,023	0,551	0,540	0,545	0,522	0,946	0,501	3
	0,460	0,028	0,460	0,460	0,460	0,432	0,893	0,401	4
	0,720	0,016	0,706	0,720	0,713	0,698	0,972	0,700	5
	0,686	0,021	0,636	0,686	0,660	0,642	0,954	0,643	6
	0,878	0,009	0,827	0,878	0,851	0,844	0,994	0,908	7
	0,200	0,017	0,304	0,200	0,241	0,225	0,843	0,205	8
	0,286	0,020	0,345	0,286	0,313	0,291	0,891	0,275	9
	0,250	0,028	0,325	0,250	0,283	0,251	0,724	0,226	10
	0,250	0,024	0,343	0,250	0,289	0,263	0,791	0,254	11
	0,510	0,018	0,595	0,510	0,549	0,530	0,953	0,549	12
	0,673	0,012	0,761	0,673	0,714	0,701	0,982	0,811	13
	0,245	0,027	0,316	0,245	0,276	0,246	0,728	0,177	14
	0,163	0,029	0,222	0,163	0,188	0,155	0,715	0,152	15
	0,882	0,012	0,804	0,882	0,841	0,833	0,992	0,886	16
Weighted Avg.	**0,544**	**0,064**	**0,529**	**0,544**	**0,530**	**0,481**	**0,853**	**0,499**	

=== Confusion Matrix ===

```
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q  <-- classified as
157  1  5  3  2  4  6  2  4  1 11  7  6  3  7 10  2 |  a = 0
  1 45  0  0  1  0  0  0  0  0  0  0  1  0  0  0  1 |  b = 1
 18  0 11  3  1  2  1  2  0  1  1  1  1  1  1  1  5 |  c = 2
  5  1  0 27  0  1  1  0  7  0  0  3  1  2  1  1  0 |  d = 3
  8  0  1  0 23  0  1  0  0 11  1  2  1  0  1  1  0 |  e = 4
  9  0  0  0  0 36  0  1  1  0  1  0  0  0  0  2  0 |  f = 5
  6  0  0  0  2  0 35  0  0  0  1  2  1  2  2  0  0 |  g = 6
  1  0  0  0  0  0  0 43  0  1  2  1  0  0  0  1  0 |  h = 7
  8  0  2 11  0  2  3  0  7  0  0  1  0  0  1  0  0 |  i = 8
  3  0  0  0 17  0  1  1  0 10  2  0  0  1  0  0  0 |  j = 9
 17  0  1  3  1  2  0  1  2  2 13  1  0  1  5  3  0 |  k = 10
 17  0  3  0  0  2  4  0  0  0  1 12  3  0  2  3  1 |  l = 11
 14  0  0  0  2  0  1  0  0  0  1  0 25  0  3  2  1 |  m = 12
  8  0  0  1  0  1  0  0  0  0  2  0  0 35  2  2  1 |  n = 13
 23  0  2  0  0  1  0  1  2  1  3  2  0  0 12  2  0 |  o = 14
 24  0  4  1  1  0  1  1  0  1  0  3  3  1  1  8  0 |  p = 15
  1  0  2  0  0  0  1  0  0  1  1  0  0  0  0  0 45 |  q = 16
```

=== Run information ===

Scheme: weka.classifiers.trees.J48 -R -N 8 -Q 1 -M 2 -J
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

Number of Leaves : 100

Size of the tree : 199

Time taken to build model: 0.58 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**891**	**89.1 %**
Incorrectly Classified Instances	**109**	**10.9 %**
Kappa statistic	0.8795	
Mean absolute error	0.013	
Root mean squared error	0.0964	
Relative absolute error	12.1102 %	
Root relative squared error	41.6634 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

	TP Rate	**FP Rate**	**Precision**	**Recall**	**F-Measure**	**MCC**	**ROC Area**	**PRC Area**	**Class**
	0,983	0,035	0,894	0,983	0,936	0,917	0,978	0,891	0
	0,980	0,002	0,960	0,980	0,970	0,968	0,988	0,920	1
	0,920	0,001	0,979	0,920	0,948	0,946	0,960	0,924	2
	0,980	0,009	0,845	0,980	0,907	0,905	0,986	0,856	3
	0,800	0,023	0,645	0,800	0,714	0,702	0,958	0,738	4
	0,860	0,000	1,000	0,860	0,925	0,924	0,930	0,867	5
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
	1,000	0,001	0,980	1,000	0,990	0,989	0,999	0,980	7
	0,286	0,012	0,455	0,286	0,351	0,342	0,841	0,435	8
	0,143	0,008	0,385	0,143	0,208	0,218	0,974	0,439	9
	0,885	0,000	1,000	0,885	0,939	0,938	0,942	0,891	10
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
	0,857	0,000	1,000	0,857	0,923	0,922	0,929	0,864	12
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
	0,755	0,017	0,698	0,755	0,725	0,711	0,972	0,764	14
	0,980	0,008	0,857	0,980	0,914	0,912	0,989	0,960	15
	0,980	0,003	0,943	0,980	0,962	0,960	0,998	0,938	16
Weighted Avg.	**0,891**	**0,012**	**0,881**	**0,891**	**0,881**	**0,875**	**0,971**	**0,871**	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
227	2	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	a = 0
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
1	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	3	c = 2
0	0	0	49	0	0	0	0	1	0	0	0	0	0	0	0	0	d = 3
2	0	0	0	40	0	0	0	0	8	0	0	0	0	0	0	0	e = 4
7	0	0	0	0	43	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h = 7
1	0	0	9	0	0	0	0	10	0	0	0	0	0	15	0	0	i = 8
0	0	0	0	22	0	0	0	0	5	0	0	0	0	0	8	0	j = 9
6	0	0	0	0	0	0	0	0	0	46	0	0	0	0	0	0	k = 10
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l = 11
7	0	0	0	0	0	0	0	0	0	0	0	42	0	0	0	0	m = 12
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n = 13
2	0	0	0	0	0	0	0	10	0	0	0	0	0	37	0	0	o = 14
0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	48	0	p = 15
0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	50	q = 16

=== Run information ===

Scheme: weka.classifiers.trees.SimpleCart -M 2.0 -N 9 -A -C 1.0 -S 1
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

CART Decision Tree

Number of Leaf Nodes: 104

Size of the Tree: 207

Time taken to build model: 5.73 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**955**	**95.5 %**
Incorrectly Classified Instances	**45**	**4.5 %**
Kappa statistic	0.9504	
Mean absolute error	0.0056	
Root mean squared error	0.0714	
Relative absolute error	5.2753 %	
Root relative squared error	30.865 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,983	0,014	0,954	0,983	0,968	0,958	0,992	0,965	0
0,959	0,002	0,959	0,959	0,959	0,957	0,978	0,898	1
0,960	0,000	1,000	0,960	0,980	0,979	0,989	0,974	2
0,900	0,006	0,882	0,900	0,891	0,885	0,947	0,799	3
0,880	0,003	0,936	0,880	0,907	0,903	0,939	0,848	4
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	5
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
0,686	0,007	0,774	0,686	0,727	0,719	0,853	0,570	8
0,857	0,006	0,833	0,857	0,845	0,839	0,940	0,742	9
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	10
0,979	0,000	1,000	0,979	0,989	0,989	0,990	0,980	11
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	12
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
0,959	0,007	0,870	0,959	0,913	0,909	0,978	0,919	14
0,918	0,002	0,957	0,918	0,938	0,935	0,958	0,882	15
0,941	0,001	0,980	0,941	0,960	0,958	0,999	0,962	16
Weighted Avg.	**0,955**	**0,005**	**0,955**	**0,955**	**0,955**	**0,951**	**0,980**	**0,931**

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
227	2	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	a = 0
2	47	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
1	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	1	c = 2
0	0	0	45	0	0	0	0	5	0	0	0	0	0	0	0	0	d = 3
0	0	0	0	44	0	0	0	0	6	0	0	0	0	0	0	0	e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6

```
  0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0 |  h = 7
  2  0  0  6  0  0  0  0 24  0  0  0  0  0  3  0  0 |  i = 8
  0  0  0  0  3  0  0  0  0 30  0  0  0  0  0  2  0 |  j = 9
  0  0  0  0  0  0  0  0  0  0 52  0  0  0  0  0  0 |  k = 10
  0  0  0  0  0  0  0  0  0  0  0 47  0  0  1  0  0 |  l = 11
  0  0  0  0  0  0  0  0  0  0  0  0 49  0  0  0  0 |  m = 12
  0  0  0  0  0  0  0  0  0  0  0  0  0 52  0  0  0 |  n = 13
  1  0  0  0  0  0  0  0  1  0  0  0  0  0 47  0  0 |  o = 14
  2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  2 45  0 |  p = 15
  3  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 48 |  q = 16
```

Scheme: weka.classifiers.meta.AdaBoostM1 -P 90 -S 1 -I 60 -W weka.classifiers.trees.J48 -- -R -N 8 -Q 1 -M 2 -J

Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
J48 pruned tree
Time taken to build model: 24.41 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	951	95.1 %
Incorrectly Classified Instances	49	4.9 %
Kappa statistic	0.9458	
Mean absolute error	0.0057	
Root mean squared error	0.0754	
Relative absolute error	5.3713 %	
Root relative squared error	32.6143 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,987	0,035	0,894	0,987	0,938	0,920	0,983	0,912	0
	0,959	0,002	0,959	0,959	0,959	0,957	1,000	0,996	1
	0,940	0,000	1,000	0,940	0,969	0,968	1,000	1,000	2
	0,960	0,006	0,889	0,960	0,923	0,920	0,998	0,926	3
	0,980	0,001	0,980	0,980	0,980	0,979	1,000	0,999	4
	0,860	0,000	1,000	0,860	0,925	0,924	0,948	0,889	5
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
	1,000	0,001	0,980	1,000	0,990	0,989	0,999	0,964	7
	0,629	0,002	0,917	0,629	0,746	0,752	0,923	0,742	8
	0,971	0,001	0,971	0,971	0,971	0,970	1,000	0,998	9
	0,885	0,000	1,000	0,885	0,939	0,938	0,960	0,920	10
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
	0,857	0,000	1,000	0,857	0,923	0,922	0,887	0,866	12
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
	0,918	0,006	0,882	0,918	0,900	0,895	0,990	0,852	14
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	15
	1,000	0,003	0,944	1,000	0,971	0,970	1,000	0,998	16
Weighted Avg.	0,951	0,009	0,953	0,951	0,950	0,945	0,982	0,941	

=== Confusion Matrix ===

```
  a  b  c  d  e  f  g  h  i  j  k  l  m  n  o  p  q  <-- classified as
228  2  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0 |  a = 0
  2 47  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 |  b = 1
  0  0 47  0  0  0  0  0  0  0  0  0  0  0  0  0  3 |  c = 2
  0  0  0 48  0  0  0  0  2  0  0  0  0  0  0  0  0 |  d = 3
  0  0  0  0 49  0  0  0  0  1  0  0  0  0  0  0  0 |  e = 4
  7  0  0  0  0 43  0  0  0  0  0  0  0  0  0  0  0 |  f = 5
```

```
 0  0  0  0  0  0 51  0  0  0  0  0  0  0  0  0  0 |  g = 6
 0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0 |  h = 7
 1  0  0  6  0  0  0  0 22  0  0  0  0  0  6  0  0 |  i = 8
 0  0  0  0  1  0  0  0  0 34  0  0  0  0  0  0  0 |  j = 9
 6  0  0  0  0  0  0  0  0  0 46  0  0  0  0  0  0 |  k = 10
 0  0  0  0  0  0  0  0  0  0  0 48  0  0  0  0  0 |  l = 11
 7  0  0  0  0  0  0  0  0  0  0  0 42  0  0  0  0 |  m = 12
 0  0  0  0  0  0  0  0  0  0  0  0  0 52  0  0  0 |  n = 13
 4  0  0  0  0  0  0  0  0  0  0  0  0  0 45  0  0 |  o = 14
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 49  0 |  p = 15
 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 51 |  q = 16
```

=== Run information ===

Scheme: weka.classifiers.meta.AdaBoostM1 -P 100 -S 1 -I 60 -W weka.classifiers.trees.SimpleCart -- -M 2.0 -N 9 -A -C 1.0 -S 1
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

AdaBoostM1: Base classifiers and their weights:

CART Decision Tree

Number of Leaf Nodes: 101

Size of the Tree: 201

Weight: 2.86

Number of performed Iterations: 60

Time taken to build model: 238.99 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**966**	**96.6 %**
Incorrectly Classified Instances	34	3.4 %
Kappa statistic	0.9625	
Mean absolute error	0.0041	
Root mean squared error	0.0608	
Relative absolute error	3.8416 %	
Root relative squared error	26.2754 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,991	0,017	0,946	0,991	0,968	0,959	0,996	0,991	0
0,980	0,002	0,960	0,980	0,970	0,968	0,999	0,972	1
0,980	0,000	1,000	0,980	0,990	0,989	1,000	1,000	2
0,960	0,005	0,906	0,960	0,932	0,929	0,999	0,988	3
0,920	0,006	0,885	0,920	0,902	0,897	0,999	0,978	4
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	5
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7

	0,771	0,002	0,931	0,771	0,844	0,843	0,996	0,927	8
	0,743	0,002	0,929	0,743	0,825	0,825	0,997	0,937	9
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	10
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	12
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
	0,898	0,001	0,978	0,898	0,936	0,934	0,990	0,954	14
	0,959	0,003	0,940	0,959	0,949	0,947	0,992	0,975	15
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	16
Weighted Avg.	**0,966**	**0,005**	**0,966**	**0,966**	**0,965**	**0,962**	**0,998**	**0,987**	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a = 0
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
1	0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c = 2
1	0	0	48	0	0	0	0	1	0	0	0	0	0	0	0	0	d = 3
2	0	0	0	46	0	0	0	0	2	0	0	0	0	0	0	0	e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h = 7
2	0	0	5	0	0	0	0	27	0	0	0	0	0	1	0	0	i = 8
0	0	0	0	6	0	0	0	0	26	0	0	0	0	0	3	0	j = 9
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k = 10
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l = 11
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m = 12
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n = 13
4	0	0	0	0	0	0	0	1	0	0	0	0	0	44	0	0	o = 14
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	47	0	p = 15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	q = 16

Scheme: weka.classifiers.meta.MultiBoostAB -C 3 -P 100 -S 1 -I 60 -W weka.classifiers.rules.MODLEM -- -RT 1 -CM 1 -CS 6 -AS 0
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

MultiBoostAB: Base classifiers and their weights:

Rule 1. (pmax_pos_c7 < 12.95)&(pcomp_c6 >= 95.25)&(pmi_c5 >= 14.35)&(rpm_c1 >= 95.25) => (faultys = 0) (16/16, 7.11%)
Rule 2. (pmax_pos_c2 >= 14.35)&(power_c2 < 1591.5)&(rpm_c4 >= 95.95)&(exhaust_gass_temp_c6 < 395.5)&(power_c5 >= 1588.5) => (faultys = 0) (21/21, 9.33%)
Rule 3. (pcomp_c7 >= 95.65)&(ignition_c3 >= 3.05)&(power_c1 < 1604.5) => (faultys = 0) (13/13, 5.78%)
Rule 4. (pmax_pos_c3 >= 14.55)&(exhaust_gass_temp_c4 >= 358.5)&(power_c7 < 1600.5)&(pcomp_c1 >= 93.85) => (faultys = 0) (12/12, 5.33%)
Rule 5. (exhaust_gass_temp_c7 >= 387.5)&(exhaust_gass_temp_c3 < 343.5)&(rpm_c2 < 96.55)&(pcomp_c7 >= 93.25)&(power_c5 >= 1593.5) => (faultys = 0) (16/16, 7.11%)
Rule 6. (power_c5 < 1586.5)&(pmax_pos_c1 >= 14.35)&(power_c1 < 1601.5) => (faultys = 0) (9/9, 4%)
Rule 7. (exhaust_gass_temp_c2 >= 405.5)&(pmax_c_c5 < 18.55)&(power_c5 < 1604.5)&(pmi_c3 >= 14.25) => (faultys = 0) (15/15, 6.67%)
Rule 8. (exhaust_gass_temp_c6 < 331.5)&(power_c5 < 1591.5)&(pmax_c_c6 >= 19.15) => (faultys = 0) (9/9, 4%)
Rule 9. (pmax_c_c4 < 18.05)&(pmax_pos_c2 >= 14.45)&(pmi_c2 >= 14.25) => (faultys = 0) (14/14, 6.22%)
Rule 10. (pmax_c_c1 >= 21.05)&(power_c2 >= 1606.5) => (faultys = 0) (5/5, 2.22%)
Rule 11. (pmax_pos_c4 < 12.85)&(exhaust_gass_temp_c3 < 344.5)&(power_c1 < 1599.5) => (faultys = 0) (6/6, 2.67%)
Rule 12. (pmax_pos_c7 >= 14.45)&(exhaust_gass_temp_c3 >= 383.5)&(pmax_c_c3 < 20.05)&(rpm_c5 < 96.85)&(pcomp_c7 < 94.85) => (faultys = 0) (13/13, 5.78%)

Rule 13. (pcomp_c7 < 93.15)&(pmi_c1 < 14.35)&(pmax_pos_c7 >= 13.25)&(pmax_c_c1 >= 17.65) => (faultys = 0) (8/8, 3.56%)
Rule 14. (exhaust_gass_temp_c2 < 332.5)&(pmax_c_c4 >= 19.85)&(power_c7 < 1594.5) => (faultys = 0) (7/7, 3.11%)
Rule 125. (rpm_c6 >= 106.35) => (faultys = 16) (5/5, 12.5%)
Rule 126. (rpm_c5 >= 105.25) => (faultys = 16) (4/4, 10%)
Rule 127. (rpm_c1 >= 107.4) => (faultys = 16) (2/2, 5%)
Rule 128. (rpm_c2 >= 105.2) => (faultys = 16) (1/1, 2.5%)

Number of Rules: 128

Weight: 2.66

Number of performed Iterations: 60

Time taken to build model: 13.73 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**925**	**92.5 %**
Incorrectly Classified Instances	75	7.5 %
Kappa statistic	0.9173	
Mean absolute error	0.0088	
Root mean squared error	0.0928	
Relative absolute error	8.1776 %	
Root relative squared error	40.1155 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,991	0,016	0,950	0,991	0,970	0,962	0,997	0,985	0
0,980	0,002	0,960	0,980	0,970	0,968	0,999	0,978	1
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	2
0,940	0,003	0,940	0,940	0,940	0,937	0,999	0,968	3
0,860	0,019	0,705	0,860	0,775	0,766	0,988	0,815	4
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	5
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
0,486	0,004	0,810	0,486	0,607	0,617	0,983	0,758	8
0,229	0,012	0,400	0,229	0,291	0,284	0,964	0,441	9
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	10
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	12
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
0,857	0,016	0,737	0,857	0,792	0,783	0,987	0,806	14
0,796	0,009	0,813	0,796	0,804	0,794	0,989	0,873	15
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	16
Weighted Avg.	**0,925**	**0,007**	**0,919**	**0,925**	**0,919**	**0,916**	**0,996**	**0,941**

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	| a = 0
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	| b = 1
0	0	50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	| c = 2
0	0	0	47	0	0	0	0	3	0	0	0	0	0	0	0	0	| d = 3
2	0	0	0	43	0	0	0	0	5	0	0	0	0	0	0	0	| e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	| f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	| g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	| h = 7

```
  0   0   0   3   0   0   0   0 17   0   0   0   0   0 15   0   0 |  i = 8
  0   0   0   0 18   0   0   0   0   8   0   0   0   0   0   9   0 |  j = 9
  0   0   0   0   0   0   0   0   0   0 52   0   0   0   0   0   0 |  k = 10
  0   0   0   0   0   0   0   0   0   0   0 48   0   0   0   0   0 |  l = 11
  0   0   0   0   0   0   0   0   0   0   0   0 49   0   0   0   0 |  m = 12
  0   0   0   0   0   0   0   0   0   0   0   0   0 52   0   0   0 |  n = 13
  6   0   0   0   0   0   0   0   1   0   0   0   0   0 42   0   0 |  o = 14
  3   0   0   0   0   0   0   0   0   7   0   0   0   0   0 39   0 |  p = 15
  0   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 51 |  q = 16
```

Scheme: weka.classifiers.meta.MultiBoostAB -C 6 -P 90 -S 1 -I 60 -W weka.classifiers.trees.J48 -- -R -N 8 -Q 1 -M 2 -J
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57

Time taken to build model: 16.3 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	952	95.2 %
Incorrectly Classified Instances	48	4.8 %
Kappa statistic	0.9469	
Mean absolute error	0.0061	
Root mean squared error	0.0748	
Relative absolute error	5.701 %	
Root relative squared error	32.3411 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,987	0,035	0,894	0,987	0,938	0,920	0,987	0,945	0
	0,980	0,002	0,960	0,980	0,970	0,968	1,000	0,995	1
	0,960	0,000	1,000	0,960	0,980	0,979	1,000	1,000	2
	0,960	0,005	0,906	0,960	0,932	0,929	0,997	0,895	3
	0,960	0,000	1,000	0,960	0,980	0,979	0,999	0,987	4
	0,860	0,000	1,000	0,860	0,925	0,924	0,940	0,893	5
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
	1,000	0,001	0,980	1,000	0,990	0,989	0,999	0,973	7
	0,657	0,002	0,920	0,657	0,767	0,771	0,909	0,733	8
	0,943	0,002	0,943	0,943	0,943	0,941	0,999	0,937	9
	0,885	0,000	1,000	0,885	0,939	0,938	0,974	0,931	10
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
	0,857	0,000	1,000	0,857	0,923	0,922	0,940	0,874	12
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
	0,918	0,006	0,882	0,918	0,900	0,895	0,981	0,832	14
	1,000	0,002	0,961	1,000	0,980	0,979	1,000	0,999	15
	1,000	0,001	0,981	1,000	0,990	0,990	1,000	1,000	16
Weighted Avg.	0,952	0,009	0,954	0,952	0,951	0,946	0,985	0,945	

=== Confusion Matrix ===

```
  a   b   c   d   e   f   g   h   i   j   k   l   m   n   o   p   q   <-- classified as
228   2   0   0   0   0   0   1   0   0   0   0   0   0   0   0   0 |  a = 0
  1  48   0   0   0   0   0   0   0   0   0   0   0   0   0   0   0 |  b = 1
  1   0  48   0   0   0   0   0   0   0   0   0   0   0   0   0   1 |  c = 2
  0   0   0  48   0   0   0   0   2   0   0   0   0   0   0   0   0 |  d = 3
  0   0   0   0  48   0   0   0   0   2   0   0   0   0   0   0   0 |  e = 4
  7   0   0   0   0  43   0   0   0   0   0   0   0   0   0   0   0 |  f = 5
```

```
  0  0  0  0  0  0 51  0  0  0  0  0  0  0  0  0  0 |  g = 6
  0  0  0  0  0  0  0 49  0  0  0  0  0  0  0  0  0 |  h = 7
  1  0  0  5  0  0  0  0 23  0  0  0  0  0  6  0  0 |  i = 8
  0  0  0  0  0  0  0  0  0 33  0  0  0  0  0  2  0 |  j = 9
  6  0  0  0  0  0  0  0  0  0 46  0  0  0  0  0  0 |  k = 10
  0  0  0  0  0  0  0  0  0  0  0 48  0  0  0  0  0 |  l = 11
  7  0  0  0  0  0  0  0  0  0  0  0 42  0  0  0  0 |  m = 12
  0  0  0  0  0  0  0  0  0  0  0  0  0 52  0  0  0 |  n = 13
  4  0  0  0  0  0  0  0  0  0  0  0  0  0 45  0  0 |  o = 14
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 49  0 |  p = 15
  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0 51 |  q = 16
```

=== Run information ===

Scheme: weka.classifiers.meta.MultiBoostAB -C 3 -P 100 -S 1 -I 60 -W weka.classifiers.trees.SimpleCart -- -M 2.0 -N 9 -A -C 1.0 -S 1
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57
Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

MultiBoostAB: Base classifiers and their weights:

CART Decision Tree

Number of Leaf Nodes: 89

Size of the Tree: 177

Weight: 1.49

Number of performed Iterations: 60

Time taken to build model: 224.86 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**960**	**96 %**
Incorrectly Classified Instances	40	4 %
Kappa statistic	0.9558	
Mean absolute error	0.0047	
Root mean squared error	0.0651	
Relative absolute error	4.3911 %	
Root relative squared error	28.1395 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,991	0,025	0,923	0,991	0,956	0,943	0,995	0,990	0
1,000	0,002	0,961	1,000	0,980	0,979	0,999	0,972	1
0,960	0,000	1,000	0,960	0,980	0,979	1,000	1,000	2
0,980	0,007	0,875	0,980	0,925	0,922	0,999	0,978	3
0,920	0,004	0,920	0,920	0,920	0,916	0,999	0,987	4
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	5
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
0,629	0,000	1,000	0,629	0,772	0,788	0,993	0,912	8

	0,829	0,002	0,935	0,829	0,879	0,876	0,997	0,945	9
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	10
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	12
	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
	0,878	0,004	0,915	0,878	0,896	0,891	0,997	0,960	14
	0,918	0,002	0,957	0,918	0,938	0,935	0,983	0,969	15
	0,961	0,000	1,000	0,961	0,980	0,979	1,000	1,000	16
Weighted Avg.	**0,960**	**0,007**	**0,962**	**0,960**	**0,959**	**0,955**	**0,997**	**0,986**	

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	a = 0
0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	b = 1
2	0	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	c = 2
1	0	0	49	0	0	0	0	0	0	0	0	0	0	0	0	0	d = 3
2	0	0	0	46	0	0	0	0	2	0	0	0	0	0	0	0	e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	h = 7
2	0	0	7	0	0	0	0	22	0	0	0	0	0	4	0	0	i = 8
0	0	0	0	4	0	0	0	0	29	0	0	0	0	0	2	0	j = 9
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	k = 10
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	l = 11
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	m = 12
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	n = 13
6	0	0	0	0	0	0	0	0	0	0	0	0	0	43	0	0	o = 14
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	45	0	p = 15
2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	49	q = 16

Scheme: weka.classifiers.meta.Decorate -E 15 -R 3.0 -S 1 -I 60 -W weka.classifiers.trees.J48 -- -R -N 8 -Q 1 -M 2 -J
Relation: MAN_7S60MC_C7
Instances: 1000
Attributes: 57

Number of Leaves : 213

Size of the tree : 425

Number of classifier in the ensemble: 15

Time taken to build model: 84.94 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances	**952**	**95.2 %**
Incorrectly Classified Instances	48	4.8 %
Kappa statistic	0.9471	
Mean absolute error	0.0416	
Root mean squared error	0.1094	
Relative absolute error	38.8852 %	
Root relative squared error	47.324 %	
Total Number of Instances	1000	

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class

0,991	0,010	0,966	0,991	0,979	0,972	0,999	0,998	0
0,980	0,002	0,960	0,980	0,970	0,968	1,000	0,994	1
0,920	0,000	1,000	0,920	0,958	0,957	0,999	0,971	2
0,960	0,013	0,800	0,960	0,873	0,869	0,996	0,908	3
0,860	0,012	0,796	0,860	0,827	0,818	0,995	0,914	4
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	5
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	6
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	7
0,571	0,002	0,909	0,571	0,702	0,713	0,988	0,739	8
0,686	0,007	0,774	0,686	0,727	0,719	0,991	0,766	9
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	10
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	11
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	12
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	13
0,939	0,002	0,958	0,939	0,948	0,946	1,000	0,997	14
0,939	0,000	1,000	0,939	0,968	0,967	1,000	1,000	15
1,000	0,004	0,927	1,000	0,962	0,961	1,000	0,998	16
Weighted Avg.	**0,952**	**0,004**	**0,953**	**0,952**	**0,950**	**0,948**	**0,999**	**0,971**

=== Confusion Matrix ===

a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	<-- classified as
229	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	| a = 0
1	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	| b = 1
0	0	46	0	0	0	0	0	0	0	0	0	0	0	0	0	4	| c = 2
0	0	0	48	0	0	0	0	2	0	0	0	0	0	0	0	0	| d = 3
0	0	0	0	43	0	0	0	0	7	0	0	0	0	0	0	0	| e = 4
0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0	0	| f = 5
0	0	0	0	0	0	51	0	0	0	0	0	0	0	0	0	0	| g = 6
0	0	0	0	0	0	0	49	0	0	0	0	0	0	0	0	0	| h = 7
1	0	0	12	0	0	0	0	20	0	0	0	0	0	0	2	0	| i = 8
0	0	0	0	11	0	0	0	0	24	0	0	0	0	0	0	0	| j = 9
0	0	0	0	0	0	0	0	0	0	52	0	0	0	0	0	0	| k = 10
0	0	0	0	0	0	0	0	0	0	0	48	0	0	0	0	0	| l = 11
0	0	0	0	0	0	0	0	0	0	0	0	49	0	0	0	0	| m = 12
0	0	0	0	0	0	0	0	0	0	0	0	0	52	0	0	0	| n = 13
3	0	0	0	0	0	0	0	0	0	0	0	0	0	46	0	0	| o = 14
3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	46	0	| p = 15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	51	| q = 16

ΠΑΡΑΡΤΗΜΑ 2

Τα κυριότερα μεγέθη λειτουργιών που μας δίνουν δεδομένα για την απόδοση μιας μηχανής είναι:

Ενεργός Ισχύς - Power (Kw ή Bhp)

Η ισχύς που αναπτύσσεται από τη μηχανή πλην την ισχύ που χάνεται λόγω τριβών κ.α.

Εδώ μελετάμε τις διακυμάνσεις της ισχύος με σκοπό την ισοφόρτηση της μηχανής.

Δεν θέλουμε μεγάλες διαφορές ώστε να μην δημιουργούν άνισες δυνάμεις και αντίστοιχες ροπές, στο σύστημα μετάδοσης και τον στροφαλοφόρο άξονα.

Δημιουργούν άνισες δυνάμεις και αντίστοιχες ροπές στο σύστημα μετάδοσης και τον στροφαλοφόρο άξονα τα αίτια εστιάζουν στο καύσιμο, την προπορεία καυσίμου, τα μεταλλικά μέρη (χιτώνια, έμβολα, ελατήρια).

Ταχύτητα Μηχανής - Στροφές ανά λεπτό (Rpm)

Η ταχύτητα περιστροφής του στροφαλοφόρου άξονα, εκφρασμένη σε στροφές ανά λεπτό (rpm).

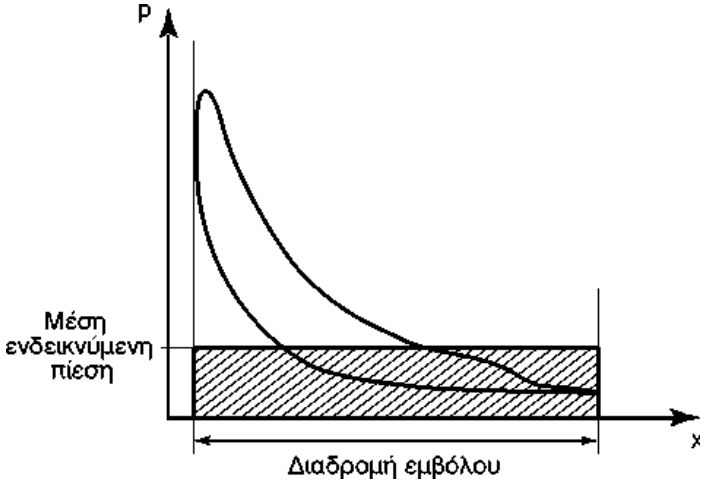
Μικρή πτώση των στροφών. (> 5%)

Μεγάλη πτώση στροφών. (> 10%)

Δεν επιθυμούμε αστάθεια στροφών στην μηχανή ιδιαίτερα όταν λαμβάνουμε στοιχεία απόδοσής της. Για πιθανή αστάθεια τα αίτια εστιάζουν στο ρυθμιστή στροφών governor και στο αντλητικό σύστημα τροφοδοσίας και εγχύσεως καυσίμου.

Μέση Ενδεικνύμενη Πίεση του κυλίνδρου - Pi(m) (bar)

Η **μέση ενδεικνυόμενη πίεση** χαρακτηρίζει την ποιότητα λειτουργίας του κινητήρα, που είναι ανεξάρτητη του μεγέθους του κινητήρα. Στο διάγραμμα p-x αποτελεί το ύψος ορθογωνίου παραλληλογράμμου, το οποίο έχει εμβαδόν ίσο με το ενδεικνυόμενο έργο του κυλίνδρου, ενώ η βάση του παραλληλογράμμου ισούται με τη διαδρομή του εμβόλου



Η **Μέση ενδεικτική πίεση** δεν θα πρέπει να διαφέρει από κύλινδρο σε κύλινδρο περισσότερο από 1 bar, (ανεξαρτήτως φορτίου).

- Η μείωση των παραμέτρων αυτών οφείλεται κυρίως στην αύξηση των τριβών.
- Σε ανωμαλία στον εγχυτή καυσίμου ή σε άλλο εξάρτημα του συστήματος έγχυσης καυσίμου.

Αν είναι **υψηλότερη** τότε τα αίτια που ελέγχουμε είναι:

- Μπορεί να προέρχεται από αυξημένη ποσότητα καυσίμου σε συγκεκριμένο κύλινδρο.(**Θα συνδυάζετε με υψηλή θερμοκρασία καυσαερίων**) Παίρνουμε αυξημένη ισχύ στον συγκεκριμένο κύλινδρο, ενώ η κατανάλωση αυξάνει υπέρμετρα.
- Αυξημένη προπορεία. (**Θα συνδυάζετε με χαμηλή θερμοκρασία καυσαερίων**) Μπορεί να προκαλεί ανεπιθύμητες, πρόωρες καύσεις πυρήνων καυσίμου.

Αν είναι **χαμηλότερη** τότε τα αίτια που ελέγχουμε είναι:

- Ελλιπείς ποσότητα καυσίμου. Δεν αποδίδετε ή πλήρης ισχύς του κυλίνδρου. (**Θα συνδυάζετε με χαμηλή θερμοκρασία καυσαερίων**).
- Βραδυπορεία εγχύσεως καυσίμου, πιθανότατα κακός χρονισμός συστήματος εγχύσεως, πιθανή δυσλειτουργία αντλίας-εγχυτήρων. (**Θα συνδυάζετε με υψηλή θερμοκρασία καυσαερίων**), δεν θα πάρουμε την πλήρη ισχύ του κυλίνδρου, η καύση είναι κακή και πιθανών να έχουμε αυξημένη ρύπανση της μηχανής.

Πίεση κατά την συμπίεση (compression pressure) - Pcomp (bar)

Η πίεση αυτή είναι η πίεση που αναπτύσσεται στον θάλαμο καύσης στο τέλος της συμπίεσης και χωρίς ανάφλεξη του καυσίμου μείγματος.

Η μέτρηση της πίεσης στο θάλαμο καύσης κατά την συμπίεση παρέχει ενδείξεις για την ποιότητα της μηχανής και την κατάσταση της. Πάντως η πίεση εξαρτάται και από τα χαρακτηριστικά λειτουργίας των βαλβίδων.

Η **Πίεση συμπίεσης** δεν πρέπει να διαφοροποιείται μεταξύ των κυλίνδρων περισσότερο από ±3 bar.

Αν είναι **υψηλότερη** από την κανονική (δεδομένη από τον κατασκευαστή) τότε τα αίτια που ελέγχουμε είναι:

- Μπορεί να προέρχεται από αυξημένη συγκέντρωση εξανθρακώματος στην κεφαλή του εμβόλου.
- Από λεπτότερη προσθήκη παρεμβύσματος, μεταξύ σώματος κυλίνδρου και πώματος.
- Αυξημένη ποσότητα - παροχή λαδιού στο έμβολο, ή αυξημένα κυλινδρέλαια σε περίπτωση δίχρονης μηχανής.
- Μεγαλύτερη σε πάχος προσθήκη διωστήρα,(Για συγκεκριμένου τύπου Μ.Ε.Κ.) *Μείωση διακένου συμπίεσης - καύσης.
- Κακό χρονισμό της βαλβίδας εξαγωγής ή βλάβη του συστήματος μετάδοσης, (υδραυλικού ή μηχανικού ανάλογα).

Αν είναι **χαμηλότερη** τότε τα αίτια που ελέγχουμε είναι:

- Φθαρμένα ελατήρια εμβόλων.
- Μειωμένη παροχή λαδιού.
- Παχύτερη προσθήκη καπακιού.
- Η βαλβίδα εξαγωγής διαρρέει ή μένει ανοικτή για μεγαλύτερο διάστημα σε μοίρες στροφάλου. Μεγάλες ελευθερίες μεταξύ χιτωνίου και εμβόλου.
- Λεπτότερη προσθήκη διωστήρα.
- Αύξηση διακένου συμπίεσης - καύσης.

Θέση Εκδήλωσης Μέγιστης Πίεσης Καύσης - Pmax position (μοίρες - degrees)

Ο παράγοντας **θέσης της μέγιστης πίεσης** στον θάλαμο καύσης (location of peak pressure - LPP) βοηθά την κατανόηση του φαινομένου. Σε ιδανικές συνθήκες, η θέσης της μέγιστης πίεσης στον θάλαμο καύσης πρέπει να εμφανίζεται 14 μοίρες μετά την έναυση της καύσης ή αλλιώς 14 μοίρες μετά το Άνω Νεκρό Σημείο (ΑΝΣ). Σε ιδανικές συνθήκες και ανάλογα με τον σχεδιασμό του θαλάμου καύσης, η έναυση πρέπει να ξεκινά περίπου 20 μοίρες πριν το ΑΝΣ και να έχουμε την μέγιστη πίεση 14 μοίρες μετά το ΑΝΣ. Η θέσης της μέγιστης πίεσης στον θάλαμο καύσης είναι ένας μηχανικός συντελεστής όπως ο κινητήρας είναι μια μηχανική συσκευή. Το έμβολο ταξιδεύει πάνω και κάτω και αν η μέγιστη πίεση είναι πολύ νωρίς ή έχει μεγάλη καθυστέρηση τότε δεν παράγεται το μέγιστο έργο. Για αυτό το λόγο η θέση μέγιστης πίεσης είναι πάντα 14 μοίρες μετά το ΑΝΣ. Κατανοώντας καλύτερα την θέσης της μέγιστης πίεσης στον θάλαμο καύσης, πριν από αυτή υπάρχει μια αύξηση

πίεσης στον θάλαμο καύσης ενώ μετά από αυτή την θέση η πίεση έχει φθίνουσα πορεία όσο το έμβολο ταξιδεύει προς τα κάτω και ανοίγουν οι βαλβίδες εξόδου.

Η **θέση εκδήλωσης μέγιστης πίεσης καύσης** δεν πρέπει να είναι μεγαλύτερη από 20 μοίρες μεταξύ των κυλίνδρων.

Αν είναι **υψηλότερη** από την κανονική (δεδομένη από τον κατασκευαστή) τότε τα αίτια που ελέγχουμε είναι:

- Έλεγχος μετακαύσης.
- Έλεγχος μοίρες στροφάλου.

Μέγιστη Πίεση Καύσεως - Pmax – combustion (bar)

Η **Μέγιστη Πίεση Καύσεως** έχουμε όταν η μάζα του καυσίμου έχει αναφλεγή και έχει δημιουργηθεί η δύναμη η οποία άνα την επιφάνεια που παρουσιάζει το έμβολο θα δόση μηχανικό έργο.

Η **Μέγιστη πίεση καύσεως** δεν πρέπει να διαφοροποιείται μεταξύ των κυλίνδρων περισσότερο από ±2 bar.

Αν είναι **υψηλότερη** από την κανονική (δεδομένη από τον κατασκευαστή) τότε τα αίτια που ελέγχουμε είναι:

- Την πίεση πριν ανοίξουν οι βαλβίδες εξαγωγής.
- Θερμική καταπόνηση.
- Έλεγχος χρονισμού μηχανής φάσεων καύσης από το διάγραμμα.

Αν είναι **χαμηλότερη** τότε τα αίτια που ελέγχουμε είναι:

- Αδυναμία υπερπλήρωσης μηχανής
- Ή χαμηλή κινητική ενέργεια προς την έξοδο καυσαερίων

Γωνία εγχύσεως καυσίμου - Injection angle (σε μοίρες - degrees)

Ο κατασκευαστής της κάθε μηχανής δίνει την **γωνία έναρξης της εγχύσεως** που είναι τιμή στάνταρ, όμως αυτό αλλάζει ανάλογα με το καύσιμο.

Ο χρόνος εγχύσεως έχει άμεση σχέση με την γωνία στροφάλου ως προς την στιγμή έναρξης της έγχυσης στην μηχανή η έγχυση ξεκινά πριν το έμβολο φτάσει στο ΑΝΣ, άρα το κομβίο του στροφάλου είναι 18 μοίρες πριν το 0 που είναι το ΑΝΣ, αν η έγχυση ξεκινήσει πριν τις 18 μοίρες είναι προπορεία, αν η έγχυση ξεκινήσει μετά τις 18 μοίρες είναι βραδυπορεία.

Η **γωνία έναρξης της εγχύσεως καυσίμου** δεν πρέπει να διαφοροποιείται μεταξύ των κυλίνδρων περισσότερο από ±0,7 μοίρες.

Αν είναι **μεγαλύτερη** τότε τα αίτια που ελέγχουμε είναι:

- Έχουμε αυξημένη προπορεία πιθανότητα κρουστικής καύσης από πρόωρη ανάφλεξη.

Αν είναι **μικρότερη** τότε τα αίτια που ελέγχουμε είναι:

- Έχουμε βραδυπορεία καυσίμου και θα παρουσιαστεί εμφανής καθυστέρηση ανάφλεξης του καυσίμου. Καθυστέρηση ανάφλεξης **(ignition delay)** είναι το χρονικό διάστημα ή ο αντίστοιχος αριθμός μοιρών περιστροφής του στροφαλοφόρου άξονα, μεταξύ έναρξης εισαγωγής και ανάφλεξης του μείγματος.
- Πιθανότητα ανεπιθύμητων μετακαύσεων στην φάση της εκτόνωσης των καυσαερίων.

Θερμοκρασία καυσαερίων - Exhaust gas temperature (^{o}C)

Η **Θερμοκρασία καυσαερίων**, είναι πολύ σημαντικό κριτήριο, λόγο των θερμικών καταπονήσεων που υποβάλει τα μεταλλικά τμήματα της μηχανής και την αυξημένη επικινδυνότητα αναφλέξεως συγκεντρώσεων εύφλεκτων ρύπων.

Τέλος για διαφοροποιήσεις της **θερμοκρασίας των καυσαερίων** πράγμα που αποτελεί ανασταλτικό παράγοντα για την καλή λειτουργία της μηχανής.

Για τη θερμοκρασία εξόδου των καυσαερίων δεν πρέπει να διαφοροποιείται μεταξύ των κυλίνδρων περισσότερο από ±30 ^{o}C, με ανώτατο όριο διαφοράς τούς 50 ^{o}C.

Αύξηση της θερμοκρασίας των καυσαερίων τότε τα αίτια που ελέγχουμε είναι:

- Αν έχουμε καθαρό ψυγείο αέρα (air coolers)
- Αν η θερμοκρασία του ψυκτικού κυκλώματος είναι σωστή΄
- Αν η πίεση στο ψυγείο έχει ανέβει, που σημαίνει ότι ο αέρας δεν ψύχεται και διατηρεί αυξημένο όγκο.
- Την θερμοκρασία των κύριων ψυγείων θαλάσσης (central cooling system).
- Την εναλλαγή του αέρα, μέσω των ανεμιστήρων και την πίεση του μηχανοστασίου (forced draft fans).
- Το φορτίο της μηχανής.

Πτώση της θερμοκρασίας των καυσαερίων τότε τα αίτια που ελέγχουμε είναι:

- Ελαττωμένη ποσότητα καύσιμου
- Αυξημένη πορεία εκχύσεως καύσιμου

ΠΑΡΑΡΤΗΜΑ 3

Οι νέες τεχνολογίες προσφέρουν όλο και περισσότερες δυνατότητες στο να πλησιάσει στην πράξη. Μία από αυτές τις δυνατότητες είναι η προσομοίωση διαφόρων λειτουργιών του πλοίου σε προσομοιωτές.

Η προσομοίωση της λειτουργίας ενός κινητήρα, αν και πολλές φορές είναι ιδιαίτερα απλοϊκή, έχει κάποια πολύ σημαντικά πλεονεκτήματα. Μέσω αυτής, είναι δυνατή η πραγματοποίηση παραμετρικών μελετών σε μικρότερο χρόνο και με χαμηλότερο κόστος σε σύγκριση με τη διεξαγωγή μετρήσεων στο πλοίο. Επιτρέπουν τη μελέτη της επίδρασης πληθώρας παραμέτρων που επηρεάζουν τη λειτουργία ενός κινητήρα και δίνουν τη δυνατότητα μεμονωμένης εξέτασης της κάθε διεργασίας που λαμβάνει χώρα σε αυτόν.

Στην παρούσα διπλωματική εργασία πραγματοποιείται με τη χρήση του προσομοιωτή μηχανοστασίου πλοίου της Kongsberg (Full Mission Engine Room Simulator) της Ακαδημίας Εμπορικού Ναυτικού Ασπροπύργου.

Η απεικόνιση μέσω προσομοίωσης της συνολικής συμπεριφοράς της μηχανής, με τη διεξαγωγή διαφόρων σεναρίων δυσλειτουργιών – ανωμαλιών κινητήρα, διενεργήθηκαν λήψεις και εισαγωγή των δεδομένων αυτών στη βάση δεδομένων που χρησιμοποιήθηκε στη διεξαγωγή των πειραμάτων με τους αλγορίθμους μηχανικής μάθησης.

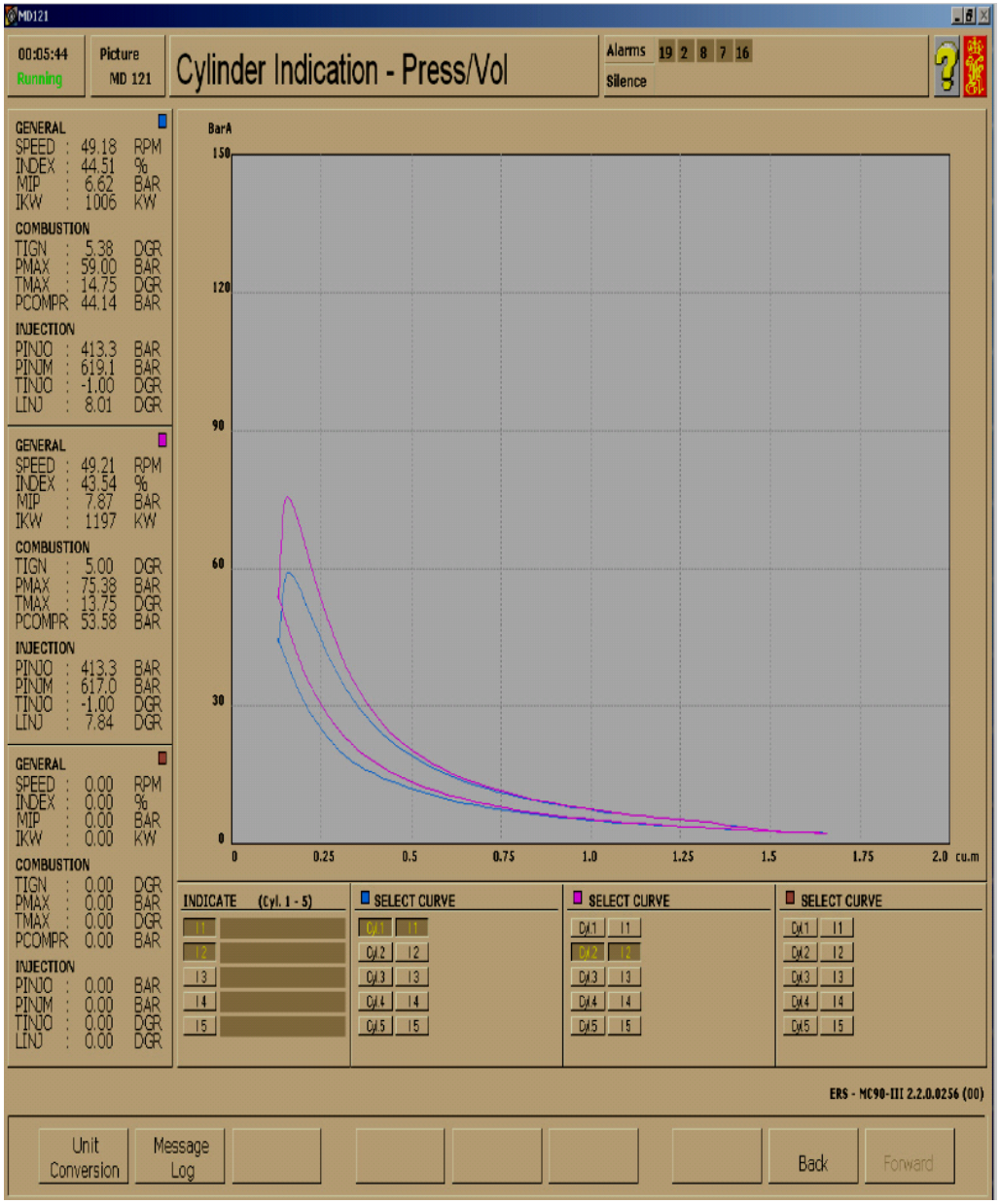
NEPTUNE Engine Room Simulator - Kongsberg Maritime

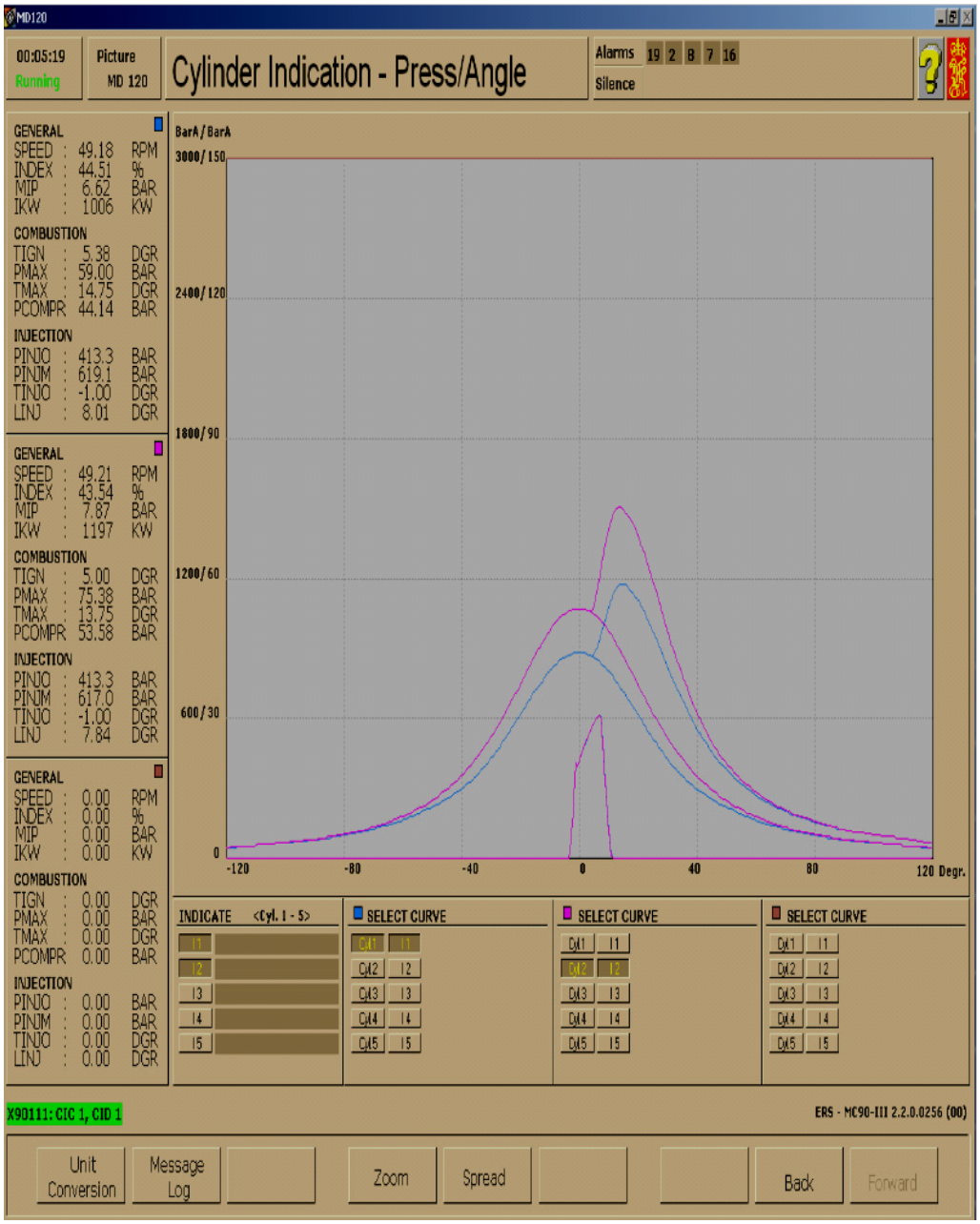
Η Kongsberg Maritime (KM) είναι μια νορβηγική εταιρία η οποία ασχολείται με συστήματα για την τοποθέτηση, την τοπογραφία, την πλοήγηση και αυτοματοποίηση σε εμπορικά πλοία και υπεράκτιες εγκαταστάσεις.

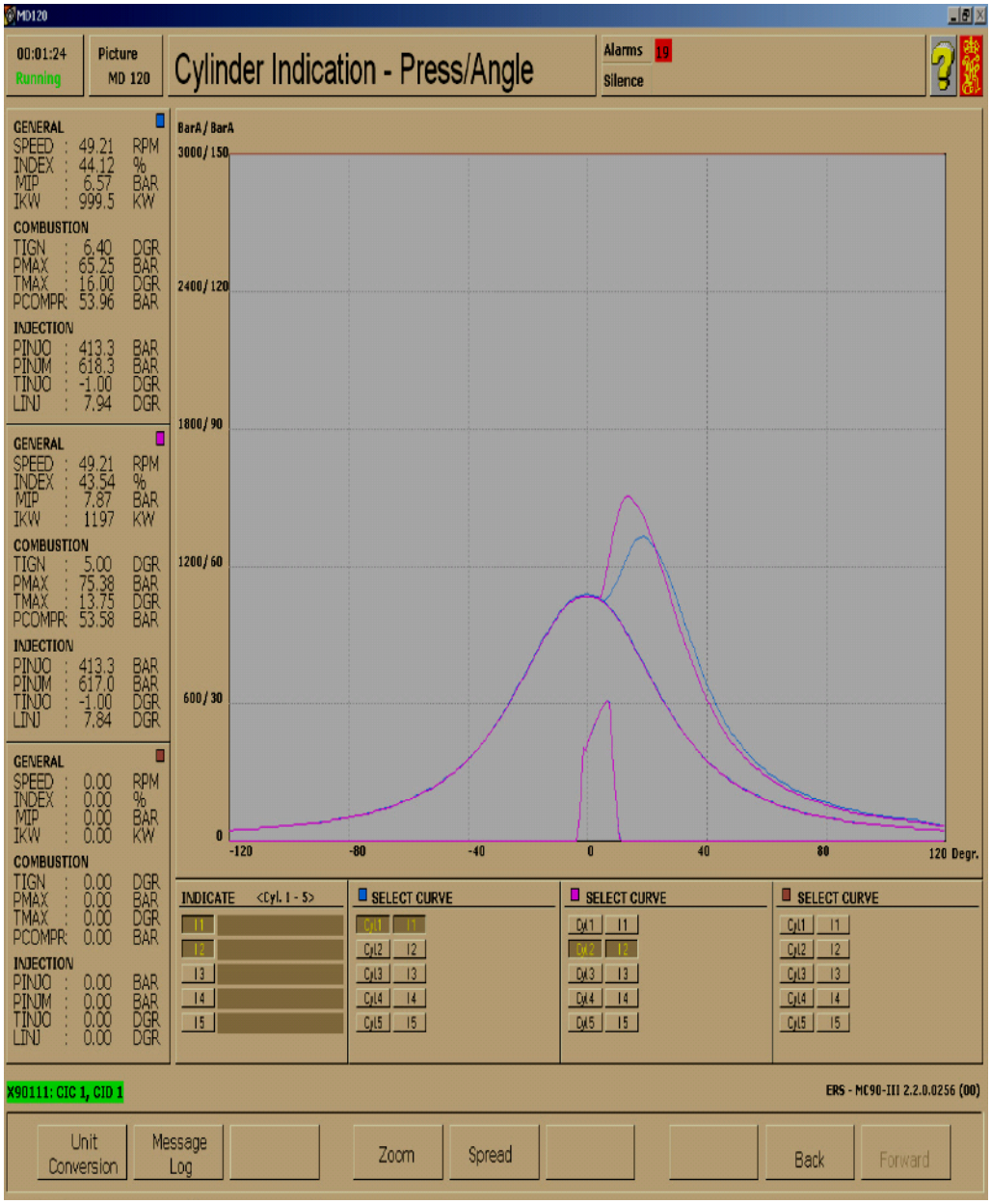
Ένα από τα προϊόντα της εταιρίας είναι και ο προσομοιωτής μηχανοστασίου NEPTUNE και αποτελεί έναν από τους πιο προηγμένους προσομοιωτές που διατίθενται σήμερα.

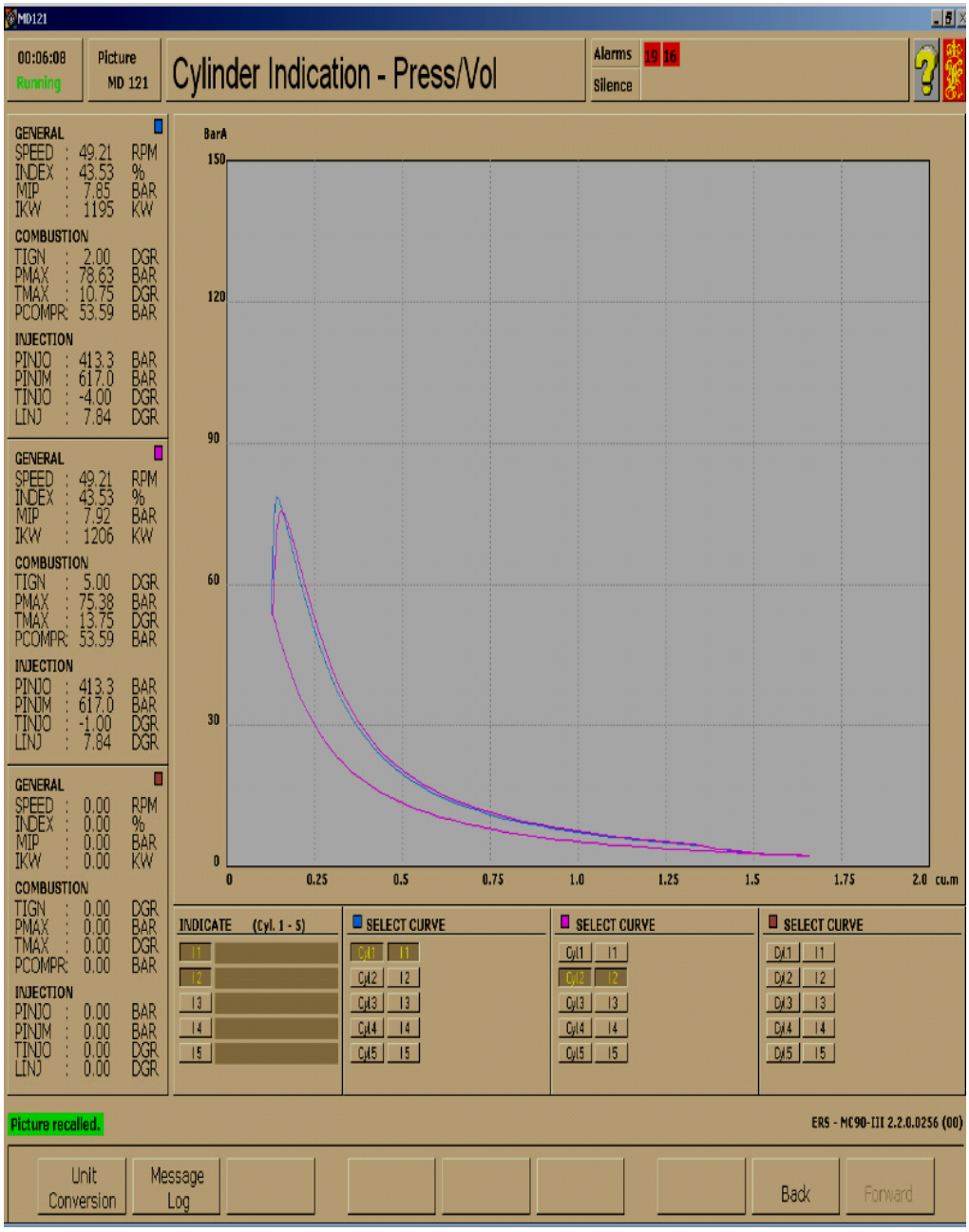
Είναι αρκετά ευέλικτος και παρέχει μία μεγάλη ποικιλία τύπων μηχανών. Το πλήρες σύστημα περιλαμβάνει δωμάτιο μηχανοστασίου (engine room), δωμάτιο ελέγχου μηχανοστασίου (engine control room) καθώς και δωμάτιο του εκπαιδευτή (instructor room). Η αρχιτεκτονική του συστήματος είναι αρκετά ευέλικτη και μπορεί να χρησιμοποιηθεί σε ένα ευρύ φάσμα διαφορετικών διεπαφών (interfaces), δίνοντας έτσι τη δυνατότητα στους μηχανικούς να εκπαιδεύονται σε πανομοιότυπο εξοπλισμό με αυτόν που αργότερα θα χρησιμοποιούν στη δουλειά τους.

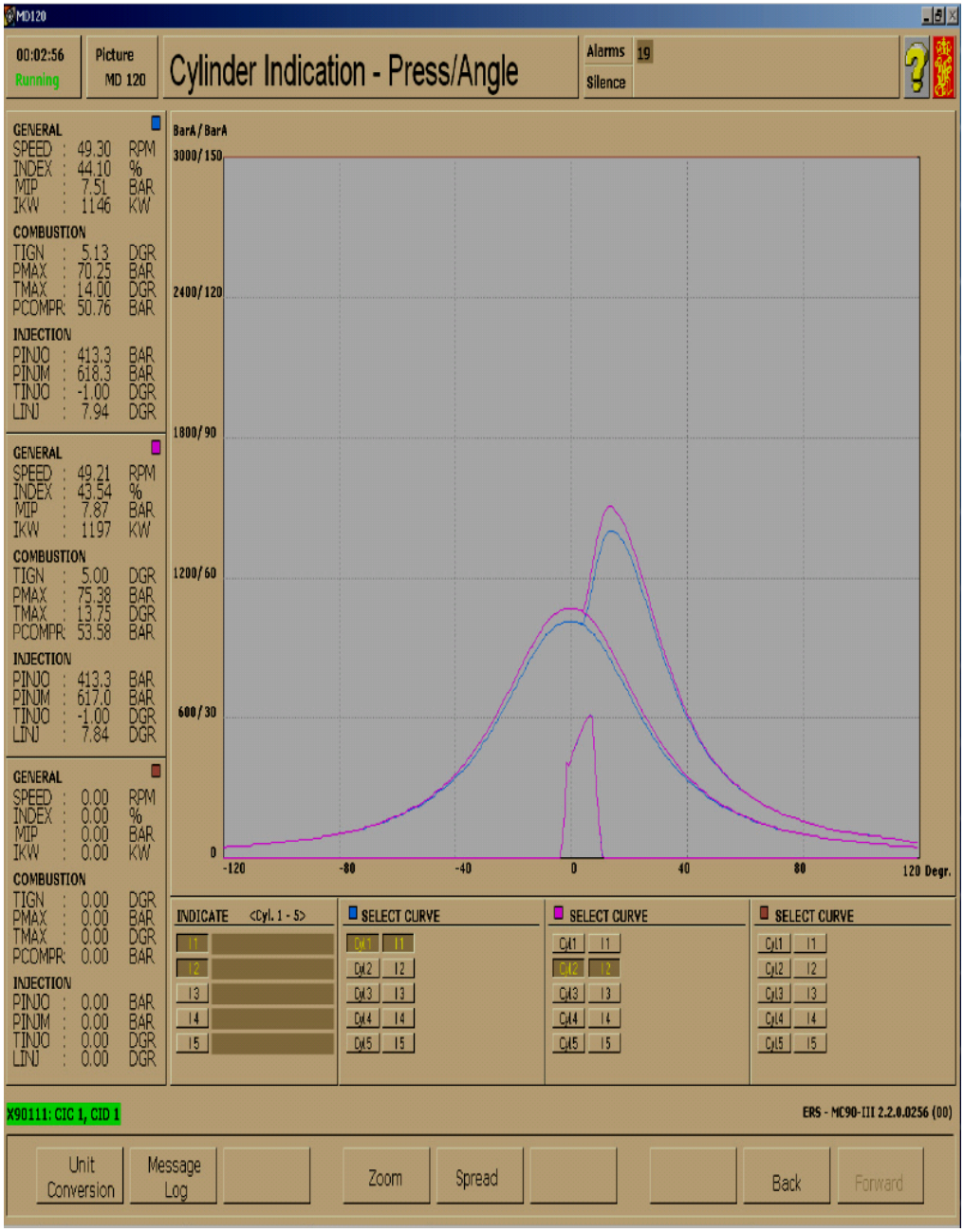
Στιγμιότυπα βλαβών στο προσομοιωτή μηχανοστασίου


MD121
00:05:44
Running
Picture
MD 121
Cylinder Indication - Press/Vol
Alarms 19 2 8 7 16
Silence
GENERAL
SPEED : 49.18 RPM
INDEX : 44.51 %
MIP : 6.62 BAR
IKW : 1006 KW
COMBUSTION
TIGN : 5.38 DGR
PMAX : 59.00 BAR
TMAX : 14.75 DGR
PCOMPR 44.14 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 619.1 BAR
TINJO : -1.00 DGR
LINJ : 8.01 DGR
GENERAL
SPEED : 49.21 RPM
INDEX : 43.54 %
MIP : 7.87 BAR
IKW : 1197 KW
COMBUSTION
TIGN : 5.00 DGR
PMAX : 75.38 BAR
TMAX : 13.75 DGR
PCOMPR 53.58 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 617.0 BAR
TINJO : -1.00 DGR
LINJ : 7.84 DGR
GENERAL
SPEED : 0.00 RPM
INDEX : 0.00 %
MIP : 0.00 BAR
IKW : 0.00 KW
COMBUSTION
TIGN : 0.00 DGR
PMAX : 0.00 BAR
TMAX : 0.00 DGR
PCOMPR 0.00 BAR
INJECTION
PINJO : 0.00 BAR
PINJM : 0.00 BAR
TINJO : 0.00 DGR
LINJ : 0.00 DGR
BarA
150
120
90
60
30
0
0 0.25 0.5 0.75 1.0 1.25 1.5 1.75 2.0 cu.m
INDICATE (Cyl. 1 - 5)
SELECT CURVE
SELECT CURVE
SELECT CURVE
ERS - MC90-III 2.2.0.0256 (00)
Unit Conversion
Message Log
Back
Forward


MD120
00:05:19
Running
Picture
MD 120
Cylinder Indication - Press/Angle
Alarms 19 2 8 7 16
Silence
GENERAL
SPEED : 49.18 RPM
INDEX : 44.51 %
MIP : 6.62 BAR
IKW : 1006 KW
COMBUSTION
TIGN : 5.38 DGR
PMAX : 59.00 BAR
TMAX : 14.75 DGR
PCOMPR 44.14 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 619.1 BAR
TINJO : -1.00 DGR
LINJ : 8.01 DGR
GENERAL
SPEED : 49.21 RPM
INDEX : 43.54 %
MIP : 7.87 BAR
IKW : 1197 KW
COMBUSTION
TIGN : 5.00 DGR
PMAX : 75.38 BAR
TMAX : 13.75 DGR
PCOMPR 53.58 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 617.0 BAR
TINJO : -1.00 DGR
LINJ : 7.84 DGR
GENERAL
SPEED : 0.00 RPM
INDEX : 0.00 %
MIP : 0.00 BAR
IKW : 0.00 KW
COMBUSTION
TIGN : 0.00 DGR
PMAX : 0.00 BAR
TMAX : 0.00 DGR
PCOMPR 0.00 BAR
INJECTION
PINJO : 0.00 BAR
PINJM : 0.00 BAR
TINJO : 0.00 DGR
LINJ : 0.00 DGR
BarA / BarA
3000/150
2400/120
1800/90
1200/60
600/30
0
-120
-80
-40
0
40
80
120 Degr.
INDICATE <Cyl. 1 - 5>
I1
I2
I3
I4
I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
X90111: CIC 1, CID 1
ERS - MC90-III 2.2.0.0256 (00)
Unit Conversion
Message Log
Zoom
Spread
Back
Forward


MD120
00:01:24
Running
Picture
MD 120
Cylinder Indication - Press/Angle
Alarms 19
Silence
GENERAL
SPEED : 49.21 RPM
INDEX : 44.12 %
MIP : 6.57 BAR
IKW : 999.5 KW
COMBUSTION
TIGN : 6.40 DGR
PMAX : 65.25 BAR
TMAX : 16.00 DGR
PCOMPR 53.96 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 618.3 BAR
TINJO : -1.00 DGR
LINJ : 7.94 DGR
GENERAL
SPEED : 49.21 RPM
INDEX : 43.54 %
MIP : 7.87 BAR
IKW : 1197 KW
COMBUSTION
TIGN : 5.00 DGR
PMAX : 75.38 BAR
TMAX : 13.75 DGR
PCOMPR 53.58 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 617.0 BAR
TINJO : -1.00 DGR
LINJ : 7.84 DGR
GENERAL
SPEED : 0.00 RPM
INDEX : 0.00 %
MIP : 0.00 BAR
IKW : 0.00 KW
COMBUSTION
TIGN : 0.00 DGR
PMAX : 0.00 BAR
TMAX : 0.00 DGR
PCOMPR 0.00 BAR
INJECTION
PINJO : 0.00 BAR
PINJM : 0.00 BAR
TINJO : 0.00 DGR
LINJ : 0.00 DGR
BarA / BarA
3000 / 150
2400 / 120
1800 / 90
1200 / 60
600 / 30
0
-120
-80
-40
0
40
80
120 Degr.
INDICATE <Cyl. 1 - 5>
SELECT CURVE
SELECT CURVE
SELECT CURVE
X90111: CIC 1, CID 1
ERS - MC90-III 2.2.0.0256 (00)
Unit Conversion
Message Log
Zoom
Spread
Back
Forward


MD121
00:06:08
Running
Picture
MD 121
Cylinder Indication - Press/Vol
Alarms 19 16
Silence
GENERAL
SPEED : 49.21 RPM
INDEX : 43.53 %
MIP : 7.85 BAR
IKW : 1195 KW
COMBUSTION
TIGN : 2.00 DGR
PMAX : 78.63 BAR
TMAX : 10.75 DGR
PCOMPR: 53.59 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 617.0 BAR
TINJO : -4.00 DGR
LINJ : 7.84 DGR
GENERAL
SPEED : 49.21 RPM
INDEX : 43.53 %
MIP : 7.92 BAR
IKW : 1206 KW
COMBUSTION
TIGN : 5.00 DGR
PMAX : 75.38 BAR
TMAX : 13.75 DGR
PCOMPR: 53.59 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 617.0 BAR
TINJO : -1.00 DGR
LINJ : 7.84 DGR
GENERAL
SPEED : 0.00 RPM
INDEX : 0.00 %
MIP : 0.00 BAR
IKW : 0.00 KW
COMBUSTION
TIGN : 0.00 DGR
PMAX : 0.00 BAR
TMAX : 0.00 DGR
PCOMPR: 0.00 BAR
INJECTION
PINJO : 0.00 BAR
PINJM : 0.00 BAR
TINJO : 0.00 DGR
LINJ : 0.00 DGR
BarA
150
120
90
60
30
0
0
0.25
0.5
0.75
1.0
1.25
1.5
1.75
2.0 cu.m
INDICATE (Cyl. 1 - 5)
I1
I2
I3
I4
I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
Picture recalled.
ER5 - MC90-III 2.2.0.0256 (00)
Unit Conversion
Message Log
Back
Forward


MD120
00:02:56
Running
Picture
MD 120
Cylinder Indication - Press/Angle
Alarms 19
Silence
GENERAL
SPEED : 49.30 RPM
INDEX : 44.10 %
MIP : 7.51 BAR
IKW : 1146 KW
COMBUSTION
TIGN : 5.13 DGR
PMAX : 70.25 BAR
TMAX : 14.00 DGR
PCOMPR 50.76 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 618.3 BAR
TINJO : -1.00 DGR
LINJ : 7.94 DGR
GENERAL
SPEED : 49.21 RPM
INDEX : 43.54 %
MIP : 7.87 BAR
IKW : 1197 KW
COMBUSTION
TIGN : 5.00 DGR
PMAX : 75.38 BAR
TMAX : 13.75 DGR
PCOMPR 53.58 BAR
INJECTION
PINJO : 413.3 BAR
PINJM : 617.0 BAR
TINJO : -1.00 DGR
LINJ : 7.84 DGR
GENERAL
SPEED : 0.00 RPM
INDEX : 0.00 %
MIP : 0.00 BAR
IKW : 0.00 KW
COMBUSTION
TIGN : 0.00 DGR
PMAX : 0.00 BAR
TMAX : 0.00 DGR
PCOMPR 0.00 BAR
INJECTION
PINJO : 0.00 BAR
PINJM : 0.00 BAR
TINJO : 0.00 DGR
LINJ : 0.00 DGR
BarA/BarA
3000/150
2400/120
1800/90
1200/60
600/30
0
-120
-80
-40
0
40
80
120 Degr.
INDICATE <Cyl. 1 - 5>
I1
I2
I3
I4
I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
SELECT CURVE
Cyl1 I1
Cyl2 I2
Cyl3 I3
Cyl4 I4
Cyl5 I5
X90111: CIC 1, CID 1
ERS - MC90-III 2.2.0.0256 (00)
Unit Conversion
Message Log
Zoom
Spread
Back
Forward