ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

Τομέας Συστημάτων Παραγωγής

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΑΤΡΙΒΗ

ΘΕΜΑ:

“Βελτίωση Επιχειρησιακών Διαδικασιών με χρήση τεχνικών Εξόρυξης Γνώσης”

Τριμελής Επιτροπή: Ν. Μπιλάλης (Επιβλέπων Καθηγητής)
Ε. Γρηγορούδης
Β. Μουστάκης

Παπάζογλου Μιχάλης

ΧΑΝΙΑ 2008

Περιεχόμενα

Εισαγωγή

Στη σημερινή εποχή της πληροφορίας παρατηρείται μια αλματώδης αύξηση των δεδομένων που αποθηκεύονται σε ηλεκτρονική μορφή. Τα δεδομένα αυτά, στο σύνολο των περιπτώσεων, αποτελούν τα "ιστορικά αρχεία" μιας οποιασδήποτε διαδικασίας-ενέργειας-διεργασίας που έχει υλοποιηθεί σε παρελθοντικό χρόνο και καταγράφεται ηλεκτρονικά. Ως ακατέργαστα ιστορικά δεδομένα προτρέπουν κάθε ενδιαφερόμενο μελετητή-αναλυτή να εκμαιεύσει τις χρήσιμες πληροφορίες που βρίσκονται "κρυμμένες" μέσα στο σύνολο των δεδομένων. Η εφαρμογή της εξορυγμένης γνώσης σε θεωρητικές πρακτικές είναι ικανή να οδηγήσει το μελετητή-αναλυτή σε ένα σύνολο ενεργειών που θα αποφέρουν την εξέλιξη και τη βελτιστοποίηση των οιοσδήποτε διεργασιών και διαδικασιών που τον αφορούν.

Στο πεδίο ενδιαφέροντος της παρούσας εργασίας, την εκμαίευση της πληροφορίας αντιπροσωπεύουν οι τεχνικές εξόρυξης γνώσης ενώ, τις θεωρητικές πρακτικές αντιπροσωπεύουν τα 7 νέα εργαλεία ποιότητας, ως κύριο συστατικό της Διοίκησης Ολικής Ποιότητας (Δ.Ο.Π.).

Η παρούσα διατριβή ασχολείται με τη διερεύνηση της σχέσης μεταξύ των 7 νέων εργαλείων ποιότητας, ως βασικό συστατικό της Δ.Ο.Π. και των τεχνικών εξόρυξης γνώσης. Ειδικότερα, επιχειρείται η υλοποίησης μιας αυτοματοποιημένης εφαρμογής που βασίζεται στην τροφοδότηση με συγκεκριμένες πληροφορίες των 7 νέων εργαλείων ποιότητας. Οι πληροφορίες προέρχονται από τα αποτελέσματα των τεχνικών εξόρυξης γνώσης σε πραγματικά δεδομένα με τελικό σκοπό την ανάδειξη των πηγών των τυχόν προβλημάτων και την παροχή πιθανών λύσεων που θα οδηγήσουν στη βελτιστοποίηση των διαδικασιών.

Όσον αφορά στην οργάνωση του κειμένου, αρχικά αναλύονται τα 7 νέα εργαλεία ποιότητας, ως κύριο συστατικό της Διοίκησης Ολικής Ποιότητας (Δ.Ο.Π.). Στην συνέχεια, αναπτύσσεται μια εισαγωγή στις τεχνικές εξόρυξης γνώσης και έπεται το κεφάλαιο που άπτεται της εφαρμογής όπου αναλύονται τα πραγματικά δεδομένα της μελέτης περίπτωσης, περιγράφονται τα τεχνικά εργαλεία με τα οποία υλοποιήθηκε, επεξηγείται ο τρόπος επιλογής των χρησιμοποιούμενων τεχνικών εξόρυξης γνώσης και η ανάλυση αυτών και πραγματοποιείται μια περιήγηση στις δυνατότητες της εφαρμογής. Ακολουθεί το κεφάλαιο των αποτελεσμάτων της εφαρμογής όπου παρουσιάζονται τα αποτελέσματα των τεχνικών εξόρυξης γνώσης και των 7 νέων εργαλείων ποιότητας και η εργασία κλείνει με την ανάπτυξη των συμπερασμάτων που προκύπτουν από τα αποτελέσματα της εφαρμογής.

Κεφάλαιο 1

Εργαλεία και Τεχνικές για την Βελτίωση Διαδικασιών

1.1 Εισαγωγή

Η Βελτίωση Διαδικασιών σε οργανισμούς, επιχειρήσεις αλλά και σε οποιοδήποτε διοικούμενο σύστημα αφορά σε μια σειρά δράσεων με σκοπό την αναγνώριση, την ανάλυση και τη βελτίωση υπαρχουσών διαδικασιών. Αυτές οι δράσεις συχνά εντάσσονται σε μια συγκεκριμένη μεθοδολογία ή στρατηγική που στοχεύει στη δημιουργία επιτυχημένων αποτελεσμάτων σε οργανισμούς και επιχειρήσεις.

Οι κυριότερες από αυτές τις μεθοδολογίες είναι οι εξής:

- Benchmarking
- Business Process Improvement
- Business Process Reengineering
- Capability Maturity Model Integration/Capability Maturity Model
- Hoshin Kanri
- ISO 9000
- Just In Time Manufacturing
- Lean Manufacturing
- Performance Improvement
- Process Management
- Process Improvement and Management (PI&M)
- Six Sigma
- Theory of Constraints
- Total Quality Management
- Trillium Model
- Twelve leverage points

Η καταλληλότητα κάθε μεθοδολογίας διαφέρει ανάλογα με τον οργανισμό, καθώς εξαρτάται κυρίως από τον τύπο των διαδικασιών που χρησιμοποιούνται σε αυτούς. Συνήθως, ένας οργανισμός χρησιμοποιεί δύο ή και περισσότερες από αυτές, αποσκοπώντας σε μια πληρέστερη εικόνα των διαδικασιών και των πιθανών τους βελτιώσεων. Επίσης, αρκετές μεθοδολογίες χρησιμοποιούνται ως υποσύνολα σε άλλες ευρύτερες μεθοδολογίες.

Μια από τις πιο αποδεκτές στρατηγικές είναι αυτή της Διοίκησης Ολικής Ποιότητας (Total Quality Management), η οποία εμπεριέχει μια σειρά από εργαλεία και τεχνικές που υλοποιούν τις προτεινόμενες διοικητικές πρακτικές της Διοίκησης Ολικής Ποιότητας, που εν τέλει οδηγούν στη ζητούμενη βελτίωση των διαδικασιών.

1.2 Διοίκηση Ολικής Ποιότητας

Η έννοια της Διοίκησης Ολικής Ποιότητας αναπτύχθηκε στη δεκαετία του 1950 όταν Ιαπωνικοί οργανισμοί συμβουλεύτηκαν τον W. Edwards Deming, έναν Αμερικάνο στατιστικολόγο και θεωρητικό της Διοίκησης Επιχειρήσεων, για να τους βοηθήσει στην ανάπτυξη των αποδυναμωμένων από το Δεύτερο Παγκόσμιο Πόλεμο Ιαπωνικών επιχειρήσεων. Υλοποιώντας τις αρχές του Deming για τη Διοίκηση Ολικής Ποιότητας, οι Ιαπωνικές επιχειρήσεις γνώρισαν μια πρωτοφανή οικονομική ανάπτυξη. Αργότερα, στη δεκαετία του 1980, όταν οι Ηνωμένες Πολιτείες Αμερικής παρατήρησαν μείωση του μεριδίου τους στην παγκόσμια αγορά σε σχέση με την Ιαπωνία, επανεκτίμησαν τη θεωρία του W. Edwards Deming και όρισαν το βιβλίο του *Out of Crisis* [Deming, 1986] ως σύμβολο της πορείας των Η.Π.Α. προς την ανάκαμψη της οικονομίας τους. Έκτοτε και με τη συνδρομή και άλλων ειδημόνων, όπως του Joseph Juran και του Philip Crosby (κυρίως με τα βιβλία τους *Juran on Planning for Quality* [Juran, 1988] και *Quality Is Free: The Art of Making Quality Certain* [Crosby, 1979] αντίστοιχα), οι θεωρίες, τα μοντέλα και τα εργαλεία της Διοίκησης Ολικής Ποιότητας εξελιχθήκαν ραγδαία. Σήμερα η Διοίκηση Ολικής Ποιότητας χρησιμοποιείται ευρύτατα στις επιχειρήσεις, στην εκπαίδευση, στη δημόσια διοίκηση και στους οργανισμούς.

Μια προσέγγιση των εννοιών των συστατικών λέξεων, και πιο συγκεκριμένα ο ρόλος του επιθέτου "ολική" στο ουσιαστικό "ποιότητα", θα βοηθούσε σε μια σαφέστερη εικόνα της ίδιας της μεθοδολογίας. Δεχόμενοι τους ορισμούς της ποιότητας που αναφέρουν ότι ποιότητα σημαίνει προϊόν απαλλαγμένο από ελαττώματα, συμμόρφωση στις προ-καθορισμένες προδιαγραφές και εκπλήρωση των αναγκών του καταναλωτή, η έννοια "ολική" δεσμεύει το σύνολο των διαδικασιών, των προϊόντων/υπηρεσιών και ατόμων που εμπλέκονται στις ενέργειες ενός οργανισμού με τις αρχές της ποιότητας [Groockock, 1980], [Zaludova, 1981], [Thoday, 1981]. Η εφαρμογή της είναι δυνατή σε όλες τις δραστηριότητες ενός οργανισμού από την έρευνα και την ανάπτυξη, το σχεδιασμό, την κατασκευή μέχρι το μάρκετινγκ και την οικονομική διοίκηση.

Όλες οι πρακτικές εφαρμογές της Διοίκησης Ολικής Ποιότητας έχουν ως σημείο αναφοράς τον καταναλωτή. Η σχέση ανάμεσα στον οργανισμό και τον καταναλωτή θεωρείται διαδραστική και αμφίδρομη. Ο καταναλωτής είναι αυτός που καθορίζει την ποιότητα μέσω των αναγκών του [Feigenbaum, 1991]. Η Δ.Ο.Π. παρέχει το πλαίσιο βάσει του οποίου το προϊόν μετασχηματίζεται σύμφωνα με τις ανάγκες και τη νοοτροπία του καταναλωτή, ύστερα από διαδικασίες αναζήτησης, συλλογής και ανάλυσης πληροφοριών που πηγάζουν από αυτόν.

Τα κυρία βήματα για τη μετεξέλιξη των διαδικασιών και των προϊόντων με βάση τις απαιτήσεις του πελάτη καλούνται Προγραμματισμός Παραγωγής και, συμφωνά με τον «οδικό χάρτη του προγραμματισμού ποιότητας» του Juran, αναπτύσσονται ως εξής [Juran, 1988]:

- Προσδιορισμός των πελατών.
- Καθορισμός των αναγκών τους.
- Μετάφραση των αναγκών στην εσωτερική διοικητική «γλώσσα» του οργανισμού.
- Ανάπτυξη προϊόντος που ανταποκρίνεται στις ανάγκες αυτές.
- Βελτιστοποίηση των προδιαγραφών του προϊόντος για την εκπλήρωση των αναγκών του οργανισμού αλλά και του πελάτη.
- Ανάπτυξη διαδικασίας που είναι ικανή να παράγει το ζητούμενο προϊόν.
- Βελτιστοποίηση της διαδικασίας.
- Απόδειξη ότι η διαδικασία μπορεί να παράγει το προϊόν σε συνθήκες παραγωγής του οργανισμού.
- Υλοποίηση της διαδικασίας.

1.3 Πρακτικές και Εργαλεία της Δ.Ο.Π.

Το σύνολο των διοικητικών πρακτικών και των τεχνικών εργαλείων της Δ.Ο.Π. συνιστά ένα πλαίσιο το οποίο υιοθετείται από όλες τις διαδικασίες του οργανισμού που εφαρμόζεται.

Οι διοικητικές πρακτικές της Δ.Ο.Π. περιγράφονται συνοπτικά παρακάτω [Rao, et al, 1996]:

Εστίαση στον καταναλωτή: Η ΔΟΠ θέτει τον καταναλωτή ως το κύριο ρυθμιστή του προϊόντος και των διαδικασιών του οργανισμού διαμέσου των αναγκών του και της κουλτούρας του. Επίσης, κατατάσσει τον πελάτη σε δυο κατηγορίες:

εσωτερικό (internal), εννοώντας τον επόμενο εργαζόμενο στην αλυσίδα παραγωγής και εξωτερικό (external), εννοώντας τον τελικό καταναλωτή.

Ικανοποίηση του καταναλωτή: Πρώτιστος στόχος είναι η ικανοποίηση των αναγκών του πελάτη είτε είναι εσωτερικός είτε εξωτερικός καταναλωτής.

Ολική συμμετοχή: Προτρέπει τον εργαζόμενο να είναι εξειδικευμένος στον τομέα εργασίας του και να αυτό-αξιολογείται καθώς και να προτείνει τρόπους βελτίωσης της διαδικασίας στην οποία συμμετέχει.

Συνεχής Βελτίωση: Η ατέρμονη επικέντρωση και βελτίωση των διαδικασιών.

Ευρύ φάσμα εφαρμοσιμοτητάς: Η στρατηγική της Δ.Ο.Π. μπορεί να εφαρμοστεί σε οποιοδήποτε οργανισμό, επιχείρηση και γενικά σε οποιοδήποτε διοικούμενο σύστημα.

Παραγωγικότητα: Δέσμευση ότι για το προϊόν/υπηρεσία θα χρησιμοποιηθεί ιδανικά το ανθρώπινο δυναμικό και ότι το προϊόν θα είναι έτοιμο εγκαίρως.

Καινοτομία: Νέες ιδέες και εφαρμογές για την ικανοποίηση γνωστών αλλά και τυχόν νέων αναγκών του καταναλωτή.

Επίλυση προβλημάτων: Η Δ.Ο.Π. καλεί τους οργανισμούς να διοχετεύσουν το δικαίωμα της επίλυσης του προβλήματος σε κάθε εργαζόμενο και να τον προτρέψουν να το αναλύσει και να προτείνει λύση.

Για την υλοποίηση των παραπάνω τεχνικών η Δ.Ο.Π. προτείνει μια σειρά από εργαλεία τα οποία περιέχουν τεχνικές, διαγράμματα, χάρτες και μεθόδους που συνεισφέρουν στη βελτίωση της ποιότητας και των διαδικασιών [Tague, 2005]:

Λειτουργική ανάπτυξη ποιότητας (Quality Function Deployment -QFD): Αφορά πρωτίστως στη σχεδίαση του προϊόντος και στηρίζεται στο σχεδιασμό του προϊόντος με βάση τις απαιτήσεις των πελατών και των προδιαγραφών που προκύπτουν από αυτές.

Σχεδιασμός Πειραμάτων: Αναφέρεται στο σχεδιασμό πειραμάτων του Taguchi [Pham & Oztemel, 1996] και συνιστά στην ελαχιστοποίηση των διαφορών του καταναλωτή και του προϊόντος κατά τις διαδικασίες ανάμεσα στο σχεδιασμό και την παραγωγή του προϊόντος (Πλήρης μηχανική ανάπτυξη, Λειτουργία).

Έλεγχος Ανταγωνιστικότητας με χρήση δεικτών (Benchmarking): Ορισμός και αποδοχή δεικτών από ανταγωνιστικούς οργανισμούς, με απώτερο σκοπό την αξιολόγηση των προϊόντων τους και τις μετρήσιμες απαιτήσεις του καταναλωτή.

Στατιστικός Έλεγχος Διαδικασιών: Μεθοδολογία πρόληψης και αποφυγής προβλημάτων κατά την παραγωγική διαδικασία, βασιζομένη σε στατιστικά εργαλεία.

Τα 7 κλασσικά εργαλεία ποιότητας: Αποτελούν τα 7 κύρια στατιστικά εργαλεία που χρησιμοποιεί ο Στατιστικός Έλεγχος Διαδικασιών [Montgomery, 1997]:

- **Διαγράμματα ελέγχου**: Γραφική αναπαράσταση ενός ποιοτικού χαρακτηριστικού του προϊόντος σε σχέση με το χρόνο.
- **Δελτίο αναφοράς - ελέγχου (Check sheet)**: Φόρμα συμπλήρωσης και συλλογής δεδομένων.
- **Ιστόγραμμα**: Γραφική αναπαράσταση της συχνότητας εμφάνισης των διαφορετικών τιμών μιας μεταβλητής ποιότητας.
- **Διάγραμμα Pareto**: Διάγραμμα κατανομής συχνοτήτων των ιδιοτήτων κάποιων δεδομένων, ταξινομημένα ανά κατηγορία.
- **Διάγραμμα αιτίου-αποτελέσματος**: Γραφική μεθοδολογία αποτύπωσης των πιθανών αιτιών των προβλημάτων.
- **Διάγραμμα συγκέντρωσης ελαττωματικών**: Γραφική αποτύπωση του προϊόντος, ή μέρος αυτού, όπου σηματοδοτούνται τα διακριτά μέρη του και τοποθετούνται τα προβλήματά του πάνω στα μέρη που ευθύνονται για την εμφάνισή του.
- **Διάγραμμα διασποράς**: Διάγραμμα που αναδεικνύει τη συσχέτιση μεταξύ δυο μεταβλητών ή χαρακτηριστικών του προϊόντος.

Τα 7 νέα εργαλεία ποιότητας: Βασικά εργαλεία της Δ.Ο.Π. για τη βελτίωση ποιότητας που αποσκοπούν στην ανάδειξη των προβλημάτων των διαδικασιών και την εύρεση πιθανών λύσεων. Τα 7 νέα εργαλεία θα τα αναλύσουμε περαιτέρω στην επόμενη παράγραφο.

1.4 Τα 7 νέα εργαλεία ποιότητας

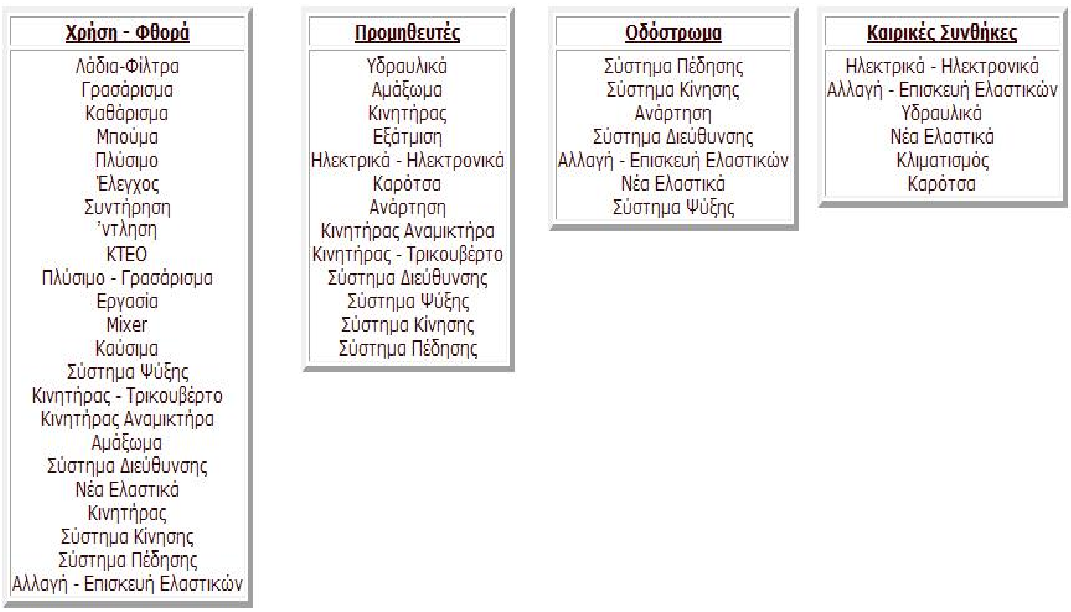
Τα 7 νέα εργαλεία προσφέρουν καλύτερο πλαίσιο διοίκησης ποιότητας και είναι περισσότερο "ποιοτικά" από τα προηγούμενα, καθώς προσανατολίζονται στην επισήμανση των προβλημάτων και των πηγών τους και στην ανάδειξη του συνόλου των λύσεων που ενδείκνυνται για τα συγκεκριμένα προβλήματα [Kolarik, 1995]. Ο παραπάνω προσανατολισμός των 7 νέων εργαλείων ποιότητας επιτυγχάνει τον συμβατό συνδυασμό με τις τεχνικές εξόρυξης γνώσης καθώς τα κυρίως αποτελέσματα των τεχνικών εξόρυξης γνώσης (κανόνες) είναι ικανά να συναποτελέσουν την πληροφορία πάνω στην οποία θα δημιουργηθούν τα 7 νέα εργαλεία ποιότητας. Ακολουθεί μια αναλυτικότερη περιγραφή τους:

Διάγραμμα Συγγένειας (Affinity Diagram): Αφορά στην οργάνωση μεγάλων ποσοτήτων δεδομένων σε ομάδες σύμφωνα με κάποια μορφή συγγένειας [Kanji & Asher, 1996]. Η ομαδοποίηση προσδίδει δομή σε ένα μεγάλο και πολύπλοκο θέμα, ταξινομεί το θέμα σε κατηγορίες και επιφέρει τον προσδιορισμό του προβλήματος [Dahlgaard, et al, 1998].

Τα διαγράμματα συγγένειας βοηθούν στην αναγνώριση του προβλήματος, καθώς οργανώνουν μια γενική ιδέα-αντίληψη για το πρόβλημα, χρησιμοποιώντας γεγονότα, απόψεις και ιδέες που έχουν προηγουμένως συλλεγχθεί και οργανωθεί.

Για την ανάπτυξή του, απαιτείται “καταιγισμός ιδεών” με σκοπό την παραγωγή ενός γραφικού αποτελέσματος. Ακολουθούν τα βασικά στάδια παραγωγής του:

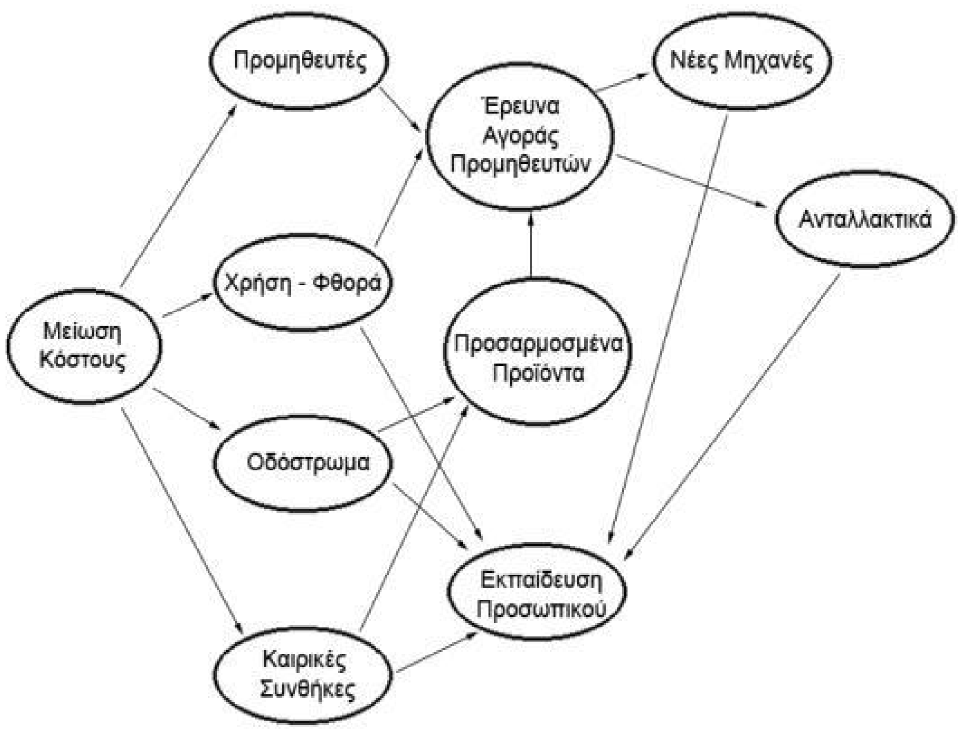
- Επιλογή θέματος ή σκοπού (πρόβλημα).
- Συλλογή δεδομένων (brainstorming).
- Μεταφορά των δεδομένων σε κάρτες.
- Ταξινόμηση των καρτών σε λογικές ομάδες (post-it).
- Ονοματολογία ομάδων καρτών.
- Διαγραμματική απεικόνιση του διαγράμματος.
- Παρουσίαση αποτελεσμάτων.



Εικόνα 1.1: Διάγραμμα Συγγένειας

Διάγραμμα Συσχετίσεων (Relations Diagram): Σκοπός του είναι η αναγνώριση, κατανόηση και απλοποίηση σύνθετων σχέσεων αιτίου-αποτελέσματος (causal relationships) σε ένα πολύπλοκο πρόβλημα, καθορίζοντας έτσι τους σημαντικούς παράγοντες για την επίλυση του [Dale, 1994].

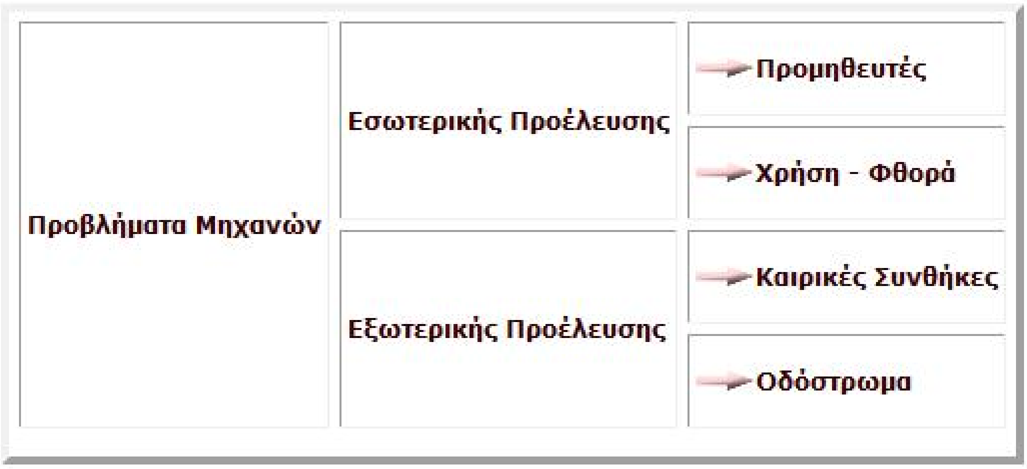
Το διάγραμμα συσχετίσεων χρησιμοποιείται για να διαπιστώσουμε ποιοι παράγοντες συνδέονται με ένα πρόβλημα. Πιο συγκεκριμένα, χρησιμοποιείται για να διακρίνουμε αλληλεπιδράσεις και υποπροβλήματα και μας βοηθά να τα προσδιορίσουμε και να τα συνδέσουμε υπό το πρίσμα της σχέσης αιτίου-αποτελέσματος και επίσης καθορίζει τα αίτια και τα αποτελέσματα των υποπροβλημάτων μιας γενικότερης κατάστασης.



Εικόνα 1.2: Διάγραμμα Συσχετίσεων

Συστηματικό Διάγραμμα (Systematic Diagram): Το συστηματικό διάγραμμα είναι μια γραφική ιεραρχική αναπαράσταση των προαπαιτουμένων βημάτων προς εκπλήρωση ενός στόχου ή μιας εργασίας (διάγραμμα δέντρου) [Dale, 1994].

Σκοπός του είναι η ανάπτυξη μιας ακολουθίας βημάτων, τα οποία εξαρτώνται το ένα από το άλλο και έτσι συνθέτουν την επίλυση του προβλήματος [Mizuno, 1988] και συνεισφέρει στο να αποσυνθέσουμε τον τελικό στόχο σε μικρότερους μέχρι να φτάσουμε σε ειδικές ενέργειες. Επίσης, δίνει τη δυνατότητα αποδόμησης ενός γενικότερου προβλήματος σε ειδικότερα, βοηθώντας στην εμβάθυνση των αιτιών των προβλημάτων.



Εικόνα 1.3: Συστηματικό Διάγραμμα

Διάγραμμα Μήτρας (Matrix Diagram): Το διάγραμμα μήτρας στοχεύει στην αποσαφήνιση των σχέσεων μεταξύ αποτελεσμάτων και αιτίων ή μεταξύ σκοπών και μεθόδων και την ανάδειξη της σημασία κάθε σχέσης [Dale, 1994].

Ταξινομεί τις ενέργειες κατά προτεραιότητα. Με τη χρήση του διαγράμματος μήτρας βελτιώνεται η αποτελεσματικότητα του συστήματος αξιολόγησης της ποιότητας, ενδυναμώνεται το σύστημα διασφάλισης ποιότητας συνδέοντας επίπεδα ποιότητας με διαφορετικές λειτουργίες ελέγχου και ανιχνεύονται τα αίτια μη συμμορφώσεων στην παραγωγική διαδικασία [Mizuno, 1988]. Οργανώνουμε, δηλαδή, συστηματικά τις πληροφορίες σε μορφή πινάκων, έτσι ώστε να μπορούν να συγκριθούν με βάση κάποια χαρακτηριστικά, τα οποία θα μας βοηθήσουν να κάνουμε συγκρίσεις ή επιλογές.

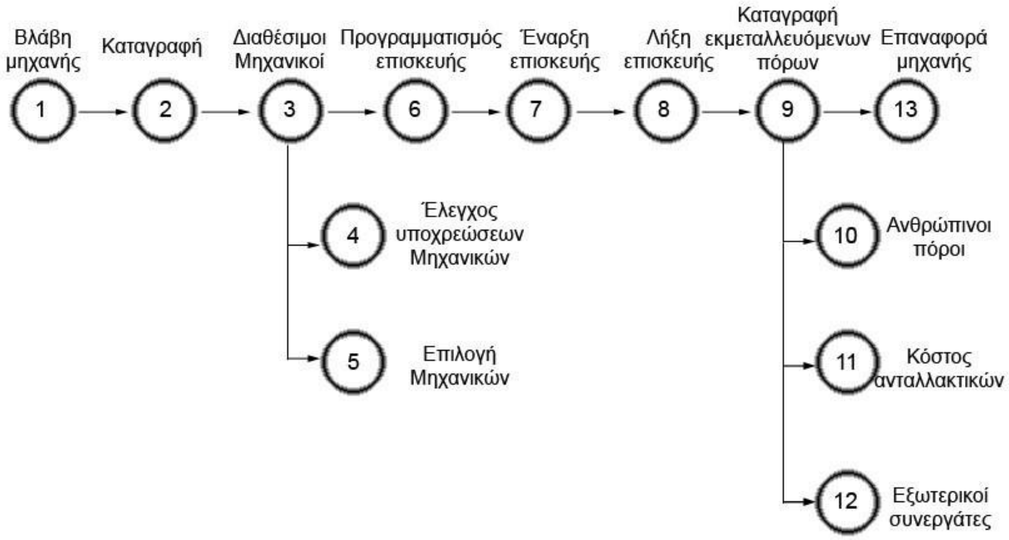
Το διάγραμμα μήτρας μπορεί να διαφωτίσει τις σχέσεις μεταξύ 2 ή περισσοτέρων παραγόντων και να παρουσιάσει μια οπτικοποιημένη μορφή των πιθανών συνδέσεων αυτών.

MATRIX DIAGRAM (L - Diagram) *ΚΑΤΗΓΟΡΙΑ ΒΛΑΒΗΣ/ΓΕΝΙΚΗ ΛΥΣΗ*	Προσαρμοσμένα Προϊόντα	Εκπαίδευση Προσωπικού	Έρευνα Αγοράς Προμηθευτών
Χρήση - Φθορά	-	✓	✓
Οδόστρωμα	✓	✓	-
Καιρικές Συνθήκες	✓	✓	-
Προμηθευτές	-	-	✓

Εικόνα 1.4: Διάγραμμα Μήτρας

Διάγραμμα Βέλους (Arrow Diagram): Το διάγραμμα βέλους χρησιμοποιείται για την ανάπτυξη του καλύτερου προγραμματισμού έργων και των κατάλληλων ελέγχων προκειμένου να επιτευχθούν οι στόχοι του έργου [Kanji & Asher 1996]. Ουσιαστικά, παρέχει τη δυνατότητα άμεσης ανάλυσης του χρονικού προγραμματισμού των διαδικασιών και της διαδοχικής ροής των αποτελεσμάτων τους.

Το διάγραμμα βέλους καταδεικνύει οπτικοποιημένα τον απαιτούμενο χρόνο για την επίλυση ενός προβλήματος και τις εργασίες που μπορούν να γίνουν παράλληλα, ενώ βοηθάει στην απεικόνιση της διαδοχής ενεργειών και γεγονότων [Lindsay & Petrick 1997].

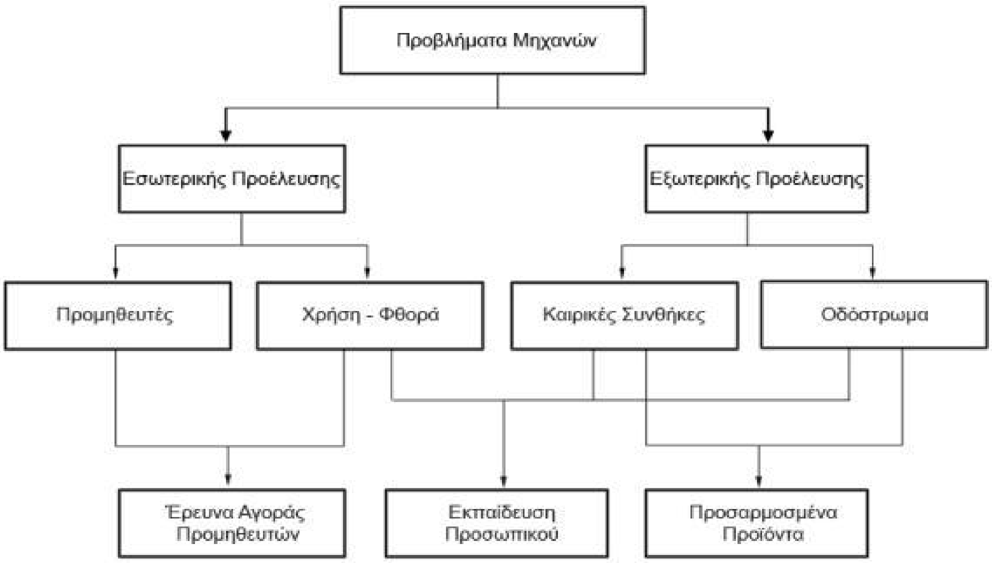


Εικόνα 1.5: Διάγραμμα Βέλους

Διάγραμμα Προγράμματος Απόφασης Διαδικασιών (Process decision program chart - PDPC): Το διάγραμμα προγράμματος απόφασης διαδικασιών βοηθάει στην εστίαση των πιθανών διαδοχών οι οποίες θα μας οδηγήσουν στην επίλυση του προβλήματος [Kanji & Asher 1996].

Χρησιμοποιείται κυρίως για το σχεδιασμό καινούργιων ή ανανεωμένων ενεργειών οι οποίες είναι πολύπλοκες και καθορίζει ποιες διαδικασίες πρέπει να χρησιμοποιηθούν με βάση διαδοχικά γεγονότα και πιθανές επιπτώσεις [Lindsay & Petrick 1997].

Αρχικά, παρουσιάζεται το γενικό πρόβλημα και στη συνεχεία αποδομείται μέχρι το ειδικότερο επίπεδο ανάλυσης των προβλημάτων και στο τέλος επιχειρείται μια σύνδεση μεταξύ των ειδικών προβλημάτων με τις πιθανές λύσεις που μπορούν να περιορίσουν τα προβλήματα.



Εικόνα 1.6: Διάγραμμα Προγράμματος Απόφασης Διαδικασιών

Ανάλυση Δεδομένων Διαγράμματος Μήτρας (Matrix Data Analysis): Σκοπός της ανάλυση δεδομένων διαγράμματος μήτρας είναι η ποσοτικοποίηση των δεδομένων του διαγράμματος μήτρας χρησιμοποιώντας μεθοδολογίες ανάλυσης δεδομένων.

Η ανάλυση δεδομένων διαγράμματος μήτρας δίνει την επιπλέον δυνατότητα, σε σχέση με το διάγραμμα μήτρας, της παρουσίασης ποσοτικών πληροφοριών που αφορούν τις διακριτές συνδέσεις μεταξύ δυο ή περισσοτέρων παραγόντων.

MATRIX DIAGRAM (L - Diagram) ***ΚΑΤΗΓΟΡΙΑ ΒΛΑΒΗΣ/ΓΕΝΙΚΗ ΛΥΣΗ***	**Προσαρμοσμένα Προϊόντα**	**Εκπαίδευση Προσωπικού**	**'Ερευνα Αγοράς Προμηθευτών**
Χρήση - Φθορά	-	3242	3958
Οδόστρωμα	698	247	
Καιρικές Συνθήκες	802	1017	-
Προμηθευτές	-	-	2513

Εικόνα 1.7: Ανάλυση Δεδομένων Διαγράμματος Μήτρας

Κεφάλαιο 2

Εξόρυξη Γνώσης

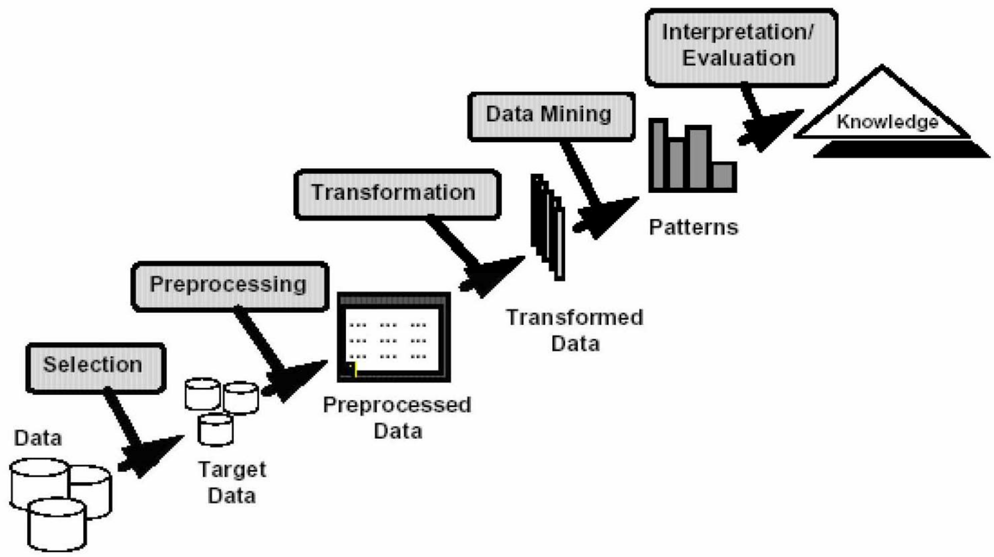
2.1 Εισαγωγή

Πέραν του ανθρώπινου δυναμικού, η πληροφορία αποτελεί τον πλέον πολύτιμο πόρο της σύγχρονης επιχείρησης. Ενώ η πληροφορία συνιστά καθοριστικό παράγοντα για την επίτευξη των επιχειρησιακών στόχων, συχνά παραμένει αποθηκευμένη σε βάσεις δεδομένων χωρίς να επιδέχεται επεξεργασίες ανάλυσης [Witten & Frank, 2005]. Η ντετερμινιστική διαδικασία αναγνώρισης έγκυρων, καινοτόμων, ενδεχομένως χρήσιμων και εν τέλει κατανοητών προτύπων στα δεδομένα και την πληροφορία ονομάζεται Ανακάλυψη Γνώσης από Βάσεις Δεδομένων (Knowledge Discovery from Databases-KDD) [Piatetsky-Shapiro & Frawley & Matheus, 1991].

Η Ανακάλυψη Γνώσης από Βάσεις Δεδομένων αναφέρεται στη διαδικασία αποκάλυψης ή παραγωγής λειτουργικής γνώσης που είναι κρυμμένη μέσα στα δεδομένα και δεν μπορεί να εξαχθεί με ευκολία από τον άνθρωπο. Η εξόρυξη δεδομένων/γνώσης/πληροφορίας (Data Mining), παρόλο που ουσιαστικά ταυτίζεται με την ανακάλυψη γνώσης από βάσεις δεδομένων, διαφέρει στο ότι αναφέρεται στις επί μέρους τεχνικές για την ανακάλυψη γνώσης.

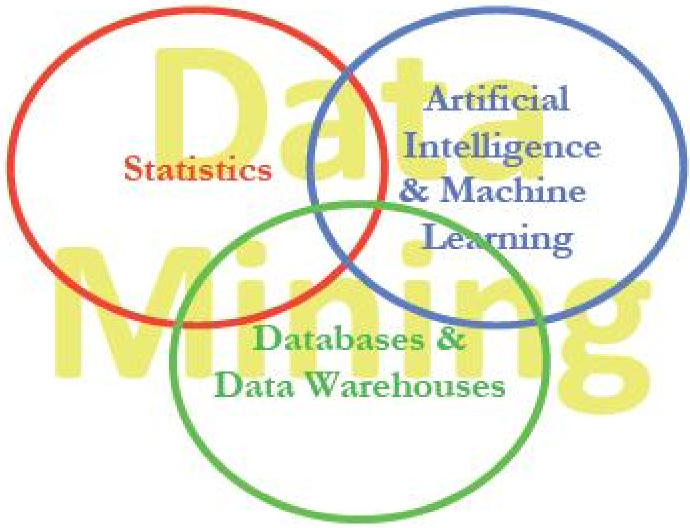
Επομένως, η εξόρυξη γνώσης αφορά στο στάδιο της ανακάλυψης γνώσης που αποτελείται από εφαρμοσμένες υπολογιστικές τεχνικές οι οποίες, κάτω από αποδεκτές προϋποθέσεις και περιορισμούς, παράγουν έναν αριθμό προτύπων και μοντέλων που πηγάζουν από τα δεδομένα [Fayyad & Piatetsky-Shapiro & Smyth & Ramasami, 1996].

Ο εντοπισμός προτύπων σε δεδομένα επιχειρήσεων δεν είναι ένα καινούργιο φαινόμενο. Ήταν, παραδοσιακά, ευθύνη των αναλυτών που γενικά χρησιμοποιούσαν στατιστικές μεθόδους. Τα τελευταία χρόνια, όμως, με την εξάπλωση των Η/Υ και των δικτύων, έχουν δημιουργηθεί τεράστιες βάσεις δεδομένων, τα δεδομένα των οποίων απαιτούν νέες τεχνικές για την ανάλυση τους και η εξόρυξη γνώσης καλύπτει αυτό ακριβώς το πεδίο [Bose & Mahapatra, 2001].



Εικόνα 2.1: Διαδικασία εξόρυξης πληροφορίας και γνώσης από τα δεδομένα

Το ερευνητικό πεδίο της εξόρυξης γνώσης αποτελείται από την τομή μεθόδων και εργαλείων που πηγάζουν από την στατιστική, τη μηχανική μάθηση και τις βάσεις & αποθήκες δεδομένων.



Εικόνα 2.2: Η τομή των ερευνητικών πεδίων, από όπου δημιουργείται η Εξόρυξη Γνώσης

Βέβαια, παρόλο που η εξόρυξη γνώσης χρησιμοποιεί πολλούς αλγορίθμους της Μηχανικής Μάθησης, υπάρχουν διακριτές διαφορές μεταξύ τους:

- Ο στόχος της Μηχανικής Μάθησης είναι η εύρεση τεχνικών μάθησης ως μίμηση της ανθρώπινης συμπεριφοράς (δηλαδή η παραγωγή προγραμμάτων που μαθαίνουν), ενώ της Εξόρυξης Γνώσης η ανακάλυψη πληροφορίας που θα είναι χρήσιμη στον άνθρωπο.

- Η Εξόρυξη Γνώσης χρησιμοποιεί μεγαλύτερες, ετερογενείς και «ακατέργαστες» βάσεις δεδομένων, ενώ η Μηχανική Μάθηση μικρότερες, ομογενείς και κατάλληλης μορφής.

Η χρήση της Εξόρυξης Γνώσης είναι ευρεία και παρουσιάζει αυξητικές τάσεις. Μπορεί να εφαρμοσθεί σε οποιοδήποτε οργανισμό ο οποίος συλλέγει δεδομένα και αναζητάει χρήσιμες πληροφορίες από την ανάλυση τους. Έχει χρησιμοποιηθεί στην ιατρική, στη βιομηχανία τροφίμων, στην ανάπτυξη λογισμικών, στην οικονομία, στον τραπεζικό-πιστωτικό τομέα, στις πωλήσεις-μάρκετινγκ, στις υπηρεσίες ίντερνετ, στη βελτιστοποίηση ποιότητας και σε άλλα. Επίσης, έχει γνωστοποιηθεί μια πλειάδα μελετών εφαρμογής (case studies) από γνωστούς οργανισμούς και επιχειρήσεις που, με τη βοήθεια τεχνικών εξόρυξης γνώσης, κατάφεραν να βελτιώσουν σημαντικά την ανάπτυξη τους, όπως για παράδειγμα η VISA, American Express, Softmap, Walmart, Victoria Secret.

Η εξόρυξη γνώσης, ειδικότερα, χρησιμοποιεί αλγόριθμους εντοπισμού προτύπων (patterns) και κανονικοτήτων σε δεδομένα, με σκοπό την κατανόηση και επεξήγηση των προτύπων, την ανάλυση των δεδομένων και την πρόβλεψη μελλοντικών πληροφοριών.

2.2 Μορφή Δεδομένων

Η κατανόηση και η επιλογή των δεδομένων τα οποία θα εισαχθούν στους αλγόριθμους Εξόρυξης Γνώσης είναι ένα πολύ σημαντικό στάδιο καθώς, μια τυχόν ασάφεια σε αυτά, παρεμποδίζει την τελική μετάφραση των αποτελεσμάτων και δυσχεραίνει την δημιουργία απτών αποτελεσμάτων για τα δεδομένα.

Εικόνα 2.3: Σχηματική αναπαράσταση της διαδικασίας της Εξόρυξης Γνώσης

Υπάρχουν τρεις βασικές συνιστώσες στα δεδομένα εισόδου στην Εξόρυξη Γνώσης:

Αντίληψη (concept): Το αντικείμενο της μάθησης.

Υπόδειγμα (instance): Το χωριστό και ανεξάρτητο παράδειγμα ενός concept.

Σύνολο Υποδειγμάτων (Dataset): Το σύνολο των δεδομένων.

Χαρακτηριστικό (attribute): Η μετρήσιμη συνιστώσα ενός υποδείγματος.

attributes

Outlook	Temperature	Humidity	Windy	**Play**
Sunny	85	85	False	**No**
Sunny	80	90	True	**No**
Overcast	83	86	False	**Yes**
Rainy	75	80	False	**Yes**
...	...	...	...	**...**

instances

Εικόνα 2.4: Παράδειγμα δεδομένων εισόδου

Πιο συγκεκριμένα, οι πιθανοί τύποι των χαρακτηριστικών (attributes) ανάλογα με το περιεχόμενο τους είναι οι παρακάτω:

Ονομαστικά (nominal): Οι τιμές των ονομαστικών χαρακτηριστικών είναι διακριτά σύμβολα και αποτελούν περιγραφή ή όνομα. Τα ονομαστικά χαρακτηριστικά καλούνται επίσης ρητά (categorical), απαριθμημένα (enumerated) ή διακριτά (discrete).

Τακτικά (ordinal): Στα τακτικά χαρακτηριστικά ορίζεται διάταξη στις τιμές αλλά δεν ορίζεται απόσταση μεταξύ τους.

Περιοδικά (Interval): Οι τιμές των περιοδικών χαρακτηριστικών δεν είναι μόνο διατεταγμένες αλλά μετρούνται σε σταθερές και ισαπέχουσες μονάδες. Ωστόσο, το άθροισμα ή το γινόμενο δεν έχει έννοια καθώς δεν έχει οριστεί το σημείο μηδέν.

Αναλογικά (Ratio): Πρόκειται για χαρακτηριστικά που διαφέρουν από τα περιοδικά μόνο ως προς τον ορισμό του σημείου μηδέν.

Από τους παραπάνω τύπους, στις περισσότερες των περιπτώσεων χρησιμοποιούνται μόνο δυο: τα ονομαστικά και τα τακτικά χαρακτηριστικά.

Κατά την είσοδο των δεδομένων πρέπει να καθοριστεί το είδος της μάθησης που επιθυμείται να προκύψει. Υπάρχουν τέσσερις τύποι:

Ταξινόμηση (Classification): Πραγματοποιείται ταξινόμηση υποδείγματος σε προκαθορισμένη τάξη.

Συσχέτιση (Association): Στοχεύει στην ανακάλυψη συσχετίσεων μεταξύ των διαφόρων χαρακτηριστικών .

Ομαδοποίηση (Clustering): Αποσκοπεί στην εύρεση ομάδων αντικειμένων με υψηλό βαθμό ομοιότητας.

Αριθμητική Πρόβλεψη (Numeric Prediction): Όμοια με την ταξινόμηση αλλά ο τύπος της τάξης είναι αριθμητικός.

Μετά την επιλογή των δεδομένων προς εξόρυξη και προτού αυτά εισέρθουν στην βασική αλγοριθμική διαδικασία, απαιτείται ένα πρώτο στάδιο επεξεργασίας τους. Η προεπεξεργασία μετασχηματίζει τα δεδομένα σε μια μορφή κατάλληλη για τον επιλεγμένο αλγόριθμο μάθησης. Παρακάτω αναφέρονται οι κύριες τεχνοτροπίες της προεπεξεργασίας δεδομένων:

Επιλογή χαρακτηριστικών: Στις περισσότερες εφαρμογές, ο αριθμός των χαρακτηριστικών είναι υπερβολικά μεγάλος για τον αποτελεσματικό χειρισμό τους από τα σχήματα εκμάθησης ενώ, τα περισσότερα από αυτά, είναι ευκρινώς μη συσχετιζόμενα ή περιττά. Κατά συνέπεια, πρέπει να επιλεγεί ένα υποσύνολο των δεδομένων προς χρήση στη διαδικασία εκμάθησης. Βέβαια, οι μέθοδοι εκμάθησης εκ φύσεως επιτελούν τη διαδικασία επιλογής των κατάλληλων χαρακτηριστικών και απόρριψης των υπολοίπων, ωστόσο η αποδοτικότητά τους, μπορεί συχνά να βελτιωθεί μέσω της προεπιλογής.

Διακριτοποίηση χαρακτηριστικών: Απολύτως αναγκαία στην περίπτωση που κάποια χαρακτηριστικά είναι αριθμητικά αλλά η επιλεγμένη μέθοδος εκμάθησης μπορεί να χειριστεί μόνο ρητά (categorical) χαρακτηριστικά. Ακόμα και οι μέθοδοι που μπορούν να χειριστούν αριθμητικά χαρακτηριστικά πολλές φορές παράγουν καλύτερα αποτελέσματα ή λειτουργούν ταχύτερα αν τα χαρακτηριστικά έχουν διακριτοποιηθεί. Το αντίστροφο πρόβλημα, στο οποίο ρήτα χαρακτηριστικά μετατρέπονται σε αριθμητικά, προκύπτει επίσης, αν και λιγότερο συχνά.

Μετασχηματισμός δεδομένων: Η μορφή των δεδομένων, είτε η φύση του αλγορίθμου εκμάθησης, συχνά υποδεικνύουν διάφορους μετασχηματισμούς των δεδομένων (μαθηματικοί, βασισμένοι στη γνώση πεδίου, λογικοί μετασχηματισμοί, αλλαγή δομής/μορφής δεδομένων). Ο μετασχηματισμός δεδομένων είναι αναγκαίος για την ανάδειξη ιδιαιτεροτήτων ή διαφορετικών γωνιών θέασης του συνόλου των δεδομένων.

Καθαρισμός δεδομένων: Τα ακάθαρτα δεδομένα συνιστούν σημαντικό πρόβλημα καθώς καταδικάζουν το αποτέλεσμα σε "θόρυβο".

2.3 Μορφή Εξαγόμενων

Οι τεχνικές Μηχανικής Μάθησης παρέχουν διάφορες δομικές περιγραφές των εξαγόμενων και καθεμία από αυτές υπαγορεύει το είδος του αλγορίθμου που πρέπει να χρησιμοποιηθεί για το σχηματισμό της από τα δεδομένα. Οι κυρίες μορφές εξαγόμενων είναι:

- **Πίνακες απόφασης (decision tables)**: Αποτελούν την απλούστερη μορφή εξαγόμενων.
- **Δένδρα απόφασης (decision trees)**: Παράγονται με την μέθοδο διαίρει και βασίλευε (divide & conquer).
- **Κανόνες απόφασης (decision rules) - Κανόνες ταξινόμησης (classification rules)**: Αποτελούν την πιο δημοφιλή μορφή εξαγόμενου. Είναι ισοδύναμοι με τα δένδρα απόφασης αλλά διαφέρει ο τρόπος αναπαράστασης τους.
- **Κανόνες συσχέτισης (association rules)**: Ειδοποιός (και μόνη ουσιαστική) διαφορά τους από τους κανόνες ταξινόμησης είναι ότι

δύνανται να προβλέψουν την τιμή κάθε χαρακτηριστικού ή συνδυασμού χαρακτηριστικών και όχι μόνο της τάξης.

- **Κανόνες με εξαιρέσεις (rules with exceptions)**: Προέκταση των κανόνων ταξινόμησης με την προσθήκη εξαιρέσεων.
- **Κανόνες με συσχετίσεις (rules involving relations)**: Βασίζεται στην σύγκριση των χαρακτηριστικών μεταξύ τους και προσφέρει μια αρτιότερη γενίκευση στα δεδομένα.
- **Γραμμική παλινδρόμηση (linear regression)**: Χρησιμοποιείται για την ανάλυση αριθμητικών δεδομένων και βασίζεται στις εξισώσεις παλινδρόμησης.
- **Δένδρα για αριθμητική πρόβλεψη (trees for numeric prediction)**: Το αντίστοιχο δένδρο απόφασης αλλά για αριθμητικά δεδομένα.
- **Απεικόνιση με βάση υποδείγματα (instance-based representation)**: Χρησιμοποιείται αρχικά μια ομάδα υποδειγμάτων και με βάση αυτά το υπόδειγμα ταξινομείται το κάθε νέο παράδειγμα.
- **Ομάδες (clusters)**: Χρησιμοποιείται όταν το ζητούμενο είναι η ομαδοποίηση και όχι η ταξινόμηση.
- **Πιθανοκρατικό γραφικό μοντέλο (probabilistic graphical model)**: Αποτυπώνει γραφικά τη σύνδεση μεταξύ των χαρακτηριστικών και της τάξης με κόμβους και με τόξα.

2.4 Κατηγορίες Αλγόριθμων

Υπάρχει ένα πλήθος μεθόδων εξαγωγής δομικών περιγραφών και προτύπων από τα δεδομένα. Η επιλογή των κατάλληλων μεθόδων εξαρτάται τόσο από τον τύπο των δεδομένων, όσο και από τη μεταξύ τους σχέση, το επιθυμητό αποτέλεσμα και την επιθυμητή μορφή εξαγωγής τους.

Στις περισσότερες περιπτώσεις συνιστάται ο πειραματισμός με αρκετούς αλγόριθμους, έτσι ώστε να βρεθεί ο κατάλληλος που απαντά στις περισσότερες από τις απαιτήσεις.

Ακολουθεί μια λίστα με τις πλέον δημοφιλείς κατηγορίες αλγορίθμων και μεθόδων που χρησιμοποιούνται για την εξόρυξη γνώσης και την ανάλυση πληροφορίας με τη δημιουργία κανόνων ταξινόμησης (classification rules).

- Decision tree method.

- Rule based method.
- Neural Networks.
- Bayesian Networks.
- Dependency Networks.
- Support Vector Machine.
- Rough Set.
- Genetic Algorithm.
- Fuzzy Logic.
- Bayesian Networks.
- Artificial Neural Networks.
- Quadratic Classifier.
- Radial Basis Function Networks.
- Inductive Logic Programming.

Από τις παραπάνω μεθόδους για την παραγωγή κανόνων ταξινόμησης θα αναλύσουμε περαιτέρω τις κατηγορίες μεθόδων Rough Set, Bayesian Networks, Rule based method και την κατηγορία παραγωγής association rules καθώς αυτές τις κατηγορίες χρησιμοποιούμε στην παρούσα εργασία. Η επιλογή των παραπάνω μεθόδων βασίζεται στα κριτήρια με τα οποία έγινε η διαλογή των επιλεγμένων αλγορίθμων που βασίζονται στους μεθόδους αυτούς, κριτήρια τα οποία αναπτύσσονται παρακάτω στην παράγραφο 3.4.

2.4.1 Rough Set

Η τεχνική Rough Set βασίζεται στη μαθηματική σύλληψη του Pawlak Zdzislaw στις αρχές του 1980 [Pawlak, 1988], [Pawlak, et al, 1991]. Παρέχει ορισμούς και αλγορίθμους για την εύρεση των χαρακτηριστικών (attributes) που διαχωρίζουν τη μια κλάση από την άλλη και παράγει προτασιακούς (propositional) κανόνες ταξινόμησης με χαρακτηριστική συντακτική και σημασιολογική ευκολία.

Το πλαίσιο Rough Set στηρίζεται στη μείωση των διακριτών υποσυνόλων που υπάρχουν σε ένα ευρύτερο σύνολο που δημιουργούν τα δεδομένα και τη μετατροπή τους σε κανόνες.

Η αποδοχή του πλαισίου Rough Set είναι αρκετά ευρεία καθώς παράγει κανόνες σε σύντομο αλγοριθμικό χρόνο και επιπλέον, έχει το προνόμιο της συντακτικής ευκολίας που βοηθά το χρήστη να αναγνωρίζει γρήγορα τους κανόνες.

Υπάρχουν αρκετές μέθοδοι αλγορίθμων που χρησιμοποιούν Rough Set, όπως για παράδειγμα οι γενετικοί αλγόριθμοι (genetic algorithm), οι εξαντλητικοί αλγόριθμοι (exhaustive algorithm), ο αλγόριθμος Johnson (Johnson's Algorithm), ο αλγόριθμος του Holte (Holte's Algorithm) και οι αλγόριθμοι κάλυψης (covering algorithm). Επίσης υπάρχουν και αρκετά λογισμικά που προσφέρουν εξόρυξη δεδομένων με τεχνικές Rough Set, αναφέροντας ενδεικτικά τα Rosetta, RSES, Datalogic και Grobian.

2.4.2 Bayesian Networks

Τα Bayesian Networks ή Belief Networks είναι πιθανοκρατικά γραφικά μοντέλα που αναπαριστούν ένα σύνολο μεταβλητών και τις πιθανοκρατικές μεταξύ τους σχέσεις. Παράγουν άκυκλους γράφους με κόμβους και τόξα όπου κάθε κόμβος αντιπροσωπεύει μια μεταβλητή και τα τόξα που ενώνουν τους κόμβους δηλώνουν την εξάρτηση μεταξύ των κόμβων.

Τα Bayesian Networks έχουν την δυνατότητα να παράγουν κανόνες συσχέτισης, συσχετίζοντας του κόμβους μεταξύ τους, καθώς και κανόνες ταξινόμησης, ορίζοντας τον κόμβο “στόχο” (target node). Το αποτέλεσμα των Bayesian Networks εκτός από το γράφο με τις συνδέσεις είναι και οι ποσοτικοποιημένες εξαρτήσεις των κόμβων.

Οι κυριότεροι αλγόριθμοι που υλοποιούν Bayesian Networks είναι οι Naïve Bayes, Sons & Spouses, Markov Blanket για κανόνες ταξινόμησης και οι SopLEQ, Taboo για κανόνες συσχέτισης. Μερικά από τα γνωστότερα λογισμικά παραγωγής Bayesian Networks, είναι τα BayesiaLab, Genie, Netica, BNsoftware, Bayesian belief network software, JBNC και άλλα.

2.4.3 Rule based method – Classification Rules

Οι κανόνες ταξινόμησης σκοπεύουν να ανακαλύψουν ένα σύνολο κανόνων σε “ακατέργαστα” δεδομένα χρησιμοποιώντας μια τεχνική ιεραρχημένων κλάσεων και κληρονομικότητας καθώς “σαρώνουν” τα δεδομένα. Στηρίζονται στους αλγορίθμους δένδρων απόφασης (decision trees) και διαφέρουν ως προς την απεικόνιση των κανόνων (με μορφή κανόνων αντί με μορφή δένδρου).

Γνωστότεροι αλγόριθμοι κανόνων ταξινόμησης είναι οι Prism, NNge, ConjuctiveRule, OneR, Part, M5Rules, Jrip και άλλοι ενώ, αρκετά δημοφιλή λογισμικά που παρέχουν κανόνες ταξινόμησης είναι τα Weka Suite και Compumine Rule Discovery.

2.4.4 Rule based method – Association Rules

Οι κανόνες συσχέτισης διαφοροποιούνται ως προς τους κανόνες ταξινόμησης μόνο ως προς το ότι, αντί να ταξινομούν τα χαρακτηριστικά σε μια τάξη, τα συσχετίζουν μεταξύ τους.

Οι κυριότεροι αλγόριθμοι που παράγουν κανόνες συσχέτισης είναι οι Tertius και Aprioi ενώ, γνωστά λογισμικά που παρέχουν κανόνες συσχέτισης είναι το Weka Suite, το CBA και το DM-II system.

2.5 Εξόρυξη γνώσης στη βελτιστοποίηση διαδικασιών

Η χρήση τεχνικών εξόρυξης γνώσης για τη βελτίωση των διαδικασιών από τους οργανισμούς, είναι ιδιαίτερα εκτεταμένη την τελευταία δεκαετία. Σε αποθηκευμένα δεδομένα προσπαθούν να αναλύσουν την πληροφορία και να προβούν σε συμπεράσματα που αφορούν στον καταναλωτή, τον εργαζόμενο, μια ομάδα εργαζομένων και τους συνεργάτες του. Αξιολογώντας τα αποτελέσματα οι οργανισμοί παρεμβαίνουν και βελτιώνουν τις διαδικασίες.

Είναι γνωστές αρκετές δημοσιευμένες μελέτες με αντικείμενο την εξόρυξη γνώσης για τη βελτίωση ποιότητας σε διαφόρους οργανισμούς, π.χ. σε αυτοκινητοβιομηχανίες, υπηρεσίες υγειάς, τηλεπικοινωνίες και σε άλλους. Αλλά δεν υπάρχουν εργασίες όπου χρησιμοποιείται η Διοίκηση Ολικής Ποιότητας ως μεθοδολογία βελτίωσης ποιότητας σε συνδυασμό με τις τεχνικές εξόρυξης γνώσης. Το ίδιο ισχύει και για τα 7 νέα εργαλεία ποιότητας (που αποτελούν βασικό συστατικό της Δ.Ο.Π.) για τα οποία δεν υπάρχουν έρευνες για το πώς μπορούνε να συνδυαστούν με τις τεχνικές εξόρυξης γνώσης με σκοπό τη βελτιστοποίηση ποιότητας.

Στην παρούσα, λοιπόν, εργασία χρησιμοποιούμε τα 7 νέα εργαλεία ποιότητας τα οποία δημιουργούνται δυναμικά με βάση τους παραγομένους κανόνες που προκύπτουν από την εξόρυξη των ακατέργαστων δεδομένων μιας κατασκευαστικής εταιρείας. Με γνώμονα τα εργαλεία αυτά και τις οδηγίες της Δ.Ο.Π. προβαίνουμε σε προτάσεις για τη βελτίωση ποιότητας της εταιρείας και των διαδικασιών της.

Κεφάλαιο 3

Εφαρμογή

3.1 Εισαγωγή

Στο παρόν κεφάλαιο περιγράφεται η εφαρμογή που αναπτύχθηκε με σκοπό τη δημιουργία ενός δυναμικού συστήματος παραγωγής των 7 νέων εργαλείων ποιότητας από κανόνες που προκύπτουν από την εξόρυξη γνώσης σε δεδομένα μιας κατασκευαστικής εταιρείας που βρίσκονται αποθηκευμένα σε μια βάση δεδομένων.

Αρχικά, περιγράφονται τα εργαλεία με τα οποία υλοποιήθηκε η εφαρμογή. Στη συνέχεια, παρουσιάζονται τα δεδομένα της κατασκευαστικής εταιρείας και ο τρόπος αποθήκευσης τους σε μια βάση δεδομένων. Ακολουθεί η περιγραφή εκείνων των επιλεγμένων τεχνικών εξόρυξης γνώσης που χρησιμοποιούνται για την ανάλυση των εκλεγόμενων υποσυνόλων των δεδομένων προς ανάλυση και το κεφάλαιο κλείνει με μια αναλυτική παρουσίαση της εφαρμογής και των δυνατοτήτων της.

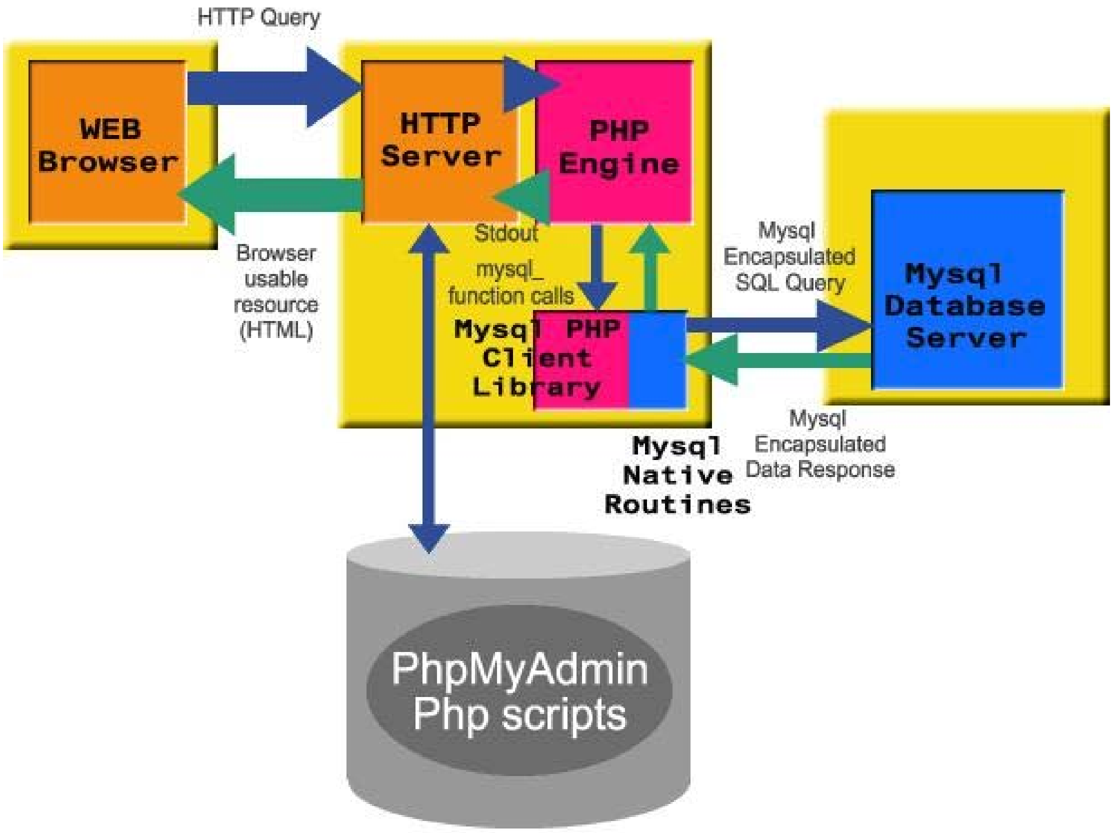
3.2 Εργαλεία Εφαρμογής

Η εφαρμογή είναι υλοποιημένη με την ιδία τεχνική που υλοποιούνται οι εφαρμογές βάσεων δεδομένων μέσω διαδικτύου (web database applications). Η επικοινωνία με τη βάση δεδομένων γίνεται μέσω φυλλομετρητή ιστοσελίδων (web browser). Ο χρήστης πληκτρολογεί στο web browser τη διεύθυνση όπου βρίσκεται το αρχικό αρχείο (html,php,asp,jsp κ.α.) εκκίνησης της εφαρμογής. Στη συνέχεια, δημιουργείται αίτημα ανάγνωσης του αρχείου από τον Web browser σε μορφή πρωτόκολλου μεταφοράς υπερκειμένου (HyperText Transfer Protocol, HTTP) το οποίο στέλνεται στο διακομιστή (server). Το λογισμικό που βρίσκεται εγκαταστημένο στον εξυπηρετητή (http server) αναζητά το αρχείο, το εντοπίζει και το στέλνει πίσω στον browser του πελάτη (client) μέσω του http πρωτοκόλλου. Ο browser στον υπολογιστή του client δέχεται την απάντηση και την "ανοίγει" στο χρηστή στη μορφή της περιγραφικής γλωσσάς HTML (HyperText Markup Language).

Στις περιπτώσεις όπου μέσα στο αρχείο που έχει ζητηθεί από τον browser του client περιέχεται αίτημα για σύνδεση με βάση δεδομένων, τότε δημιουργείται καινούργιο αίτημα από τον http server (μέσω μιας δυναμικής γλωσσάς προγραμματισμού και επικοινωνίας) προς τον database management system (λογισμικό αποθήκευσης και διαχείρισης βάσεων δεδομένων που αναλαμβάνει και την περάτωση των αιτήσεων που αποστέλλονται προς την βάση δεδομένων). Ο database server εκτελεί τις εντολές που περιέχονται στο αίτημα σε μορφή Structured Query Language (SQL) (εντολή *insert* για εισαγωγή νέων εγγραφών, εντολή *update* για ανανέωση υπαρχουσών αλλαγών, εντολή *delete* για διαγραφή εγγραφών και εντολή *select* για ανάκτηση εγγραφών) τις αποστέλλει στον http server, τις ενσωματώνει στο αιτούμενο αρχείο και το στέλνει με http πρωτόκολλο στον browser του client.

Στην παρούσα εφαρμογή ο http server, ο database server και ο browser του client βρίσκονται σε ένα υπολογιστή. Το μόνο που αλλάζει σε σύγκριση με τις εφαρμογές που βρίσκονται στο internet είναι ότι στο browser πληκτρολογείται αρχικά η διεύθυνση localhost (και ακολουθεί η διεύθυνση που βρίσκεται το αρχείο). Αυτό σημαίνει ουσιαστικά ότι η αίτηση στέλνεται μεν κανονικά με το ίδιο πρωτόκολλο, αλλά στον ίδιο υπολογιστή.

Στην εφαρμογή επιλέχθηκε ο Apache HTTP server (http://httpd.apache.org/) για http server, η MySQL (http://mysql.com/) για το λογισμικό διαχείρισης της βάσης δεδομένων (database management system DBMS) και η php (http://php.net/) ως μια γλώσσα προγραμματισμού για τη δημιουργία σελίδων web με δυναμικό περιεχόμενο.



Εικόνα 3.1: Η αλληλεπίδραση μεταξύ MySQL, Apache Server και php γλωσσά προγραμματισμού

Η χρησιμοποίηση των 3 αυτών εργαλείων MySQL, Apache και PHP είναι ένας αρκετά διαδεδομένος συνδυασμός καθώς, αφενός πρόκειται για τρία λογισμικά που διανέμονται δωρεάν και αφετέρου έχουν επιτύχει αρκετά υψηλές αποδόσεις στη λειτουργία τους. Συν τοις άλλης, η ευρεία αποδοχή τους έχει δημιουργήσει ένα μεγάλο δίκτυο υποστήριξης για τους προγραμματιστές για την εύκολη, λειτουργική και ασφαλή ανάπτυξη των εφαρμογών, ενώ δημοσιεύονται συνεχώς βελτιωμένες εκδόσεις και από τους τρεις οργανισμούς ανάπτυξης των λογισμικών αυτών.

Εκτός από τα προηγούμενα λογισμικά χρησιμοποιείται επίσης η τεχνολογία Flash (http://www.adobe.com/products/flash/) για την καλύτερη οπτικοποίηση των αποτελεσμάτων της εφαρμογής και η γλώσσα προγραμματισμού javascript (client-side scripting language) για να δίνει δυναμικότητα σε μια σελίδα χωρίς να χρειαστεί να σταλεί request στο server.

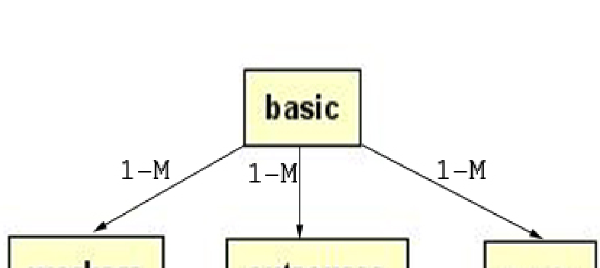
3.3 Μελέτη Περίπτωσης

Τα δεδομένα για τον πειραματισμό των μεθόδων εξόρυξης γνώσης ανήκουν στην εταιρεία ΛΑΤΟΜΙΚΗ Α.Ε.. Η ΛΑΤΟΜΙΚΗ Α.Ε. είναι μια εταιρεία που δραστηριοποιείται στο χώρο των κατασκευών δημόσιων έργων, των λατομείων και της παραγωγής και διανομής οπλισμένου σκυροδέματος. Εδρεύει στο Νομό Ηρακλείου αλλά δραστηριοποιείται σε όλο το νησί τα τελευταία είκοσι χρόνια και παρουσιάζεται στις πρώτες θέσεις ανάμεσα στις Κρητικές εταιρείες στον τομέα αυτό τόσο σε αναγνωσιμότητα, όσο και σε κύκλο εργασιών.

Από το 2005 και μετά, η εταιρεία διατηρεί μια ηλεκτρονική βάση δεδομένων όπου καταγράφεται οποιοδήποτε υφιστάμενο πρόβλημα στα μηχανήματα της εταιρείας. Ως μηχανήματα θεωρούνται όλοι οι τύποι οχημάτων που ανήκουν και χρησιμοποιούνται από την εταιρεία αλλά και τα μηχανήματα που χρησιμοποιούνται για την παραγωγή σκυροδέματος. Το σύνολο των καταχωρημένων περιστατικών είναι 12.477 για το χρονικό διάστημα μεταξύ 03-01-2005 και 08-11-2007.

Η βάση δεδομένων καταχώρησης προβλημάτων μηχανών ονομάζεται “ιστορικό μηχανής” και περιγράφεται συνοπτικά στη συνέχεια: Για κάθε βλάβη στις μηχανές καταγράφονται τα χαρακτηριστικά της βλάβης (όνομα μηχανής , τύπος μηχανής, χειριστής μηχανής , ημερομηνία βλάβης, προέλευση βλάβης, λύσεις κ.α.), οι εργαζόμενοι της εταιρείας που διόρθωσαν την βλάβη (ονοματεπώνυμο, ώρες εργασίας κ.α.), τα ανταλλακτικά που χρησιμοποιήθηκαν για την επισκευή (τύπος ανταλλακτικού, ονομασία, κόστος, τεμάχια κ.α.) και οι εξωτερικοί συνεργάτες που συνέβαλαν, είτε προσφέροντας ανταλλακτικά που δεν υπήρχαν στην αποθήκη της εταιρείας, είτε προσφέροντας εξειδικευμένο ανθρώπινο δυναμικό (όνομα συνεργάτη, κόστος για εργασία, κόστος για ανταλλακτικά).

Για την καταχώρηση των δεδομένων αυτών δημιουργήθηκε μια βάση δεδομένων στη MySQL με ένα βασικό πίνακα (table) που ονομάστηκε *basic* και τρεις πίνακες (*workers*, *parts*, *outsource* για την αποθήκευση των πληροφοριών που αφορούν στους εργαζόμενους, στα ανταλλακτικά και στους εξωτερικούς συνεργάτες αντίστοιχα), που συνδέονται σχεσιακά (relational) με τον πίνακα *basic* με σχέση 1-Μ (ένα προς πολλά).



Εικόνα 3.2: Οι πίνακες και οι σχέσεις μεταξύ τους

Ακολουθεί ο κατάλογος των πεδίων για κάθε πίνακα ξεχωριστά:

Κύριος πίνακας *basic* (αποθήκευση των πληροφοριών που αφορούν στη βλάβη της μηχανής):

- ΜΗΧΑΝΗ: Το όνομα της μηχανής (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΚΩΔΙΚΟΣ ΜΗΧΑΝΗΣ: Ο κωδικός της μηχανής (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΣΥΝΟΛΙΚΟ ΚΟΣΤΟΣ: Το συνολικό κόστος της βλάβης (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: πληκτρολόγηση).
- ΤΥΠΟΣ ΠΕΡΙΣΤΑΤΙΚΟΥ: Το είδος του περιστατικού (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΚΑΤΗΓΟΡΙΑ ΜΗΧΑΝΗΣ: Η κατηγορία στην οποία ανήκει η μηχανή (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΚΩΔΙΚΟΣ ΒΛΑΒΗΣ: Η κωδική ονομασία της βλάβης (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΚΑΤΗΓΟΡΙΑ ΒΛΑΒΗΣ: Η κατηγορία στην οποία ανήκει η βλάβη (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΠΡΟΕΛΕΥΣΗ ΒΛΑΒΗΣ: Η προέλευση της βλάβης (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΓΕΝΙΚΗ ΛΥΣΗ: Η γενική προτεινόμενη λύση για την βλάβη (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΕΙΔΙΚΗ ΛΥΣΗ: Η ειδική λύση που εφαρμόζεται μετά την επιλογή της γενικής λύσης (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΕΠΙΠΛΕΟΝ ΛΥΣΕΙΣ: Οι επιπλέον λύσεις που έπονται της ειδικής λύσης (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: επιλογή από κατάλογο).
- ΠΕΡΙΓΡΑΦΗ: Περιγραφή του περιστατικού (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΕΠΙΣΚΕΥΗ: Η περιγραφή της επισκευής που εφαρμόστηκε (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΚΟΣΤΟΣ ΕΡΓΑΣΙΑΣ ΤΕΧΝΙΤΩΝ: Το κόστος της εργασίας των τεχνικών που εργάζονται μέσα στην εταιρεία (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΑΙΤΩΝ: Ο εργαζόμενος που έκανε την αίτηση βλάβης (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- Α/Α: Ο αύξων αριθμός της βλάβης (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: αυτόματη εισαγωγή).

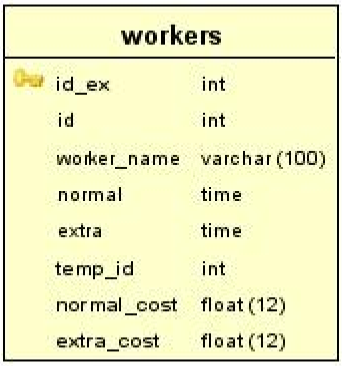
- ΚΑΤΑΓΡΑΦΗ ΠΕΡΙΣΤΑΤΙΚΟΥ: Η ημερομηνία και η ώρα καταγραφής της βλάβης (τύπος πεδίου: ημερομηνία – ώρα , τρόπος καταχώρησης: πληκτρολόγηση).
- ΕΝΑΡΞΗ ΕΠΙΣΚΕΥΗΣ: Η ημερομηνία και η ώρα της έναρξης της επισκευής (τύπος πεδίου: ημερομηνία – ώρα, τρόπος καταχώρησης: πληκτρολόγηση).
- ΛΗΞΗ ΕΠΙΣΚΕΥΗΣ: Η ημερομηνία και η ώρα λήξης της επισκευής (τύπος πεδίου: ημερομηνία – ώρα, τρόπος καταχώρησης: πληκτρολόγηση).
- ΔΙΑΡΚΕΙΑ ΕΠΙΣΚΕΥΗΣ: Η διάρκεια σε ώρες της επισκευής (τύπος πεδίου: ώρα , τρόπος καταχώρησης: πληκτρολόγηση).
- ΧΙΛΙΟΜΕΤΡΑ ΜΗΧΑΝΗΣ (ΓΙΑ ΤΡΟΧΟΦΟΡΑ): Τα χιλιόμετρα του τροχοφόρου τη στιγμή της βλάβης (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: πληκτρολόγηση).

basic	
id	int
number_up	int
type	varchar (100)
name	varchar (100)
full_name	varchar (100)
family	varchar (100)
time_find	datetime
start_fix	datetime
end_fix	datetime
duration	time
klm	bigint
person	varchar (100)
code	varchar (100)
description	varchar (100)
fixing	text (65535)
cost	float (12)
total	float (12)
part_temp	float (12)
out_temp	float (12)
time_test	int
origin	varchar (100)
category	varchar (100)
solution1	varchar (100)
solution2	varchar (100)
solution3	varchar (100)

Εικόνα 3.3: Ο πίνακας basic με τα πεδία και τους τύπους των πεδίων στη μορφή που τα παρουσιάζει η MySQL

Πίνακας *workers* (αποθήκευση των πληροφοριών που αφορούν στους εργαζόμενους της εταιρείας που ανέλαβαν την επισκευή της βλάβης):

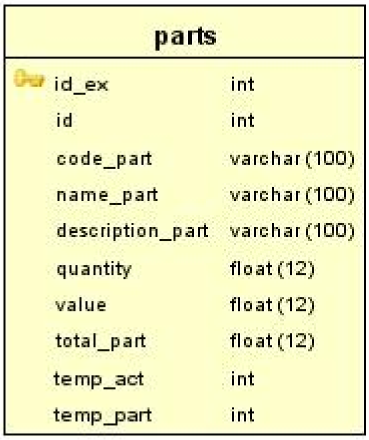
- ΟΝΟΜΑ ΤΕΧΝΙΤΗ: Το όνομα του εργαζόμενου στην εταιρεία τεχνίτη (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΚΑΝΟΝΙΚΗ ΑΠΑΣΧΟΛΗΣΗ: Η διάρκεια σε ώρες που κατανάλωσε ο εργαζόμενος στα πλαίσια του ημερήσιου ωραρίου για την επισκευή της συγκεκριμένης βλάβης (τύπος πεδίου: ώρα, τρόπος καταχώρησης: πληκτρολόγηση).
- ΥΠΕΡΩΡΙΕΣ: Η διάρκεια σε ώρες που κατανάλωσε ο εργαζόμενος πέρα από τα πλαίσια του ημερήσιου ωραρίου (υπερωρίες) για την επισκευή της συγκεκριμένης βλάβης (τύπος πεδίου: ώρα, τρόπος καταχώρησης: πληκτρολόγηση).



Εικόνα 3.4: Ο πίνακας workers με τα πεδία και τους τύπους των πεδίων στη μορφή που τα παρουσιάζει η MySQL

Πίνακας *parts* (αποθήκευση των πληροφοριών που αφορούν στα ανταλλακτικά που κατείχε η εταιρεία και χρησιμοποιήθηκαν για την επισκευή της βλάβης):

- ΚΩΔΙΚΟΣ: Ο κωδικός του ανταλλακτικού (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΟΝΟΜΑΣΙΑ: Η ονομασία του ανταλλακτικού (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΠΟΣΟΤΗΤΑ: Ο αριθμός τεμαχίων του ανταλλακτικού (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: πληκτρολόγηση).
- ΑΞΙΑ ΜΟΝΑΔΑΣ: Η χρηματική αξία του κάθε τεμαχίου (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: πληκτρολόγηση).
- ΣΥΝΟΛΟ: Το συνολικό κόστος του ανταλλακτικού [ποσότητα επί την αξία μονάδας] (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: αυτόματη).

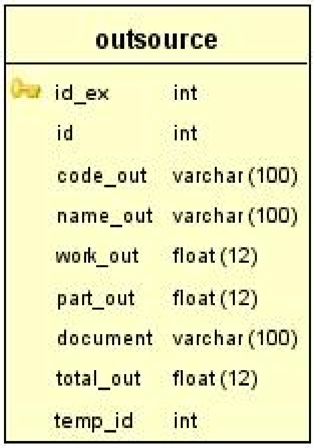


parts	
id_ex	int
id	int
code_part	varchar (100)
name_part	varchar (100)
description_part	varchar (100)
quantity	float (12)
value	float (12)
total_part	float (12)
temp_act	int
temp_part	int

Εικόνα 3.5: Ο πίνακας parts με τα πεδία και τους τύπους των πεδίων στη μορφή που τα παρουσιάζει η MySQL

Πίνακας *outsource* (αποθήκευση των πληροφοριών που αφορούν στους εξωτερικούς συνεργάτες που συμμετείχαν στην επισκευή της βλάβης είτε με την παροχή ανταλλακτικών είτε με την παροχή εξειδικευμένου προσωπικού):

- ΚΩΔΙΚΟΣ ΣΥΝΕΡΓΑΤΗ : Ο κωδικός του συνεργάτη (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).
- ΕΡΓΑΤΙΚΑ: Το χρηματικό κόστος που χρέωσε την εταιρεία ο συνεργάτης για την παροχή ανθρώπινου δυναμικού (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: πληκτρολόγηση).
- ΑΝΤΑΛΛΑΚΤΙΚΑ: Το χρηματικό κόστος με το οποίο χρέωσε την εταιρεία ο συνεργάτης για τα ανταλλακτικά (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: πληκτρολόγηση).
- ΣΥΝΟΛΟ: Το σύνολο του κόστους [εργατικά συν ανταλλακτικά] (τύπος πεδίου: αριθμός, τρόπος καταχώρησης: αυτόματος).
- ΠΑΡΑΣΤΑΤΙΚΟ: Ο κωδικός του τιμολογίου πληρωμής του συνεργάτη (τύπος πεδίου: κείμενο, τρόπος καταχώρησης: πληκτρολόγηση).



Εικόνα 3.6: Ο πίνακας outsource με τα πεδία και τους τύπους των πεδίων στη μορφή που τα παρουσιάζει η MySQL.

3.4 Επιλεγμένοι Αλγόριθμοι

Πριν αναφερθούν οι αλγόριθμοι που επιλέχθηκαν για την ανάλυση των δεδομένων, θα παρουσιαστούν μερικά χαρακτηριστικά των δεδομένων σύμφωνα με την ορολογία και τους κανόνες της εξόρυξης γνώσης.

Τα υπό επεξεργασία δεδομένα ανήκουν στην κατηγορία ονομαστικά (nominal), δηλαδή είναι περιγραφή ή όνομα. Επίσης, ο αριθμός των χαρακτηριστικών (attributes) που συσχετίζονται για να εισέλθουν ως είσοδος στους αλγόριθμους εξόρυξης γνώσης είναι πάντα 2. Αυτό γίνεται είτε “φυσικά”, όταν δηλαδή θέλουμε να συσχετίσουμε μόνο 2 χαρακτηριστικά, είτε “τεχνητά”, όταν δηλαδή μετασχηματίζουμε τα δεδομένα διαιρώντας ένα υπόδειγμα (instance) σε περισσότερα έως ότου τα συσχετιζόμενα χαρακτηριστικά να είναι 2. Για παράδειγμα, εάν έχουμε ένα υπόδειγμα με 4 χαρακτηριστικά (a, b, c, d), τότε δημιουργούνται 3 υποδείγματα, a – b, b – c, c – d. Προϋπόθεση για αυτόν το μετασχηματισμό είναι η διατήρηση της σημασίας και της έννοιας των συσχετιζόμενων δεδομένων και των εν δυνάμει κανόνων που περιέχονται μέσα στα δεδομένα αυτά. Επομένως, χρησιμοποιήσαμε το συγκεκριμένο μετασχηματισμό διασφαλίζοντας ότι μένει απολύτως ακέραιο το νόημα των δεδομένων το όποιο επιτυγχάνεται εξαιτίας του ότι κάθε χαρακτηριστικό εξαρτάται από ένα μόνο χαρακτηριστικό ή/και επηρεάζει μονό ένα. Εκτός από τον προαναφερθέντα μετασχηματισμό χρησιμοποιήθηκε και κωδικοποίηση των δεδομένων (μετατροπή των τιμών των χαρακτηριστικών από ονομαστικά σε αριθμητικά-αναλογικά) σε όποιες περιπτώσεις το απαιτούσαν οι αλγόριθμοι.

Έχοντας συγκεκριμενοποιήσει τα δεδομένα και τους πιθανούς τους μετασχηματισμούς και κωδικοποιήσεις, αναζητήσαμε τους κατάλληλους αλγορίθμους οι οποίοι θα μας οδηγούσαν στα επιθυμητά αποτελέσματα. Η μοναδική αλλά αναγκαία προϋπόθεση για να επιλεχθεί ένας αλγόριθμος ήταν να παράγει όλους τους διακριτούς κανόνες που περιέχονταν στα δεδομένα. Αναλυτικότερα, αν στις 12.477 εγγραφές (περιπτώσεις βλαβών) που είναι καταχωρημένες στη βάση δεδομένων κάποιος αλγόριθμος αδυνατεί να βρει έναν κανόνα ο οποίος να αντιπροσωπεύει έστω και 1 εγγραφή στις 12477, τότε αυτός ο αλγόριθμος αποκλείεται από τη διαδικασία. Επιλέχθηκαν λοιπόν οι αλγόριθμοι οι οποίοι παρήγαγαν όλους τους κανόνες που περιέχονται στα επιλεγμένα προς ανάλυση δεδομένα της βάσης δεδομένων, ακόμα και αν οι κανόνες αφορούσαν μόνο σε ελάχιστες περιπτώσεις στη βάση δεδομένων. Ο λόγος που τέθηκε αυτή η προϋπόθεση ήταν η επιθυμία να στραφεί ο προσανατολισμός του τύπου του αλγορίθμου προς τη συνολική ανακάλυψη των “κρυμμένων” κανόνων στα δεδομένα, με μελλοντική σκέψη την εφαρμογή των αλγορίθμων σε δεδομένα μηδενικής ανοχής λάθους.

Συν τοις άλλης, θετικό παράγοντα στην επιλογή του αλγορίθμου είχε η ευκρίνεια των κανόνων στην “έξοδο” του κάθε αλγορίθμου αλλά και η παροχή της επιπρόσθετης πληροφορίας που αφορούσε στον αριθμό των περιπτώσεων που συμφωνούν με τον κάθε κανόνα.

Με βάση, λοιπόν, την παραπάνω αναγκαία προϋπόθεση επιλεχθήκαν οι κάτωθι αλγόριθμοι:

- Ο αλγόριθμος Tertius, που παράγει κανόνες συσχέτισης (association rules) από το λογισμικό Weka Suite.
- Ο αλγόριθμος Prism, που παράγει κανόνες ταξινόμησης (classifications rules) από το λογισμικό Weka Suite.
- Ο αλγόριθμος Naive Bayes, που χρησιμοποιείται στην κατηγορία Bayesian Networks από το λογισμικό BayesiaLab.
- Ο αλγόριθμος Holte’s 1R, που χρησιμοποιείται στην κατηγορία Rough Set από το λογισμικό Rosetta.
- Ο αλγόριθμος covering algorithm, που χρησιμοποιείται στην κατηγορία Rough Set από το λογισμικό RSES2 (RSES version: 2).

3.4.1 Αλγόριθμος Tertius

Ο αλγόριθμος Tertius παράγει κανόνες συσχέτισης και προσφέρεται μέσα από το λογισμικό Weka Suite.

Ο αλγόριθμος Tertius υλοποιεί μια εξερεύνηση κανόνων στα δεδομένα με unsupervised τεχνική (χωρίς, δηλαδή, να προσδιορίζεται από το χρήστη το αποτέλεσμα του κάθε κανόνα) και με first order μέθοδο (μια μέθοδο συμπερασμάτων όπου χρησιμοποιούνται μεταβλητές αντί των πραγματικών τιμών, με σκοπό την εύρεση ομοιοτήτων και ιδιοτήτων), στηριζόμενος σε υποθέσεις [Flach & Lachiche, 1999a]. Ο κώδικας του Tertius προσφέρεται δωρεάν για ερευνητικούς σκοπούς [Flach & Lachiche, 1999b].

Ακολουθεί ένα παράδειγμα εξόδου του αλγόριθμου Tertius της σουίτας Weka.

```
=== Run information ===

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: systematic
Instances: 12477
Attributes: 2
category
origin
=== Associator model (full training set) ===


Tertius
=======

1. /* 0,556518 0,000000 */ category = Χρήση_-_Φθορά ==> origin = Εσωτερικής_Προέλευσης
2. /* 0,508036 0,000000 */ category = Καιρικές_Συνθήκες ==> origin = Εξωτερικής_Προέλευσης
3. /* 0,320688 0,000000 */ category = Οδόστρωμα ==> origin = Εξωτερικής_Προέλευσης
4. /* 0,267796 0,000000 */ category = Προμηθευτές ==> origin = Εσωτερικής_Προέλευσης
5. /* 0,000000 0,075739 */ category = Οδόστρωμα ==> FALSE
6. /* 0,000000 0,145788 */ category = Καιρικές_Συνθήκες ==> FALSE
7. /* 0,000000 0,201411 */ category = Προμηθευτές ==> FALSE
8. /* 0,000000 0,221528 */ TRUE ==> origin = Εσωτερικής_Προέλευσης
9. /* 0,000000 0,577062 */ category = Χρήση_-_Φθορά ==> FALSE
10. /* 0,000000 0,778472 */ TRUE ==> origin = Εξωτερικής_Προέλευσης

Number of hypotheses considered: 14
Number of hypotheses explored: 14
Time: 00 min 00 s 156 ms
```

Στην πρώτη σειρά της παραγράφου *Run information* αναφέρονται τα εξής:

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0.

Αυτή είναι η κωδικοποιημένη παρουσίαση των επιλογών παραμετροποίησης που έχει ο χρήστης πριν τρέξει τον αλγόριθμο. Επιγραμματικά αναφέρουμε τις ερμηνείες των παραμετροποιήσεων αυτών.

classIndex – Η θέση του χαρακτηριστικού που έχει οριστεί ως τάξη.

classification – Επιλογή εύρεσης κανόνων ταξινόμησης.

confirmationThreshold – Ελάχιστο όριο επιβεβαίωσης των κανόνων.

confirmationValues – Αριθμός των καλύτερων τιμών επιβεβαίωσης προς εύρεση.

frequencyThreshold – Ελάχιστο ποσοστό των περιπτώσεων που καλύπτονται από τους κανόνες.

hornClauses – Επιστρόφη των κανόνων κατά τους οποίους ισχύει ένας κανόνας για κάθε τάξη.

missingValues – Επιλογή του τρόπου αντιμετώπισης των χαμένων τιμών (missing values).

negation – Αναφέρεται στον τρόπο αναίρεσης ενός κανόνα.

noiseThreshold – Ανώτατο όριο του ποσοστού αναίρεσης που μπορεί να έχει ένας κανόνας για να θεωρηθεί ικανοποιητικός.

numberLiterals – Ο συνολικός αριθμός των χαρακτηριστικών που δύναται να εμφανίζονται σε ένα κανόνα.

repeatLiterals – Επιτρέπει την επανάληψη των χαρακτηριστικών.

rocAnalysis – Υπολογισμός της ανάλυσης ROC [Receiver Operating Characteristic] (Απεικονίζει την απόδοση ενός ταξινομητή).

valuesOutput – Δίνει την δυνατότητα οπτικοποίησης των καλύτερων και των χειρότερων τιμών κατά την διάρκεια της επεξεργασίας.

Στην επόμενη γραμμή αναφέρεται το όνομα του αρχείου (Relation: systematic), ακολουθεί ο αριθμός των υποδειγμάτων (Instances: 12.477), ο αριθμός των χαρακτηριστικών (Attributes: 2) και η λίστα με τα χαρακτηριστικά (category origin).

Στην παράγραφο *tertius* παρουσιάζεται η λίστα των κανόνων που ανακάλυψε ο αλγόριθμος. Για παράδειγμα, ο πρώτος κανόνας αναφέρει:

1. /* 0,556518 0,000000 */ category = Χρήση_-_Φθορά ==> origin = Εσωτερικής_Προέλευσης

Ο κανόνας αυτός μεταφράζεται ως εξής: *Αν η τιμή του χαρακτηριστικού category είναι ίση με <u>Χρήση_-_Φθορά</u> τότε η τιμή του χαρακτηριστικού origin είναι <u>Εσωτερικής_Προέλευσης</u>.*

Ο πρώτος από τους δυο αριθμούς που παρουσιάζονται, αναφέρεται στο δείκτη που δείχνει την τιμή επιβεβαίωσης του κανόνα, ενώ ο δεύτερος αριθμός, αναφέρεται στο δείκτη που καταδεικνύει τη συχνότητα εμφάνισης αναιρετικών προς το συγκεκριμένο κανόνα υποδειγμάτων.

Αφού παρουσιαστούν όλοι οι κανόνες συσχέτισης (σε αύξουσα σειρά με βάση το δείκτη επιβεβαίωσης του κανόνα), ακολουθεί ο αριθμός των υποθέσεων που δημιούργησε και ερεύνησε ο αλγόριθμος κατά τη διάρκεια της λειτουργίας του (Number of hypotheses considered: 14 Number of hypotheses explored: 14). Τα αρχείο εξόδου του Tertius κλείνει με την παρουσίαση του χρόνου επεξεργασίας του αλγορίθμου (Time: 00 min 00 s 156 ms).

Στον αλγόριθμο Tertius η μόνη τροποποίηση που ήταν απαραίτητη στα δεδομένα εισόδου ήταν η αντικατάσταση του χαρακτήρα του κενού (space character) με το χαρακτήρα underscore (_).

3.4.2 Αλγόριθμος Prism

Ο αλγόριθμος Prism παράγει κανόνες ταξινόμησης και προσφέρεται μέσα από το λογισμικό Weka Suite.

Ο αλγόριθμος Prism πρωτοπαρουσιάστηκε από τον Jadia Cendrowska το 1987 [Cendrowska, 1987]. Θεωρείται μετεξέλιξη του γνωστού αλγορίθμου παραγωγής δένδρων αποφάσεων (decision trees) ID3 [Quinlan, 1986]. Η βασική διαφορά με τον ID3 είναι ότι ο PRISM παράγει κατευθείαν κανόνες ταξινόμησης και όχι δένδρα αποφάσεων που τροποποιούνται σε κανόνες ταξινόμησης. Παρακάτω παρουσιάζεται ο ψευδοαλγόριθμος του PRISM.

For i=1 to n
repeat until all instances of class i have been removed
    (1) Calculate *the probability of occurrence of class i for each attribute-value pair.*
    (2) Select *the attribute-value pair with maximum probability and create a subset of the training set comprising all instances with the selected combination.*
    (3) Repeat *(1) and (2) for this subset until it contains only instances of class i. The induced rule is then the conjunction of all the attribute-value pairs selected in creating this subset.*
    (4) Remove *all instances covered by this rule from training set.*

Ακολουθεί ένα παράδειγμα εξόδου του αλγόριθμου Prism της σουίτας Weka.

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: systematic
Instances: 12477
Attributes: 2
category
origin
Test mode: 10-fold cross-validation

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

Prism rules

If category = Χρήση_-_Φθορά then Εσωτερικής_Προέλευσης
If category = Προμηθευτές then Εσωτερικής_Προέλευσης
If category = Καιρικές_Συνθήκες then Εξωτερικής_Προέλευσης
If category = Οδόστρωμα then Εξωτερικής_Προέλευσης

Time taken to build model: 0.14 seconds

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

Correctly Classified Instances 12477 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 12477

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
1 0 1 1 1 Εσωτερικής_Προέλευσης
1 0 1 1 1 Εξωτερικής_Προέλευσης

=== Confusion Matrix ===

a b <-- classified as
9713 0 | a = Εσωτερικής_Προέλευσης
0 2764 | b = Εξωτερικής_Προέλευσης

Στην παράγραφο *Run information* αναφέρεται το όνομα του αρχείου (Relation: systematic), ακολουθεί ο αριθμός των υποδειγμάτων (Instances: 12.477), ο αριθμός των χαρακτηριστικών (Attributes: 2), η λίστα με τα χαρακτηριστικά (category origin) και ο τρόπος εκπαίδευσης του αλγορίθμου (Test mode: 10-fold cross-validation). Το Weka Suite προσφέρει 4 τρόπους εκπαίδευσης του Prism: training set (θεώρηση ενός υποσυνόλου των δεδομένων ως σύνολο εκπαίδευσης), supplied test set (εφοδιασμός του αλγορίθμου με ένα εξωτερικό σύνολο δεδομένων για εκπαίδευση), cross-validation (διασταυρωμένη επικύρωση), percentage split (διαχωρισμός του συνόλου των υποδειγμάτων σε υποδείγματα προς εκπαίδευση και υποδείγματα προς έλεγχο).

Πιο συγκεκριμένα, για τον τρόπο εκπαίδευσης cross-validation (τον οποίο και χρησιμοποιήσαμε στην παρούσα εργασία) επεξηγούμε ότι διασπάει τα δεδομένα σε k ισομεγέθη υποσύνολα (το k στο παράδειγμα είναι 10 και καθορίζεται από το χρήστη), χρησιμοποιεί το κάθε υποσύνολο διαδοχικά για έλεγχο και τα υπόλοιπα δεδομένα για εκπαίδευση και στο τέλος διασταυρώνει τα αποτελέσματα. Σημειώνουμε ότι συνήθως ο αριθμός k επιλέγεται να ισούται με 10 καθώς πειραματικά έχει αποδειχτεί πως πρόκειται για τη βέλτιστη επιλογή για λήψη αξιόπιστης εκτίμησης, ανεξαρτήτως του τύπου των δεδομένων.

Στην παράγραφο *Prism rules* παρουσιάζονται οι κανόνες που έχουν προκύψει. Οι κανόνες χαρακτηρίζονται από ευκολία κατανόησης καθώς προσφέρουν συντακτική βοήθεια. Για παράδειγμα, ο πρώτος κανόνας αναφέρει: If category = Χρήση_-_Φθορά then Εσωτερικής_Προέλευσης. Είναι εύκολο να συμπεράνουμε το νόημα αυτού του κανόνα. *Αν η κατηγορία ισούται με Χρήση -Φθορά τότε η τάξη* (στο παράδειγμα η τάξη είναι το χαρακτηριστικό *origin*) *έχει τιμή Εσωτερικής Προέλευσης*.

Στην παράγραφο *summary* παρουσιάζονται διάφορα στατιστικά στοιχεία και δείκτες σφάλματος. Αρχικά, παρουσιάζεται ο απόλυτος αριθμός και το ποσοστό των σωστών προβλέψεων του αλγορίθμου (Correctly Classified Instances 12477 100 %) ενώ ακολουθούν ο απόλυτος αριθμός και το ποσοστό των λάθων προβλέψεων του αλγορίθμου (Incorrectly Classified Instances 0 0 %), ο στατιστικός δείκτης kappa stastistic (Kappa statistic 1) που δείχνει το βαθμό συμφωνίας των παραγόμενων κανόνων με την πραγματικότητα, μια σειρά δεικτών σφάλματος (Mean absolute error 0,Root mean squared error 0,Relative absolute error 0 %,Root relative squared error 0 %) και ο συνολικός αριθμός υποδειγμάτων (Total Number of Instances 12477).

Στην παράγραφο *Detailed Accuracy By Class* παρουσιάζεται ένας πίνακας που δείχνει διάφορους δείκτες για το ποσοστό επιτυχίας ταξινόμησης κάθε κλάσης. Κατά σειρά εμφανίζονται οι εξής δείκτες: το ποσοστό αληθών θετικών προτάσεων στο δείγμα (TP Rate), το ποσοστό ψευδών θετικών προτάσεων στο δείγμα (FP Rate), το κλάσμα αληθών θετικών δια το άθροισμα αληθών θετικών συν ψευδών θετικών (Precision), το κλάσμα αληθών θετικών δια το άθροισμα αληθών θετικών συν ψευδών αρνητικών [FN Rate] (Recall) και ο αρμονικός μέσος του *precision* και του *recall* (*F-measure)* που ισούται με 2TP/(2TP+FP+FN) (F-Measure). Η τελευταία στήλη του πίνακα είναι το όνομα της τάξης (Class).

Στην τελευταία παράγραφο παρουσιάζεται ο πίνακας σύγχυσης (confusion matrix), ο οποίος δείχνει σε απόλυτους αριθμούς τις επιδόσεις του αλγορίθμου, αναφέροντας πόσα από τα υποδείγματα κατάφερε να κατατάξει σωστά και πόσα όχι και την τάξη στην οποία κατέταξε τα λανθασμένα.

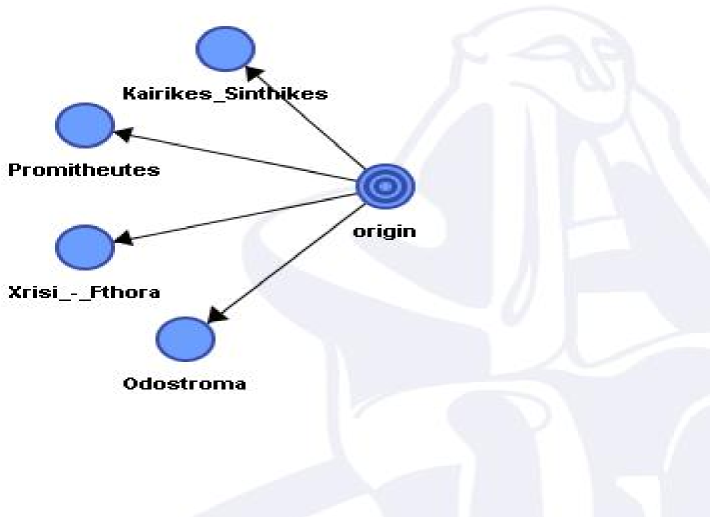
Στον αλγόριθμο Prism όπως και στον Tertius αντικαταστήθηκε ο χαρακτήρας του κενού (space character) με το χαρακτήρα underscore (_) στα υποδείγματα εισόδου.

3.4.3 Αλγόριθμος Naive Bayes

Ο αλγόριθμος Naive Bayes παράγει κανόνες ταξινόμησης μέσω της μεθόδου Bayesian networks και προσφέρεται μέσα από το λογισμικό BayesiaLab.

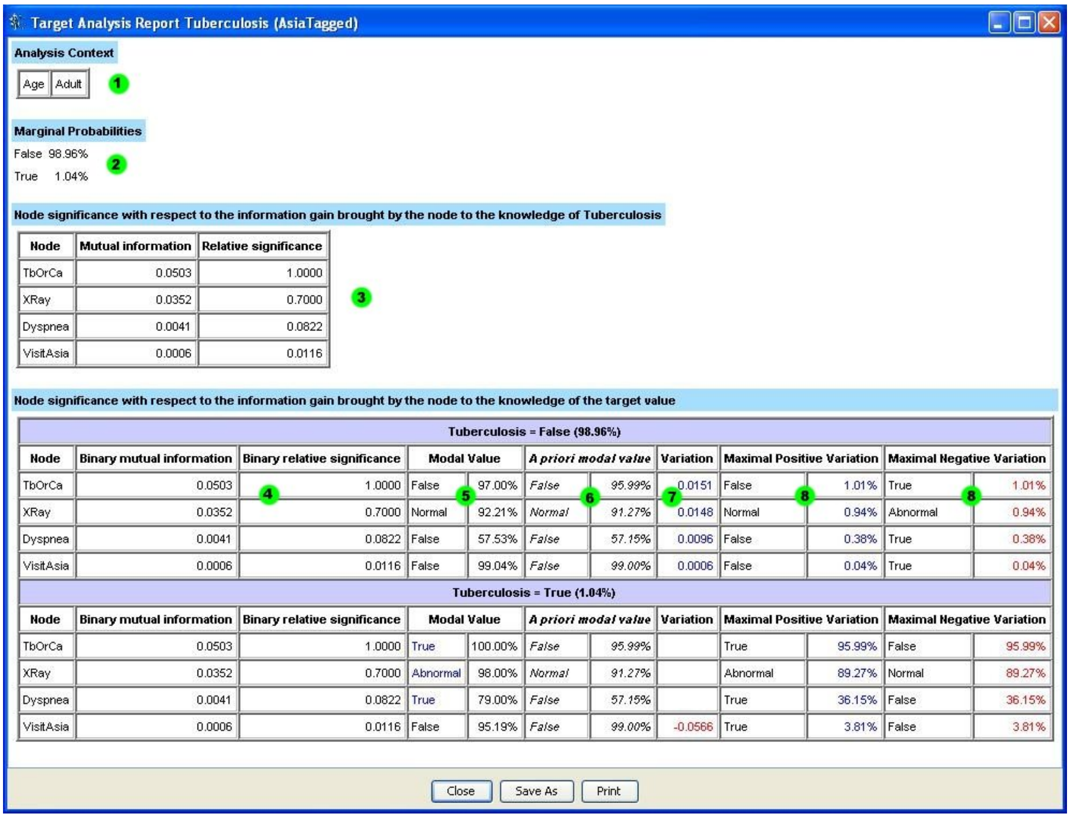
Ο αλγόριθμος Naive Bayes είναι ένας απλός πιθανοκρατικός αλγόριθμος ταξινόμησης. Χρησιμοποιείται από το λογισμικό BayesiaLab και απαιτεί να έχει καθοριστεί ένας κόμβος "στόχος" (target node) ο οποίος θεωρείται "πατέρας" όλων των κόμβων (parent node), προσδίδοντας ανεξαρτησία μεταξύ των υπόλοιπων κόμβων. Τα προνόμια του Naive Bayes είναι η ευρωστία του και ο γρήγορος χρόνος απόκρισης .

Ακολουθεί ένα παράδειγμα εξόδου του αλγόριθμου Naive Bayes του λογισμικού BayesiaLab.



Εικόνα 3.7: Ο παραγόμενος γράφος του Naïve Bayes στο BayesiaLab

Στην εικόνα 7 απεικονίζεται ο παραγόμενος γράφος του αλγορίθμου Naïve Bayes στο λογισμικό BayesiaLab. Είναι εμφανής ο κόμβος "στόχος" "origin" (target node) με βάση τον οποίο γίνεται η ταξινόμηση.



Target Analysis Report Tuberculosis (AsiaTagged)

Analysis Context

Age	Adult

1

Marginal Probabilities

False 98.96%

True 1.04%

2

Node significance with respect to the information gain brought by the node to the knowledge of Tuberculosis

Node	Mutual information	Relative significance
TbOrCa	0.0503	1.0000
XRay	0.0352	0.7000
Dyspnea	0.0041	0.0822
VisitAsia	0.0006	0.0116

3

Node significance with respect to the information gain brought by the node to the knowledge of the target value

Tuberculosis = False (98.96%)											
Node	Binary mutual information	Binary relative significance	Modal Value		*A priori modal value*		Variation	Maximal Positive Variation		Maximal Negative Variation	
TbOrCa	0.0503	1.0000	False	97.00%	*False*	*95.99%*	0.0151	False	1.01%	True	1.01%
XRay	0.0352	0.7000	Normal	92.21%	*Normal*	*91.27%*	0.0148	Normal	0.94%	Abnormal	0.94%
Dyspnea	0.0041	0.0822	False	57.53%	*False*	*57.15%*	0.0096	False	0.38%	True	0.38%
VisitAsia	0.0006	0.0116	False	99.04%	*False*	*99.00%*	0.0006	False	0.04%	True	0.04%
Tuberculosis = True (1.04%)											
Node	Binary mutual information	Binary relative significance	Modal Value		*A priori modal value*		Variation	Maximal Positive Variation		Maximal Negative Variation	
TbOrCa	0.0503	1.0000	True	100.00%	*False*	*95.99%*		True	95.99%	False	95.99%
XRay	0.0352	0.7000	Abnormal	98.00%	*Normal*	*91.27%*		Abnormal	89.27%	Normal	89.27%
Dyspnea	0.0041	0.0822	True	79.00%	*False*	*57.15%*		True	36.15%	False	36.15%
VisitAsia	0.0006	0.0116	False	95.19%	*False*	*99.00%*	-0.0566	True	3.81%	False	3.81%

4 5 6 7 8 8

Close Save As Print

Εικόνα 3.8: Η παραγόμενη αναλυτική αναφορά του target node του Naïve Bayes στο BayesiaLab (Target Node Analysis)

Στην εικόνα 3.8 απεικονίζεται η ανάλυση των σχέσεων και των εξαρτήσεων του target node με τους υπολοίπους κόμβους. Ακολουθεί αναλυτική περιγραφή των δεικτών που απαρτίζουν την ανάλυση αυτή (στην εικόνα 3.8 υπάρχουν αριθμητικά σημάδια με βάση τα οποία γίνεται η επεξήγηση):

1. **Context of the analysis**: Περιγραφή των παρατηρώμενων (observed) μεταβλητών.

2. **Marginal probability distribution**: Κατανομή των πιθανοτήτων του target node.

3. **Node relative significance with respect to the information gain brought by the node to the knowledge of the target node**: Η λίστα των κόμβων παιδιών (child nodes) οι οποίοι προσφέρουν γνώση στο target node.

4. **Node relative significance with respect to the information gain brought by the node to the knowledge of the target value**: Για κάθε τιμή του target node παρουσιάζεται η λίστα των κόμβων που συσχετίζονται με το target node, ταξινομημένη αντιστρόφως με βάση τη σχετική συνεισφορά τους στη γνώση του target node.

5. **Modal Value**: Για κάθε συστατικό της λίστας γίνεται αναγορά στην πιο πιθανή τιμή του κόμβου ως προς τη συγκεκριμένη τιμή του target node και η πιθανότητα εμφάνισης αυτής της τιμής.

6. **A priori Modal Value:** Για κάθε συστατικό της λίστας η περιγραφή της πιο πιθανής τιμής όταν ο target node είναι μη παρατηρήσιμος (unobserved).

7. **Variation**: Η μέτρηση που καταδεικνύει την απόκλιση ανάμεσα στην a priori Modal Value και τη Modal Value. Η εξίσωση που χρησιμοποιείται για τον υπολογισμό της απόκλισης είναι η εξής: -log2(P(X=modal value)) +log2(P(X=modal value|Target=observed value)).

8. **Maximal positive/negative variation**: Οι μετρήσεις που καταδεικνύουν τις πιθανές τιμές του κόμβου (modalities) που έχουν τη μεγαλύτερη απήχηση (θετική ή αρνητική).

Αναλύοντας την έξοδο του συστήματος παρατηρούμε ότι δεν παρουσιάζονται πουθενά ευκρινώς οι κανόνες αλλά υπάρχει πληθώρα στατιστικών μετρήσεων και δεικτών. Αναλύοντας τις στατιστικές μετρήσεις που προσφέρονται παρατηρήσαμε ότι υπάρχει μια μέτρηση που με την κατάλληλη μετατροπή μπορεί να μας δώσει τους κανόνες.

Αυτή η μέτρηση είναι η Modal Value. Πριν αναφερθούμε όμως στη διαδικασία παραγωγής κανόνων μέσα από την μέτρηση Modal Value θα αναλύσουμε τη διαδικασία κωδικοποίησης των δεδομένων εισόδου για το BayesiaLab η οποία ακολουθήθηκε στην παρούσα διπλωματική.

Μετά από πολλές δοκιμαστικές εναλλακτικές κωδικοποιήσεις που εφαρμόσαμε στα δεδομένα μας πρώτου αυτά μπουν ως είσοδος στο BayesiaLab, παρατηρήσαμε πως μόνο ένας τρόπος κωδικοποίησης ήταν αυτός που πέτυχε την εύρεση των κανόνων στο σύνολο τους (η απαραίτητη προϋπόθεση αποδοχής του αλγορίθμου για τα δεδομένα μας). Η κωδικοποίηση αυτή φαίνεται καθαρά από το παράδειγμα που ακολουθεί: έστω ότι έχουμε να συσχετίσουμε 2 πεδία (χαρακτηριστικά - attributes) τα a και b και ότι το χαρακτηριστικό b είναι target node. Τότε για κάθε διακριτή τιμή του χαρακτηριστικού a δημιουργείται και ένας κόμβος παιδί (child node) που ονομάζεται με το όνομα της τιμής του χαρακτηριστικού. Δηλαδή αν το χαρακτηριστικό a περιέχει μέσα στο πλήθος των υποδειγμάτων 3 διακριτές τιμές (π.χ. a_value1, a_value2 και a_value3) τότε δημιουργούμε τρία child nodes (a_value1, a_value2 και a_value3) αντί για ένα (a). Οι τιμές των υποδειγμάτων για κάθε ένα από του καινούργιους κόμβους παίρνουν τιμές 0 ή 1 ανάλογα με το αν η τιμή του πρώην κόμβου συμπίπτει με το όνομα του καινούργιου κόμβου.

Στον ακόλουθο πίνακα φαίνονται τα δεδομένα πριν την κωδικοποίηση.

Attributes	**a**	**b**
Instance 1	a_value1	B_value1
Instance 2	a_value2	B_value2
Instance 3	a_value3	B_value2
Instance 4	a_value1	B_value1
Instance 5	a_value2	B_value1

Πίνακας 3.1: Τα δεδομένα πριν την απαιτούμενη κωδικοποίηση για το BayesiaLab

Στον επόμενο πίνακα φαίνονται τα δεδομένα μετά την κωδικοποίηση.

Attributes	**a_value1**	**a_value2**	**a_value3**	**b**
Instance 1	1	0	0	B_value1
Instance 2	0	1	0	B_value2
Instance 3	0	0	1	B_value2
Instance 4	1	0	0	B_value1
Instance 5	0	1	0	B_value1

Πίνακας 3.2: Τα δεδομένα μετά την κωδικοποίηση για το BayesiaLab

Εξηγώντας λοιπόν την παραπάνω κωδικοποίηση μπορούμε να περιγράψουμε πώς από το δείκτη Modal Value μπορούν να προκύψουν οι κανόνες ταξινόμησης. Ο δείκτης modal value αναφέρει για κάθε child node που συσχετίζεται με μια συγκείμενη τιμή του target node την πιθανή τιμή του και την πιθανότητα της τιμής αυτής. Επειδή όμως το σύνολο τιμών του κάθε child node σύμφωνα με την κωδικοποίηση μας είναι 0 και 1 {0,1}, συνεπάγεται ότι εκτός από τους κανόνες ταξινόμησης, μπορούμε να βρούμε και τον απόλυτο αριθμό ικανοποίησης του κανόνα από τα υποδείγματα. Δηλαδή, αν στον πίνακα Modal Value έχουμε τις τιμές 0 και

100,00% σημαίνει ότι η συγκεκριμένη τιμή του child node δεν συσχετίζεται με τη συγκεκριμένη τιμή του target node. Αν όμως ο δείκτης modal value αναφέρει 0 και 75,00%, αυτό σημαίνει ότι η συγκεκριμένη τιμή του child node συσχετίζεται με τη συγκεκριμένη τιμή του target node κατά 25,00% (100,00% - 75,00%), διότι στις περιπτώσεις που δεν είναι 0 η τιμή του child node, τότε σύμφωνα με το πεδίο ορισμού του θα είναι 1. Συν τοις άλλης, γνωρίζοντας το συνολικό αριθμό υποδειγμάτων, μπορούμε να βρούμε τον αριθμό των υποδειγμάτων που συμφωνούν με τον κανόνα ως εξής: Αν *X* ο συνολικός αριθμός υποδειγμάτων, Y% το ποσοστό συμμετοχής της συγκεκριμένης τιμής του target node στο συνολικό αριθμό (παράγεται από το BayesiaLab) και *Z%* το ποσοστό των τιμών 1 του child node στο συνολικό αριθμό υποδειγμάτων που ορίζει το ποσοστό *Y%*, τότε ο αριθμός *V* των υποδειγμάτων που ικανοποιούνται με τον κανόνα είναι: *V=X*Y/100*(Z/100)* (Αν μας δίνεται το ποσοστό των τιμών 0, τότε για να παράγουμε το ποσοστό των τιμών 1 *Z%* απλά το αφαιρούμε από το 100).

Κατά αυτόν λοιπόν τον τρόπο παράγονται οι κανόνες μέσα από τα αποτελέσματα του αλγορίθμου Naive Bayes στο λογισμικό BayesiaLab.

Αναφέρουμε ότι στον αλγόριθμο Naive Bayes εκτός από την προαναφερθείσα κωδικοποίηση, τροποποιήσαμε περαιτέρω τα δεδομένα μετατρέποντας όλους τους ελληνικούς χαρακτήρες που υπήρχαν στα υποδείγματα εισόδου σε λατινικούς και αντικαθιστώντας το χαρακτήρα του κενού (space character) με το χαρακτήρα underscore (_).

3.4.4 Αλγόριθμος Holte's 1R

Ο αλγόριθμος Holte's 1R παράγει κανόνες ταξινόμησης μέσω της μεθόδου Rough Set και προσφέρεται μέσα από το λογισμικό Rosetta.

Ο αλγόριθμος Holte's 1R είναι ένας αλγόριθμος ταξινόμησης που βασίζεται στην παραλλαγή του αλγορίθμου 1R όπως παρουσιάζεται από τον Robert C. Holte [Holte, 1993].

Ακολουθεί ένα παράδειγμα εξόδου του αλγόριθμου Holte's 1R του λογισμικού Rosetta.

```
% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:57:54 by mike.
%
% 1R
% 4 rules.

category(Προμηθευτές) => origin(Εσωτερικής_Προέλευσης)
Supp. (LHS) = [2513 object(s)]
Supp. (RHS) = [2513 object(s)]
```

Acc. (RHS) = [1]
Cov. (LHS) = [0.201411]
Cov. (RHS) = [0.258725]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

category(Καιρικές_Συνθήκες) => origin(Εξωτερικής_Προέλευσης)
Supp. (LHS) = [1819 object(s)]
Supp. (RHS) = [1819 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.145788]
Cov. (RHS) = [0.658104]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

category(Οδόστρωμα) => origin(Εξωτερικής_Προέλευσης)
Supp. (LHS) = [945 object(s)]
Supp. (RHS) = [945 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0757394]
Cov. (RHS) = [0.341896]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

category(Χρήση_-_Φθορά) => origin(Εσωτερικής_Προέλευσης)
Supp. (LHS) = [7200 object(s)]
Supp. (RHS) = [7200 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.577062]
Cov. (RHS) = [0.741275]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

Στις γραμμές του παραπάνω παραδείγματος αναφέρονται στοιχειά για την ημερομηνία εξαγωγής του αρχείου εξόδου, τον επιλεγμένο αλγόριθμο και τον αριθμό κανόνων που παράχθηκαν (% Rules/patterns generated by ROSETTA., % Exported 2007.12.23 17:57:54 by mike.,%,% 1R,% 4 rules.). Ακολουθούν οι κανόνες και μια λίστα στατιστικών μετρήσεων για κάθε κανόνα. Ο κανόνας στα αριστερά αναφέρει την τιμή του χαρακτηριστικού και δεξιά την τιμή της τάξης για τη συγκεκριμένη τιμή (category(Προμηθευτές)=>origin(Εσωτερικής_Προέλευσης)).

Οι στατιστικοί δείκτες σημαντικότητας που αναφέρονται στη λίστα που ακολουθεί για κάθε κανόνα, σημαίνουν τα εξής:

- **Support (LHS)**: Support of Left Hand Side. Υποστήριξη του αριστερού μέρους του κανόνα.
- **Support (RHS)**: Support of Right Hand Side. Υποστήριξη του δεξιού μέρους του κανόνα.
- **Accuracy (RHS)**: Η ακρίβεια του κανόνα.

- **Coverage (LHS)**: Το ποσοστό κάλυψης του αριστερού μέρους του κανόνα.
- **Coverage (RHS)**: Το ποσοστό κάλυψης του δεξιού μέρους του κανόνα.
- **Stability (LHS)**: Η σταθερότητα του αριστερού μέρους του κανόνα.
- **Stability (RHS)**: Η σταθερότητα του δεξιού μέρους του κανόνα.

Και στον αλγόριθμο Holte's 1R είναι απαραίτητη η αντικατάσταση του χαρακτήρα του κενού (space character) με το χαρακτήρα underscore (_) στα δεδομένα εισόδου.

3.4.5 Covering Algorithm

Αναφερόμαστε σε έναν αλγόριθμο τύπου covering (covering algorithm) ο οποίος παράγει κανόνες ταξινόμησης μέσω της μεθόδου Rough Set και προσφέρεται μέσα από το λογισμικό RSES2 (RSES version:2).

Ο αλγόριθμος covering είναι ένας αλγόριθμος ταξινόμησης και βασίζεται στη δημοσίευση του J. Bazan [Bazan et al. ,2000], [Bazan et al. ,2003]. Σε γενικές γραμμές ο covering algorithm αναπτύσσει μια "κάλυψη" στο σύνολο των δεδομένων που συνδέονται με τον κανόνα. Πιο αναλυτικά λειτουργεί ως εξής:

1. Ξεκινάει με ένα κενό κάλυμμα (cover).
2. Διαλέγει ένα υπόδειγμα.
3. Βρίσκει το σύνολο των συνδυαστικών εκφράσεων (conjunctive expression) που καλύπτουν τον κανόνα.
4. Διαλέγει την καλύτερη έκφραση συμφωνά με ένα κριτήριο.
5. Προσθέτει την έκφραση αυτή στο κάλυμμα (cover).
6. Συνεχίζει από το βήμα 2 έως το τέλος των υποδειγμάτων.

Ακολουθεί ένα παράδειγμα εξόδου του αλγόριθμου covering του λογισμικού RSES2.

```
RULE_SET rses.txt
ATTRIBUTES 2
category symbolic
origin symbolic
DECISION_VALUES 2
Εσωτερικής_Προέλευσης
```

Εξωτερικής_Προέλευσης
RULES 4
(category=Χρήση_-_Φθορά)=>(origin=Εσωτερικής_Προέλευσης[7200]) 7200
(category=Προμηθευτές)=>(origin=Εσωτερικής_Προέλευσης[2513]) 2513
(category=Καιρικές_Συνθήκες)=>(origin=Εξωτερικής_Προέλευσης[1819]) 1819
(category=Οδόστρωμα)=>(origin=Εξωτερικής_Προέλευσης[945]) 945

Στην πρώτη γραμμή (RULE_SET rses.txt) αναφέρεται το όνομα του αρχείου εξόδου του συστήματος, στη δεύτερη ακολουθεί ο αριθμός των χαρακτηριστικών (ATTRIBUTES 2), στη συνεχεία τα ονόματα των χαρακτηριστικών και οι τύποι αυτών [εδώ το symbolic αναφέρεται στα ονομαστικά] (category symbolic, origin symbolic), ενώ ακολουθούν οι διακριτές τιμές της τάξης [decision value] (DECISION_VALUES2, Εσωτερικής_Προέλευσης, Εξωτερικής_Προέλευσης). Ακολουθούν οι κανόνες και ο συνολικός αριθμός αυτών, οι οποίοι αποτελούνται από το αριστερό μέρος ((category=Χρήση_-_Φθορά)) όπου αναφέρεται η τιμή του χαρακτηριστικού και το δεξί μέρος όπου αναφέρεται η επιλεγμένη τιμή απόφασης (decision value) για τη συγκεκριμένη τιμή ((origin=Εσωτερικής_Προέλευσης[7200]) 7200) και ο αριθμός υποδειγμάτων που το επαληθεύουν, ενώ ο τελευταίος αριθμός αναφέρεται στο συνολικό αριθμό υποδειγμάτων που συναντάται η συγκεκριμένη τιμή του χαρακτηριστικού του αριστερού μέρους.

Όπως και στους υπόλοιπους επιλεγμένους αλγόριθμους, έτσι και στον covering algorithm ήταν απαραίτητη η τροποποίηση στα δεδομένα εισόδου αντικαθιστώντας το χαρακτήρα του κενού (space character) με το χαρακτήρα underscore (_).

3.5 Περιγραφή Εφαρμογής

Η εφαρμογή, η οποία ονομάζεται “ιστορικό μηχανής”, σκοπό έχει την δυναμική ανάπτυξη των 7 νέων εργαλείων ποιότητας με βάση τους κανόνες που παράγονται από τους επιλεγμένους αλγορίθμους οι οποίοι αναζητούν γνώση μέσα σε καθορισμένα δεδομένα από το σύνολο των δεδομένων της μελέτης περίπτωσης. Επίσης, παρέχεται η δυνατότητα πρόβλεψης σε νέα δεδομένα συμφωνά με τους παραγόμενους κανόνες και η δυναμική στατιστική απεικόνιση των υποδειγμάτων που ικανοποιούν τους κανόνες. Από τα αποτελέσματα της εφαρμογής αυτής θα γίνει εφικτή η εξαγωγή συμπερασμάτων που θα αφορούν τόσο στη μελέτη περίπτωσης (μέσω των 7 νέων εργαλείων και της πρόβλεψης), όσο και στην κριτική των επιλεγμένων τεχνικών εξόρυξης γνώσης που χρησιμοποιήθηκαν.

Αρχικά, στην εφαρμογή δίνεται ένας κατάλογος επιλογής της επιθυμητής από το χρηστή ενέργειας. Η πρώτη επιλογή αφορά στην αναζήτηση πληροφοριών από τη βάση δεδομένων όπου βρίσκονται αποθηκευμένα όλα τα δεδομένα της μελέτης περίπτωσης και οι υπόλοιπες έξι επιλογές αφορούν στην δημιουργία των 6 ποιοτικών εργαλείων από τα 7 νέα εργαλεία ποιότητας μέσω των κανόνων που ανακαλύπτονται από τους επιλεγμένους αλγορίθμους εξόρυξης γνώσης.

Πρέπει να σημειώσουμε ότι το έβδομο από τα 7 εργαλεία ποιότητας , η Ανάλυση Δεδομένων Διαγράμματος Μήτρας (Matrix Data Analysis), δε δημιουργείται δυναμικά απευθείας από τα δεδομένα, αλλά είναι παράγωγο του διαγράμματος μήτρας (Matrix Diagram) στις περιπτώσεις όπου έχουμε ποσοτικά δεδομένα.

Εικόνα 3.9: Το μενού επιλογών της εφαρμογής

Παρακάτω ακολουθεί αναλυτική περιγραφή για κάθε μια από τις ενέργειες του μενού επιλογών.

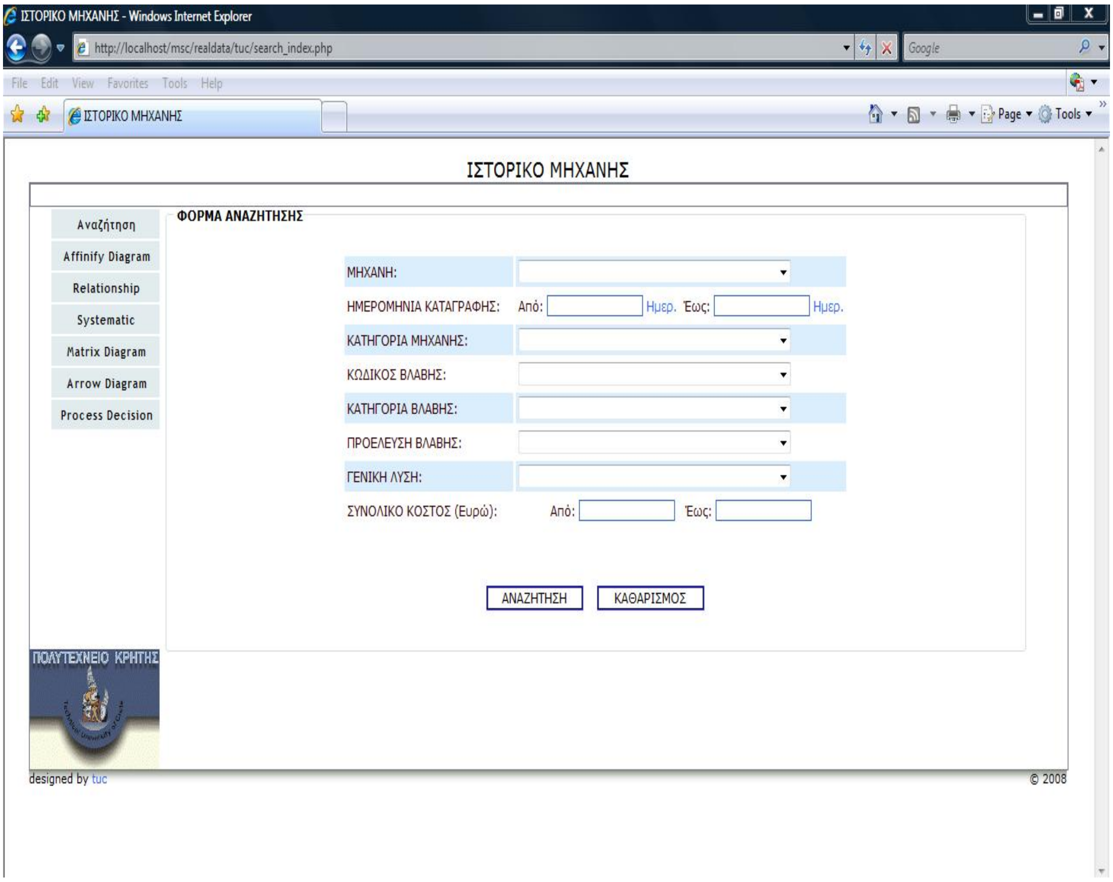
3.5.1 Αναζήτηση Δεδομένων

Η πρώτη επιλογή από το μενού είναι αυτή της Αναζήτησης Δεδομένων. Με την επιλογή της Αναζήτησης ο χρήστης οδηγείται στη φόρμα αναζήτησης, όπου του δίνεται η δυνατότητα δημιουργίας ενός συνδυαστικού ερωτήματος SQL προκειμένου να οδηγηθεί ο χρήστης στην εγγραφή την οποία αναζητά. Ο κατάλογος των πεδίων με τα οποία ο χρήστης μπορεί να δημιουργήσει το ερώτημα της συνδυαστικής αναζήτησης είναι (σε όλα τα πεδία ακολουθεί διπλά ο κατάλογος με τις διακριτές τιμές των πεδίων που βρίσκονται στη βάση δεδομένων):

- ΜΗΧΑΝΗ
- ΗΜΕΡΟΜΗΝΙΑ ΚΑΤΑΓΡΑΦΗΣ (Από – Έως)
- ΚΑΤΗΓΟΡΙΑ ΜΗΧΑΝΗΣ
- ΚΩΔΙΚΟΣ ΒΛΑΒΗΣ

- ΚΑΤΗΓΟΡΙΑ ΒΛΑΒΗΣ
- ΠΡΟΕΛΕΥΣΗ ΒΛΑΒΗΣ
- ΓΕΝΙΚΗ ΛΥΣΗ
- ΣΥΝΟΛΙΚΟ ΚΟΣΤΟΣ (Ευρώ) (Από – Έως)

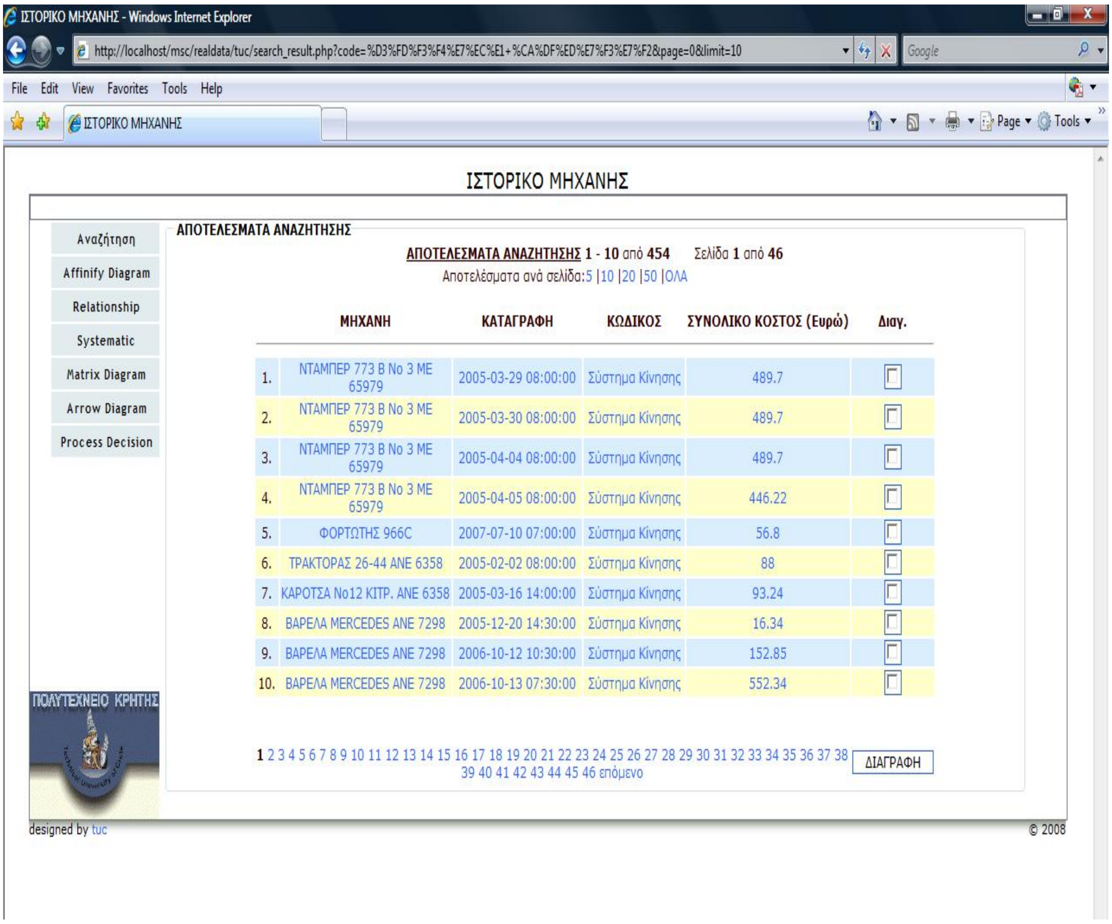
Μόλις επιλέξει ο χρηστής τα πεδία και τις τιμές με βάση τις οποίες επιθυμεί να γίνει η αναζήτηση, κλικάρει στο κουμπί *Αναζήτηση* για να οδηγηθεί στη σελίδα αποτελεσμάτων της αναζήτησης. Με το κουμπί *Καθαρισμός* καθαρίζονται όλες οι επιλογές και ο χρήστης μπορεί να κάνει καινούργια αναζήτηση.



Εικόνα 3.10: Η φόρμα αναζήτησης

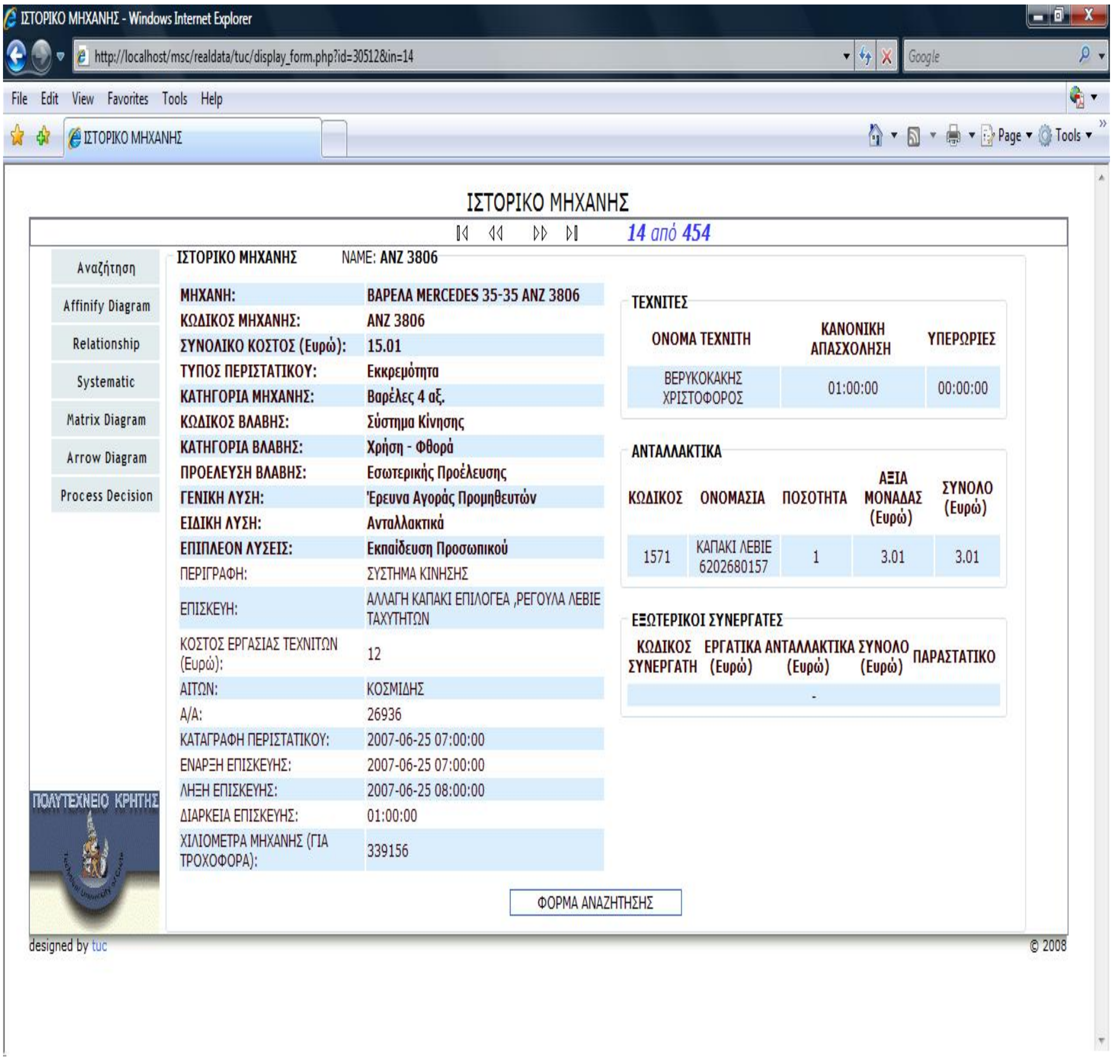
Όταν ο χρήστης πατήσει το κουμπί αναζήτηση οδηγείται σε μια νέα σελίδα όπου του παρουσιάζονται τα αποτελέσματα, δηλαδή οι εγγραφές οι οποίες συμφωνούν με τα κριτήρια τα οποία έθεσε, ταξινομημένα σύμφωνα με το όνομα της μηχανής που υπέστη τη βλάβη, την ημερομηνία της βλάβης και το συνολικό κόστος επισκευή της, με τη σειρά που αναφέρεται παραπάνω. Στην φόρμα αποτελεσμάτων παρουσιάζονται μερικά από τα βασικά χαρακτηριστικά (ΜΗΧΑΝΗ, ΚΑΤΑΓΡΑΦΗ, ΚΩΔΙΚΟΣ, ΣΥΝΟΛΙΚΟ ΚΟΣΤΟΣ (Ευρώ)) των εγγραφών που βρίσκονται στη λίστα αποτελεσμάτων της αναζήτησης.

Επίσης, στη φόρμα αναγράφεται ο συνολικός αριθμός των αποτελεσμάτων, ενώ δίνεται η δυνατότητα επιλογής του αριθμού της λίστας των εγγραφών που επιθυμεί να βλέπει ο χρήστης σε κάθε σελίδα, καθώς και η δυνατότητα διαγραφής μια εγγραφής. Εάν ο χρήστης κλικάρει σε ένα από τα χαρακτηριστικά μιας εγγραφής, τότε θα οδηγηθεί σε μια νέα σελίδα οπού παρουσιάζονται τα πλήρη στοιχειά αυτής της εγγραφής. Τέλος, στο κάτω μέρος της σελίδας υπάρχει κουμπί το οποίο επιστρέφει το χρήστη στην φόρμα αναζήτησης για δημιουργία καινούργιας αναζήτησης .



Εικόνα 3.11: Η φόρμα αποτελεσμάτων

Με το πάτημα του συνδέσμου πάνω σε ένα από τα χαρακτηριστικά των εγγραφών, ο χρήστης οδηγείται σε μια νέα σελίδα όπου αναφέρονται όλα τα πεδία της εγγραφής τα οποία παρουσιάζονται στην παράγραφο *Μελέτη Περίπτωσης* (*παρ. 3.3*). Εκτός από τα πλήρη στοιχειά της εγγραφής δίνεται η δυνατότητα στο χρήστη να περιηγηθεί στην επόμενη/προηγούμενη/πρώτη/τελευταία εγγραφή της λίστας των αποτελεσμάτων της αναζήτησης μέσω ειδικών συνδέσμων. Επίσης, στο κάτω μέρος της σελίδας υπάρχει κουμπί για την επιστροφή στη φόρμα αναζήτησης. Εκτός από αυτά, και στις τρεις προαναφερθείσες φόρμες υπάρχει σταθερά στα αριστερά της οθόνης το αρχικό μενού επιλογών (εικόνα 3.9).



Εικόνα 3.12: Η φόρμα απεικόνισης του συνόλου των στοιχειών μιας εγγραφής

3.5.2 Affinity Diagram

Όταν ο χρήστης επιλέξει το σύνδεσμο *Affinity Diagram* οδηγείται σε μια σελίδα όπου του παρουσιάζονται αρχικά τα χαρακτηριστικά (attributes) τα οποία θα συσχετίσουν οι αλγόριθμοι εξόρυξης γνώσης, δηλαδή τα πεδία αυτά από την βάση δεδομένων τα οποία επιθυμούμε να συσχετίσουμε.

Στην περίπτωση της εφαρμογής για την παραγωγή του Affinity Diagram θα συσχετίσουμε τα πεδία *Κωδικός Βλάβης* και *Κατηγορία Βλάβης*. Στη σελίδα αυτή φαίνονται επίσης και οι διακριτές τιμές των δυο χαρακτηριστικών αυτών από το σύνολο της βάσης δεδομένων. Παρακάτω στην εφαρμογή, ακολουθούν τα αρχεία εισόδου για τους επιλεγμένους αλγορίθμους Εξόρυξης Γνώσης. Τέσσερα από τα πέντε λογισμικά (Rosetta, RSES2, WEKA-Association Rules, WEKA-Classification Rules) χρησιμοποιούν κοινό αρχείο εισόδου με κοινή κωδικοποίηση (αντικατάσταση του χαρακτήρα space με τον χαρακτήρα underscore). Το πέμπτο λογισμικό χρησιμοποιεί την κωδικοποίηση και το μετασχηματισμό που αναφέρεται στην παράγραφο 3.4.3 (Αλγόριθμος Naive Bayes), δηλαδή αλλαγή του αριθμού των χαρακτηριστικών-κόμβων με βάση τις διακριτές τιμές του αρχικού κόμβου παιδιού σε νέο πεδίο ορισμού {0,1}, μεταφορά των ελληνικών χαρακτήρων σε λατινικούς και αντικατάσταση των χαρακτήρων space με χαρακτήρες underscore. Τα αρχεία εισόδου φαίνονται μέσω του προγράμματος excel αν ενεργοποιηθούν τα αντίστοιχα links.

A4 fx Ηλεκτρικά_-_Ηλεκτρονικά

	A	B
1	**code**	**category**
2	Υδραυλικά	Προμηθευτές
3	Υδραυλικά	Προμηθευτές
4	Ηλεκτρικά_-_Ηλεκτρονικά	Προμηθευτές
5	Αμάξωμα	Προμηθευτές
6	Ηλεκτρικά_-_Ηλεκτρονικά	Προμηθευτές
7	Ηλεκτρικά_-_Ηλεκτρονικά	Προμηθευτές
8	Υδραυλικά	Προμηθευτές
9	Ηλεκτρικά_-_Ηλεκτρονικά	Προμηθευτές
10	Υδραυλικά	Προμηθευτές
11	Υδραυλικά	Προμηθευτές
12	Υδραυλικά	Προμηθευτές
13	Υδραυλικά	Προμηθευτές
14	Υδραυλικά	Προμηθευτές
15	Αμάξωμα	Προμηθευτές
16	Αμάξωμα	Προμηθευτές
17	Αμάξωμα	Προμηθευτές
18	Αμάξωμα	Προμηθευτές
19	Σύστημα_Κίνησης	Προμηθευτές
20	Υδραυλικά	Προμηθευτές
21	Υδραυλικά	Προμηθευτές
22	Υδραυλικά	Προμηθευτές
23	Σύστημα_Κίνησης	Προμηθευτές
24	Υδραυλικά	Προμηθευτές
25	Υδραυλικά	Προμηθευτές
26	Υδραυλικά	Προμηθευτές
27	Ηλεκτρικά_-_Ηλεκτρονικά	Προμηθευτές

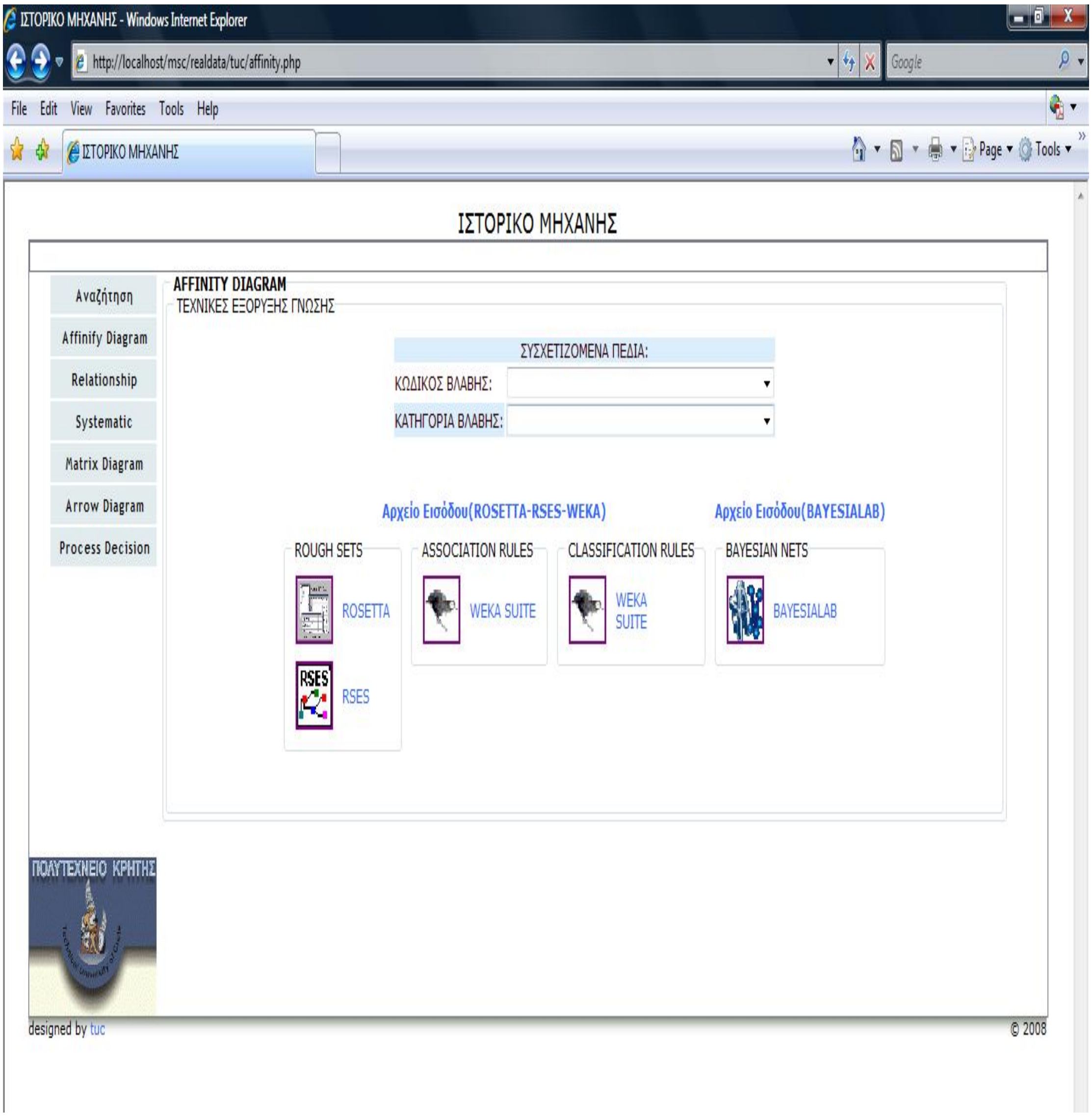
Εικόνα 3.13: Δείγμα του κοινού αρχείου εισόδου που χρησιμοποιούν 4 από τα 5 λογισμικά

AD1 category

	T	U	V	W	X	Y	Z	AA	AB	AC	AD
1	Sintirisi	Sistima_Pedisis	Sistima_kinisis	Sistima_Diefthinsis	Sistima_Psiksis	Idravlika	Eksatmisi	Ergasia	Antlisi	Elegchos	category
2	0	0	0	0	0	1	0	0	0	0	Promitheutes
3	0	0	0	0	0	1	0	0	0	0	Promitheutes
4	0	0	0	0	0	0	0	0	0	0	Promitheutes
5	0	0	0	0	0	0	0	0	0	0	Promitheutes
6	0	0	0	0	0	0	0	0	0	0	Promitheutes
7	0	0	0	0	0	0	0	0	0	0	Promitheutes
8	0	0	0	0	0	1	0	0	0	0	Promitheutes
9	0	0	0	0	0	0	0	0	0	0	Promitheutes
10	0	0	0	0	0	1	0	0	0	0	Promitheutes
11	0	0	0	0	0	1	0	0	0	0	Promitheutes
12	0	0	0	0	0	1	0	0	0	0	Promitheutes
13	0	0	0	0	0	1	0	0	0	0	Promitheutes
14	0	0	0	0	0	1	0	0	0	0	Promitheutes
15	0	0	0	0	0	0	0	0	0	0	Promitheutes
16	0	0	0	0	0	0	0	0	0	0	Promitheutes
17	0	0	0	0	0	0	0	0	0	0	Promitheutes
18	0	0	0	0	0	0	0	0	0	0	Promitheutes
19	0	0	1	0	0	0	0	0	0	0	Promitheutes
20	0	0	0	0	0	1	0	0	0	0	Promitheutes
21	0	0	0	0	0	1	0	0	0	0	Promitheutes
22	0	0	0	0	0	1	0	0	0	0	Promitheutes
23	0	0	1	0	0	0	0	0	0	0	Promitheutes
24	0	0	0	0	0	1	0	0	0	0	Promitheutes
25	0	0	0	0	0	1	0	0	0	0	Promitheutes
26	0	0	0	0	0	1	0	0	0	0	Promitheutes
27	0	0	0	0	0	0	0	0	0	0	Promitheutes

Εικόνα 3.14: Δείγμα του αρχείου εισόδου που χρησιμοποιεί το BayesiaLab

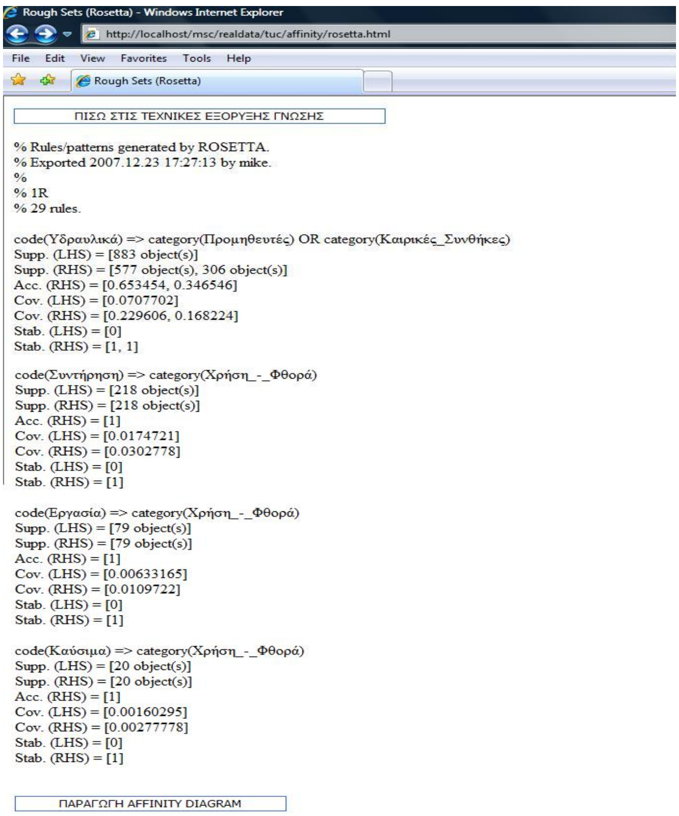
Κάτω από τους συνδέσμους που οδηγούν στα αρχεία εισόδου βρίσκονται 5 σύνδεσμοι που οδηγούν στην παρουσίαση των αποτελεσμάτων των αλγορίθμων εξόρυξης γνώσης που εφαρμόστηκαν για την ανάλυση των δεδομένων. Πρώτα παρουσιάζονται οι σύνδεσμοι των τεχνικών Rough Set με τα λογισμικά Rosetta και RSES2, στη συνέχεια τα αποτελέσματα των Association Rules και Classification Rules του λογισμικού Weka και έπεται ο σύνδεσμος για την παρουσίαση των αποτελεσμάτων του BayesiaLab. Το κεντρικό μενού επιλογών βρίσκεται πάντα στα αριστερά της οθόνης της εφαρμογής.



Εικόνα 3.15: Σελίδα Affinity Diagram

Μόλις επιλέξουμε το link *Rosetta* οδηγούμαστε στη παρουσίαση των αποτελεσμάτων του αλγόριθμου Holte's 1R του λογισμικού Rosetta και δυο συνδέσμων, έναν για επιστροφή στην κεντρική σελίδα του Affinity Diagram (*Πίσω στις Τεχνικές Εξόρυξης Γνώσης*) και έναν άλλο (Παραγωγή Affinity Diagram) για την ενεργοποίηση ενός αλγορίθμου δυναμικής παραγωγής του Affinity Diagram (*Παραγωγή Affinity Diagram*). Ο αλγόριθμος αυτός διαβάζει το αρχείο εξόδου του λογισμικού Rosetta γραμμή προς γραμμή και συγκρατεί εκείνες τις γραμμές που αναφέρονται στους κανόνες που ανακαλύφθηκαν καθώς και τις γραμμές που αναγράφονται οι αριθμοί των υποδειγμάτων που ικανοποιούν τους κανόνες αυτούς, ενώ αποκλείει όλες τις υπόλοιπες. Ο αλγόριθμος, μόλις διαβάσει όλο το αρχείο, αποθηκεύει τους κανόνες σε έναν πίνακα μέσα στη βάση δεδομένων. Στον πίνακα αυτόν αποθηκεύονται όλες οι διακριτές σχέσεις των χαρακτηριστικών τα οποία συσχετίζονται και ο αριθμός υποδειγμάτων που ικανοποιούν το συγκεκριμένο κανόνα συσχέτισης. Τα τρία πεδία του πίνακα είναι: *παιδί* (αποθήκευση του αριστερού μέρους του κανόνα), *πατέρας* (αποθήκευση των διακριτών τιμών του δεξιού μέρους του κανόνα), χαρακτηριστικά (ο αριθμός των υποδειγμάτων). Τα πεδία αυτά "γεμίζουν" από τον αλγόριθμο ανάγνωσης του αρχείου εξόδου του αλγόριθμου εξόρυξης γνώσης.

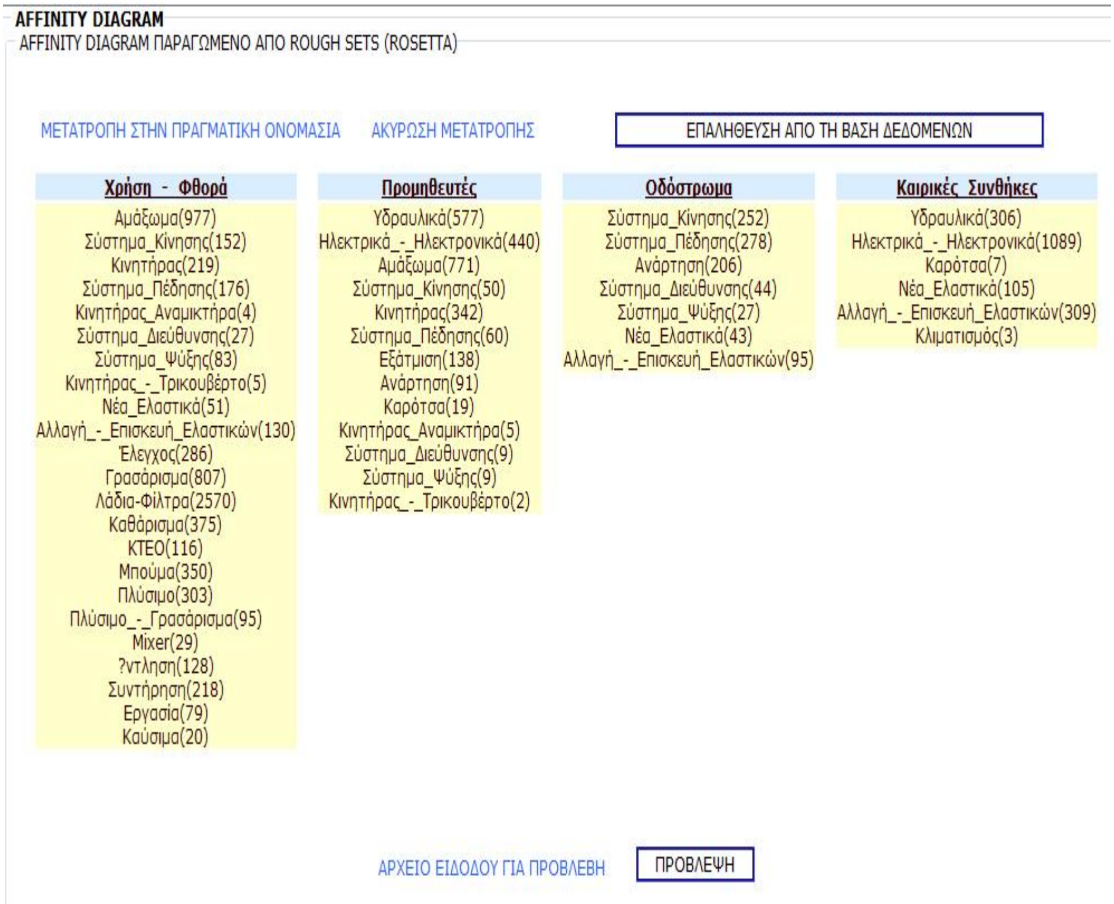
Εικόνα 3.16: Ο πίνακας αποθήκευσης των παραγομένων κανόνων στη μορφή που τα παρουσιάζει η MySQL



Εικόνα 3.17: Σελίδα εμφάνισης των κανόνων του Rosetta με το κουμπί για την έναρξη της διαδικασίας παραγωγής του Affinity Diagram

Μόλις ο αλγόριθμος διαβάσει όλο το αρχείο εξόδου και αποθηκεύσει τους κανόνες στη βάση δεδομένων, μπορούμε να δημιουργήσουμε το πρώτο από τα 7 νέα εργαλεία, το Affinity Diagram, βάσει αυτών των κανόνων. Συμφωνά με τη θεωρία του Affinity Diagram, αναζητούμε την κατηγοριοποίηση διάφορων χαρακτηριστικών.

Στη μελέτη περίπτωσης μας θα κατηγοριοποιήσουμε το χαρακτηριστικό *Κωδικό Βλάβης* στην τάξη του. Ο τρόπος που δημιουργείται το δυναμικό Affinity Diagram είναι ο εξής: Με ένα SQL ερώτημα (query) ανακτούμε τις διακριτές τιμές του πεδίου “πατέρα” και δημιουργούμε τόσες στήλες όσο και οι τιμές που ανακτήσαμε. Σε κάθε στήλη μπαίνουν όσες τιμές από το πεδίο “παιδί” συνδέονται με την αντίστοιχη στήλη. Αυτό επιτυγχάνεται με το συνδυασμό των προγραμματιστικών γλωσσών SQL query, php και html. Στην επόμενη εικόνα φαίνεται το παραγόμενο affinity diagram.



Εικόνα 3.18: Παραγόμενο Affinity Diagram από τους κανόνες που ανακάλυψε ο αλγόριθμος του Rosetta

Εκτός από το Affinity Diagram, η εφαρμογή δίνει την δυνατότητα μετατροπής των δεδομένων στην αρχική τους μορφή (πριν την κωδικοποίηση) και επίσης τη δυνατότητα σύγκρισης με το αντίστοιχο affinity diagram που προκύπτει κατευθείαν από τη βάση δεδομένων (με SQL query). Εκτός από αυτά δίνεται η δυνατότητα πρόβλεψης σε καινούργια δεδομένα σύμφωνα με τους αποθηκευμένους πια κανόνες.

ΠΡΟΒΛΕΨΗ

ΠΡΟΒΛΕΨΗ

Είσοδος	Πρόβλεψη
Σύστημα Πέδησης	*Χρήση - Φθορά* (34.24%) *Οδόστρωμα* (54.09%) *Προμηθευτές* (11.67%)
Σύστημα Κίνησης	*Χρήση - Φθορά* (33.48%) *Οδόστρωμα* (55.51%) *Προμηθευτές* (11.01%)
Ανάρτηση	*Οδόστρωμα* (69.36%) *Προμηθευτές* (30.64%)
Κινητήρας Αναμικτήρα	*Χρήση - Φθορά* (44.44%) *Προμηθευτές* (55.56%)
Κινητήρας - Τρικουβέρτο	*Χρήση - Φθορά* (71.43%) *Προμηθευτές* (28.57%)
Σύστημα Διεύθυνσης	*Χρήση - Φθορά* (33.75%) *Οδόστρωμα* (55%) *Προμηθευτές* (11.25%)
Αλλαγή - Επισκευή Ελαστικών	*Χρήση - Φθορά* (24.34%) *Οδόστρωμα* (17.79%) *Καιρικές Συνθήκες* (57.87%)
Υδραυλικά	*Καιρικές Συνθήκες* (34.65%) *Προμηθευτές* (65.35%)
Νέα Ελαστικά	*Χρήση - Φθορά* (25.63%) *Οδόστρωμα* (21.61%) *Καιρικές Συνθήκες* (52.76%)
Κλιματισμός	*Καιρικές Συνθήκες* (100%)
Λάδια-Φίλτρα	*Χρήση - Φθορά* (100%)
Γρασάρισμα	*Χρήση - Φθορά* (100%)

Εικόνα 3.19: Αποτελέσματα πρόβλεψης

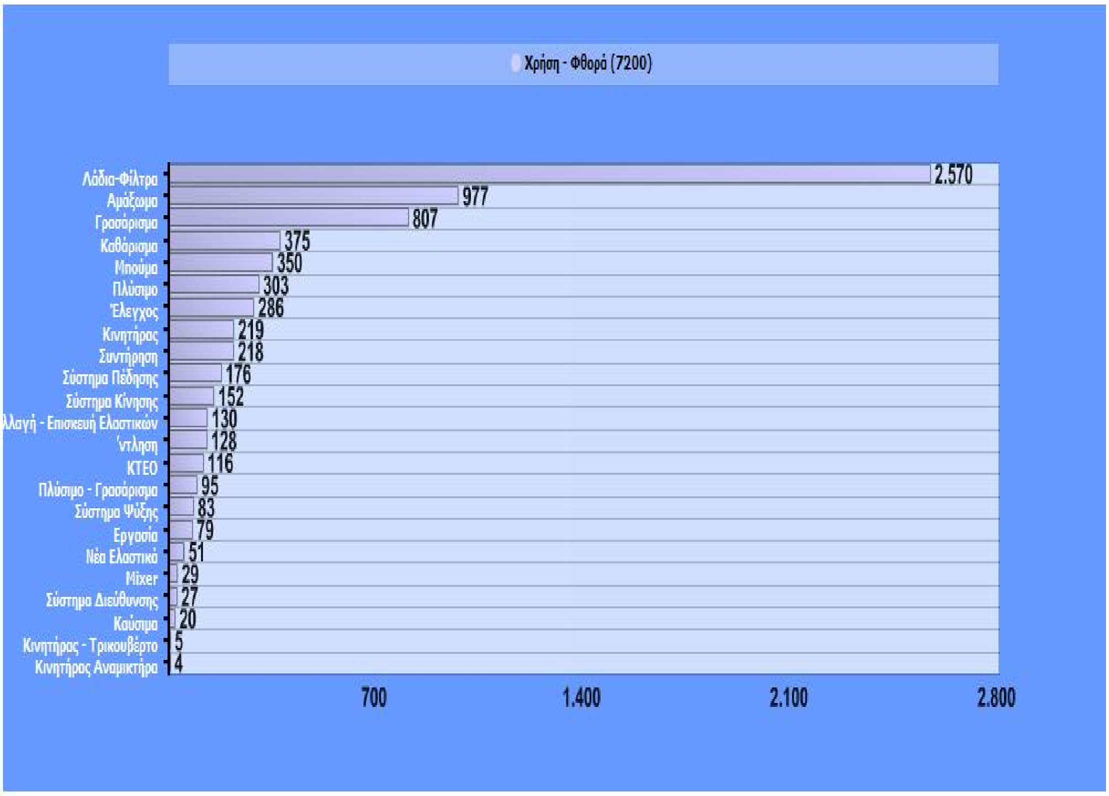
AFFINITY DIAGRAM

AFFINITY DIAGRAM ΠΡΟΕΡΧΟΜΕΝΟ ΑΠΟ ΤΗΝ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ

Χρήση - Φθορά	Οδόστρωμα	Καιρικές Συνθήκες	Προμηθευτές
Λάδια-Φίλτρα(2570)	Σύστημα Πέδησης(278)	Ηλεκτρικά - Ηλεκτρονικά(1089)	Αμάξωμα(771)
Αμάξωμα(977)	Σύστημα Κίνησης(252)	Αλλαγή - Επισκευή Ελαστικών(309)	Υδραυλικά(577)
Γρασάρισμα(807)	Ανάρτηση(206)	Υδραυλικά(306)	Ηλεκτρικά - Ηλεκτρονικά(440)
Καθάρισμα(375)	Αλλαγή - Επισκευή Ελαστικών(95)	Νέα Ελαστικά(105)	Κινητήρας(342)
Μπούμα(350)	Σύστημα Διεύθυνσης(44)	Καρότσα(7)	Εξάτμιση(138)
Πλύσιμο(303)	Νέα Ελαστικά(43)	Κλιματισμός(3)	Ανάρτηση(91)
Έλεγχος(286)	Σύστημα Ψύξης(27)		Σύστημα Πέδησης(60)
Κινητήρας(219)			Σύστημα Κίνησης(50)
Συντήρηση(218)			Καρότσα(19)
Σύστημα Πέδησης(176)			Σύστημα Διεύθυνσης(9)
Σύστημα Κίνησης(152)			Σύστημα Ψύξης(9)
Αλλαγή - Επισκευή Ελαστικών(130)			Κινητήρας Αναμικτήρα(5)
'ντληση(128)			Κινητήρας - Τρικουβέρτο(2)
ΚΤΕΟ(116)			
Πλύσιμο - Γρασάρισμα(95)			
Σύστημα Ψύξης(83)			
Εργασία(79)			
Νέα Ελαστικά(51)			
Mixer(29)			
Σύστημα Διεύθυνσης(27)			
Καύσιμα(20)			
Κινητήρας - Τρικουβέρτο(5)			
Κινητήρας Αναμικτήρα(4)			

Εικόνα 3.20: Παραγόμενο Affinity Diagram από τη βάση δεδομένων

Τέλος παρουσιάζεται και μια διαγραμματική μορφή των κανόνων (με ποσοτικές πληροφορίες).



Εικόνα 3.21: Στατιστικά κανόνων

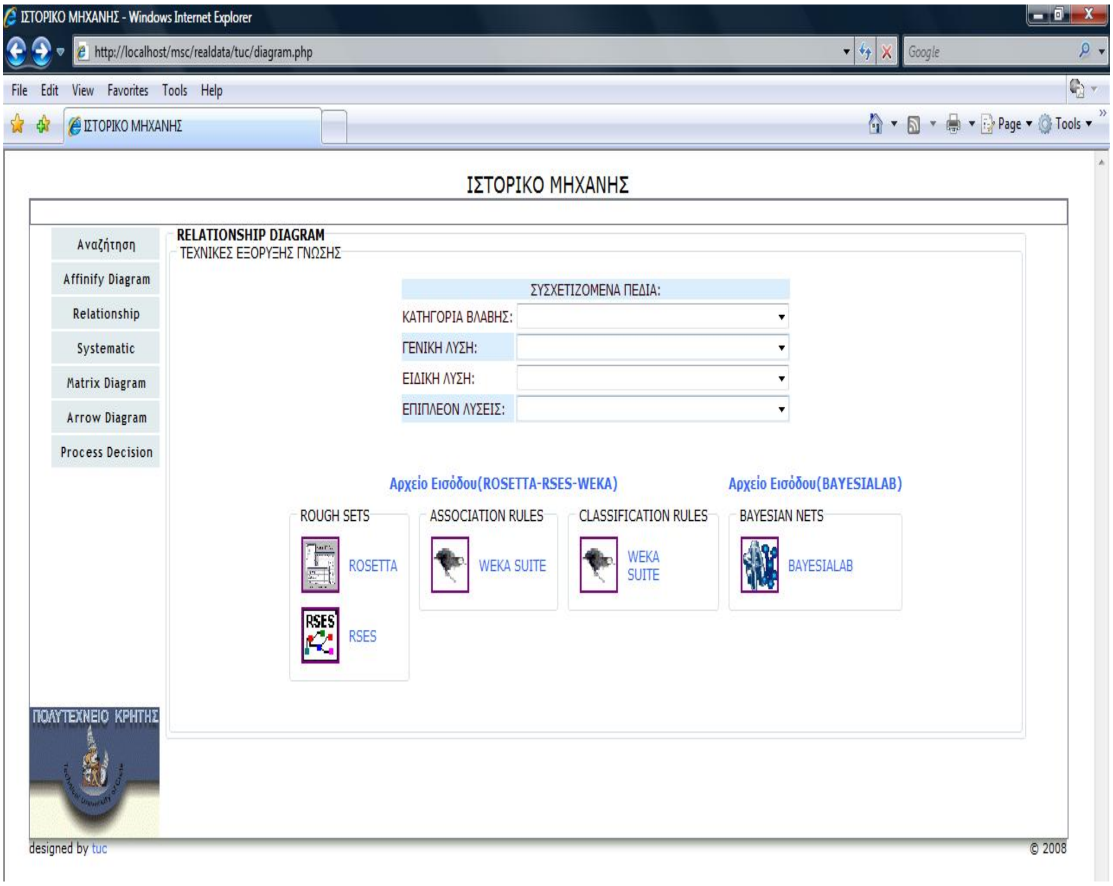
Η παρόμοια διαδικασία ακολουθείται και στις περιπτώσεις όπου ο χρήστης δεν επιλέξει το λογισμικό Rosetta αλλά ένα εκ των υπολοίπων. Αναλυτικότερα, ο χρήστης ανοίγει το αρχείο εξόδου του αντίστοιχου λογισμικού (τα αρχεία εξόδου του κάθε λογισμικού περιγράφονται στην παράγραφο 3.4), στη συνεχεία με εντολή του χρήστη ενεργοποιείται ο αλγόριθμος “σάρωσης” (αποκλειστικός για κάθε λογισμικό) του αρχείου αυτού και αναδεικνύονται οι κανόνες που αναγράφονται σε αυτό και τελικά αποθηκεύονται στη βάση δεδομένων και χρησιμοποιούνται για τη δημιουργία του Affinity Diagram, των στατιστικών διαγραμμάτων και της διαδικασίας πρόβλεψης.

Πρέπει να σημειωθεί ότι για την περίπτωση του αρχείου εξόδου του BayesiaLab δεν έχει διατηρηθεί η μορφή που παρουσιάζεται στην παράγραφο 3.4.3 αλλά έχει κρατηθεί μόνο ο πίνακας modal value όπου όπως αναφέρουμε στην παρ. 3.4.3 μας οδηγεί στους κανόνες που αναζητούμε.

3.5.3 Relationship Diagram

Αν ο χρήστης επιλέξει το σύνδεσμο *Relationship*, οδηγείται στη σελίδα παραγωγής του εργαλείου ποιότητας Relationship Diagram. Όπως και στο Affinity

Diagram, ακολουθείται η ίδια διαδικασία και προσφέρονται οι ίδιες δυνατότητες για τα ανάλογα δεδομένα οι οποίες είναι προσαρμοσμένες στο Relationship Diagram.



Εικόνα 3.22: Σελίδα Relationship Diagram

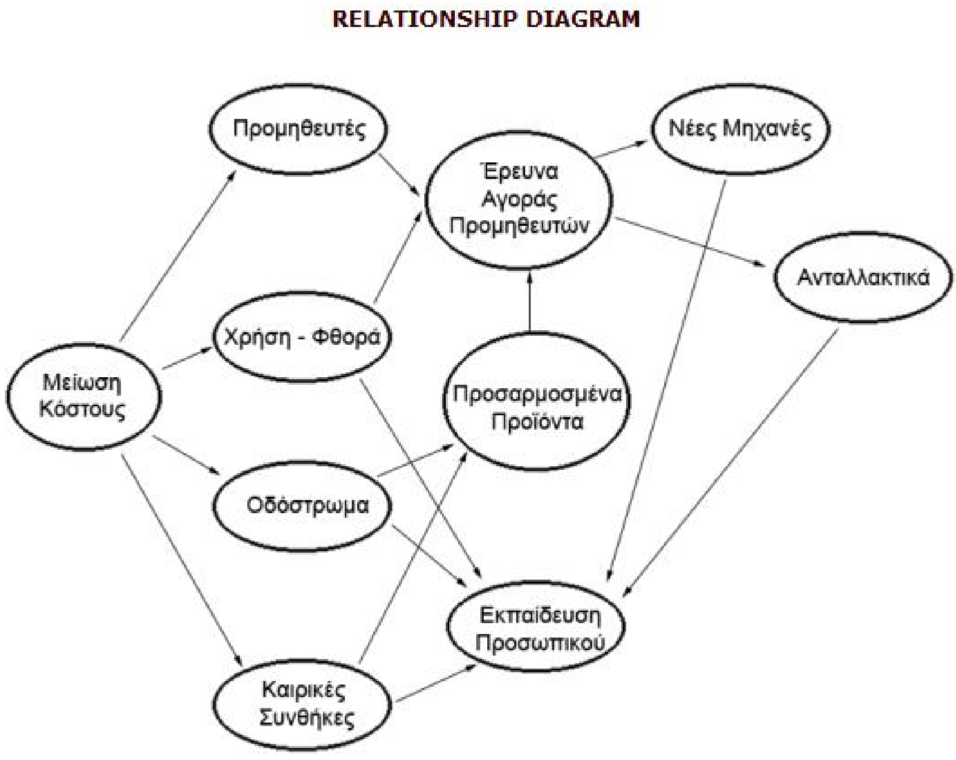
Στην περίπτωση μας, τα δεδομένα τα οποία θα συσχετίσουμε είναι η *Κατηγορία Βλάβης*, η *Γενική Λύση*, η *Ειδική Λύση* και οι *Επιπλέον Λύσεις* των οποίων οι διακριτές τιμές παρουσιάζονται στην προαναφερόμενη σελίδα. Πρέπει να σημειωθεί ότι στα δεδομένα αυτά ακολουθήθηκε ο μετασχηματισμός που αναφέρεται στην παράγραφο *3.4*. Στο μετασχηματισμό αυτό προσθέθηκε και η κωδικοποίηση των δεδομένων με τον ίδιο τρόπο που εφαρμόστηκε η κωδικοποίηση και στο Affinity Diagram.

Ακολουθώντας την ιδία διαδικασία με το Affinity Diagram καταλήγουμε στα αποτελέσματα των τεχνικών εξόρυξης γνώσης επάνω στα μετασχηματισμένα-κωδικοποιημένα δεδομένα στα οποία εφαρμόζονται οι αλγόριθμοι σάρωσης των αρχείων εξόδων και συλλέγονται οι κανόνες σε μορφή “πατέρα” - “παιδιού” οπού και αποθηκεύονται στη βάση δεδομένων στον κατάλληλο πίνακα (table).

Για κάθε μια από τις τεχνικές εξόρυξης γνώσης παρουσιάζονται τα αποτελέσματα τα οποία αποτελούνται από την απεικόνιση των κανόνων σε μορφή στηλών με σχέση “πατέρα” - “παιδιού”, από το δυναμικά δημιουργημένου από τους κανόνες relationship diagram, από το ισοδύναμο relationship diagram στην κλασσική του μορφή και από την απεικόνιση των κανόνων σε διαγράμματα. Επίσης, όπως και στο Affinity Diagram, προσφέρεται η δυνατότητα αποκωδικοποίησης των αποτελεσμάτων, η δυνατότητα πρόβλεψης σε καινούργια δεδομένα και η δυνατότητα σύγκρισης με τους κανόνες που επιστρέφει η βάση δεδομένων μέσω ενός SQL ερωτήματος.

Ο αλγόριθμος ο οποίος δημιουργεί το δυναμικό relationship diagram λειτουργεί ως εξής: Από τους παραγόμενους κανόνες που αποθηκεύονται στη βάση δεδομένων ανακτά όλες τις τιμές του πεδίου “παιδί” οι οποίες δεν είναι και δεν υπάρχουν καθόλου στο πεδίο “πατέρας”. Επομένως, στους κόμβους αυτούς δεν καταλήγει κανένας άλλος, αλλά αντίθετα οι ίδιοι καταλήγουν σε άλλους. Αυτοί οι κόμβοι δημιουργούν την πρώτη στήλη. Στη συνεχεία ανακτώνται οι τιμές των “πατεράδων” που έχουν ως “παιδιά” τους κόμβους της πρώτης στήλης. Αυτές οι τιμές δημιουργούν καινούργιους κόμβους εκτός και αν έχει ήδη χρησιμοποιηθεί σαν κόμβος ένας από αυτούς επομένως δεν δημιουργείται. Αναδρομικά, ακολουθείται η ιδία διαδικασία έως ότου εξεταστεί και η τελευταία σχέση “πατέρα” - “παιδιού”.

Εικόνα 3.23: Δυναμικά παραγόμενο Relationship Diagram

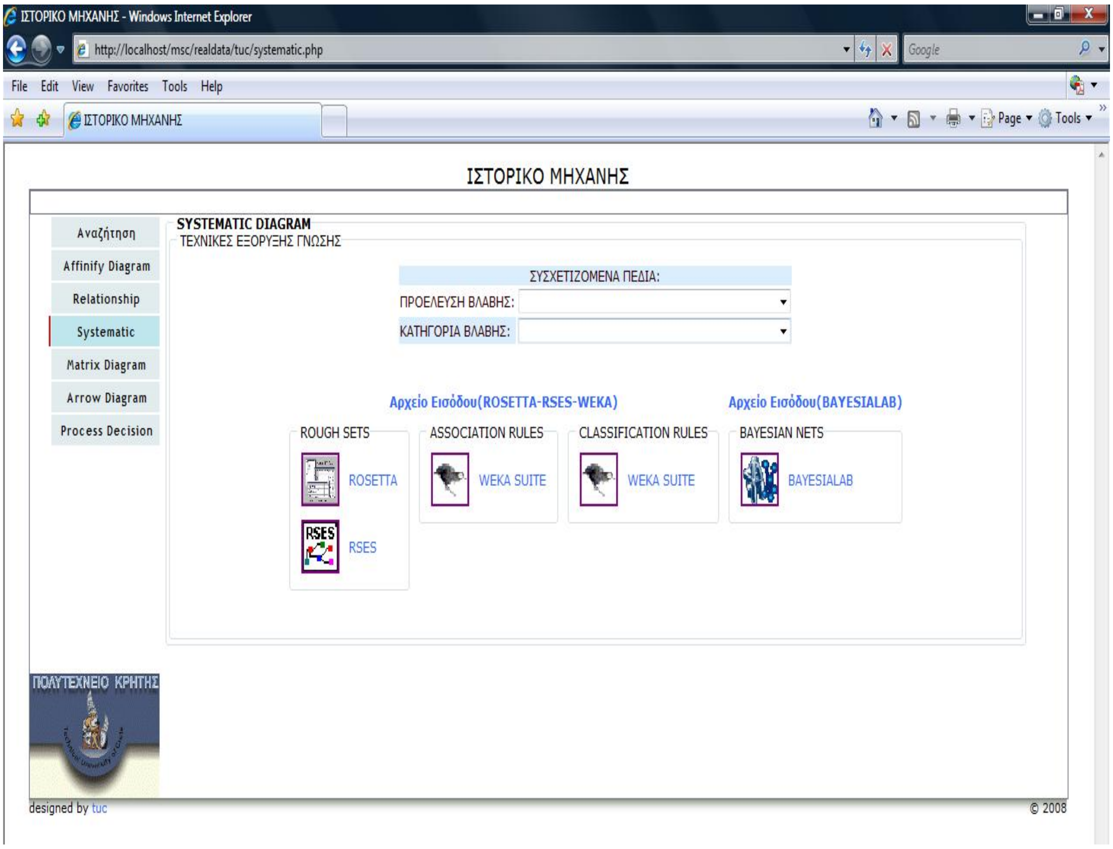


Εικόνα 3.24: Ισοδύναμο "κλασσικό" Relationship Diagram

3.5.4 Systematic Diagram

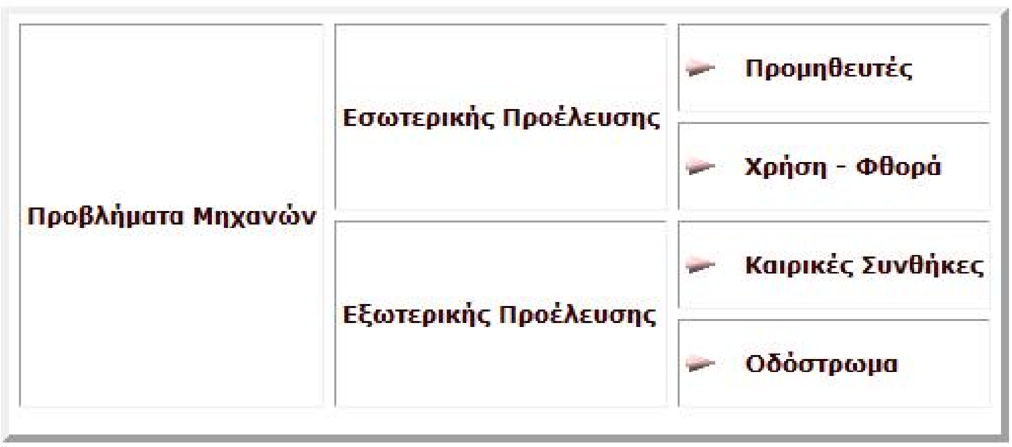
Με την επιλογή του συνδέσμου *Systematic, ο χρηστής* οδηγείται στη σελίδα παραγωγής του εργαλείου ποιότητας Systematic Diagram. Όπως και στην παραγωγή των προηγουμένων εργαλείων ποιότητας, ακολουθείται η ιδία διαδικασία και προσφέρονται οι ίδιες δυνατότητες.

Αναφέρεται ότι τα πεδία τα οποία συσχετίζονται για την παραγωγή κανόνων είναι η *Προέλευση Βλάβης* και η *Κατηγορία Βλάβης* τα οποία κωδικοποιούνται σύμφωνα με τις προηγούμενες περιπτώσεις.



Εικόνα 3.25: Σελίδα Systematic Diagram

Όσον αφορά στον αλγόριθμο ο οποίος δημιουργεί δυναμικά το Systematic Diagram, αυτός λειτουργεί ως εξής: Ανακτά από το πεδίο "πατέρας" τις διακριτές τιμές, τις τοποθετεί σε μια λίστα και στη συνέχεια ανακτά την σχέση του κάθε "πατέρα" με τα "παιδιά" του.

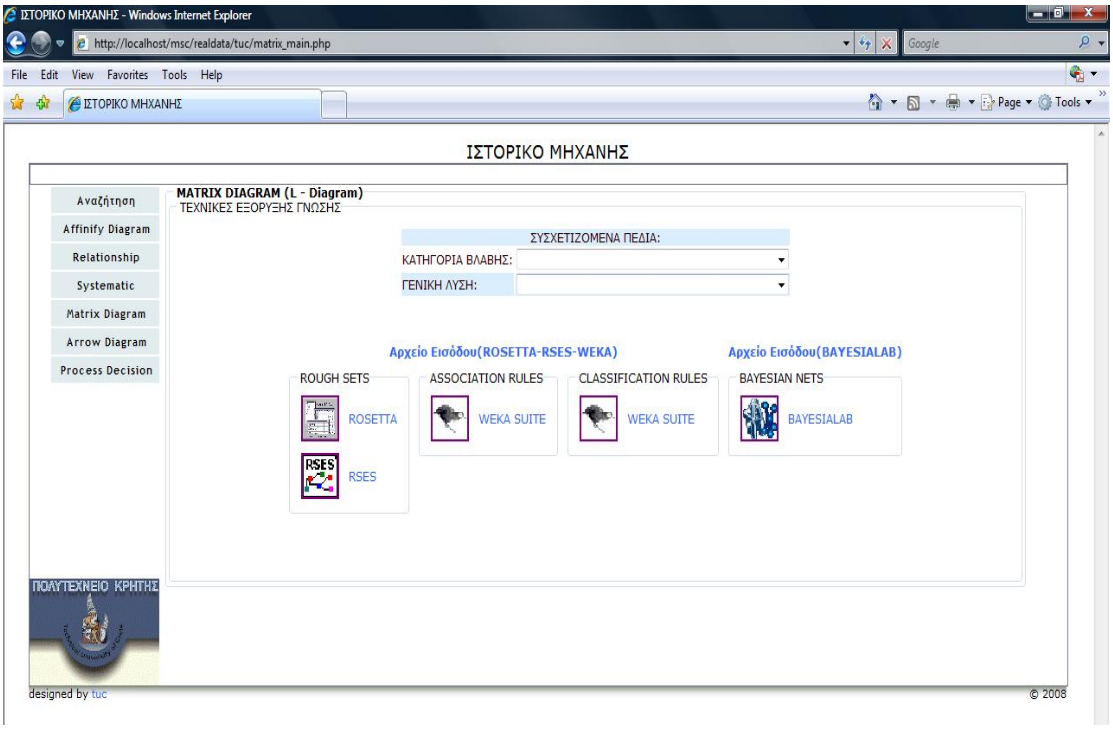


Εικόνα 3.26: Δυναμικά παραγόμενο Systematic Diagram

3.5.5 Matrix Diagram

Με την επιλογή του συνδέσμου *Matrix Diagram, ο χρήστης* οδηγείται στη σελίδα παραγωγής του εργαλείου ποιότητας Matrix Diagram. Όπως και στην παραγωγή των προηγουμένων εργαλείων ποιότητας, ακολουθείται η ίδια διαδικασία και προσφέρονται οι ίδιες δυνατότητες.

Τα δεδομένα τα οποία συσχετίζονται είναι η *Κατηγορία Βλάβης* και η *Γενική Λύση* .



Εικόνα 3.27: Σελίδα Matrix Diagram

Αναφορικά με τον αλγόριθμο ο οποίος δημιουργεί δυναμικά το Matrix Diagram (το οποίο είναι L-Diagram, δηλαδή συσχετίζει 2 πεδία), αυτός λειτουργεί ως εξής: Ανακτά όλες τις διακριτές τιμές από το πεδίο "πατέρας" και το πεδίο "παιδί" και τοποθετεί τις πρώτες σε οριζόντια στήλη και τις δεύτερες σε κάθετη. Έτσι δημιουργείται ένας δυσδιάστατος πίνακας, οπού το κάθε στοιχειό του σημειώνεται ή όχι με tick (✓), ανάλογα με το αν υπάρχει σύνδεση του συγκεκριμένου ζευγαριού τιμών "πατέρας" – "παιδί" στη βάση δεδομένων.

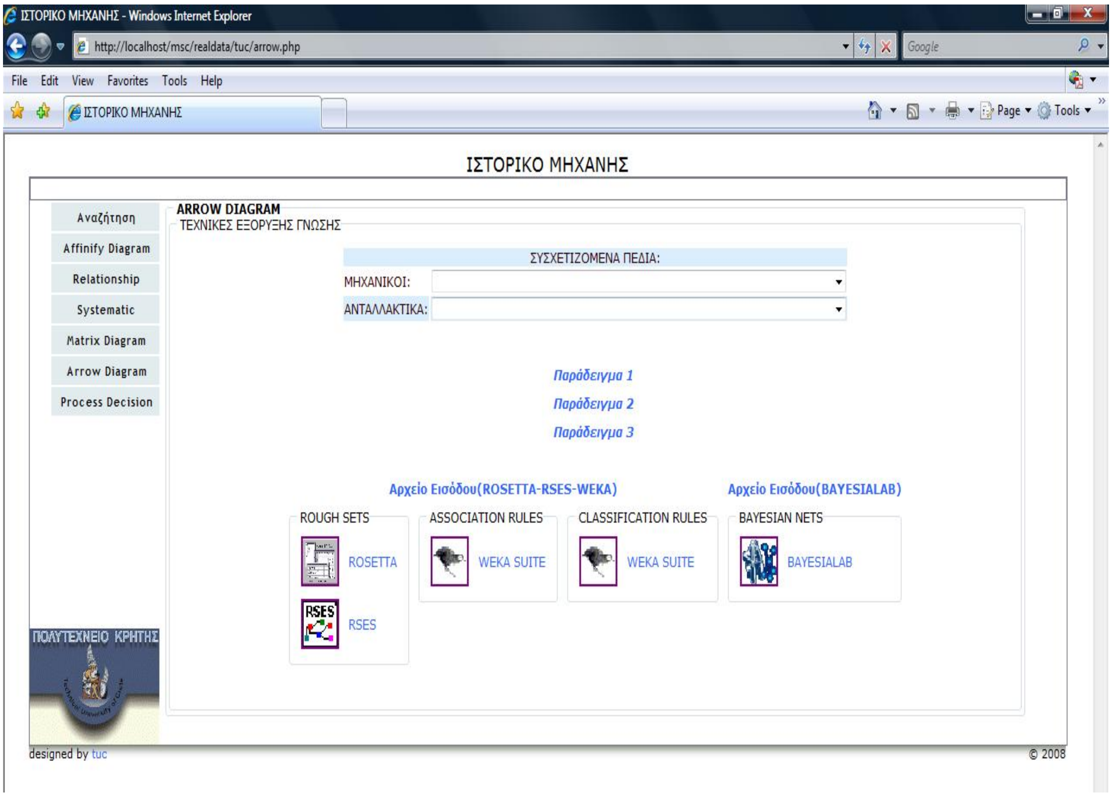
MATRIX DIAGRAM (L - Diagram) *ΚΑΤΗΓΟΡΙΑ ΒΛΑΒΗΣ/ΓΕΝΙΚΗ ΛΥΣΗ*	Προσαρμοσμένα Προϊόντα	Εκπαίδευση Προσωπικού	Έρευνα Αγοράς Προμηθευτών
Χρήση - Φθορά	-	✓ (3242)	✓ (3958)
Οδόστρωμα	✓ (698)	✓ (247)	-
Καιρικές Συνθήκες	✓ (802)	✓ (1017)	-
Προμηθευτές	-	-	✓ (2513)

Εικόνα 3.28: Δυναμικά παραγόμενο Matrix Diagram

3.5.6 Arrow Diagram

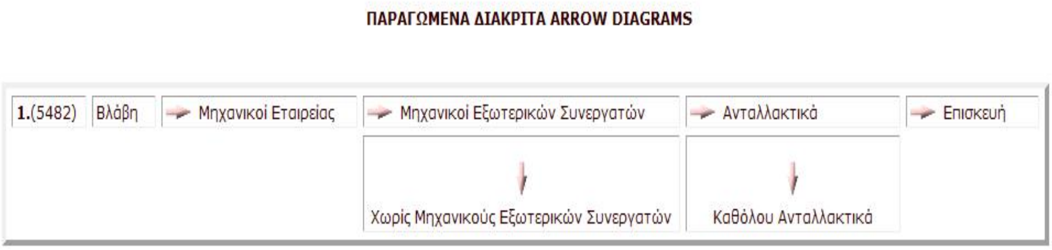
Με την επιλογή της σύνδεσης Arrow Diagram, ο χρήστης κατευθύνεται στη σελίδα παραγωγής του Arrow Diagram από τους κανόνες που ανακαλύπτουν οι τεχνικές εξόρυξης γνώσης. Και εδώ ακολουθείται η ίδια διαδικασία και προσφέρονται οι ίδιες δυνατότητες όπως και στα υπόλοιπα εργαλεία.

Τα δεδομένα τα οποία συσχετίζονται είναι οι *Μηχανικοί* και τα *Ανταλλακτικά*. Η ιδιαιτερότητα των δεδομένων αυτών είναι ότι δεν προέρχονται κατευθείαν από τη βάση δεδομένων αλλά προκύπτουν από αυτή και συγκεκριμένα από τους σχεσιακούς πίνακες της βάσης δεδομένων *outsource* και *parts* (παράγραφος 3.3). Ο στόχος είναι να μελετήσουμε τη διαδικασία που εμπλέκει τους μηχανικούς και τα ανταλλακτικά (εσωτερικά και εξωτερικά), παρόλο που τα δεδομένα αυτά δεν υπάρχουν αυτούσια για κάθε εγγραφή, αλλά προκύπτουν από τη σχέση των προαναφερόμενων πινάκων σε σχέση με το κεντρικό πίνακα *basic*. Επομένως, δημιουργώντας ένα ερώτημα SQL δημιουργήσαμε ένα αρχείο εισόδου το οποίο ισοδυναμεί με τις συνδέσεις των εμπλεκομένων πινάκων. Για να γίνει πιο κατανοητή η δημιουργία του αρχείου προσφέρονται από την κεντρική σελίδα του Arrow Diagram 3 παραδείγματα μέσω συνδέσμων σε συγκεκριμένες εγγραφές στη βάση δεδομένων.



Εικόνα 3.29: Σελίδα Arrow Diagram

Στην προκειμένη περίπτωση ο αλγόριθμος δημιουργεί τόσα arrow diagrams όσες και οι διακριτές συσχετίσεις των δυο πεδίων. Διαβάζοντας λοιπόν όλους τους κανόνες που είναι αποθηκευμένοι μετά το σάρωμα των αρχείων εξόδων, ο αλγόριθμος παραθέτει στη σειρά τις διακριτές σχέσεις των 2 πεδίων και στο τέλος δημιουργεί και το συνολικό arrow diagram όπου βρίσκονται όλες οι πιθανές συνδέσεις.



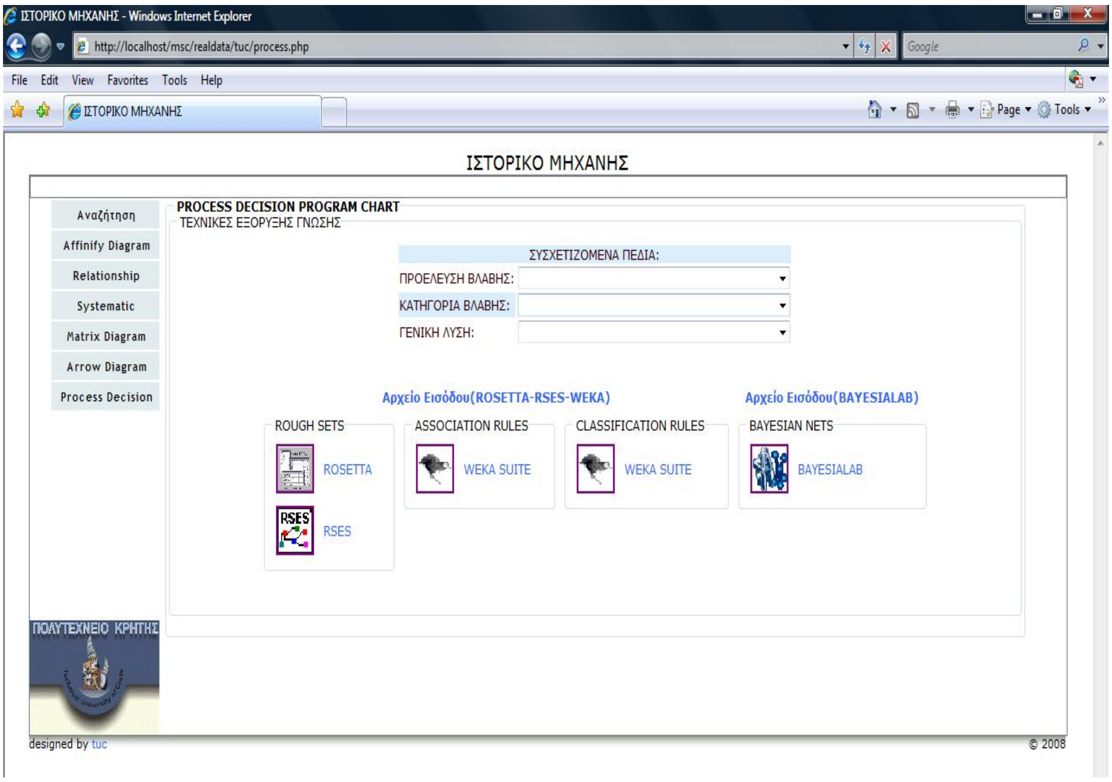
Εικόνα 3.30: Δυναμικά παραγόμενο διακριτό Arrow Diagram

Εικόνα 3.31: Δυναμικά παραγόμενο συνολικό Arrow Diagram

3.5.7 Process Decision Program Chart

Όταν ο χρήστης επιλέξει τη σύνδεση Process Decision οδηγείται στην σελίδα παραγωγής του Process Decision Program Chart. Χωρίς διαφοροποιήσεις από τα υπόλοιπα, ακολουθείται η ίδια διαδικασία και προσφέρονται οι ίδιες δυνατότητες.

Τα δεδομένα τα οποία συσχετίζονται είναι η *Προέλευση Βλάβης*, η *Κατηγορία Βλάβης* και η *Γενική Λύση*.



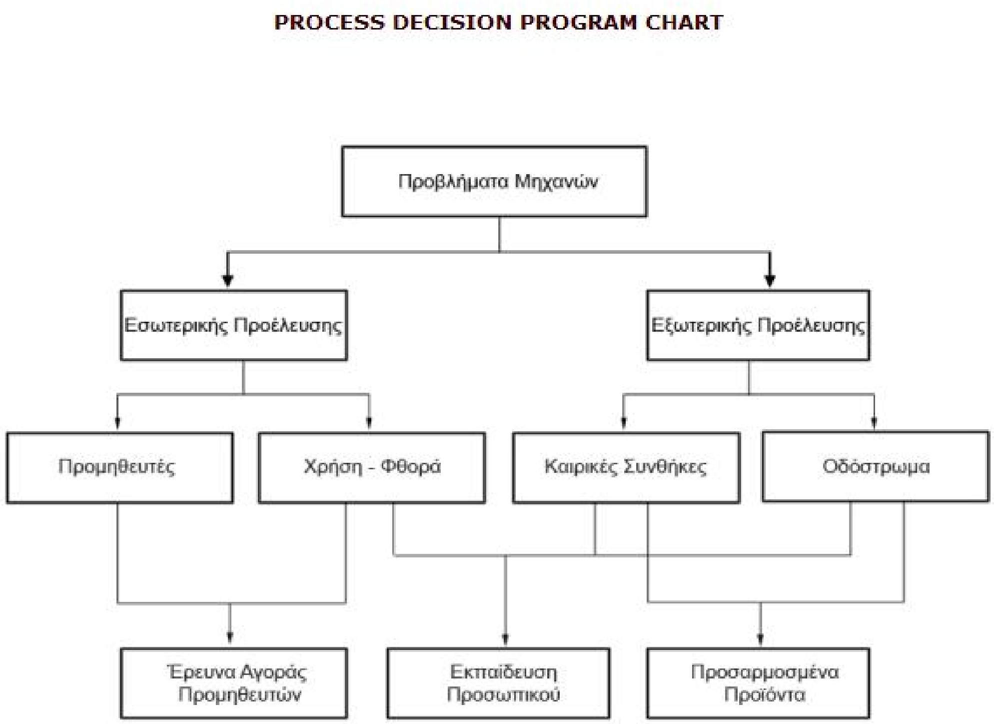
Εικόνα 3.32: Σελίδα Process Decision Program Chart

Ο αλγόριθμος παραγωγής του Process Decision Program Chart λειτουργεί ως εξής: ο αλγόριθμος ανακτά τις τιμές από το πεδίο "παιδί" οι οποίες δεν παρουσιάζονται στο πεδίο "πατέρας" και τις τοποθετεί στη σειρά. Στη συνέχεια βρίσκει τους "πατεράδες" των τιμών αυτών και τους τοποθετεί σε καινούργια σειρά στη σωστή θέση, ούτως ώστε να είναι εμφανείς οι συνδέσεις. Η ιδία διαδικασία ακολουθείται έως ότου να τελειώσουν όλοι οι κανόνες.

ΙΣΟΔΥΝΑΜΟ ΔΥΝΑΜΙΚΟ PROCESS DECISION PROGRAM CHART

Προβλήματα Μηχανών			
Εσωτερικής Προέλευσης		Εξωτερικής Προέλευσης	
Προμηθευτές	Χρήση - Φθορά	Καιρικές Συνθήκες	Οδόστρωμα
1.Έρευνα Αγοράς Προμηθευτών	1.Εκπαίδευση Προσωπικού 2.Έρευνα Αγοράς Προμηθευτών	1.Προσαρμοσμένα Προϊόντα 2.Εκπαίδευση Προσωπικού	1.Προσαρμοσμένα Προϊόντα 2.Εκπαίδευση Προσωπικού

Εικόνα 3.33: Δυναμικά παραγόμενο Process Decision Program Chart



Εικόνα 3.34: Ισοδύναμο "κλασσικό" Process Decision Program Chart

Κεφάλαιο 4

Αποτελέσματα

4.1 Εισαγωγή

Στο παρόν κεφάλαιο παρουσιάζεται το σύνολο των αποτελεσμάτων της εφαρμογής. Πιο συγκεκριμένα, παρουσιάζονται τα αποτελέσματα των 5 αλγορίθμων εξόρυξης γνώσης για κάθε νέο εργαλείο ποιότητας, τα δυναμικά παραγόμενα 7 νέα εργαλεία ποιότητας που προκύπτουν από την εξόρυξη γνώσης των 5 αλγόριθμων αλλά και τα αποτελέσματα των προβλέψεων σε νέα δεδομένα.

4.2 Affinity Diagram

Στην παράγραφο αυτή παρουσιάζονται τα αποτελέσματα που αφορούν τη δυναμική δημιουργία του Affinity Diagram μέσω των ανακαλυπτόμενων κανόνων των 5 χρησιμοποιούμενων λογισμικών. Αρχικά παρουσιάζονται τα αποτελέσματα των αλγορίθμων και στη συνέχεια το Affinity Diagram που προκύπτει για το κάθε ένα από αυτά. Τέλος, παρουσιάζονται και τα αποτελέσματα της πρόβλεψης σε ένα σύνολο νέων δεδομένων.

4.2.1 Rough Set - Rosetta

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα (τα οποία παρουσιάζονται στην παράγραφο 3.5.2) στην περίπτωση επιλογής του Rosetta ως λογισμικού εξόρυξης γνώσης είναι το εξής:

```
% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:27:13 by mike.
%
% 1R
% 29 rules.

code(Υδραυλικά) => category(Προμηθευτές) OR category(Καιρικές_Συνθήκες)
Supp. (LHS) = [883 object(s)]
Supp. (RHS) = [577 object(s), 306 object(s)]
Acc. (RHS) = [0.653454, 0.346546]
Cov. (LHS) = [0.0707702]
Cov. (RHS) = [0.229606, 0.168224]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Ηλεκτρικά_-_Ηλεκτρονικά) => category(Προμηθευτές) OR category(Καιρικές_Συνθήκες)
Supp. (LHS) = [1529 object(s)]
Supp. (RHS) = [440 object(s), 1089 object(s)]
Acc. (RHS) = [0.28777, 0.71223]
Cov. (LHS) = [0.122545]
Cov. (RHS) = [0.17509, 0.598681]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Αμάξωμα) => category(Προμηθευτές) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [1748 object(s)]
Supp. (RHS) = [771 object(s), 977 object(s)]
Acc. (RHS) = [0.441076, 0.558924]
Cov. (LHS) = [0.140098]
Cov. (RHS) = [0.306805, 0.135694]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Σύστημα_Κίνησης) => category(Προμηθευτές) OR category(Οδόστρωμα) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [454 object(s)]
Supp. (RHS) = [50 object(s), 252 object(s), 152 object(s)]
Acc. (RHS) = [0.110132, 0.555066, 0.334802]
Cov. (LHS) = [0.036387]
Cov. (RHS) = [0.0198965, 0.266667, 0.0211111]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1]

code(Κινητήρας) => category(Προμηθευτές) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [561 object(s)]
Supp. (RHS) = [342 object(s), 219 object(s)]
Acc. (RHS) = [0.609626, 0.390374]
Cov. (LHS) = [0.0449627]
Cov. (RHS) = [0.136092, 0.0304167]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Σύστημα_Πέδησης) => category(Προμηθευτές) OR category(Οδόστρωμα) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [514 object(s)]
Supp. (RHS) = [60 object(s), 278 object(s), 176 object(s)]
Acc. (RHS) = [0.116732, 0.540856, 0.342412]
Cov. (LHS) = [0.0411958]
Cov. (RHS) = [0.0238758, 0.29418, 0.0244444]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1]

code(Εξάτμιση) => category(Προμηθευτές)
Supp. (LHS) = [138 object(s)]
Supp. (RHS) = [138 object(s)]
Acc. (RHS) = [1]
```

Cov. (LHS) = [0.0110604]
Cov. (RHS) = [0.0549144]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Ανάρτηση) => category(Προμηθευτές) OR category(Οδόστρωμα)
Supp. (LHS) = [297 object(s)]
Supp. (RHS) = [91 object(s), 206 object(s)]
Acc. (RHS) = [0.306397, 0.693603]
Cov. (LHS) = [0.0238038]
Cov. (RHS) = [0.0362117, 0.217989]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Καρότσα) => category(Προμηθευτές) OR category(Καιρικές_Συνθήκες)
Supp. (LHS) = [26 object(s)]
Supp. (RHS) = [19 object(s), 7 object(s)]
Acc. (RHS) = [0.730769, 0.269231]
Cov. (LHS) = [0.00208383]
Cov. (RHS) = [0.00756068, 0.00384827]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Κινητήρας_Αναμικτήρα) => category(Προμηθευτές) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [9 object(s)]
Supp. (RHS) = [5 object(s), 4 object(s)]
Acc. (RHS) = [0.555556, 0.444444]
Cov. (LHS) = [0.000721327]
Cov. (RHS) = [0.00198965, 0.000555556]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Σύστημα_Διεύθυνσης) => category(Προμηθευτές) OR category(Οδόστρωμα) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [80 object(s)]
Supp. (RHS) = [9 object(s), 44 object(s), 27 object(s)]
Acc. (RHS) = [0.1125, 0.55, 0.3375]
Cov. (LHS) = [0.0064118]
Cov. (RHS) = [0.00358138, 0.0465608, 0.00375]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1]

code(Σύστημα_Ψύξης) => category(Προμηθευτές) OR category(Οδόστρωμα) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [119 object(s)]
Supp. (RHS) = [9 object(s), 27 object(s), 83 object(s)]
Acc. (RHS) = [0.0756303, 0.226891, 0.697479]
Cov. (LHS) = [0.00953755]
Cov. (RHS) = [0.00358138, 0.0285714, 0.0115278]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1]

code(Κινητήρας_-_Τρικουβέρτο) => category(Προμηθευτές) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [7 object(s)]
Supp. (RHS) = [2 object(s), 5 object(s)]
Acc. (RHS) = [0.285714, 0.714286]
Cov. (LHS) = [0.000561032]
Cov. (RHS) = [0.000795862, 0.000694444]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

code(Νέα_Ελαστικά) => category(Καιρικές_Συνθήκες) OR category(Οδόστρωμα) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [199 object(s)]
Supp. (RHS) = [105 object(s), 43 object(s), 51 object(s)]
Acc. (RHS) = [0.527638, 0.21608, 0.256281]

Cov. (LHS) = [0.0159493]
Cov. (RHS) = [0.057724, 0.0455026, 0.00708333]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1]

code(Αλλαγή_-_Επισκευή_Ελαστικών) => category(Καιρικές_Συνθήκες) OR category(Οδόστρωμα) OR category(Χρήση_-_Φθορά)
Supp. (LHS) = [534 object(s)]
Supp. (RHS) = [309 object(s), 95 object(s), 130 object(s)]
Acc. (RHS) = [0.578652, 0.177903, 0.243446]
Cov. (LHS) = [0.0427987]
Cov. (RHS) = [0.169874, 0.100529, 0.0180556]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1]

code(Κλιματισμός) => category(Καιρικές_Συνθήκες)
Supp. (LHS) = [3 object(s)]
Supp. (RHS) = [3 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.000240442]
Cov. (RHS) = [0.00164926]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Έλεγχος) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [286 object(s)]
Supp. (RHS) = [286 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0229222]
Cov. (RHS) = [0.0397222]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Γρασάρισμα) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [807 object(s)]
Supp. (RHS) = [807 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.064679]
Cov. (RHS) = [0.112083]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Λάδια-Φίλτρα) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [2570 object(s)]
Supp. (RHS) = [2570 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.205979]
Cov. (RHS) = [0.356944]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Καθάρισμα) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [375 object(s)]
Supp. (RHS) = [375 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0300553]
Cov. (RHS) = [0.0520833]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(ΚΤΕΟ) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [116 object(s)]
Supp. (RHS) = [116 object(s)]
Acc. (RHS) = [1]

Cov. (LHS) = [0.00929711]
Cov. (RHS) = [0.0161111]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Μπούμα) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [350 object(s)]
Supp. (RHS) = [350 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0280516]
Cov. (RHS) = [0.0486111]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Πλύσιμο) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [303 object(s)]
Supp. (RHS) = [303 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0242847]
Cov. (RHS) = [0.0420833]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Πλύσιμο_-_Γρασάρισμα) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [95 object(s)]
Supp. (RHS) = [95 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.00761401]
Cov. (RHS) = [0.0131944]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Mixer) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [29 object(s)]
Supp. (RHS) = [29 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.00232428]
Cov. (RHS) = [0.00402778]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(?ντληση) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [128 object(s)]
Supp. (RHS) = [128 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0102589]
Cov. (RHS) = [0.0177778]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Συντήρηση) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [218 object(s)]
Supp. (RHS) = [218 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0174721]
Cov. (RHS) = [0.0302778]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Εργασία) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [79 object(s)]
Supp. (RHS) = [79 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.00633165]

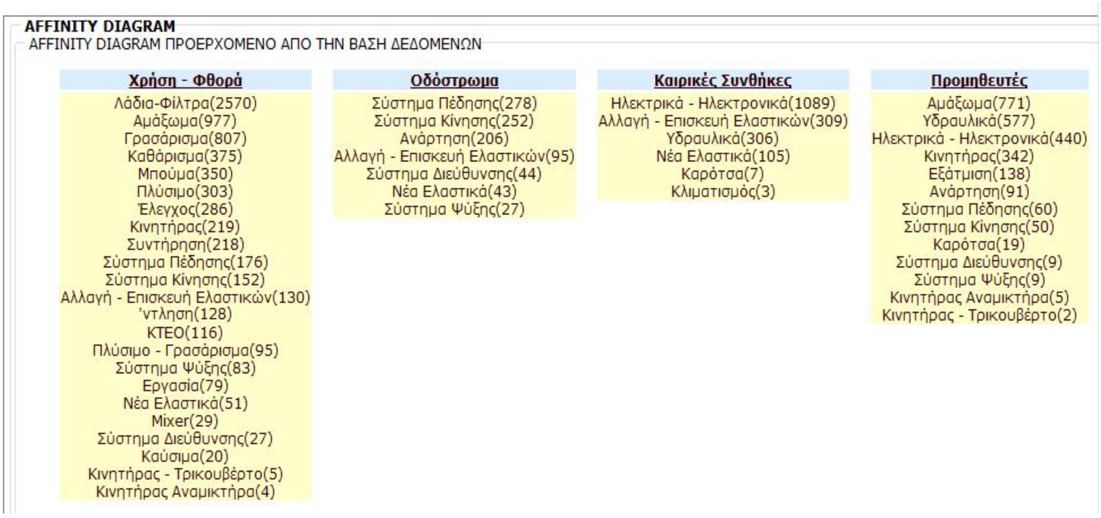
Cov. (RHS) = [0.0109722]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

code(Καύσιμα) => category(Χρήση_-_Φθορά)
Supp. (LHS) = [20 object(s)]
Supp. (RHS) = [20 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.00160295]
Cov. (RHS) = [0.00277778]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

Σε αυτά τα δεδομένα ενεργοποιείται ο αλγόριθμος “σάρωσης” ο οποίος εκμαιεύει τους κανόνες που βρίσκονται στο αρχείο εξόδου (σχετικά με τη διαδικασία βλ. παράγραφο 3.5.2). Οι κανόνες αυτοί αποθηκεύονται σε ένα πίνακα στη βάση δεδομένων και δημιουργούν το παρακάτω δυναμικό Affinity Diagram.

Εικόνα 4.1: Παραγόμενο Affinity Diagram από τους κανόνες που ανακάλυψε ο αλγόριθμος του Rosetta

Τα αποτελέσματα του αλγορίθμου επιβεβαιώνονται από τα αποτελέσματα ενός συγκεκριμενοποιημένου ερωτήματος (query) στη βάση δεδομένων. Επιβεβαιώνονται τόσο ποιοτικά (κανόνες), όσο και ποσοτικά (ο αριθμός των υποδειγμάτων που ικανοποιούν τον κανόνα).



Εικόνα 4.2: Παραγόμενο Affinity Diagram από την βάση δεδομένων

Επίσης, εφαρμόζουμε τους κανόνες σε νέα δεδομένα στα οποία επιθυμούμε να προβλέψουμε την κατηγορία τους. Ακολουθεί το αρχείο εισόδου των δεδομένων.

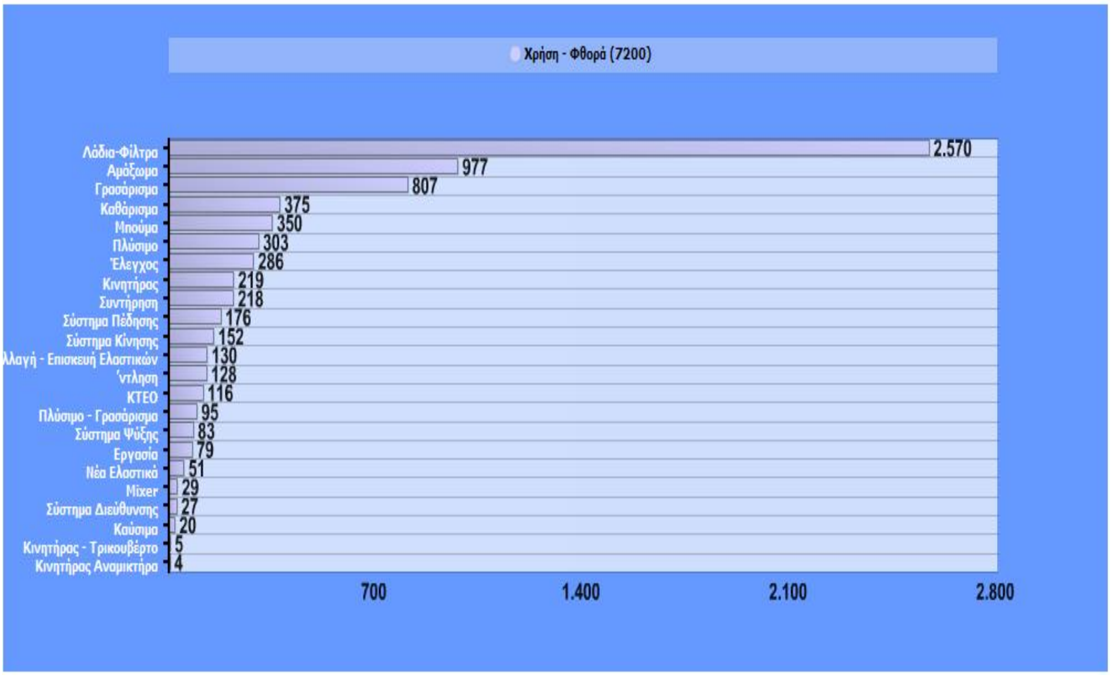
Σύστημα Πέδησης
Σύστημα Κίνησης
Ανάρτηση
Κινητήρας Αναμικτήρα
Κινητήρας - Τρικουβέρτο
Σύστημα Διεύθυνσης
Αλλαγή - Επισκευή Ελαστικών
Υδραυλικά
Νέα Ελαστικά
Κλιματισμός
Λάδια-Φίλτρα
Γρασάρισμα
Καθάρισμα
Μπούμα
Πλύσιμο
Κινητήρας Αναμικτήρα
Σύστημα Κίνησης
Σύστημα Κίνησης
Σύστημα Κίνησης
Σύστημα Κίνησης

Παίρνοντας λοιπόν τα παραπάνω δεδομένα ως είσοδο και εφαρμόζοντας σε αυτά τους αποθηκευμένους κανόνες παράγεται η κάτωθι πρόβλεψη:

Είσοδος (Κωδικός Βλάβης)	**Πρόβλεψη (Κατηγορία Βλάβης)**
Σύστημα Πέδησης	*Χρήση - Φθορά* (34.24%) *Οδόστρωμα* (54.09%) *Προμηθευτές* (11.67%)

Σύστημα Κίνησης	*Χρήση - Φθορά* (33.48%) *Οδόστρωμα* (55.51%) *Προμηθευτές* (11.01%)
Ανάρτηση	*Οδόστρωμα* (69.36%) *Προμηθευτές* (30.64%)
Κινητήρας Αναμικτήρα	*Χρήση - Φθορά* (44.44%) *Προμηθευτές* (55.56%)
Κινητήρας - Τρικουβέρτο	*Χρήση - Φθορά* (71.43%) *Προμηθευτές* (28.57%)
Σύστημα Διεύθυνσης	*Χρήση - Φθορά* (33.75%) *Οδόστρωμα* (55%) *Προμηθευτές* (11.25%)
Αλλαγή - Επισκευή Ελαστικών	*Χρήση - Φθορά* (24.34%) *Οδόστρωμα* (17.79%) *Καιρικές Συνθήκες* (57.87%)
Υδραυλικά	*Καιρικές Συνθήκες* (34.65%) *Προμηθευτές* (65.35%)
Νέα Ελαστικά	*Χρήση - Φθορά* (25.63%) *Οδόστρωμα* (21.61%) *Καιρικές Συνθήκες* (52.76%)
Κλιματισμός	*Καιρικές Συνθήκες* (100%)
Λάδια-Φίλτρα	*Χρήση - Φθορά* (100%)
Γρασάρισμα	*Χρήση - Φθορά* (100%)
Καθάρισμα	*Χρήση - Φθορά* (100%)
Μπούμα	*Χρήση - Φθορά* (100%)
Πλύσιμο	*Χρήση - Φθορά* (100%)
Κινητήρας Αναμικτήρα	*Χρήση - Φθορά* (44.44%) *Προμηθευτές* (55.56%)
Σύστημα Κίνησης	*Χρήση - Φθορά* (33.48%) *Οδόστρωμα* (55.51%) *Προμηθευτές* (11.01%)
Σύστημα Κίνησης	*Χρήση - Φθορά* (33.48%) *Οδόστρωμα* (55.51%) *Προμηθευτές* (11.01%)
Σύστημα Κίνησης	*Χρήση - Φθορά* (33.48%) *Οδόστρωμα* (55.51%) *Προμηθευτές* (11.01%)
Σύστημα Κίνησης	*Χρήση - Φθορά* (33.48%) *Οδόστρωμα* (55.51%) *Προμηθευτές* (11.01%)

Επίσης παράγονται δυναμικά από τους κανόνες και τα στατιστικά διαγράμματα των κανόνων.



Εικόνα 4.3: Δείγμα από τα στατιστικά των κανόνων

4.2.2 Rough Set – RSES2

Το αρχείο εξόδου του RSES2 είναι το εξής:

RULE_SET rses
ATTRIBUTES 2
code symbolic
category symbolic
DECISION_VALUES 4
Προμηθευτές
Καιρικές_Συνθήκες
Οδόστρωμα
Χρήση_-_Φθορά
RULES 29
(code=Λάδια-Φίλτρα)=>(category=Χρήση_-_Φθορά[2570]) 2570
(code=Αμάξωμα)=>(category={Προμηθευτές[771],Χρήση_-_Φθορά[977]}) 1748
(code=Ηλεκτρικά_-_Ηλεκτρονικά)=>(category={Προμηθευτές[440],Καιρικές_Συνθήκες[1089]}) 1529
(code=Υδραυλικά)=>(category={Προμηθευτές[577],Καιρικές_Συνθήκες[306]}) 883
(code=Γρασάρισμα)=>(category=Χρήση_-_Φθορά[807]) 807
(code=Κινητήρας)=>(category={Προμηθευτές[342],Χρήση_-_Φθορά[219]}) 561
(code=Αλλαγή_-_Επισκευή_Ελαστικών)=>(category={Καιρικές_Συνθήκες[309],Οδόστρωμα[95],Χρήση_-_Φθορά[130]}) 534
(code=Σύστημα_Πέδησης)=>(category={Προμηθευτές[60],Οδόστρωμα[278],Χρήση_-_Φθορά[176]}) 514
(code=Σύστημα_Κίνησης)=>(category={Προμηθευτές[50],Οδόστρωμα[252],Χρήση_-_Φθορά[152]}) 454
(code=Καθάρισμα)=>(category=Χρήση_-_Φθορά[375]) 375

(code=Μπούμα)=>(category=Χρήση_-_Φθορά[350]) 350
(code=Πλύσιμο)=>(category=Χρήση_-_Φθορά[303]) 303
(code=Ανάρτηση)=>(category={Προμηθευτές[91],Οδόστρωμα[206]}) 297
(code=Έλεγχος)=>(category=Χρήση_-_Φθορά[286]) 286
(code=Συντήρηση)=>(category=Χρήση_-_Φθορά[218]) 218
(code=Νέα_Ελαστικά)=>(category={Καιρικές_Συνθήκες[105],Οδόστρωμα[43],Χρήση_-_Φθορά[51]}) 199
(code=Εξάτμιση)=>(category=Προμηθευτές[138]) 138
(code=?ντληση)=>(category=Χρήση_-_Φθορά[128]) 128
(code=Σύστημα_Ψύξης)=>(category={Προμηθευτές[9],Οδόστρωμα[27],Χρήση_-_Φθορά[83]}) 119
(code=ΚΤΕΟ)=>(category=Χρήση_-_Φθορά[116]) 116
(code=Πλύσιμο_-_Γρασάρισμα)=>(category=Χρήση_-_Φθορά[95]) 95
(code=Σύστημα_Διεύθυνσης)=>(category={Προμηθευτές[9],Οδόστρωμα[44],Χρήση_-_Φθορά[27]}) 80
(code=Εργασία)=>(category=Χρήση_-_Φθορά[79]) 79
(code=Mixer)=>(category=Χρήση_-_Φθορά[29]) 29
(code=Καρότσα)=>(category={Προμηθευτές[19],Καιρικές_Συνθήκες[7]}) 26
(code=Καύσιμα)=>(category=Χρήση_-_Φθορά[20]) 20
(code=Κινητήρας_Αναμικτήρα)=>(category={Προμηθευτές[5],Χρήση_-_Φθορά[4]}) 9
(code=Κινητήρας_-_Τρικουβέρτο)=>(category={Προμηθευτές[2],Χρήση_-_Φθορά[5]}) 7
(code=Κλιματισμός)=>(category=Καιρικές_Συνθήκες[3]) 3

Όπως και στην περίπτωση του Rosetta, ενεργοποιείται ο αλγόριθμος σάρωσης των αποτελεσμάτων του RSES2 ο οποίος παράγει τους κανόνες και στη συνεχεία ενεργοποιείται ο επόμενος αλγόριθμος που δημιουργεί δυναμικά το Affinity Diagram. Το σύνολο των αποτελεσμάτων (κανόνες, Affinity Diagram, πρόβλεψη) που προκύπτουν από το RSES2, ταυτίζονται απολύτως με τα αποτελέσματα του Rosetta.

4.2.3 Association Rules – Weka Suite

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα στην περίπτωση του Weka Suite (Association Rules) είναι το εξής:

=== Run information ===

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: affinity
Instances: 12477
Attributes: 2
code
result
=== Associator model (full training set) ===

Tertius
=======

1. /* 0,418751 0,000000 */ code = Λάδια-Φίλτρα ==> result = Χρήση_-_Φθορά
2. /* 0,317162 0,035265 */ code = Ηλεκτρικά_-_Ηλεκτρονικά ==> result = Καιρικές_Συνθήκες
3. /* 0,198170 0,000000 */ code = Γρασάρισμα ==> result = Χρήση_-_Φθορά
4. /* 0,176537 0,024525 */ code = Υδραυλικά ==> result = Προμηθευτές
5. /* 0,150835 0,078304 */ code = Αμάξωμα ==> result = Προμηθευτές
6. /* 0,127072 0,000000 */ code = Καθάρισμα ==> result = Χρήση_-_Φθορά
7. /* 0,122237 0,000000 */ code = Μπούμα ==> result = Χρήση_-_Φθορά
8. /* 0,122002 0,018915 */ code = Σύστημα_Πέδησης ==> result = Οδόστρωμα
9. /* 0,119797 0,018033 */ code = Αλλαγή_-_Επισκευή_Ελαστικών ==> result = Καιρικές_Συνθήκες
10. /* 0,119508 0,017552 */ code = Κινητήρας ==> result = Προμηθευτές

11. /* 0,116464 0,016190 */ code = Σύστημα_Κίνησης ==> result = Οδόστρωμα
12. /* 0,116425 0,007293 */ code = Ανάρτηση ==> result = Οδόστρωμα
13. /* 0,112775 0,000000 */ code = Πλύσιμο ==> result = Χρήση_-_Φθορά
14. /* 0,109215 0,000000 */ code = Έλεγχος ==> result = Χρήση_-_Φθορά
15. /* 0,103731 0,000000 */ code = Εξάτμιση ==> result = Προμηθευτές
16. /* 0,094048 0,000000 */ code = Συντήρηση ==> result = Χρήση_-_Φθορά
17. /* 0,076625 0,046245 */ code = Υδραυλικά ==> result = Καιρικές_Συνθήκες
18. /* 0,070515 0,000000 */ code = ?ντληση ==> result = Χρήση_-_Φθορά
19. /* 0,066902 0,000000 */ code = ΚΤΕΟ ==> result = Χρήση_-_Φθορά
20. /* 0,060161 0,000000 */ code = Πλύσιμο_-_Γρασάρισμα ==> result = Χρήση_-_Φθορά
21. /* 0,059072 0,007534 */ code = Νέα_Ελαστικά ==> result = Καιρικές_Συνθήκες
22. /* 0,054572 0,000000 */ code = Εργασία ==> result = Χρήση_-_Φθορά
23. /* 0,049230 0,087281 */ code = Ηλεκτρικά_-_Ηλεκτρονικά ==> result = Προμηθευτές
24. /* 0,042796 0,002885 */ code = Σύστημα_Διεύθυνσης ==> result = Οδόστρωμα
25. /* 0,032368 0,000000 */ code = Mixer ==> result = Χρήση_-_Φθορά
26. /* 0,028191 0,000561 */ code = Καρότσα ==> result = Προμηθευτές
27. /* 0,027442 0,035185 */ code = Αλλαγή_-_Επισκευή_Ελαστικών ==> result = Οδόστρωμα
28. /* 0,026734 0,000000 */ code = Καύσιμα ==> result = Χρήση_-_Φθορά
29. /* 0,021024 0,016510 */ code = Ανάρτηση ==> result = Προμηθευτές
30. /* 0,020983 0,012503 */ code = Νέα_Ελαστικά ==> result = Οδόστρωμα
31. /* 0,019309 0,002885 */ code = Σύστημα_Ψύξης ==> result = Χρήση_-_Φθορά
32. /* 0,016945 0,007374 */ code = Σύστημα_Ψύξης ==> result = Οδόστρωμα
33. /* 0,014540 0,000000 */ code = Κλιματισμός ==> result = Καιρικές_Συνθήκες
34. /* 0,010905 0,000321 */ code = Κινητήρας_Αναμικτήρα ==> result = Προμηθευτές
35. /* 0,006366 0,001523 */ code = Καρότσα ==> result = Καιρικές_Συνθήκες
36. /* 0,005076 0,000160 */ code = Κινητήρας_-_Τρικουβέρτο ==> result = Χρήση_-_Φθορά
37. /* 0,002283 0,000401 */ code = Κινητήρας_-_Τρικουβέρτο ==> result = Προμηθευτές
38. /* 0,000000 0,000240 */ code = Κλιματισμός ==> FALSE
39. /* 0,000000 0,000561 */ code = Κινητήρας_-_Τρικουβέρτο ==> FALSE
40. /* 0,000000 0,000721 */ code = Κινητήρας_Αναμικτήρα ==> FALSE
41. /* 0,000000 0,001603 */ code = Καύσιμα ==> FALSE
42. /* 0,000000 0,002084 */ code = Καρότσα ==> FALSE
43. /* 0,000000 0,002324 */ code = Mixer ==> FALSE
44. /* 0,000000 0,006332 */ code = Εργασία ==> FALSE
45. /* 0,000000 0,006412 */ code = Σύστημα_Διεύθυνσης ==> FALSE
46. /* 0,000000 0,007614 */ code = Πλύσιμο_-_Γρασάρισμα ==> FALSE
47. /* 0,000000 0,009297 */ code = ΚΤΕΟ ==> FALSE
48. /* 0,000000 0,009538 */ code = Σύστημα_Ψύξης ==> FALSE
49. /* 0,000000 0,010259 */ code = ?ντληση ==> FALSE
50. /* 0,000000 0,011060 */ code = Εξάτμιση ==> FALSE
51. /* 0,000000 0,015949 */ code = Νέα_Ελαστικά ==> FALSE
52. /* 0,000000 0,017472 */ code = Συντήρηση ==> FALSE
53. /* 0,000000 0,022922 */ code = Έλεγχος ==> FALSE
54. /* 0,000000 0,023804 */ code = Ανάρτηση ==> FALSE
55. /* 0,000000 0,024285 */ code = Πλύσιμο ==> FALSE
56. /* 0,000000 0,028052 */ code = Μπούμα ==> FALSE
57. /* 0,000000 0,030055 */ code = Καθάρισμα ==> FALSE
58. /* 0,000000 0,036387 */ code = Σύστημα_Κίνησης ==> FALSE
59. /* 0,000000 0,041196 */ code = Σύστημα_Πέδησης ==> FALSE
60. /* 0,000000 0,042799 */ code = Αλλαγή_-_Επισκευή_Ελαστικών ==> FALSE
61. /* 0,000000 0,044963 */ code = Κινητήρας ==> FALSE
62. /* 0,000000 0,064679 */ code = Γρασάρισμα ==> FALSE
63. /* 0,000000 0,070770 */ code = Υδραυλικά ==> FALSE
64. /* 0,000000 0,122545 */ code = Ηλεκτρικά_-_Ηλεκτρονικά ==> FALSE
65. /* 0,000000 0,140098 */ code = Αμάξωμα ==> FALSE
66. /* 0,000000 0,205979 */ code = Λάδια-Φίλτρα ==> FALSE
67. /* 0,000000 0,422938 */ TRUE ==> result = Χρήση_-_Φθορά
68. /* 0,000000 0,798589 */ TRUE ==> result = Προμηθευτές
69. /* 0,000000 0,854212 */ TRUE ==> result = Καιρικές_Συνθήκες
70. /* 0,000000 0,924261 */ TRUE ==> result = Οδόστρωμα
71. /* -0,005574 0,000401 */ code = Κινητήρας_Αναμικτήρα ==> result = Χρήση_-_Φθορά
72. /* -0,008581 0,005690 */ code = Σύστημα_Διεύθυνσης ==> result = Προμηθευτές
73. /* -0,013797 0,061794 */ code = Αμάξωμα ==> result = Χρήση_-_Φθορά
74. /* -0,015060 0,008816 */ code = Σύστημα_Ψύξης ==> result = Προμηθευτές

75. /* -0,023488 0,032380 */ code = Σύστημα_Κίνησης ==> result = Προμηθευτές
76. /* -0,023494 0,036387 */ code = Σύστημα_Πέδησης ==> result = Προμηθευτές
77. /* -0,031117 0,004248 */ code = Σύστημα_Διεύθυνσης ==> result = Χρήση_-_Φθορά
78. /* -0,067867 0,011862 */ code = Νέα_Ελαστικά ==> result = Χρήση_-_Φθορά
79. /* -0,070607 0,027410 */ code = Κινητήρας ==> result = Χρήση_-_Φθορά
80. /* -0,081122 0,024205 */ code = Σύστημα_Κίνησης ==> result = Χρήση_-_Φθορά
81. /* -0,084370 0,027090 */ code = Σύστημα_Πέδησης ==> result = Χρήση_-_Φθορά
82. /* -0,122625 0,032380 */ code = Αλλαγή_-_Επισκευή_Ελαστικών ==> result = Χρήση_-_Φθορά

Number of hypotheses considered: 149
Number of hypotheses explored: 149
Time: 00 min 00 s 858 ms

Όπως και στις παραπάνω περιπτώσεις, ενεργοποιείται ο αλγόριθμος σάρωσης των αποτελεσμάτων (στην περίπτωση μας ο αλγόριθμος σάρωσης του αρχείου εξόδου του λογισμικού Weka Suite (Association Rules)) ο οποίος παράγει τους κανόνες και στη συνέχεια ενεργοποιείται ο επόμενος αλγόριθμος που δημιουργεί δυναμικά το Affinity Diagram. Το σύνολο των αποτελεσμάτων (κανόνες, Affinity Diagram, πρόβλεψη) που προκύπτουν από το Weka Suite (Association Rules), ταυτίζονται απολύτως με τα αποτελέσματα των προηγουμένων, με τη διαφορά όμως ότι δεν υπάρχουν ποσοτικά δεδομένα.

4.2.4 Classification Rules – Weka Suite

Το αρχείο εξόδου στην περίπτωση του Weka Suite (Classification Rules) είναι το εξής:

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: affinity
Instances: 12477
Attributes: 2
code
result
Test mode: 10-fold cross-validation

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

Prism rules

If code = Εξάτμιση then Προμηθευτές
If code = Καρότσα then Προμηθευτές
If code = Υδραυλικά then Προμηθευτές
If code = Κινητήρας then Προμηθευτές
If code = Κινητήρας_Αναμικτήρα then Προμηθευτές
If code = Αμάξωμα then Προμηθευτές
If code = Ανάρτηση then Προμηθευτές
If code = Ηλεκτρικά_-_Ηλεκτρονικά then Προμηθευτές
If code = Κινητήρας_-_Τρικουβέρτο then Προμηθευτές
If code = Σύστημα_Πέδησης then Προμηθευτές
If code = Σύστημα_Διεύθυνσης then Προμηθευτές

If code = Σύστημα_Κίνησης then Προμηθευτές
If code = Σύστημα_Ψύξης then Προμηθευτές
If code = Λάδια-Φίλτρα then Χρήση_-_Φθορά
If code = Γρασάρισμα then Χρήση_-_Φθορά
If code = Καθάρισμα then Χρήση_-_Φθορά
If code = Μπούμα then Χρήση_-_Φθορά
If code = Πλύσιμο then Χρήση_-_Φθορά
If code = Έλεγχος then Χρήση_-_Φθορά
If code = Συντήρηση then Χρήση_-_Φθορά
If code = ?ντληση then Χρήση_-_Φθορά
If code = ΚΤΕΟ then Χρήση_-_Φθορά
If code = Πλύσιμο_-_Γρασάρισμα then Χρήση_-_Φθορά
If code = Εργασία then Χρήση_-_Φθορά
If code = Mixer then Χρήση_-_Φθορά
If code = Καύσιμα then Χρήση_-_Φθορά
If code = Κινητήρας_-_Τρικουβέρτο then Χρήση_-_Φθορά
If code = Σύστημα_Ψύξης then Χρήση_-_Φθορά
If code = Αμάξωμα then Χρήση_-_Φθορά
If code = Κινητήρας_Αναμικτήρα then Χρήση_-_Φθορά
If code = Κινητήρας then Χρήση_-_Φθορά
If code = Σύστημα_Πέδησης then Χρήση_-_Φθορά
If code = Σύστημα_Διεύθυνσης then Χρήση_-_Φθορά
If code = Σύστημα_Κίνησης then Χρήση_-_Φθορά
If code = Νέα_Ελαστικά then Χρήση_-_Φθορά
If code = Αλλαγή_-_Επισκευή_Ελαστικών then Χρήση_-_Φθορά
If code = Ανάρτηση then Οδόστρωμα
If code = Σύστημα_Κίνησης then Οδόστρωμα
If code = Σύστημα_Διεύθυνσης then Οδόστρωμα
If code = Σύστημα_Πέδησης then Οδόστρωμα
If code = Σύστημα_Ψύξης then Οδόστρωμα
If code = Νέα_Ελαστικά then Οδόστρωμα
If code = Αλλαγή_-_Επισκευή_Ελαστικών then Οδόστρωμα
If code = Κλιματισμός then Καιρικές_Συνθήκες
If code = Ηλεκτρικά_-_Ηλεκτρονικά then Καιρικές_Συνθήκες
If code = Αλλαγή_-_Επισκευή_Ελαστικών then Καιρικές_Συνθήκες
If code = Νέα_Ελαστικά then Καιρικές_Συνθήκες
If code = Υδραυλικά then Καιρικές_Συνθήκες
If code = Καρότσα then Καιρικές_Συνθήκες

Time taken to build model: 0.25 seconds

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

Correctly Classified Instances 8071 64.687 %
Incorrectly Classified Instances 4406 35.313 %
Kappa statistic 0.4254
Mean absolute error 0.1766
Root mean squared error 0.4202
Relative absolute error 58.9044 %
Root relative squared error 108.5446 %
Total Number of Instances 12477

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
0.999 0.387 0.395 0.999 0.566 Προμηθευτές
0.772 0.105 0.909 0.772 0.835 Χρήση_-_Φθορά
0 0 0 0 0 Οδόστρωμα
0.002 0 1 0.002 0.003 Καιρικές_Συνθήκες

=== Confusion Matrix ===

a b c d <-- classified as
2511 2 0 0 | a = Προμηθευτές
1643 5557 0 0 | b = Χρήση_-_Φθορά
807 138 0 0 | c = Οδόστρωμα
1402 414 0 3 | d = Καιρικές_Συνθήκες

Τα αποτελέσματα και οι διαδικασίες ταυτίζονται με αυτές του Weka Suite - Association Rules, εκτός από τον αλγόριθμο σάρωσης του αρχείου εξόδου ο οποίος είναι διαφορετικός για κάθε είδος λογισμικού.

4.2.5 Bayesian Networks – BayesiaLab

Από το αρχείο που εξάγει το λογισμικό BayesiaLab έχουμε κρατήσει μόνο τον πίνακα modal value που, όπως έχουμε αναφέρει, περιέχει τους επιθυμητούς κανόνες. Ακολουθεί η έξοδος του BayesiaLab.

category = Xrisi_-_Fthora (57,70%)		
Node	**Modal Value**	
Ladia-Filtra	0	64,30%
Ilektrika_-_Ilektronika	0	100,00%
Idravlika	0	100,00%
Grasarisma	0	88,79%
Anartisi	0	100,00%
Katharisma	0	94,79%
Mpouma	0	95,14%
Plysimo	0	95,79%
Elegchos	0	96,03%
Allagi_-_Episkevi_Elastikon	0	98,19%
Sintirisi	0	96,97%
Eksatmisi	0	100,00%
Antlisi	0	98,24%
KTEO	0	98,39%
Sistima_Pedisis	0	97,56%
Sistima_kinisis	0	97,89%

Plysimo-Grasarisma	0	98,68%
Ergasia	0	98,90%
Nea_Elastika	0	99,29%
Kinitiras	0	96,96%
Karotsa	0	100,00%
Mixer	0	99,60%
Kafsima	0	99,72%
Sistima_Diefthinsis	0	99,62%
Sistima_Psiksis	0	98,85%
Klimatismos	0	100,00%
Amaksoma	0	86,43%
KInitiras_Anamiktira	0	99,94%
Kinitiras_-_Trikouverto	0	99,93%
category = Promitheutes (20,14%)		
Node	**Modal Value**	
Ladia-Filtra	0	100,00%
Idravlika	0	77,08%
Amaksoma	0	69,28%
Kinitiras	0	86,39%
Eksatmisi	0	94,51%
Grasarisma	0	100,00%
Allagi_-_Episkevi_Elastikon	0	100,00%
Katharisma	0	100,00%
Mpouma	0	100,00%
Plysimo	0	100,00%
Elegchos	0	100,00%
Sintirisi	0	100,00%
Nea_Elastika	0	100,00%
Ilektrika_-_Ilektronika	0	82,49%
Antlisi	0	100,00%
KTEO	0	100,00%
Plysimo-Grasarisma	0	100,00%
Ergasia	0	100,00%
Karotsa	0	99,24%
Sistima_kinisis	0	98,01%
Sistima_Pedisis	0	97,61%

Anartisi	0	96,38%
Sistima_Psiksis	0	99,64%
Mixer	0	100,00%
Kafsima	0	100,00%
KInitiras_Anamiktira	0	99,80%
Sistima_Diefthinsis	0	99,64%
Klimatismos	0	100,00%
Kinitiras_-_Trikouverto	0	99,92%
category = Kairikes_Sinthikes (14,58%)		
Node	**Modal Value**	
Ilektrika_-_Ilektronika	1	59,87%
Ladia-Filtra	0	100,00%
Amaksoma	0	100,00%
Allagi_-_Episkevi_Elastikon	0	83,01%
Grasarisma	0	100,00%
Idravlika	0	83,18%
Kinitiras	0	100,00%
Sistima_Pedisis	0	100,00%
Nea_Elastika	0	94,23%
Sistima_kinisis	0	100,00%
Katharisma	0	100,00%
Mpouma	0	100,00%
Plysimo	0	100,00%
Anartisi	0	100,00%
Elegchos	0	100,00%
Sintirisi	0	100,00%
Eksatmisi	0	100,00%
Antlisi	0	100,00%
Sistima_Psiksis	0	100,00%
KTEO	0	100,00%
Plysimo-Grasarisma	0	100,00%
Sistima_Diefthinsis	0	100,00%
Ergasia	0	100,00%
Klimatismos	0	99,84%
Mixer	0	100,00%
Kafsima	0	100,00%

KInitiras_Anamiktira	0	100,00%
Karotsa	0	99,62%
Kinitiras_-_Trikouverto	0	100,00%
category = Odostroma (7,57%)		
Node	**Modal Value**	
Sistima_Pedisis	0	70,58%
Sistima_kinisis	0	73,33%
Anartisi	0	78,20%
Ladia-Filtra	0	100,00%
Amaksoma	0	100,00%
Ilektrika_-_Ilektronika	0	100,00%
Idravlika	0	100,00%
Grasarisma	0	100,00%
Sistima_Diefthinsis	0	95,34%
Kinitiras	0	100,00%
Allagi_-_Episkevi_Elastikon	0	89,95%
Katharisma	0	100,00%
Mpouma	0	100,00%
Plysimo	0	100,00%
Elegchos	0	100,00%
Nea_Elastika	0	95,45%
Sintirisi	0	100,00%
Sistima_Psiksis	0	97,14%
Eksatmisi	0	100,00%
Antlisi	0	100,00%
KTEO	0	100,00%
Plysimo-Grasarisma	0	100,00%
Ergasia	0	100,00%
Mixer	0	100,00%
Karotsa	0	100,00%
Kafsima	0	100,00%
KInitiras_Anamiktira	0	100,00%
Kinitiras_-_Trikouverto	0	100,00%
Klimatismos	0	100,00%

Ενώ ο κόμβος που παράγεται είναι ο εξής:

Εικόνα 4.4: Ο παραγόμενος γράφος του BayesiaLab

Όπως στις περιπτώσεις του Rosetta και του RSES2, έτσι και εδώ παράγονται τα ιδία αποτελέσματα και ακολουθούνται οι ίδιες διαδικασίες. Η μόνη διαφορά που υπάρχει είναι στα ποσοτικά δεδομένα που επιστρέφει ο αλγόριθμος ο οποίος, εξαιτίας του ότι παρουσιάζει τους αριθμούς των υποδειγμάτων που συμφωνούν με τον κάθε κανόνα σε μορφή ποσοστού, εμφανίζονται μερικές διαφορές στα ποσοτικά δεδομένα κατά τη στρογγυλοποίηση σε σχέση με αυτά του RSES2, Rosetta και με αυτά που προκύπτουν κατευθείαν από τη βάση δεδομένων. Παρόλα αυτά, η διαφορά παρουσιάζεται σε λίγες περιπτώσεις (3 στις 49) και ισούται πάντα με την μονάδα. Να σημειωθεί ότι και στην περίπτωση του BayesiaLab (όπως και σε κάθε είδος αλγορίθμου), ενεργοποιείται ειδικός αλγόριθμος "σάρωσης" του αρχείου.

4.3 Relationship Diagram

Στην παρούσα παράγραφο παρουσιάζονται τα αποτελέσματα των 5 λογισμικών εξόρυξης γνώσης σε δεδομένα τα οποία προορίζονται για τη δημιουργία ενός Relationship Diagram, καθώς και το δυναμικά δημιουργημένο Relationship Diagram. Επίσης, παραθέτονται και τα αποτελέσματα της πρόβλεψης σε νέα δεδομένα.

4.3.1 Rough Set - Rosetta

Στην περίπτωση επιλογής του λογισμικού Rosetta παράγεται το εξής αρχείο εξόδου:

% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:51:20 by mike.
%
% 1R
% 8 rules.

cause(Προμηθευτές) => result(Έρευνα_Αγοράς_Προμηθευτών)
Supp. (LHS) = [2513 object(s)]
Supp. (RHS) = [2513 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0933541]
Cov. (RHS) = [0.315268]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

cause(Έρευνα_Αγοράς_Προμηθευτών) => result(Ανταλλακτικά) OR result(Νέες_Μηχανές)
Supp. (LHS) = [6471 object(s)]
Supp. (RHS) = [5541 object(s), 930 object(s)]
Acc. (RHS) = [0.856282, 0.143718]
Cov. (LHS) = [0.240388]
Cov. (RHS) = [1, 1]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

cause(Ανταλλακτικά) => result(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [5541 object(s)]
Supp. (RHS) = [5541 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.20584]
Cov. (RHS) = [0.504783]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

cause(Νέες_Μηχανές) => result(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [930 object(s)]
Supp. (RHS) = [930 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0345481]
Cov. (RHS) = [0.0847226]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

cause(Καιρικές_Συνθήκες) => result(Εκπαίδευση_Προσωπικού) OR result(Προσαρμοσμένα_Προϊόντα)
Supp. (LHS) = [1819 object(s)]
Supp. (RHS) = [1017 object(s), 802 object(s)]
Acc. (RHS) = [0.559098, 0.440902]
Cov. (LHS) = [0.0675731]
Cov. (RHS) = [0.0926483, 0.534667]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

cause(Προσαρμοσμένα_Προϊόντα) => result(Έρευνα_Αγοράς_Προμηθευτών)
Supp. (LHS) = [1500 object(s)]
Supp. (RHS) = [1500 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0557227]
Cov. (RHS) = [0.188182]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

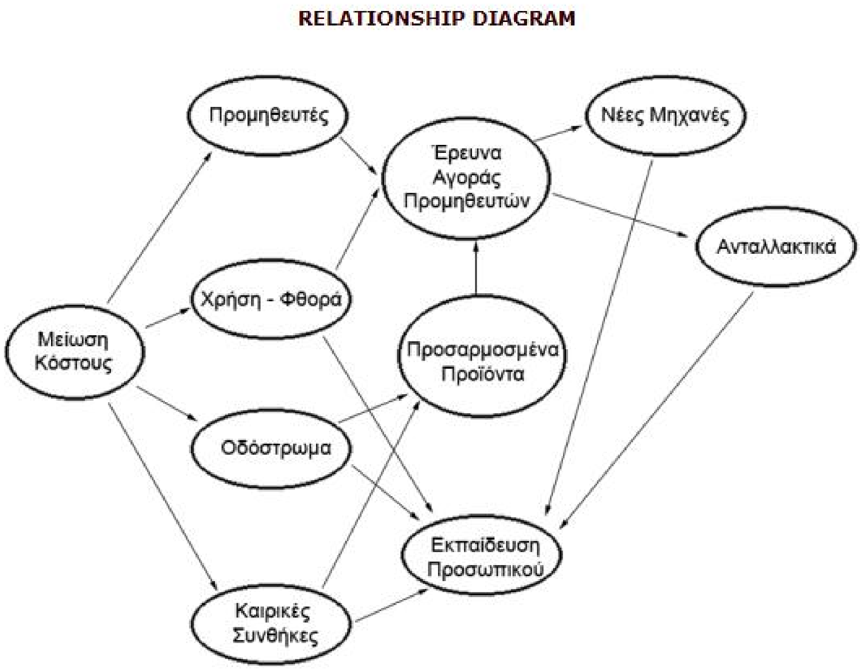
cause(Οδόστρωμα) => result(Προσαρμοσμένα_Προϊόντα) OR result(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [945 object(s)]
Supp. (RHS) = [698 object(s), 247 object(s)]
Acc. (RHS) = [0.738624, 0.261376]
Cov. (LHS) = [0.0351053]
Cov. (RHS) = [0.465333, 0.0225016]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

cause(Χρήση_-_Φθορά) => result(Έρευνα_Αγοράς_Προμηθευτών) OR result(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [7200 object(s)]
Supp. (RHS) = [3958 object(s), 3242 object(s)]
Acc. (RHS) = [0.549722, 0.450278]
Cov. (LHS) = [0.267469]
Cov. (RHS) = [0.49655, 0.295345]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

Οι κανόνες επιβεβαιώνονται από το αποτέλεσμα ενός ερωτήματος προς τη βάση δεδομένων.

Μετά την ενεργοποίηση του αλγορίθμου “σάρωσης”, συλλέγονται οι κανόνες και αποθηκεύονται στον κατάλληλο πίνακα (table) στη βάση δεδομένων. Στη συνέχεια ενεργοποιείται ο αλγόριθμος δυναμικής δημιουργίας του Relationship Diagram (παρ. 3.5.3).

Εικόνα 4.5: Δυναμικά παραγόμενο Relationship Diagram



Εικόνα 4.6: Ισοδύναμο "κλασσικό" Relationship Diagram

Στην εικόνα 4.6 παρουσιάζεται η εικόνα του ισοδύναμου Relationship Diagram αλλά στην κλασσική του μορφή. Όσον αφορά στην πρόβλεψη, θέτουμε το παρακάτω σύνολο δεδομένων προς πρόβλεψη:

Οδόστρωμα
Προμηθευτές
Χρήση - Φθορά
Καιρικές Συνθήκες
Προμηθευτές
Χρήση - Φθορά
Προμηθευτές
Οδόστρωμα
Χρήση - Φθορά
Οδόστρωμα

Από το παραπάνω σύνολο δεδομένων παράγεται η εξής πρόβλεψη:

Είσοδος (Κατηγορία Βλάβης)	**Πρόβλεψη (Γενική Λύση)**
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)
Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Καιρικές Συνθήκες	*Εκπαίδευση Προσωπικού* (55.91%) *Προσαρμοσμένα Προϊόντα* (44.09%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)
Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)
Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)
Γενική Λύση	**Ειδική Λύση**
Εκπαίδευση Προσωπικού	-
Προσαρμοσμένα Προϊόντα	*Έρευνα Αγοράς Προμηθευτών* (100%)
Έρευνα Αγοράς Προμηθευτών	*Νέες Μηχανές* (14.37%) *Ανταλλακτικά* (85.63%)
Ειδική Λύση	**Επιπλέον Λύσεις**

Νέες Μηχανές	*Εκπαίδευση Προσωπικού* (100%)
Ανταλλακτικά	*Εκπαίδευση Προσωπικού* (100%)

Όπως και στην περίπτωση του Affinity Diagram, παρέχεται η δυνατότητα της διαγραμματικής απεικόνισης των κανόνων.

4.3.2 Rough Set – RSES2

Με την επιλογή του αλγορίθμου Rough Set – RSES2 παράγονται οι ίδιοι κανόνες και αποτελέσματα και ακολουθούνται οι ίδιες διαδικασίες (με τους ανάλογους αλγορίθμους "σάρωσης"), όπως και στην περίπτωση του Rough Set – Rosetta. Ακολουθούν τα αποτελέσματα του αλγόριθμου.

```
RULE_SET rses.txt
ATTRIBUTES 2
cause symbolic
result symbolic
DECISION_VALUES 5
Έρευνα_Αγοράς_Προμηθευτών
Ανταλλακτικά
Εκπαίδευση_Προσωπικού
Νέες_Μηχανές
Προσαρμοσμένα_Προϊόντα
RULES 8
(cause=Χρήση_-_Φθορά)=>(result={Έρευνα_Αγοράς_Προμηθευτών[3958],Εκπαίδευση_Προσωπικού[3242]})
7200
(cause=Έρευνα_Αγοράς_Προμηθευτών)=>(result={Ανταλλακτικά[5541],Νέες_Μηχανές[930]}) 6471
(cause=Ανταλλακτικά)=>(result=Εκπαίδευση_Προσωπικού[5541]) 5541
(cause=Προμηθευτές)=>(result=Έρευνα_Αγοράς_Προμηθευτών[2513]) 2513
(cause=Καιρικές_Συνθήκες)=>(result={Εκπαίδευση_Προσωπικού[1017],Προσαρμοσμένα_Προϊόντα[802]}) 1819
(cause=Προσαρμοσμένα_Προϊόντα)=>(result=Έρευνα_Αγοράς_Προμηθευτών[1500]) 1500
(cause=Οδόστρωμα)=>(result={Προσαρμοσμένα_Προϊόντα[698],Εκπαίδευση_Προσωπικού[247]}) 945
(cause=Νέες_Μηχανές)=>(result=Εκπαίδευση_Προσωπικού[930]) 930
```

4.3.3 Association Rules – Weka Suite

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα στην περίπτωση του Weka Suite (Association Rules), είναι το εξής:

```
=== Run information ===

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: relationship
Instances: 26919
Attributes: 2
```

cause
result
=== Associator model (full training set) ===

Tertius
=======

1. /* 0,635546 0,034548 */ cause = Έρευνα_Αγοράς_Προμηθευτών ==> result = Ανταλλακτικά
2. /* 0,536442 0,000000 */ cause = Ανταλλακτικά ==> result = Εκπαίδευση_Προσωπικού
3. /* 0,344703 0,000000 */ cause = Προμηθευτές ==> result = Έρευνα_Αγοράς_Προμηθευτών
4. /* 0,276160 0,120435 */ cause = Χρήση_-_Φθορά ==> result = Έρευνα_Αγοράς_Προμηθευτών
5. /* 0,246956 0,000000 */ cause = Προσαρμοσμένα_Προϊόντα ==> result = Έρευνα_Αγοράς_Προμηθευτών
6. /* 0,166914 0,000000 */ cause = Νέες_Μηχανές ==> result = Εκπαίδευση_Προσωπικού
7. /* 0,160982 0,009176 */ cause = Οδόστρωμα ==> result = Προσαρμοσμένα_Προϊόντα
8. /* 0,137863 0,037780 */ cause = Καιρικές_Συνθήκες ==> result = Προσαρμοσμένα_Προϊόντα
9. /* 0,105113 0,205840 */ cause = Έρευνα_Αγοράς_Προμηθευτών ==> result = Νέες_Μηχανές
10. /* 0,063896 0,029793 */ cause = Καιρικές_Συνθήκες ==> result = Εκπαίδευση_Προσωπικού
11. /* 0,047443 0,147034 */ cause = Χρήση_-_Φθορά ==> result = Εκπαίδευση_Προσωπικού
12. /* 0,000000 0,034548 */ cause = Νέες_Μηχανές ==> FALSE
13. /* 0,000000 0,035105 */ cause = Οδόστρωμα ==> FALSE
14. /* 0,000000 0,055723 */ cause = Προσαρμοσμένα_Προϊόντα ==> FALSE
15. /* 0,000000 0,067573 */ cause = Καιρικές_Συνθήκες ==> FALSE
16. /* 0,000000 0,093354 */ cause = Προμηθευτές ==> FALSE
17. /* 0,000000 0,205840 */ cause = Ανταλλακτικά ==> FALSE
18. /* 0,000000 0,240388 */ cause = Έρευνα_Αγοράς_Προμηθευτών ==> FALSE
19. /* 0,000000 0,267469 */ cause = Χρήση_-_Φθορά ==> FALSE
20. /* 0,000000 0,592221 */ TRUE ==> result = Εκπαίδευση_Προσωπικού
21. /* 0,000000 0,703889 */ TRUE ==> result = Έρευνα_Αγοράς_Προμηθευτών
22. /* 0,000000 0,794160 */ TRUE ==> result = Ανταλλακτικά
23. /* 0,000000 0,944277 */ TRUE ==> result = Προσαρμοσμένα_Προϊόντα
24. /* 0,000000 0,965452 */ TRUE ==> result = Νέες_Μηχανές
25. /* -0,041650 0,025930 */ cause = Οδόστρωμα ==> result = Εκπαίδευση_Προσωπικού

Number of hypotheses considered: 53
Number of hypotheses explored: 53
Time: 00 min 00 s 748 ms

Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με τις παραπάνω περιπτώσεις (Rosetta, RSES2), πλην του ότι δεν υπάρχουν ποσοτικά δεδομένα.

4.3.4 Classification Rules – Weka Suite

Το αρχείο εξόδου στην περίπτωση του Weka Suite (Classification Rules) είναι το εξής:

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: relationship
Instances: 26919

Attributes: 2
cause
result
Test mode: 10-fold cross-validation

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

Prism rules

If cause = Προμηθευτές then Έρευνα_Αγοράς_Προμηθευτών
If cause = Προσαρμοσμένα_Προϊόντα then Έρευνα_Αγοράς_Προμηθευτών
If cause = Χρήση_-_Φθορά then Έρευνα_Αγοράς_Προμηθευτών
If cause = Οδόστρωμα then Προσαρμοσμένα_Προϊόντα
If cause = Καιρικές_Συνθήκες then Προσαρμοσμένα_Προϊόντα
If cause = Ανταλλακτικά then Εκπαίδευση_Προσωπικού
If cause = Νέες_Μηχανές then Εκπαίδευση_Προσωπικού
If cause = Καιρικές_Συνθήκες then Εκπαίδευση_Προσωπικού
If cause = Χρήση_-_Φθορά then Εκπαίδευση_Προσωπικού
If cause = Οδόστρωμα then Εκπαίδευση_Προσωπικού
If cause = Έρευνα_Αγοράς_Προμηθευτών then Νέες_Μηχανές
If cause = Έρευνα_Αγοράς_Προμηθευτών then Ανταλλακτικά

Time taken to build model: 0.09 seconds

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

Correctly Classified Instances 16872 62.6769 %
Incorrectly Classified Instances 10047 37.3231 %
Kappa statistic 0.5119
Mean absolute error 0.1493
Root mean squared error 0.3864
Relative absolute error 53.3654 %
Root relative squared error 103.3121 %
Total Number of Instances 26919

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
1 0.171 0.711 1 0.831 Έρευνα_Αγοράς_Προμηθευτών
1 0.05 0.543 1 0.704 Προσαρμοσμένα_Προϊόντα
0.59 0 1 0.59 0.742 Εκπαίδευση_Προσωπικού
1 0.213 0.144 1 0.251 Νέες_Μηχανές
0 0 0 0 0 Ανταλλακτικά

=== Confusion Matrix ===

a b c d e <-- classified as
7971 0 0 0 0 | a = Έρευνα_Αγοράς_Προμηθευτών
0 1500 0 0 0 | b = Προσαρμοσμένα_Προϊόντα
3242 1264 6471 0 0 | c = Εκπαίδευση_Προσωπικού
0 0 0 930 0 | d = Νέες_Μηχανές
0 0 0 5541 0 | e = Ανταλλακτικά

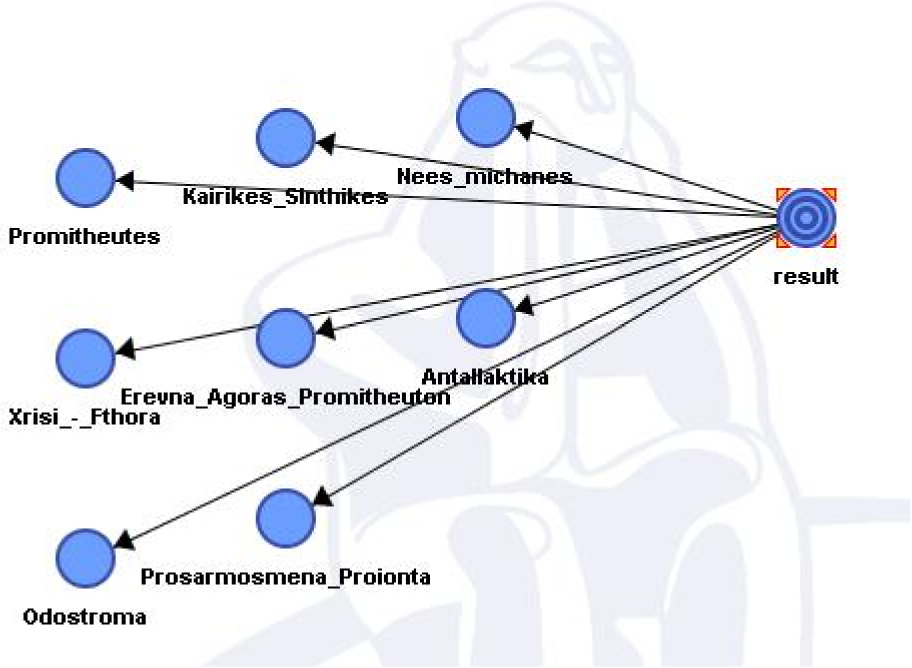
Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με την περίπτωση του Weka Suite (Association Rules).

4.3.5 Bayesian Networks – BayesiaLab

Τα αποτελέσματα του BayesiaLab είναι τα παρακάτω:

result = Ekpaideusi_Prosopikou (40,78%)		
Node	**Modal Value**	
Antallaktika	1	50,48%
Erevna_Agoras_Promitheuton	0	100,00%
Promitheutes	0	100,00%
Nees_michanes	0	91,53%
Prosarmosmena_Proionta	0	100,00%
Kairikes_Sinthikes	0	90,73%
Odostroma	0	97,75%
Xrisi_-_Fthora	0	70,47%
result = Erevna_Agoras_Promitheuton (29,61%)		
Node	**Modal Value**	
Promitheutes	0	68,47%
Erevna_Agoras_Promitheuton	0	100,00%
Antallaktika	0	100,00%
Prosarmosmena_Proionta	0	81,18%
Xrisi_-_Fthora	0	50,35%
Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%
Nees_michanes	0	100,00%
result = Antallaktika (20,58%)		
Node	**Modal Value**	
Erevna_Agoras_Promitheuton	1	100,00%
Xrisi_-_Fthora	0	100,00%
Antallaktika	0	100,00%
Promitheutes	0	100,00%
Kairikes_Sinthikes	0	100,00%
Prosarmosmena_Proionta	0	100,00%
Odostroma	0	100,00%

Nees_michanes	0	100,00%
result = Prosarmosmena_Proionta (5,57%)		
Node	**Modal Value**	
Odostroma	0	53,47%
Kairikes_Sinthikes	1	53,47%
Xrisi_-_Fthora	0	100,00%
Erevna_Agoras_Promitheuton	0	100,00%
Antallaktika	0	100,00%
Promitheutes	0	100,00%
Prosarmosmena_Proionta	0	100,00%
Nees_michanes	0	100,00%
result = Nees_michanes (3,45%)		
Node	**Modal Value**	
Erevna_Agoras_Promitheuton	1	100,00%
Xrisi_-_Fthora	0	100,00%
Antallaktika	0	100,00%
Promitheutes	0	100,00%
Kairikes_Sinthikes	0	100,00%
Prosarmosmena_Proionta	0	100,00%
Odostroma	0	100,00%
Nees_michanes	0	100,00%



Εικόνα 4.7: Ο παραγόμενος γράφος του BayesiaLab

Και σε αυτήν την περίπτωση, τα αποτελέσματα και οι διαδικασίες είναι ίδιες με τις παραπάνω περιπτώσεις στην παραγωγή του Relationship Diagram, με εξαίρεση την πιθανή απόκλιση των αριθμητικών τιμών που αναφέρεται αναλυτικά στην παράγραφο 4.2.5.

4.4 Systematic Diagram

Στην παρούσα παράγραφο παρουσιάζονται τα αποτελέσματα των 5 λογισμικών εξόρυξης γνώσης σε δεδομένα τα οποία προορίζονται για τη δημιουργία ενός Systematic Diagram, καθώς και το δυναμικά δημιουργημένο Systematic Diagram. Επίσης, παραθέτονται και τα αποτελέσματα της πρόβλεψης σε νέα δεδομένα.

4.4.1 Rough Set - Rosetta

Στην περίπτωση επιλογής του λογισμικού Rosetta, παράγεται το εξής αρχείο εξόδου:

```
% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:57:54 by mike.
%
% 1R
% 4 rules.
```

category(Προμηθευτές) => origin(Εσωτερικής_Προέλευσης)
Supp. (LHS) = [2513 object(s)]
Supp. (RHS) = [2513 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.201411]
Cov. (RHS) = [0.258725]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

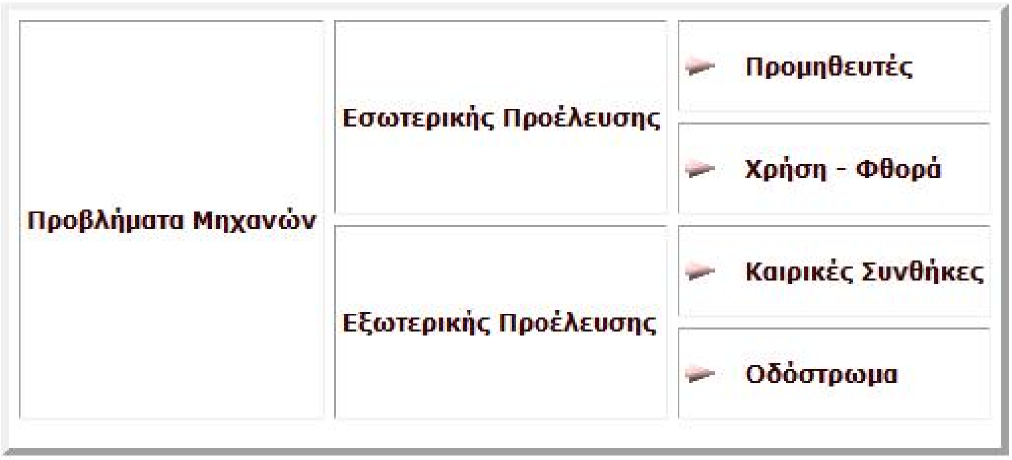
category(Καιρικές_Συνθήκες) => origin(Εξωτερικής_Προέλευσης)
Supp. (LHS) = [1819 object(s)]
Supp. (RHS) = [1819 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.145788]
Cov. (RHS) = [0.658104]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

category(Οδόστρωμα) => origin(Εξωτερικής_Προέλευσης)
Supp. (LHS) = [945 object(s)]
Supp. (RHS) = [945 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.0757394]
Cov. (RHS) = [0.341896]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

category(Χρήση_-_Φθορά) => origin(Εσωτερικής_Προέλευσης)
Supp. (LHS) = [7200 object(s)]
Supp. (RHS) = [7200 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.577062]
Cov. (RHS) = [0.741275]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

Οι κανόνες επιβεβαιώνονται από το αποτέλεσμα ενός ερωτήματος προς τη βάση δεδομένων.

Μετά την ενεργοποίηση του αλγορίθμου “σάρωσης”, συλλέγονται οι κανόνες και αποθηκεύονται στον κατάλληλο πίνακα (table) στη βάση δεδομένων. Στη συνέχεια ενεργοποιείται ο αλγόριθμος δυναμικής δημιουργίας του Systematic Diagram (παρ. 3.5.4).



Εικόνα 4.8: Δυναμικά παραγόμενο Relationship Diagram

Όσον αφορά στην πρόβλεψη, θέτουμε το παρακάτω σύνολο δεδομένων προς πρόβλεψη:

Εσωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εξωτερικής Προέλευσης

Από το παραπάνω σύνολο δεδομένων παράγεται η εξής πρόβλεψη:

Είσοδος (Προέλευση Βλάβης)	**Πρόβλεψη (Κατηγορία Βλάβης)**
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)

Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)

Όπως και στις περιπτώσεις του Affinity Diagram και του Relationship Diagram, παρέχεται η δυνατότητα της διαγραμματικής απεικόνισης των κανόνων.

4.4.2 Rough Set – RSES2

Με την επιλογή του αλγορίθμου Rough Set – RSES2 παράγονται οι ίδιοι κανόνες και αποτελέσματα και με τις ίδιες διαδικασίες (με τους ανάλογους αλγόριθμους “σάρωσης”) όπως στην περίπτωση του Rough Set – Rosetta. Ακολουθούν τα αποτελέσματα του αλγόριθμου.

```
RULE_SET rses.txt
ATTRIBUTES 2
category symbolic
origin symbolic
DECISION_VALUES 2
Εσωτερικής_Προέλευσης
Εξωτερικής_Προέλευσης
RULES 4
(category=Χρήση_-_Φθορά)=>(origin=Εσωτερικής_Προέλευσης[7200]) 7200
(category=Προμηθευτές)=>(origin=Εσωτερικής_Προέλευσης[2513]) 2513
(category=Καιρικές_Συνθήκες)=>(origin=Εξωτερικής_Προέλευσης[1819]) 1819
(category=Οδόστρωμα)=>(origin=Εξωτερικής_Προέλευσης[945]) 945
```

4.4.3 Association Rules – Weka Suite

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα στην περίπτωση του Weka Suite (Association Rules) είναι το εξής:

```
=== Run information ===
```

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: systematic
Instances: 12477
Attributes: 2
category
origin
=== Associator model (full training set) ===

Tertius
=======

1. /* 0,556518 0,000000 */ category = Χρήση_-_Φθορά ==> origin = Εσωτερικής_Προέλευσης
2. /* 0,508036 0,000000 */ category = Καιρικές_Συνθήκες ==> origin = Εξωτερικής_Προέλευσης
3. /* 0,320688 0,000000 */ category = Οδόστρωμα ==> origin = Εξωτερικής_Προέλευσης
4. /* 0,267796 0,000000 */ category = Προμηθευτές ==> origin = Εσωτερικής_Προέλευσης
5. /* 0,000000 0,075739 */ category = Οδόστρωμα ==> FALSE
6. /* 0,000000 0,145788 */ category = Καιρικές_Συνθήκες ==> FALSE
7. /* 0,000000 0,201411 */ category = Προμηθευτές ==> FALSE
8. /* 0,000000 0,221528 */ TRUE ==> origin = Εσωτερικής_Προέλευσης
9. /* 0,000000 0,577062 */ category = Χρήση_-_Φθορά ==> FALSE
10. /* 0,000000 0,778472 */ TRUE ==> origin = Εξωτερικής_Προέλευσης

Number of hypotheses considered: 14
Number of hypotheses explored: 14
Time: 00 min 00 s 156 ms

Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με τις παραπάνω περιπτώσεις (Rosetta, RSES2), πλην του ότι δεν υπάρχουν ποσοτικά δεδομένα.

4.4.4 Classification Rules – Weka Suite

Το αρχείο εξόδου στην περίπτωση του Weka Suite (Classification Rules) είναι το εξής:

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: systematic
Instances: 12477
Attributes: 2
category
origin
Test mode: 10-fold cross-validation

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

Prism rules

If category = Χρήση_-_Φθορά then Εσωτερικής_Προέλευσης
If category = Προμηθευτές then Εσωτερικής_Προέλευσης

If category = Καιρικές_Συνθήκες then Εξωτερικής_Προέλευσης
If category = Οδόστρωμα then Εξωτερικής_Προέλευσης

Time taken to build model: 0.14 seconds

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

Correctly Classified Instances 12477 100 %
Incorrectly Classified Instances 0 0 %
Kappa statistic 1
Mean absolute error 0
Root mean squared error 0
Relative absolute error 0 %
Root relative squared error 0 %
Total Number of Instances 12477

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
1 0 1 1 1 Εσωτερικής_Προέλευσης
1 0 1 1 1 Εξωτερικής_Προέλευσης

=== Confusion Matrix ===

a b <-- classified as
9713 0 | a = Εσωτερικής_Προέλευσης
0 2764 | b = Εξωτερικής_Προέλευσης

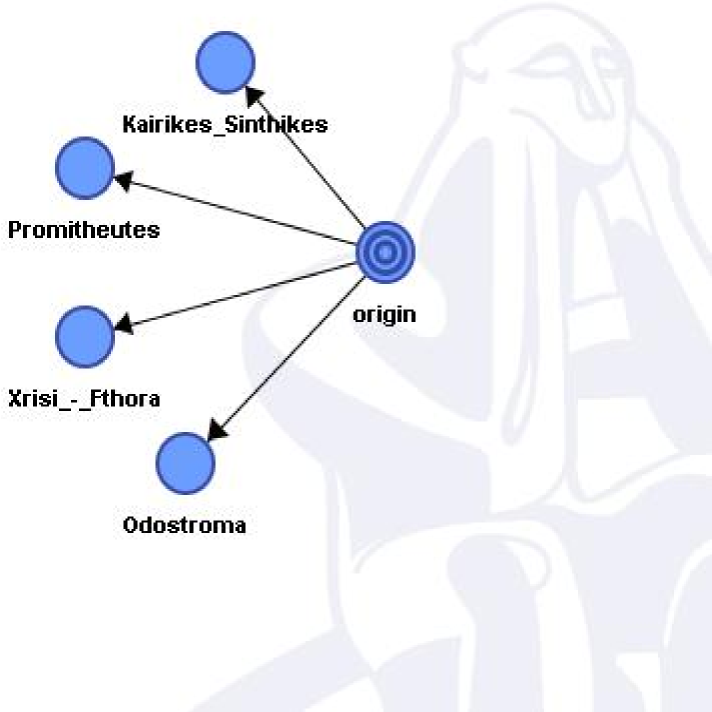
Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με την περίπτωση του Weka Suite (Association Rules).

4.4.5 Bayesian Networks – BayesiaLab

Τα αποτελέσματα του BayesiaLab είναι τα εξής:

origin = Esoterikis_Proelefsis (77,85%)		
Node	**Modal Value**	
Kairikes_Sinthikes	0	100,00%
Xrisi_-_Fthora	1	74,13%
Odostroma	0	100,00%
Promitheutes	0	74,13%
origin = Eksoterikis_Proelefsis (22,15%)		
Node	**Modal Value**	

Kairikes_Sinthikes	1	65,81%
Xrisi_-_Fthora	0	100,00%
Odostroma	0	65,81%
Promitheutes	0	100,00%



Εικόνα 4.9: Ο παραγόμενος γράφος του BayesiaLab

Και σε αυτήν την περίπτωση, τα αποτελέσματα και οι διαδικασίες είναι ίδιες με τις παραπάνω περιπτώσεις στην παραγωγή του Systematic Diagram.

4.5 Matrix Diagram

Στην παρούσα παράγραφο παρουσιάζονται τα αποτελέσματα των 5 λογισμικών εξόρυξης γνώσης σε δεδομένα τα οποία προορίζονται για τη δημιουργία ενός Matrix Diagram (L-Diagram) καθώς και το δυναμικά δημιουργημένο Matrix Diagram (L-Diagram). Επίσης, παραθέτονται και τα αποτελέσματα της πρόβλεψης σε νέα δεδομένα.

4.5.1 Rough Set - Rosetta

Στην περίπτωση επιλογής του λογισμικού Rosetta παράγεται το εξής αρχείο εξόδου:

```
% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:38:18 by mike.
%
% 1R
% 4 rules.

category(Προμηθευτές) => solution(Έρευνα_Αγοράς_Προμηθευτών)
Supp. (LHS) = [2513 object(s)]
Supp. (RHS) = [2513 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.201411]
Cov. (RHS) = [0.388348]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

category(Καιρικές_Συνθήκες) => solution(Προσαρμοσμένα_Προϊόντα) OR solution(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [1819 object(s)]
Supp. (RHS) = [802 object(s), 1017 object(s)]
Acc. (RHS) = [0.440902, 0.559098]
Cov. (LHS) = [0.145788]
Cov. (RHS) = [0.534667, 0.225699]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

category(Οδόστρωμα) => solution(Εκπαίδευση_Προσωπικού) OR solution(Προσαρμοσμένα_Προϊόντα)
Supp. (LHS) = [945 object(s)]
Supp. (RHS) = [247 object(s), 698 object(s)]
Acc. (RHS) = [0.261376, 0.738624]
Cov. (LHS) = [0.0757394]
Cov. (RHS) = [0.0548158, 0.465333]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

category(Χρήση_-_Φθορά) => solution(Εκπαίδευση_Προσωπικού) OR solution(Έρευνα_Αγοράς_Προμηθευτών)
Supp. (LHS) = [7200 object(s)]
Supp. (RHS) = [3242 object(s), 3958 object(s)]
Acc. (RHS) = [0.450278, 0.549722]
Cov. (LHS) = [0.577062]
Cov. (RHS) = [0.719485, 0.611652]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]
```

Οι κανόνες επιβεβαιώνονται από το αποτέλεσμα ενός ερωτήματος προς τη βάση δεδομένων.

Μετά την ενεργοποίηση του αλγορίθμου "σάρωσης", συλλέγονται οι κανόνες και αποθηκεύονται στον κατάλληλο πίνακα (table) στη βάση δεδομένων. Στη συνέχεια ενεργοποιείται ο αλγόριθμος δυναμικής δημιουργίας του Matrix Diagram (παρ. 3.5.5).

MATRIX DIAGRAM (L - Diagram) *ΚΑΤΗΓΟΡΙΑ ΒΛΑΒΗΣ/ΓΕΝΙΚΗ ΛΥΣΗ*	Προσαρμοσμένα Προϊόντα	Εκπαίδευση Προσωπικού	Έρευνα Αγοράς Προμηθευτών
Χρήση - Φθορά	-	✓ (3242)	✓ (3958)
Οδόστρωμα	✓ (698)	✓ (247)	-
Καιρικές Συνθήκες	✓ (802)	✓ (1017)	-
Προμηθευτές	-	-	✓ (2513)

Εικόνα 4.10: Δυναμικά παραγόμενο Matrix Diagram

Το παραπάνω Matrix Diagram μπορεί να θεωρηθεί και ως Matrix Data Analysis (Ανάλυση Δεδομένων Διαγράμματος Μήτρας), καθώς παρέχει και ποσοτικοποιημένες πληροφορίες για το διάγραμμα μήτρας.

Όσον αφορά στην πρόβλεψη, θέτουμε το παρακάτω σύνολο δεδομένων προς πρόβλεψη:

Οδόστρωμα
Προμηθευτές
Χρήση - Φθορά
Καιρικές Συνθήκες
Προμηθευτές
Χρήση - Φθορά
Προμηθευτές
Οδόστρωμα
Χρήση - Φθορά
Οδόστρωμα

Από το παραπάνω σύνολο δεδομένων παράγεται η εξής πρόβλεψη:

Είσοδος (Κατηγορία Βλάβης)	**Πρόβλεψη (Γενική Λύση)**
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)
Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Καιρικές Συνθήκες	*Εκπαίδευση Προσωπικού* (55.91%) *Προσαρμοσμένα Προϊόντα* (44.09%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)

Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)
Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)

Όπως και στις περιπτώσεις των υπολοίπων εργαλείων ποιότητας, παρέχεται η δυνατότητα της διαγραμματικής απεικόνισης των κανόνων.

4.5.2 Rough Set – RSES2

Με την επιλογή του αλγορίθμου Rough Set – RSES2 παράγονται οι ίδιοι κανόνες και αποτελέσματα και με τις ίδιες διαδικασίες (με τους ανάλογους αλγόριθμους "σάρωσης"), όπως στην περίπτωση του Rough Set – Rosetta. Ακολουθούν τα αποτελέσματα του αλγόριθμου.

```
RULE_SET rses.txt
ATTRIBUTES 2
category symbolic
solution symbolic
DECISION_VALUES 3
Έρευνα_Αγοράς_Προμηθευτών
Προσαρμοσμένα_Προϊόντα
Εκπαίδευση_Προσωπικού
RULES 4
(category=Χρήση_-
_Φθορά)=>(solution={Εκπαίδευση_Προσωπικού[3242],Έρευνα_Αγοράς_Προμηθευτών[3958]}) 7200
(category=Προμηθευτές)=>(solution=Έρευνα_Αγοράς_Προμηθευτών[2513]) 2513
(category=Καιρικές_Συνθήκες)=>(solution={Προσαρμοσμένα_Προϊόντα[802],Εκπαίδευση_Προσωπικού[1017]})
1819
(category=Οδόστρωμα)=>(solution={Εκπαίδευση_Προσωπικού[247],Προσαρμοσμένα_Προϊόντα[698]}) 945
```

4.5.3 Association Rules – Weka Suite

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα στην περίπτωση του Weka Suite (Association Rules) είναι το εξής:

```
=== Run information ===
```

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: matrix
Instances: 12477
Attributes: 2
category
solution
=== Associator model (full training set) ===

Tertius
=======

1. /* 0,452162 0,000000 */ category = Προμηθευτές ==> solution = Έρευνα_Αγοράς_Προμηθευτών
2. /* 0,244580 0,019796 */ category = Οδόστρωμα ==> solution = Προσαρμοσμένα_Προϊόντα
3. /* 0,215649 0,317224 */ category = Χρήση_-_Φθορά ==> solution = Εκπαίδευση_Προσωπικού
4. /* 0,203378 0,081510 */ category = Καιρικές_Συνθήκες ==> solution = Προσαρμοσμένα_Προϊόντα
5. /* 0,136099 0,064278 */ category = Καιρικές_Συνθήκες ==> solution = Εκπαίδευση_Προσωπικού
6. /* 0,071969 0,259838 */ category = Χρήση_-_Φθορά ==> solution = Έρευνα_Αγοράς_Προμηθευτών
7. /* 0,000000 0,075739 */ category = Οδόστρωμα ==> FALSE
8. /* 0,000000 0,145788 */ category = Καιρικές_Συνθήκες ==> FALSE
9. /* 0,000000 0,201411 */ category = Προμηθευτές ==> FALSE
10. /* 0,000000 0,481366 */ TRUE ==> solution = Έρευνα_Αγοράς_Προμηθευτών
11. /* 0,000000 0,577062 */ category = Χρήση_-_Φθορά ==> FALSE
12. /* 0,000000 0,638855 */ TRUE ==> solution = Εκπαίδευση_Προσωπικού
13. /* 0,000000 0,879779 */ TRUE ==> solution = Προσαρμοσμένα_Προϊόντα
14. /* -0,044040 0,055943 */ category = Οδόστρωμα ==> solution = Εκπαίδευση_Προσωπικού

Number of hypotheses considered: 19
Number of hypotheses explored: 19
Time: 00 min 00 s 171 ms

Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με τις παραπάνω περιπτώσεις (Rosetta, RSES2), πλην του ότι δεν υπάρχουν ποσοτικά δεδομένα.

4.5.4 Classification Rules – Weka Suite

Το αρχείο εξόδου στην περίπτωση του Weka Suite (Classification Rules) είναι το εξής:

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: matrix
Instances: 12477
Attributes: 2
category
solution
Test mode: 10-fold cross-validation

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

Prism rules

If category = Προμηθευτές then Έρευνα_Αγοράς_Προμηθευτών
If category = Χρήση_-_Φθορά then Έρευνα_Αγοράς_Προμηθευτών
If category = Οδόστρωμα then Προσαρμοσμένα_Προϊόντα
If category = Καιρικές_Συνθήκες then Προσαρμοσμένα_Προϊόντα
If category = Καιρικές_Συνθήκες then Εκπαίδευση_Προσωπικού
If category = Χρήση_-_Φθορά then Εκπαίδευση_Προσωπικού
If category = Οδόστρωμα then Εκπαίδευση_Προσωπικού

Time taken to build model: 0.03 seconds

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

Correctly Classified Instances 7971 63.8855 %
Incorrectly Classified Instances 4506 36.1145 %
Kappa statistic 0.366
Mean absolute error 0.2408
Root mean squared error 0.4907
Relative absolute error 61.6118 %
Root relative squared error 111.0081 %
Total Number of Instances 12477

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
1 0.54 0.666 1 0.8 Έρευνα_Αγοράς_Προμηθευτών
1 0.115 0.543 1 0.704 Προσαρμοσμένα_Προϊόντα
0 0 0 0 0 Εκπαίδευση_Προσωπικού

=== Confusion Matrix ===

a b c <-- classified as
6471 0 0 | a = Έρευνα_Αγοράς_Προμηθευτών
0 1500 0 | b = Προσαρμοσμένα_Προϊόντα
3242 1264 0 | c = Εκπαίδευση_Προσωπικού

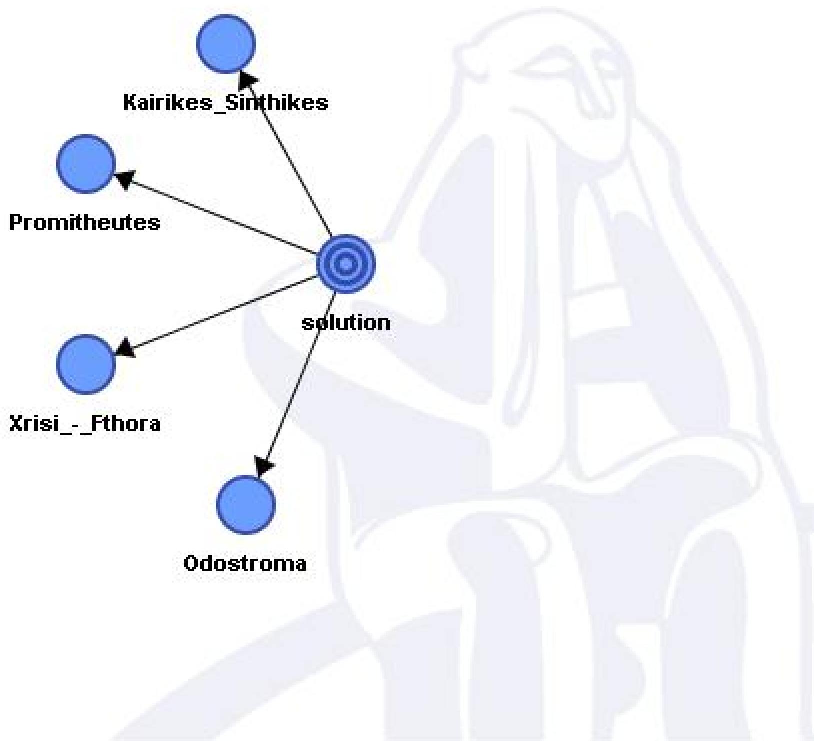
Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με την περίπτωση του Weka Suite (Association Rules).

4.5.5 Bayesian Networks – BayesiaLab

Τα αποτελέσματα του BayesiaLab είναι τα εξής:

solution = Erevna_Agoras_Promitheuton (51,86%)		
Node	**Modal Value**	
Promitheutes	0	61,17%

Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%
Xrisi_-_Fthora	1	61,17%
solution = Ekpaideusi_Prosopikou (36,11%)		
Node	**Modal Value**	
Promitheutes	0	100,00%
Xrisi_-_Fthora	1	71,95%
Kairikes_Sinthikes	0	77,43%
Odostroma	0	94,52%
solution = Prosarmosmena_Proionta (12,02%)		
Node	**Modal Value**	
Xrisi_-_Fthora	0	100,00%
Odostroma	0	53,47%
Kairikes_Sinthikes	1	53,47%
Promitheutes	0	100,00%



Εικόνα 4.11: Ο παραγόμενος γράφος του BayesiaLab

Και σε αυτήν την περίπτωση, τα αποτελέσματα και οι διαδικασίες είναι ίδιες με τις παραπάνω περιπτώσεις στην παραγωγή του Matrix Diagram.

4.6 Arrow Diagram

Στην παρούσα παράγραφο παρουσιάζονται τα αποτελέσματα των 5 λογισμικών εξόρυξης γνώσης σε δεδομένα τα οποία προορίζονται για τη δημιουργία ενός Arrow Diagram, καθώς και το δυναμικά δημιουργημένο Arrow Diagram. Επίσης, παραθέτονται και τα αποτελέσματα της πρόβλεψης σε νέα δεδομένα.

4.6.1 Rough Set - Rosetta

Στην περίπτωση επιλογής του λογισμικού Rosetta, παράγεται το εξής αρχείο εξόδου:

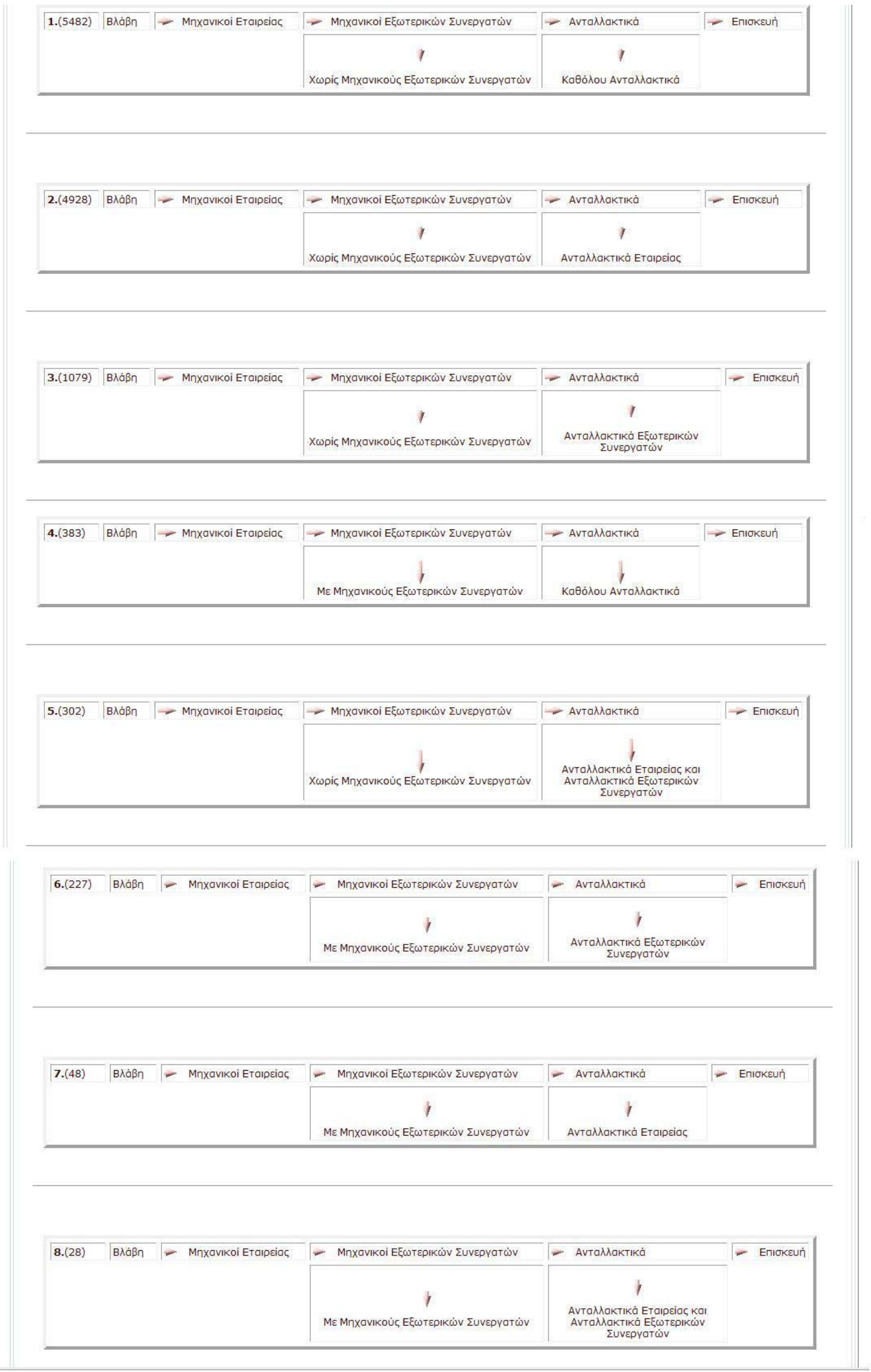
% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:34:02 by mike.
%
% 1R
% 2 rules.

Μηχανικοί(Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών) => Ανταλλακτικά(Ανταλλακτικά_Εταιρείας) OR Ανταλλακτικά(Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών) OR Ανταλλακτικά(Καθόλου_Ανταλλακτικά) OR Ανταλλακτικά(Ανταλλακτικά_Εξωτερικών_Συνεργατών)
Supp. (LHS) = [11791 object(s)]
Supp. (RHS) = [4928 object(s), 302 object(s), 5482 object(s), 1079 object(s)]
Acc. (RHS) = [0.417946, 0.0256128, 0.464931, 0.0915105]
Cov. (LHS) = [0.945019]
Cov. (RHS) = [0.990354, 0.915152, 0.934697, 0.826187]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1, 1]

Μηχανικοί(Με_Μηχανικούς_Εξωτερικών_Συνεργατών) =>
Ανταλλακτικά(Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών) OR Ανταλλακτικά(Ανταλλακτικά_Εξωτερικών_Συνεργατών) OR Ανταλλακτικά(Ανταλλακτικά_Εταιρείας) OR Ανταλλακτικά(Καθόλου_Ανταλλακτικά)
Supp. (LHS) = [686 object(s)]
Supp. (RHS) = [28 object(s), 227 object(s), 48 object(s), 383 object(s)]
Acc. (RHS) = [0.0408163, 0.330904, 0.0699708, 0.558309]
Cov. (LHS) = [0.0549812]
Cov. (RHS) = [0.0848485, 0.173813, 0.0096463, 0.0653026]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1, 1, 1]

Οι κανόνες επιβεβαιώνονται από το αποτέλεσμα ενός ερωτήματος προς τη βάση δεδομένων.

Μετά την ενεργοποίηση του αλγόριθμου “σάρωσης”, συλλέγονται οι κανόνες και αποθηκεύονται στον κατάλληλο πίνακα (table) στη βάση δεδομένων. Στη συνέχεια ενεργοποιείται ο αλγόριθμος δυναμικής δημιουργίας του Arrow Diagram (παρ. 3.5.6).



Εικόνα 4.12: Τα 8 διακριτά δυναμικά παραγόμενα Arrow Diagrams

Εικόνα 4.13: Το συνολικό δυναμικά παραγόμενο Arrow Diagram

Όσον αφορά στην πρόβλεψη, θέτουμε το παρακάτω σύνολο δεδομένων προς πρόβλεψη:

Με Μηχανικούς Εξωτερικών Συνεργατών
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών
Με Μηχανικούς Εξωτερικών Συνεργατών
Με Μηχανικούς Εξωτερικών Συνεργατών
Με Μηχανικούς Εξωτερικών Συνεργατών
Με Μηχανικούς Εξωτερικών Συνεργατών
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών

Από το παραπάνω σύνολο δεδομένων παράγεται η εξής πρόβλεψη:

Είσοδος (Μηχανικοί Εξωτερικών Συνεργατών)	**Πρόβλεψη (Ανταλλακτικά)**
Με Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (55.83%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (4.08%) *Ανταλλακτικά Εταιρείας* (7%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (33.09%)
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (46.49%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (2.56%) *Ανταλλακτικά Εταιρείας* (41.79%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (9.15%)
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (46.49%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (2.56%)

	Ανταλλακτικά Εταιρείας (41.79%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (9.15%)
Με Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (55.83%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (4.08%) *Ανταλλακτικά Εταιρείας* (7%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (33.09%)
Με Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (55.83%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (4.08%) *Ανταλλακτικά Εταιρείας* (7%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (33.09%)
Με Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (55.83%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (4.08%) *Ανταλλακτικά Εταιρείας* (7%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (33.09%)
Με Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (55.83%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (4.08%) *Ανταλλακτικά Εταιρείας* (7%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (33.09%)
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (46.49%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (2.56%) *Ανταλλακτικά Εταιρείας* (41.79%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (9.15%)
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (46.49%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (2.56%) *Ανταλλακτικά Εταιρείας* (41.79%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (9.15%)
Χωρίς Μηχανικούς Εξωτερικών Συνεργατών	*Καθόλου Ανταλλακτικά* (46.49%) *Ανταλλακτικά Εταιρείας και Ανταλλακτικά Εξωτερικών Συνεργατών* (2.56%) *Ανταλλακτικά Εταιρείας* (41.79%) *Ανταλλακτικά Εξωτερικών Συνεργατών* (9.15%)

Όπως και στις περιπτώσεις των υπολοίπων εργαλείων ποιότητας, παρέχεται η δυνατότητα της διαγραμματικής απεικόνισης των κανόνων.

4.6.2 Rough Set – RSES2

Με την επιλογή του αλγορίθμου Rough Set – RSES2 παράγονται οι ίδιοι κανόνες και αποτελέσματα και με τις ίδιες διαδικασίες (με τους ανάλογους αλγόριθμους "σάρωσης"), όπως στην περίπτωση του Rough Set – Rosetta. Ακολουθούν τα αποτελέσματα του αλγόριθμου.

RULE_SET rses.txt
ATTRIBUTES 2
Μηχανικοί symbolic
Ανταλλακτικά symbolic
DECISION_VALUES 4
Ανταλλακτικά_Εταιρείας
Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
Ανταλλακτικά_Εξωτερικών_Συνεργατών
Καθόλου_Ανταλλακτικά
RULES 2
(Μηχανικοί=Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών)=>(Ανταλλακτικά={Ανταλλακτικά_Εταιρείας[4928], Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών[302],Καθόλου_Ανταλλακτικά[5482],Ανταλλακτικά_Εξωτερικών_Συνεργατών[1079]}) 11791
(Μηχανικοί=Με_Μηχανικούς_Εξωτερικών_Συνεργατών)=>(Ανταλλακτικά={Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών[28],Ανταλλακτικά_Εξωτερικών_Συνεργατών[227],Ανταλλακτικά_Εταιρείας[48],Καθόλου_Ανταλλακτικά[383]}) 686

4.6.3 Association Rules – Weka Suite

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα στην περίπτωση του Weka Suite (Association Rules) είναι το εξής:

=== Run information ===

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: systematic
Instances: 12477
Attributes: 2
mechanics
spares
=== Associator model (full training set) ===

Tertius
=======

1. /* 0,097408 0,550052 */ mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Ανταλλακτικά_Εταιρείας
2. /* 0,072047 0,036788 */ mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Ανταλλακτικά_Εξωτερικών_Συνεργατών
3. /* 0,034274 0,024285 */ mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Καθόλου_Ανταλλακτικά
4. /* 0,004442 0,052737 */ mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
5. /* 0,000000 0,054981 */ mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών ==> FALSE
6. /* 0,000000 0,529935 */ TRUE ==> spares = Καθόλου_Ανταλλακτικά
7. /* 0,000000 0,601186 */ TRUE ==> spares = Ανταλλακτικά_Εταιρείας
8. /* 0,000000 0,895327 */ TRUE ==> spares = Ανταλλακτικά_Εξωτερικών_Συνεργατών
9. /* 0,000000 0,945019 */ mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών ==> FALSE
10. /* 0,000000 0,973551 */ TRUE ==> spares = Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
11. /* -0,020175 0,920814 */ mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών

12. /* -0,023453 0,505650 */ mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Καθόλου_Ανταλλακτικά
13. /* -0,121545 0,051134 */ mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Ανταλλακτικά_Εταιρείας
14. /* -0,168688 0,858540 */ mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών ==> spares = Ανταλλακτικά_Εξωτερικών_Συνεργατών

Number of hypotheses considered: 14
Number of hypotheses explored: 14
Time: 00 min 00 s 156 ms

Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με τις παραπάνω περιπτώσεις (Rosetta, RSES2), πλην του ότι δεν υπάρχουν ποσοτικά δεδομένα.

4.6.4 Classification Rules – Weka Suite

Το αρχείο εξόδου στην περίπτωση του Weka Suite (Classification Rules) είναι το εξής:

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: systematic
Instances: 12477
Attributes: 2
mechanics
spares
Test mode: 10-fold cross-validation

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

Prism rules

If mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών then Ανταλλακτικά_Εταιρείας
If mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών then Ανταλλακτικά_Εταιρείας
If mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών then Ανταλλακτικά_Εξωτερικών_Συνεργατών
If mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών then Ανταλλακτικά_Εξωτερικών_Συνεργατών
If mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών then Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
If mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών then Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
If mechanics = Με_Μηχανικούς_Εξωτερικών_Συνεργατών then Καθόλου_Ανταλλακτικά
If mechanics = Χωρίς_Μηχανικούς_Εξωτερικών_Συνεργατών then Καθόλου_Ανταλλακτικά

Time taken to build model: 0.03 seconds

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

Correctly Classified Instances 4976 39.8814 %

Incorrectly Classified Instances 7501 60.1186 %
Kappa statistic 0
Mean absolute error 0.3006
Root mean squared error 0.5483
Relative absolute error 98.8174 %
Root relative squared error 140.5884 %
Total Number of Instances 12477

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
1 1 0.399 1 0.57 Ανταλλακτικά_Εταιρείας
0 0 0 0 0 Ανταλλακτικά_Εξωτερικών_Συνεργατών
0 0 0 0 0 Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
0 0 0 0 0 Καθόλου_Ανταλλακτικά

=== Confusion Matrix ===

a b c d <-- classified as
4976 0 0 0 | a = Ανταλλακτικά_Εταιρείας
1306 0 0 0 | b = Ανταλλακτικά_Εξωτερικών_Συνεργατών
330 0 0 0 | c = Ανταλλακτικά_Εταιρείας_και_Ανταλλακτικά_Εξωτερικών_Συνεργατών
5865 0 0 0 | d = Καθόλου_Ανταλλακτικά

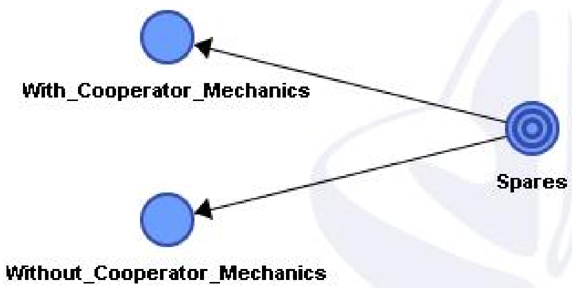
Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με την περίπτωση του Weka Suite (Association Rules).

4.6.5 Bayesian Networks – BayesiaLab

Τα αποτελέσματα του BayesiaLab είναι τα εξής:

Spares = Without_Spares (47,01%)		
Node	**Modal Value**	
Without_Cooperator_Mechanics	1	93,47%
With_Cooperator_Mechanics	0	93,47%
Spares = Company_Spares (39,88%)		
Node	**Modal Value**	
Without_Cooperator_Mechanics	1	99,04%
With_Cooperator_Mechanics	0	99,04%
Spares = Cooperator_Spares (10,47%)		
Node	**Modal Value**	
Without_Cooperator_Mechanics	1	82,62%
With_Cooperator_Mechanics	0	82,62%

Spares = Company_Spares_and_Cooperator_Spares (2,64%)		
Node	**Modal Value**	
Without_Cooperator_Mechanics	1	91,52%
With_Cooperator_Mechanics	0	91,52%



Εικόνα 4.14: Ο παραγόμενος γράφος του BayesiaLab

Και σε αυτήν την περίπτωση, τα αποτελέσματα και οι διαδικασίες είναι ίδιες με τις παραπάνω περιπτώσεις στην παραγωγή του Arrow Diagram.

4.7 Process Decision Program Chart

Στην παρούσα παράγραφο παρουσιάζονται τα αποτελέσματα των 5 λογισμικών εξόρυξης γνώσης σε δεδομένα τα οποία προορίζονται για τη δημιουργία ενός Process Decision Program Chart, καθώς και το δυναμικά δημιουργημένο Process Decision Program Chart. Επίσης, παραθέτονται και τα αποτελέσματα της πρόβλεψης σε νέα δεδομένα.

4.7.1 Rough Set - Rosetta

Στην περίπτωση επιλογής του λογισμικού Rosetta, παράγεται το εξής αρχείο εξόδου:

% Rules/patterns generated by ROSETTA.
% Exported 2007.12.23 17:42:28 by mike.
%
% 1R
% 6 rules.

cause(Εσωτερικής_Προέλευσης) => result(Προμηθευτές) OR result(Χρήση_-_Φθορά)
Supp. (LHS) = [9713 object(s)]
Supp. (RHS) = [2513 object(s), 7200 object(s)]
Acc. (RHS) = [0.258725, 0.741275]
Cov. (LHS) = [0.389236]
Cov. (RHS) = [1, 1]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

cause(Προμηθευτές) => result(Έρευνα_Αγοράς_Προμηθευτών)
Supp. (LHS) = [2513 object(s)]
Supp. (RHS) = [2513 object(s)]
Acc. (RHS) = [1]
Cov. (LHS) = [0.100705]
Cov. (RHS) = [0.388348]
Stab. (LHS) = [0]
Stab. (RHS) = [1]

cause(Εξωτερικής_Προέλευσης) => result(Καιρικές_Συνθήκες) OR result(Οδόστρωμα)
Supp. (LHS) = [2764 object(s)]
Supp. (RHS) = [1819 object(s), 945 object(s)]
Acc. (RHS) = [0.658104, 0.341896]
Cov. (LHS) = [0.110764]
Cov. (RHS) = [1, 1]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

cause(Καιρικές_Συνθήκες) => result(Προσαρμοσμένα_Προϊόντα) OR result(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [1819 object(s)]
Supp. (RHS) = [802 object(s), 1017 object(s)]
Acc. (RHS) = [0.440902, 0.559098]
Cov. (LHS) = [0.0728941]
Cov. (RHS) = [0.534667, 0.225699]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

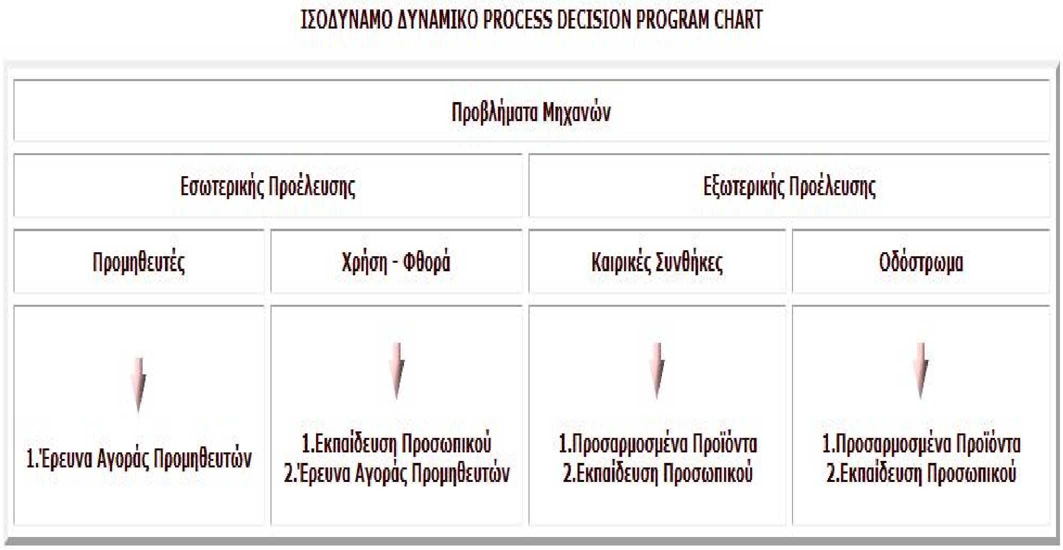
cause(Οδόστρωμα) => result(Προσαρμοσμένα_Προϊόντα) OR result(Εκπαίδευση_Προσωπικού)
Supp. (LHS) = [945 object(s)]
Supp. (RHS) = [698 object(s), 247 object(s)]
Acc. (RHS) = [0.738624, 0.261376]
Cov. (LHS) = [0.0378697]
Cov. (RHS) = [0.465333, 0.0548158]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

cause(Χρήση_-_Φθορά) => result(Εκπαίδευση_Προσωπικού) OR result(Έρευνα_Αγοράς_Προμηθευτών)
Supp. (LHS) = [7200 object(s)]
Supp. (RHS) = [3242 object(s), 3958 object(s)]
Acc. (RHS) = [0.450278, 0.549722]
Cov. (LHS) = [0.288531]

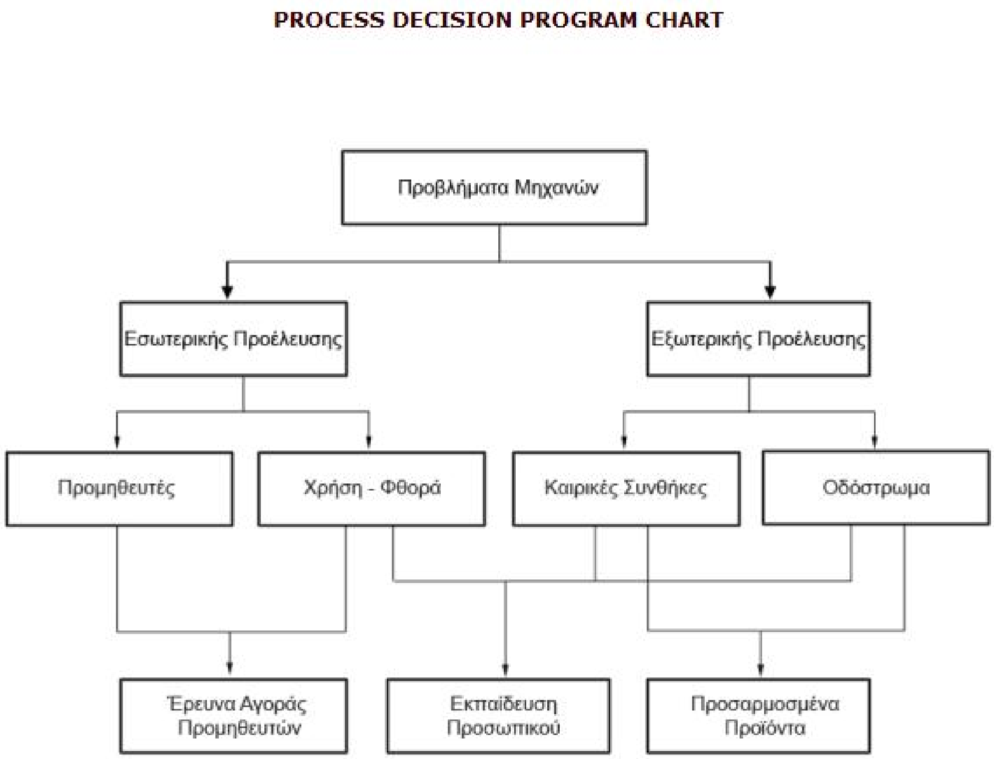
Cov. (RHS) = [0.719485, 0.611652]
Stab. (LHS) = [0]
Stab. (RHS) = [1, 1]

Οι κανόνες επιβεβαιώνονται από το αποτέλεσμα ενός ερωτήματος προς τη βάση δεδομένων.

Μετά την ενεργοποίηση του αλγόριθμου “σάρωσης” συλλέγονται οι κανόνες και αποθηκεύονται στον κατάλληλο πίνακα (table) στη βάση δεδομένων. Στη συνέχεια ενεργοποιείται ο αλγόριθμος δυναμικής δημιουργίας του Process Decision Program Chart (παρ. 3.5.7).



Εικόνα 4.15: Το δυναμικά παραγόμενο Process Decision Program Chart



Εικόνα 4.16: Ισοδύναμο “κλασσικό” Process Decision Program Chart

Όσον αφορά στην πρόβλεψη, θέτουμε το παρακάτω σύνολο δεδομένων προς πρόβλεψη:

Εσωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εξωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εσωτερικής Προέλευσης
Εξωτερικής Προέλευσης

Από το παραπάνω σύνολο δεδομένων παράγεται η εξής πρόβλεψη:

Είσοδος (Προέλευση Βλάβης)	**Πρόβλεψη (Κατηγορία Βλάβης)**
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εσωτερικής Προέλευσης	*Χρήση - Φθορά* (74.13%) *Προμηθευτές* (25.87%)
Εξωτερικής Προέλευσης	*Οδόστρωμα* (34.19%) *Καιρικές Συνθήκες* (65.81%)
Κατηγορία Βλάβης	**Γενική Λύση**
Χρήση - Φθορά	*Έρευνα Αγοράς Προμηθευτών* (54.97%) *Εκπαίδευση Προσωπικού* (45.03%)
Προμηθευτές	*Έρευνα Αγοράς Προμηθευτών* (100%)
Οδόστρωμα	*Εκπαίδευση Προσωπικού* (26.14%) *Προσαρμοσμένα Προϊόντα* (73.86%)
Καιρικές Συνθήκες	*Εκπαίδευση Προσωπικού* (55.91%) *Προσαρμοσμένα Προϊόντα* (44.09%)

Όπως και στις περιπτώσεις των υπολοίπων εργαλείων ποιότητας, παρέχεται η δυνατότητα της διαγραμματικής απεικόνισης των κανόνων.

4.7.2 Rough Set – RSES2

Με την επιλογή του αλγορίθμου Rough Set – RSES2 παράγονται οι ίδιοι κανόνες και αποτελέσματα και με τις ίδιες διαδικασίες (με τους ανάλογους αλγορίθμους "σάρωσης"), όπως και στην περίπτωση του Rough Set – Rosetta. Ακολουθούν τα αποτελέσματα του αλγόριθμου.

RULE_SET rses.txt
ATTRIBUTES 2
cause symbolic
result symbolic
DECISION_VALUES 7
Προμηθευτές
Έρευνα_Αγοράς_Προμηθευτών
Καιρικές_Συνθήκες
Προσαρμοσμένα_Προϊόντα
Εκπαίδευση_Προσωπικού
Οδόστρωμα
Χρήση_-_Φθορά
RULES 6
(cause=Εσωτερικής_Προέλευσης)=>(result={Προμηθευτές[2513],Χρήση_-_Φθορά[7200]}) 9713
(cause=Χρήση_-_Φθορά)=>(result={Εκπαίδευση_Προσωπικού[3242],Έρευνα_Αγοράς_Προμηθευτών[3958]})
7200
(cause=Εξωτερικής_Προέλευσης)=>(result={Καιρικές_Συνθήκες[1819],Οδόστρωμα[945]}) 2764
(cause=Προμηθευτές)=>(result=Έρευνα_Αγοράς_Προμηθευτών[2513]) 2513
(cause=Καιρικές_Συνθήκες)=>(result={Προσαρμοσμένα_Προϊόντα[802],Εκπαίδευση_Προσωπικού[1017]}) 1819
(cause=Οδόστρωμα)=>(result={Προσαρμοσμένα_Προϊόντα[698],Εκπαίδευση_Προσωπικού[247]}) 945

4.7.3 Association Rules – Weka Suite

Το αρχείο εξόδου που παράγεται από τα συσχετιζόμενα δεδομένα στην περίπτωση του Weka Suite (Association Rules) είναι το εξής:

=== Run information ===

Scheme: weka.associations.Tertius -K 0 -F 0.0 -C -1.0 -N 1.0 -L 2 -G 0 -S -c 0 -I 0 -p -P 0
Relation: process_decision
Instances: 24954
Attributes: 2
cause
result
=== Associator model (full training set) ===

Tertius
=======

1. /* 0,706844 0,100705 */ cause = Εσωτερικής_Προέλευσης ==> result = Χρήση_-_Φθορά
2. /* 0,375730 0,000000 */ cause = Προμηθευτές ==> result = Έρευνα_Αγοράς_Προμηθευτών
3. /* 0,337081 0,129919 */ cause = Χρήση_-_Φθορά ==> result = Έρευνα_Αγοράς_Προμηθευτών
4. /* 0,311509 0,158612 */ cause = Χρήση_-_Φθορά ==> result = Εκπαίδευση_Προσωπικού
5. /* 0,297664 0,037870 */ cause = Εξωτερικής_Προέλευσης ==> result = Καιρικές_Συνθήκες
6. /* 0,254580 0,288531 */ cause = Εσωτερικής_Προέλευσης ==> result = Προμηθευτές
7. /* 0,167867 0,009898 */ cause = Οδόστρωμα ==> result = Προσαρμοσμένα_Προϊόντα
8. /* 0,153152 0,072894 */ cause = Εξωτερικής_Προέλευσης ==> result = Οδόστρωμα
9. /* 0,149415 0,032139 */ cause = Καιρικές_Συνθήκες ==> result = Εκπαίδευση_Προσωπικού
10. /* 0,143645 0,040755 */ cause = Καιρικές_Συνθήκες ==> result = Προσαρμοσμένα_Προϊόντα
11. /* 0,021085 0,027971 */ cause = Οδόστρωμα ==> result = Εκπαίδευση_Προσωπικού
12. /* 0,000000 0,037870 */ cause = Οδόστρωμα ==> FALSE
13. /* 0,000000 0,072894 */ cause = Καιρικές_Συνθήκες ==> FALSE

14. /* 0,000000 0,100705 */ cause = Προμηθευτές ==> FALSE
15. /* 0,000000 0,110764 */ cause = Εξωτερικής_Προέλευσης ==> FALSE
16. /* 0,000000 0,288531 */ cause = Χρήση_-_Φθορά ==> FALSE
17. /* 0,000000 0,389236 */ cause = Εσωτερικής_Προέλευσης ==> FALSE
18. /* 0,000000 0,711469 */ TRUE ==> result = Χρήση_-_Φθορά
19. /* 0,000000 0,740683 */ TRUE ==> result = Έρευνα_Αγοράς_Προμηθευτών
20. /* 0,000000 0,819428 */ TRUE ==> result = Εκπαίδευση_Προσωπικού
21. /* 0,000000 0,899295 */ TRUE ==> result = Προμηθευτές
22. /* 0,000000 0,927106 */ TRUE ==> result = Καιρικές_Συνθήκες
23. /* 0,000000 0,939889 */ TRUE ==> result = Προσαρμοσμένα_Προϊόντα
24. /* 0,000000 0,962130 */ TRUE ==> result = Οδόστρωμα

Number of hypotheses considered: 55
Number of hypotheses explored: 55
Time: 00 min 00 s 686 ms

Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με τις παραπάνω περιπτώσεις (Rosetta, RSES2), πλην του ότι δεν υπάρχουν ποσοτικά δεδομένα.

4.7.4 Classification Rules – Weka Suite

Το αρχείο εξόδου στην περίπτωση του Weka Suite (Classification Rules) είναι το εξής:

=== Run information ===

Scheme: weka.classifiers.rules.Prism
Relation: process_decision
Instances: 24954
Attributes: 2
cause
result
Test mode: 10-fold cross-validation

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

Prism rules

If cause = Εσωτερικής_Προέλευσης then Προμηθευτές
If cause = Εσωτερικής_Προέλευσης then Χρήση_-_Φθορά
If cause = Εξωτερικής_Προέλευσης then Οδόστρωμα
If cause = Εξωτερικής_Προέλευσης then Καιρικές_Συνθήκες
If cause = Προμηθευτές then Έρευνα_Αγοράς_Προμηθευτών
If cause = Χρήση_-_Φθορά then Έρευνα_Αγοράς_Προμηθευτών
If cause = Οδόστρωμα then Προσαρμοσμένα_Προϊόντα
If cause = Καιρικές_Συνθήκες then Προσαρμοσμένα_Προϊόντα
If cause = Καιρικές_Συνθήκες then Εκπαίδευση_Προσωπικού
If cause = Χρήση_-_Φθορά then Εκπαίδευση_Προσωπικού
If cause = Οδόστρωμα then Εκπαίδευση_Προσωπικού

Time taken to build model: 0.25 seconds

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

Correctly Classified Instances 11429 45.8003 %
Incorrectly Classified Instances 13525 54.1997 %
Kappa statistic 0.3616
Mean absolute error 0.1549
Root mean squared error 0.3935
Relative absolute error 68.0547 %
Root relative squared error 116.6673 %
Total Number of Instances 24954

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure Class
1 0.321 0.259 1 0.411 Προμηθευτές
0 0 0 0 0 Χρήση_-_Φθορά
1 0.076 0.342 1 0.51 Οδόστρωμα
0 0 0 0 0 Καιρικές_Συνθήκες
1 0.175 0.666 1 0.8 Έρευνα_Αγοράς_Προμηθευτών
1 0.054 0.543 1 0.704 Προσαρμοσμένα_Προϊόντα
0 0 0 0 0 Εκπαίδευση_Προσωπικού

=== Confusion Matrix ===

a b c d e f g <-- classified as
2513 0 0 0 0 0 0 | a = Προμηθευτές
7200 0 0 0 0 0 0 | b = Χρήση_-_Φθορά
0 0 945 0 0 0 0 | c = Οδόστρωμα
0 0 1819 0 0 0 0 | d = Καιρικές_Συνθήκες
0 0 0 0 6471 0 0 | e = Έρευνα_Αγοράς_Προμηθευτών
0 0 0 0 0 1500 0 | f = Προσαρμοσμένα_Προϊόντα
0 0 0 0 3242 1264 0 | g = Εκπαίδευση_Προσωπικού

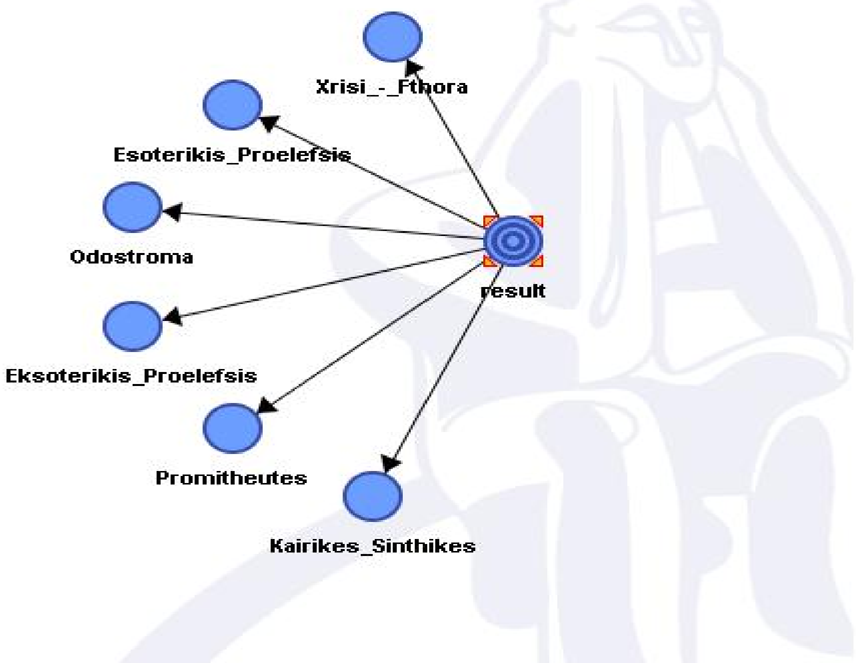
Τα αποτελέσματα που παράγονται και οι διαδικασίες που ακολουθούνται είναι ίδιες με την περίπτωση του Weka Suite (Association Rules).

4.7.5 Bayesian Networks – BayesiaLab

Τα αποτελέσματα του BayesiaLab είναι τα εξής:

result = Xrisi_-_Fthora (28,85%)		
Node	**Modal Value**	
Esoterikis_Proelefsis	1	100,00%
Xrisi_-_Fthora	0	100,00%
Eksoterikis_Proelefsis	0	100,00%
Promitheutes	0	100,00%
Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%
result = Erevna_Agoras_Promitheuton (25,93%)		
Node	**Modal Value**	
Esoterikis_Proelefsis	0	100,00%
Promitheutes	0	61,17%
Xrisi_-_Fthora	1	61,17%
Eksoterikis_Proelefsis	0	100,00%
Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%
result = Ekpaideusi_Prosopikou (18,06%)		
Node	**Modal Value**	
Esoterikis_Proelefsis	0	100,00%
Xrisi_-_Fthora	1	71,95%
Kairikes_Sinthikes	0	77,43%
Eksoterikis_Proelefsis	0	100,00%
Promitheutes	0	100,00%
Odostroma	0	94,52%
result = Promitheutes (10,07%)		
Node	**Modal Value**	
Esoterikis_Proelefsis	1	100,00%
Xrisi_-_Fthora	0	100,00%
Eksoterikis_Proelefsis	0	100,00%
Promitheutes	0	100,00%
Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%
result = Kairikes_Sinthikes (7,29%)		
Node	**Modal Value**	
Eksoterikis_Proelefsis	1	100,00%
Esoterikis_Proelefsis	0	100,00%

Xrisi_-_Fthora	0	100,00%
Promitheutes	0	100,00%
Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%
result = Prosarmosmena_Proionta (6,01%)		
Node	**Modal Value**	
Odostroma	0	53,47%
Kairikes_Sinthikes	1	53,47%
Esoterikis_Proelefsis	0	100,00%
Xrisi_-_Fthora	0	100,00%
Eksoterikis_Proelefsis	0	100,00%
Promitheutes	0	100,00%
result = Odostroma (3,79%)		
Node	**Modal Value**	
Eksoterikis_Proelefsis	1	100,00%
Esoterikis_Proelefsis	0	100,00%
Xrisi_-_Fthora	0	100,00%
Promitheutes	0	100,00%
Kairikes_Sinthikes	0	100,00%
Odostroma	0	100,00%



Εικόνα 4.17: Ο παραγόμενος γράφος του BayesiaLab

Και σε αυτήν την περίπτωση, τα αποτελέσματα και οι διαδικασίες είναι ίδιες με τις παραπάνω περιπτώσεις στην παραγωγή του Process Decision Program Chart.

Κεφάλαιο 5

Συμπεράσματα

5.1 Εισαγωγή

Σε αυτό το κεφάλαιο αναπτύσσονται τα συμπεράσματα που προκύπτουν από τα αποτελέσματα της εφαρμογής. Αρχικά, παρουσιάζονται τα συμπεράσματα που αφορούν στους επιλεγμένους αλγόριθμους εξόρυξης γνώσης, στη συνέχεια αναλύονται τα συμπεράσματα για τη μελέτη περίπτωσης, όπως προκύπτουν από τα παραγόμενα εργαλεία ποιότητας και το κεφαλαίο κλείνει με μια παράγραφο όπου αναπτύσσονται κάποιες σκέψεις για τις μελλοντικές κατευθύνσεις στο πεδίο της βελτιστοποίησης των διαδικασιών με τεχνικές εξόρυξης γνώσης.

5.2 Επιλεγμένοι Αλγόριθμοι

Όλοι οι χρησιμοποιούμενοι αλγόριθμοι εξόρυξης γνώσης κατάφεραν να εκπληρώσουν το βασικό στόχο-προϋπόθεση, δηλαδή την εύρεση όλων των πιθανών κανόνων που δυναμικά υπάρχουν μέσα στα δεδομένα. Αυτό άλλωστε ήταν και το κύριο κριτήριο επιλογής τους. Παρακάτω, αναπτύσσονται τα βασικά χαρακτηριστικά που προκύπτουν για τον καθένα από αυτούς σύμφωνα με τα αποτελέσματά τους.

Τα αποτελέσματα του λογισμικού Rosetta (Rough Set - Holte's 1R algorithm) παρουσιάζουν το σύνολο των πιθανών κανόνων και έχουν το πλεονέκτημα ότι δίνουν και ποσοτικές πληροφορίες για τον κάθε κανόνα. Επίσης, ο τρόπος που παρουσιάζονται οι κανόνες θεωρείται αρκετά ικανοποιητικός για έναν αλγόριθμο που προσπαθεί να εκμαιεύσει τους κανόνες σαρώνοντας το αρχείο εξόδου του Rosetta.

Συν τοις άλλης, δεν παρουσιάζει καθόλου περιττούς κανόνες, γεγονός που διευκολύνει ακόμη περισσότερο την εξαγωγή των κανόνων και προσφέρει και αρκετούς δείκτες σημαντικότητας για κάθε κανόνα. Τέλος, το αρχείο εισόδου δεν απαιτεί σημαντική κωδικοποίηση αλλά ούτε και μετασχηματισμούς.

Για το λογισμικό RSES2 (Rough Set - Covering algorithm) καταλήγουμε στα ίδια ακριβώς συμπεράσματα με αυτά του λογισμικού Rosetta, με την διαφορά ότι δεν παρέχει κανένα δείκτη σημαντικότητας για κάθε κανόνα.

Τα αποτελέσματα του Weka Suite - Association Rules (Tertius algorithm) παράγουν τα ιδία με αυτά του Rosetta – Rough Set, πλην της πληροφορίας για τα ποσοτικά δεδομένα. Επίσης, στον αλγόριθμο παρουσιάζονται αρκετοί περιττοί κανόνες.

Για τα αποτελέσματα του Weka Suite - Classification Rules (Prism algorithm) συμπεραίνουμε ότι παρουσιάζεται το σύνολο των κανόνων, δεν υπάρχουν περιττοί κανόνες, παρέχονται αρκετοί δείκτες σημαντικότητας για κάθε κανόνα, υπάρχει ευκολία εντοπισμού και εξαγωγής των κανόνων και το κόστος κωδικοποίησης των δεδομένων εισόδου είναι ελάχιστο. Το αρνητικό σε αυτήν την περίπτωση είναι ότι οι κανόνες δεν τροφοδοτούνται με ποσοτικές πληροφορίες.

Το λογισμικό BayesiaLab (Bayesian Networks – Naive Bayes Algorithm) παράγει μεν όλους τους κανόνες με τη συνοδεία ποσοτικής πληροφορίας και με παροχή αρκετών δεικτών σημαντικότητας του κάθε κανόνα, αλλά απαιτεί πολύπλοκη κωδικοποίηση και μετασχηματισμό των δεδομένων εισόδου και δυσκολεύει πολύ στην εξαγωγή των κανόνων μέσα από τα αρχεία εξόδου, χωρίς να αποκλείει την εμφάνιση περιττών κανόνων.

Παρακάτω ακολουθούν 2 πίνακες, ένας πίνακας που συσχετίζει τα βασικά κριτήρια αξιολόγησης των αλγορίθμων εξόρυξης γνώσης με τους επιλεγμένους αλγορίθμους της παρούσας εργασίας και ένας που κατατάσσει τους αλγόριθμους σύμφωνα με την αξιολόγηση τους.

Αλγόριθμοι / **Κριτήρια**	*Holte's1R*	*Covering*	*Tertius*	*Prism*	*Naive Bayes*
Σύνολο Κανόνων	✓	✓	✓	✓	✓
Ποσοτικές Πληροφορίες	✓	✓			✓
Απόρριψη περιττών κανόνων	✓	✓		✓	
Ευκολία εξαγωγής κανόνων	✓	✓	✓	✓	
Δείκτες σημαντικότητας	✓		✓	✓	✓
Χαμηλό κόστος κωδικοποίησης	✓	✓	✓	✓	

Πίνακας 5.1: Συσχετισμός των αλγορίθμων με τα κριτήρια αξιολόγησης αυτών

Κατάταξη Αλγορίθμων	***Αλγόριθμοι***
1.	*Holte's1R (Rosetta)*
2.	*Covering (RSES2)*
3.	*Prism (Weka Suite)*
4.	*Tertius (Weka Suite)*
5.	*Naive Bayes (BayesiaLab)*

Πίνακας 5.2: Κατάταξη αξιολόγησης αλγορίθμων

5.3 Μελέτη Περίπτωσης

Από τα αποτελέσματα των 7 νέων εργαλείων ποιότητας εξάγονται αρκετά σημαντικά συμπεράσματα τα οποία μπορούν να βοηθήσουν στη βελτίωση των διαδικασιών της εταιρείας που μελετάμε και στην περεταίρω ανάπτυξή της.

Πιο συγκεκριμένα, από τα αποτελέσματα του Affinity Diagram προκύπτει ότι οι περισσότερες βλάβες οφείλονται στην κακή χρήση των μηχανών ή στη φυσιολογική τους φθορά. Λιγότερες βλάβες οφείλονται στα "ελαττωματικά" προϊόντα των προμηθευτών, στις αντίξοες για τα μηχανήματα καιρικές συνθήκες και στο κακό οδόστρωμα. Άρα το Affinity Diagram "προτείνει" μια πιο προσεκτική χρήση των μηχανών με σκοπό τη μείωση των βλαβών.

Το Relationship Diagram καταδεικνύει ότι οι βλάβες μπορούν να μειωθούν αν δοθεί περισσότερη προσοχή στη λύση με τίτλο "Ερευνά Αγοράς Προμηθευτών" και στη λύση με τίτλο "Εκπαίδευση Προσωπικού", καθώς όλες σχεδόν οι κατηγορίες των βλαβών καταλήγουν άμεσα ή έμμεσα σε αυτές (ειδικά στην "Εκπαίδευση Προσωπικού").

Το Systematic Diagram "αποδεικνύει" ότι με την εφαρμογή των κατάλληλων λύσεων θα μειωθούν σημαντικά οι βλάβες των μηχανών, καθώς η πλειοψηφία τους οφείλεται περισσότερο σε εσωτερικούς παράγοντες (Χρήση-Φθορά, Προμηθευτές) και μπορούν να βελτιωθούν από ενέργειες της εταιρείας, παρά σε εξωτερικούς παράγοντες (Οδόστρωμα, Καιρικές Συνθήκες), όπου είναι δύσκολο να παρέμβει η εταιρεία.

Το Matrix Diagram "καταδεικνύει" ποιες λύσεις ταιριάζουν με τις υπάρχουσες κατηγορίες βλαβών, ενώ το Matrix Data Analysis (που προκύπτει αν υπάρχουν ποσοτικές πληροφορίες) ποσοτικοποιεί τις σχέσεις αυτές.

Το Arrow Diagram "παρουσιάζει" τη διαδικασία που ακολουθείται χρονικά ως προς τη συσχέτιση της προέλευσης των ανταλλακτικών που χρησιμοποιούνται για την επισκευή της βλάβης και των μηχανικών (της εταιρείας ή των εξωτερικών συνεργατών) που εργάστηκαν για τη βλάβη.

Τέλος, το Process Decision Program Chart περικλείει τα αποτελέσματα του Systematic Diagram και Matrix Diagram σε ένα κοινό διάγραμμα, παρέχοντας το συνδυασμό των αποτελεσμάτων των δυο παραπάνω εργαλείων.

Συνοψίζοντας, οδηγούμαστε στο γενικό συμπέρασμα ότι τα 7 νέα εργαλεία ποιότητας παράγουν επισημάνσεις και πορίσματα τα οποία σε συνδυασμό και με τα υπόλοιπα εργαλεία και πρακτικές της Διοίκησης Ολικής Ποιότητας (Δ.Ο.Π.), τείνουν να οδηγήσουν την υπό μελέτη εταιρεία σε περαιτέρω ανάπτυξη.

5.4 Μελλοντικές Κατευθύνσεις

Από τα αποτέλεσμα της παρούσας εργασίας προκύπτει ότι θα είχε ιδιαίτερο ενδιαφέρον η αποδοχή των τεχνικών εξόρυξης γνώσης ως μεθόδου όχι μόνο για τη δημιουργία των 7 νέων εργαλείων ποιότητας αλλά και για τη θεμελίωση των υπολοίπων εργαλείων που χρησιμοποιούνται στην Δ.Ο.Π. από τους οργανισμούς και τις εταιρείες.

Ειδικότερα, σε εταιρείες και οργανισμούς όπου υπάρχει αυτοματοποιημένο σύστημα παραγωγής, οι τεχνικές εξόρυξης γνώσης με τα εργαλεία της Δ.Ο.Π. μπορούν να συνδυαστούν σε ένα συνολικό δυναμικό σύστημα διαχείρισης και ανάλυσης των δεδομένων, το οποίο θα εξάγει σε πραγματικό χρόνο (real time) συμπεράσματα και προτάσεις για τη βελτιστοποίηση των διαδικασιών.

Βιβλιογραφία

Bazan J. G., Nguyen H. S., Nguyen S. H., Synak P., Wróblewski J. (2000). “Rough set algorithms in classification problem”. In Polkowski L., Tsumoto S., Lin T. (eds.), *Rough Set Methods and Applications*, Heidelberg Germany: Physica-Verlag, pp. 49–88.

Bazan J., Nguyen H. S., Skowron A., Szczuka M. (2003). “A View on Rough Set Concept Approximations”. In Wang G., Liu Q., Yao Y., Skowron A. (eds.), *Proceedings of R.S.F.D.G.r.C.*, China: Springer, pp. 181-188.

Bose I., Mahapatra K. R. (2001). “Business datamining – A machine learning perspective”. *Information and Management*, vol. 39, num. 3, pp. 211-225.

Cendrowska J. (1987). “PRISM: An algorithm for inducing modular rules”. *International Journal of Man-Machine Studies*, vol.27, num. 4, pp. 349-370.

Crosby P. B. (1979). *Quality Is Free: The Art of Making Quality Certain*. New York: McGraw-Hill.

Dahlgaard J. J., Kristensen K., Kanji K. (1998). *Fundamentals of total quality management*. London: Chapman & Hall.

Dale G. D. (1994). *Managing quality* (2nd Edition). London: Prentice Hall.

Deming W. E. (1986). *Out of the Crisis*. Cambridge, MA: MIT Center for Advanced Engineering Study.

Fayyad U., Piatetsky-Shapiro G., Smyth P., Ramasami U. (1996). *Advances in Knowledge Discovery and Data Mining*. Cambridge: MIT Press

Feigenbaum A. V. (1991). *Total quality control* (3rd Edition). New York: McGraw Hill.

Flach P. A. , Lachiche N. (1999a). “Confirmation-Guided Discovery of first-order rules with Tertius”. *Machine Learning*, vol. 42, nums. 1-2, pp. 61-95.

Flach P. A., Lachiche N. (1999b). *The Tertius system*. http://www.cs.bris.ac.uk/Research/MachineLearning/Tertius/.

Frawley W., Piatetsky-Shapiro G., Matheus C. (1992). "Knowledge Discovery in Databases: An Overview". *AI Magazine*, vol. 13, num. 3, pp. 57-70.

Groocock J. M. (1980). “Conformance or fitness for use?”. *EOQC Quality*, vol. 2, pp. 3-6.

Groocock J. M. (1981). “Conformance or fitness for use? (II)”. *EOQC Quality*, vol. 4, pp. 5-7.

Holte R. C. (1993). “Very Simple Classification Rules Perform Well on Most Commonly Used Datasets”. *Machine Learning*, vol. 11, pp. 63-90.

Juran J. M. (1988). *Juran on Planning for Quality*. New York: MacMillan.

Kanji K. G., Asher M. (1996). *100 methods for total quality management*. London: Sage Publications.

Kolarik J. W. (1995). *Creating Quality: Concepts, Systems, Strategies and Tools*. New York: McGraw-Hill.

Lindsay M. W., Petrick A. J. (1997). *Total quality and organization development*. Florida: St. Lucie Press.

Mizuno S. (1988). *Managing for quality improvement: The seven new QC tools*. Cambridge: Productivity Press.

Montgomery D. C. (1997). *Introduction to statistical quality control* (3rd Edition). New York: John Willey & Sons.

Pawlak Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning About Data*. Dordrecht: Kluwer Academic Publishing.

Pawlak Z. Wong, S. K. M., Ziarko W. (1988). "Rough sets: Probabilistic versus deterministic approach". *International Journal of Man-Machine Studies,* vol. 29, pp. 81–95.

Pham D. T., Oztemel E. (1996). *Intelligent Quality Systems*. London: Springer.

Piatetsky-Shapiro G., Frawley W. J., Matheus C. (1991). *Knowledge Discovery in Databases*. A.A.A.I./MIT Press

Quinlan R. (1986). “Induction of decision trees”. *Machine Learning*, vol. 1, num. 1, pp. 81-106.

Rao A., Carr L. P., Dambolena I., Kopp R. J., Martin J., Rafii F., Schlesinger P. F. (1996). *Total quality management: A cross functional perspective*. New York: John Willey & Sons.

Tague N. R. (2005). *The Quality Toolbox*. Milwaukee: ASQ Quality Press.

Thoday W. R. (1981). “Additional reflections on quality and terminology”. *EOQC Quality*, vol. 1, p. 11.

Witten I. H., Frank E. (2005). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.

Zaludova A. H. (1981). "Some reflections on quality terminology". *EOQC Quality*, vol.1, pp. 3-10.