

ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ



ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**« Ανάλυση παραγόντων αποχώρησης εργαζομένων με χρήση της μεθόδου
fs/QCA »**

Υπό: Καντουνατάκη Σοφία

Επιβλέπων Καθηγητής : Τσαφάρκης Στέλιος

ΧΑΝΙΑ 2023

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω θερμά τον επιβλέπων υποψήφιο διδάκτορα, κο Κυριακίδη Αναστάσιο, για την βοήθεια, την καθοδήγηση και την αμέριστη συμπαράσταση που μου πρόσφερε καθ' όλη την διάρκεια την συγγραφής της διπλωματικής μου εργασίας, όπως επίσης και τον κο Τσαφαράκη Στέλιο για τις συμβουλές και παρατηρήσεις του.

Επιπλέον θα ήθελα να ευχαριστήσω τους δικούς μου ανθρώπους, οι οποίοι στάθηκαν δίπλα μου και με στήριξαν καθ' όλη την διάρκεια της προσπάθειας μου δείχνοντας μου εμπιστοσύνη κατανόηση και αγάπη.

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

1. Δρ. Τσαφάρας Σ.
2. Δρ. Δούμος Μ.
3. Δρ. Ατσαλάκης Γ.

Πίνακας περιεχομένων

Περίληψη.....	6
Κεφάλαιο 1: Εισαγωγή.....	7
1.1 Δομή της διπλωματικής εργασίας	7
1.2 Κριτήρια προς αξιολόγηση.....	7
Κεφάλαιο 2: EMPLOYEE ATTRITION (Απώλεια των εργαζομένων)	9
2.1 Απώλεια των εργαζομένων	9
2.2 Παράγοντες που επηρεάζουν την οικειοθελή αποχώρηση	11
2.3 Αποχώρηση Προσωπικού VS Εναλλαγή Προσωπικού.....	11
2.4 Μείωση του ποσοστού αποχώρησης εργαζομένων.....	12
Κεφάλαιο 3: Παρουσίαση προηγούμενων μελετών employee attrition (αποχώρησης εργαζομένων).....	14
Κεφάλαιο 4: Ποιοτική συγκριτική ανάλυση της QCA (Qualitative Comparative Analysis).....	16
4.1 Ποιοτική προσέγγιση της QCA.....	16
4.2 Ποσοτική προσέγγιση της QCA.....	17
4.3 Κατηγορίες της QCA.....	17
4.4 Διαδικασία της QCA	18
Κεφάλαιο 5: Η διαδικασία της fs/QCA.....	19
5.1 Εισαγωγή.....	19
5.2 Calibration – Βαθμονόμηση	19
5.3 Πίνακας Αλήθειας	20
5.4 Κατώφλι Συχνότητας.....	21
5.5 Κατώφλι Συνέπειας	22
5.6 Ελαχιστοποίηση Πίνακα.....	22
5.7 Είδη λύσεων	23
5.8 Απλουστευτικές Υποθέσεις (Simplifying Assumptions)	24
5.9 Μέτρα προσαρμογής	25
5.10 Αναγκαιότητα και Επάρκεια	26
Κεφάλαιο 6: Association Rules.....	27
6.1 Θεωρία Κανόνων Συσχέτισης (Association Rules)	27
6.2 Αλγόριθμος Apriori.....	28
Κεφάλαιο 7: Στατιστική Ανάλυση Δεδομένων	30
7.1 Γραφήματα Κριτηρίων	30
7.2 Πίνακας Συσχετίσεων.....	35
Κεφάλαιο 8: Ανάλυση Αποτελεσμάτων fs/QCA	36
8.1 Εισαγωγή.....	36
8.2 Δεδομένα	36

8.3 Βαθμονόμηση (calibration)	37
8.4 Descriptive Statistics	37
8.5 Frequencies.....	38
8.6 Αναγκαίες συνθήκες.....	39
8.7 Ικανές Συνθήκες	39
Κεφάλαιο 9: fsQCA με χρήση Undersampling	45
9.1 Παρουσία αποτελέσματος (left)	45
9.2 Στατιστική μέθοδος X^2	49
9.2.1 Squared Cramer's Phi Coefficient (ϕ^2).....	49
9.2.2 Έλεγχος στατιστικής σχέσης μεταξύ αποχώρησης των εργαζομένων και αν είχαν εργατικό ατύχημα	50
9.2.3 Έλεγχος στατιστικής σχέσης μεταξύ αποχώρησης των εργαζομένων και αν είχαν πάρει προαγωγή τα τελευταία 5 χρόνια.....	52
9.3 Απουσία αποτελέσματος (~ left)	53
Κεφάλαιο 10: Αλγόριθμος Apriori.....	57
10.1 Εφαρμογή του Αλγορίθμου Apriori	57
Κεφάλαιο 11: Σύγκριση αποτελεσμάτων της μεθόδου fsQCA και της Apriori.....	59
Κεφάλαιο 12: Συμπεράσματα.....	61
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	63

Περίληψη

Σκοπός της παρούσας διπλωματικής εργασίας είναι η πρόταση ενός μεθοδολογικού πλαισίου που θα είναι σε θέση να εντοπίζει αιτιώδη μονοπάτια (ικανές και αναγκαίες συνθήκες) που οδηγούν στην αποχώρηση των εργαζομένων (employee attrition) μέσω της Ποιοτικής Συγκριτικής Ανάλυσης με ασαφή σύνολα (fuzzy-set Qualitative Comparative Analysis – fs/QCA), και να προτείνει πιθανούς τρόπους αποφυγής. Επίσης, εφαρμόζεται ο αλγόριθμος Apriori (ως αντιπροσωπευτικό παράδειγμα Association rules mining) για σύγκριση και έλεγχο ευρωστίας (robustness test) των αποτελεσμάτων των δυο μεθόδων.

Τα δεδομένα που πρόκειται να αναλυθούν προέρχονται από ένα τεχνητό dataset που δημιουργήθηκε για έναν διαγωνισμό της Kaggle, και αποτελούν αντιπροσωπευτικό δείγμα αντίστοιχων πραγματικών dataset. Κύριο χαρακτηριστικό του συγκεκριμένου dataset είναι η άνιση κατανομή των δύο κλάσεων, αποχωρήσαντων και μη εργαζομένων. (imbalanced class distribution).

Η μέθοδος fs/QCA βασίζεται στη θεωρία των ασαφών συνόλων και στη χρήση της Boolean άλγεβρας για να εντοπίσει αιτιώδεις συνθήκες/κριτήρια (ή συνδυασμούς αυτών), οι οποίες θα είναι αναγκαίες ή/και ικανές για την παρουσία του αποτελέσματος (αποχώρηση από την εταιρεία). Μετά την αρχική εφαρμογή της μεθόδου, θα χρησιμοποιηθεί η τεχνική Undersampling για την αντιμετώπιση της άνισης κατανομής των κλάσεων, με στόχο την βελτίωση των αποτελεσμάτων, καθώς και ο έλεγχος X^2 για την περαιτέρω ανάλυση των αναγκαίων συνθηκών που προκύπτουν.

Η εξόρυξη κανόνων συσχέτισης (association rules) χρησιμοποιεί μοντέλα μηχανικής μάθησης για να αναζητήσει σε βάσεις δεδομένων συγκεκριμένα μοτίβα. Ο αλγόριθμος Apriori (Agrawal & Srikant, 1994) που χρησιμοποιείται στην παρούσα εργασία είναι ένας από τους πρώτους και πιο δημοφιλείς αλγόριθμους εύρεσης κανόνων συσχέτισης.

Τα συμπεράσματα επικεντρώνονται στις αιτίες αποχώρησης των εργαζομένων από την εταιρεία, με την εξαγωγή κανόνων (αιτιωδών μονοπατιών), δίνοντας με αυτόν τον τρόπο τη δυνατότητα στην εκάστοτε εταιρεία να εντοπίσει τα δυνατά και τα αδύναμα σημεία της, καθώς και να προσδιορίσει αυτά που χρήζουν βελτίωσης, ώστε να αποτρέπεται η φυγή των εργαζομένων.

Κεφάλαιο 1: Εισαγωγή

1.1 Δομή της διπλωματικής εργασίας

Η παρούσα διπλωματική εργασία αποτελείται συνολικά από δώδεκα (12) κεφάλαια. Στο πρώτο κεφάλαιο γίνεται μια αναφορά στα προς αξιολόγηση κριτήρια. Στο δεύτερο κεφάλαιο περιγράφεται η έννοια του employee attrition (απώλεια των εργαζομένων) , ενώ παράλληλα αναφέρονται και εξηγούνται οι σημαντικές αιτίες που μπορεί να οδηγήσουν τους εργαζόμενους στην αποχώρηση από την εργασία τους. Επιπλέον παρουσιάζονται τρόποι αντιμετώπισης του φαινομένου. Στο τρίτο κεφάλαιο παραθέτονται προγενέστερες έρευνες αποχώρησης προσωπικού.

Στα κεφάλαια 4 και 5 γίνεται η παρουσίαση της μεθόδου QCA και της διαδικασίας fsQCA που θα εφαρμοστεί, αντίστοιχα. Στο κεφάλαιο 6 αναφερόμαστε στην μέθοδο Association rules, ενώ η παρουσίαση της στατιστικής ανάλυσης των δεδομένων γίνεται στο κεφάλαιο 7.

Στο κεφάλαιο 8 παρουσιάζεται όλη η διαδικασία της μεθόδου fs/QCA στο σύνολο δεδομένων που έχουμε, καθώς και τα αποτελέσματα στα οποία καταλήγουμε. Στο κεφάλαιο 9 γίνεται η παρουσίαση της τεχνικής Undersampling και η εφαρμογή της στο μειωμένο δείγμα μας. Ακόμα, χρησιμοποιούμε την στατιστική ανάλυση X^2 για περαιτέρω ανάλυση των αναγκαίων συνθηκών. Στο κεφάλαιο 10 αναλύεται η εφαρμογή του αλγορίθμου Apriori και παρουσιάζονται οι επικρατέστεροι κανόνες συσχέτισης.

Στα κεφάλαια 11 και 12 γίνεται σύγκριση των αποτελεσμάτων των μεθόδων fsQCA και Apriori και μια ανασκόπηση των αποτελεσμάτων του προτεινόμενου μεθοδολογικού πλαισίου.

1.2 Κριτήρια προς αξιολόγηση

Τα κριτήρια στα οποία αξιολογούνται οι εργαζόμενοι είναι συνολικά 7: satisfaction level, last evaluation, number of project, average monthly hours, years at company, work accident και promotion last 5 years, με εξαρτημένη μεταβλητή την left.

- Satisfaction level
Οι εργαζόμενοι αξιολογήθηκαν σύμφωνα με το πόσο ικανοποιημένοι είναι στην εταιρεία που εργάζονται.
- Last evaluation
Οι εργαζόμενοι αξιολογήθηκαν σύμφωνα με τι σκορ είχαν λάβει στην τελευταία αξιολόγηση που τους έγινε.
- Number of projects
Οι εργαζόμενοι αξιολογήθηκαν για το πόσες εργασίες έχουν ανατεθεί σε αυτούς.
- Average monthly hours
Οι εργαζόμενοι αξιολογήθηκαν σύμφωνα με το ποιος είναι ο μέσος όρος ωρών που εργάζονται σε έναν μήνα.

➤ Years at company

Οι εργαζόμενοι αξιολογήθηκαν ως προς τα πόσα χρόνια εργάζονται στην εταιρεία.

➤ Work accident

Οι εργαζόμενοι αξιολογήθηκαν σύμφωνα με το αν είχαν ή όχι εργατικό ατύχημα στην εταιρεία(1=ΝΑΙ, 0=ΟΧΙ)

➤ Promotion last 5 years

Οι εργαζόμενοι αξιολογήθηκαν σύμφωνα με το αν έχουν πάρει προαγωγή τα τελευταία πέντε χρόνια.

➤ Left

Οι εργαζόμενοι αξιολογήθηκαν σύμφωνα με το αν έφυγαν ή παρέμειναν στην εταιρεία (1=ΝΑΙ, 0=ΟΧΙ)

Κεφάλαιο 2: EMPLOYEE ATTRITION (Απώλεια των εργαζομένων)

2.1 Απώλεια των εργαζομένων

Η απώλεια εργαζομένων είναι η διαδικασία μέσω της οποίας οι εργαζόμενοι εγκαταλείπουν μια εταιρεία για οποιαδήποτε αιτία (εκούσια ή ακούσια), όπως συνταξιοδότηση, απόλυση, θάνατος ή παραίτηση. Το ποσοστό φθοράς υπολογίζεται διαιρώντας το ποσοστό με το οποίο οι εργαζόμενοι εγκαταλείπουν μια επιχείρηση ,με τον αριθμό εργαζομένων σε μια συγκεκριμένη χρονική περίοδο.

Τα δύο βασικά ζητήματα που απαιτούν σημαντική εξέταση είναι η αποτροπή της ενίσχυσης άλλων εταιρειών από ταλαντούχους εργαζόμενους, οι οποίοι αποχωρούν, και η μείωση των εξόδων απώλειας. Δεδομένου ότι δαπανώνται σημαντικά ποσά για την πρόσληψη και την εκπαίδευση των νέων εργαζομένων, το κόστος εναλλαγής προσωπικού αυξάνει σημαντικά τα έξοδα της εκάστοτε εταιρείας. Όταν ένας υπάλληλος αποχωρεί από έναν οργανισμό, αφήνει πίσω του σημαντικές γνώσεις για την επιχείρηση, τους πελάτες της και άλλες ευαίσθητες πληροφορίες. Οι πελάτες δένονται με τους εργαζόμενους μιας εταιρείας και δημιουργούνται δεσμοί μεταξύ τους, οπότε όταν ένας υπάλληλος αποχωρήσει αυτές οι συνδέσεις διακόπτονται με αποτέλεσμα να χαθούν πιθανοί πελάτες. Μία επιχείρηση για να πετύχει χρειάζεται πολύ χρόνο, και η εκπαίδευση νέων εργαζομένων απαιτεί εξίσου σημαντικό χρόνο και χρήμα. Επομένως, καταλήγουμε ότι εάν ένας εργαζόμενος αποχωρήσει από την εταιρεία, η εταιρεία θα υποστεί άμεση ζημία.



Εικόνα 1. Τύποι αποχώρησης των εργαζομένων

Υπάρχουν 5 τύποι αποχώρησης των εργαζομένων και θα αναλυθούν παρακάτω:

- Συνταξιοδότηση

Εάν μόνο δύο ή τρεις εργαζόμενοι αποσύρθηκαν από μία εταιρεία σε μία χρονιά, στατιστικά μιλώντας, αυτή η ομάδα εργαζομένων είναι πολύ μικρή για να συμπεριληφθεί στην απώλεια. Ωστόσο, μπορεί να προκύψει ζήτημα εάν ένα σημαντικό μέρος της ομάδας αποχωρήσει ταυτόχρονα.

- Εκούσια αποχώρηση

Οι εργαζόμενοι που επιλέγουν να εγκαταλείψουν οικειοθελώς τις θέσεις τους αντιπροσωπεύουν το πιο διαδεδομένο είδος απώλειας. Η εκούσια αποχώρηση μπορεί να συμβεί για διάφορους λόγους, οι περισσότεροι από τους οποίους βρίσκονται υπό τον έλεγχο της εταιρείας. Εργαζόμενοι με ταλέντο και υψηλή αξία θα πρέπει να αποφεύγεται να αποχωρήσουν, καθώς μπορεί να οδηγήσει σταδιακά στην μείωση της παραγωγικότητας.

- Ακούσια αποχώρηση

Σε αυτή την περίπτωση, η επιχείρηση και όχι ο εργαζόμενος είναι αυτή που ξεκινά τη διαδικασία της αποχώρησης. Για παράδειγμα, ο εργαζόμενος μπορεί να έχει εμπλακεί σε ανάρμοστη συμπεριφορά στο χώρο εργασίας, η οποία είναι μια κοινή αιτία ακούσιας απώλειας. Η αποχώρηση θα μπορούσε επίσης να οφείλεται σε διαρθρωτικά ζητήματα. Ένα κύμα ακούσιας φθοράς ακολουθεί, για παράδειγμα, ύστερα από συγχωνεύσεις και εξαγορές.

- Εσωτερική αποχώρηση

Οι εργαζόμενοι σε αυτήν την κατάσταση εγκαταλείπουν ένα τμήμα για να εργαστούν σε άλλο της ίδιας εταιρείας. Η εσωτερική αποχώρηση μπορεί να είναι επωφελής σε ορισμένες περιπτώσεις, καθώς κατευθύνει το ταλέντο σε πιο επικερδείς τομείς. Επιπλέον, εγγυάται μεγαλύτερη προσαρμογή εργασίας-εργαζομένου. Ωστόσο, αξίζει να εξεταστεί εάν σε ένα συγκεκριμένο τμήμα υπήρχε υψηλό ποσοστό αποχωρήσεων σε ένα μόνο έτος, αν χρειάζεται κάτι ακόμα η δουλειά ή αν η διοίκηση του τμήματος στερείται ικανοτήτων κλπ.

- Δημογραφική αποχώρηση

Για τις προοδευτικές επιχειρήσεις που προσπαθούν να δημιουργήσουν έναν χώρο εργασίας ίσων ευκαιριών, αυτό είναι μια μεγάλη πρόκληση. Η δημογραφική αποχώρηση εμφανίζεται όταν μια συγκεκριμένη ομάδα εργαζομένων, όπως γυναίκες, εθνικές μειονότητες, άτομα με αναπηρία, βετεράνοι ή ανώτεροι επαγγελματίες, εγκαταλείπουν τον οργανισμό σε μεγάλους αριθμούς. Προκειμένου να βρεθεί ο κύριος λόγος της απώλειας βάση δημογραφικών στοιχείων, προτού επηρεαστεί η κουλτούρα του χώρου εργασίας, πρέπει να ξεκινήσουν αμέσως έρευνες για τους υπαλλήλους.

2.2 Παράγοντες που επηρεάζουν την οικειοθελή αποχώρηση

Οι λόγοι που μπορεί κάποιος να επιλέξει να αποχωρήσει από τον χώρο εργασίας του είναι ποικίλοι. Αρχικά σημαντικό ρόλο λαμβάνουν οι προσωπικές ανάγκες του εργαζόμενου, άλλες οικογενειακές ή προσωπικές υποχρεώσεις όπως η οικογένεια, το μέρος της εργασίας και άλλες ασχολίες που δεν συμβαδίζουν με την συγκεκριμένη εργασία. Ο εργαζόμενος με το βάθος του χρόνου λογικό είναι να θέλει να αναπτυχθεί επαγγελματικά γι' αυτό τον λόγο αν δεν του δοθεί αυτή η ευκαιρία στον συγκεκριμένο χώρο εργασίας θα αναγκαστεί να την αναζητήσει κάπου αλλού. Επιπλέον οι συνθήκες εργασίας που επικρατούν πρέπει να είναι οι ιδανικές και ως προς την σχέση των εργαζόμενων μεταξύ τους αλλά και με την ηγεσία της εταιρείας καθώς και να υπάρχουν τα απαραίτητα μέσα για να εργαστούν όλοι ομαλά. Τέλος και ίσως το σημαντικότερο από όλα είναι ο εργαζόμενος να ταιριάζει με την δουλειά του και κατ' επέκταση και η δουλειά να ταιριάζει σε αυτόν ώστε να μπορεί να αποδώσει σε αυτά που θα του ζητηθούν.



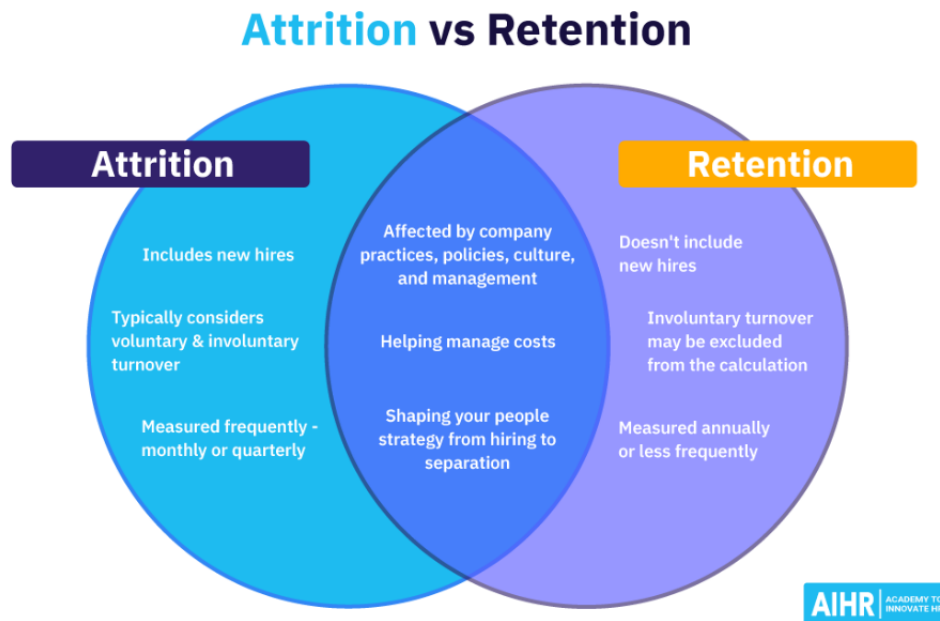
Εικόνα 2. Παράγοντες που επηρεάζουν την αποχώρηση των εργαζομένων

2.3 Αποχώρηση Προσωπικού VS Εναλλαγή Προσωπικού

Σύμφωνα με όσα αναφέρθηκαν παραπάνω για την αποχώρηση του προσωπικού θα γίνει σύγκριση με την εναλλαγή του προσωπικού. Ως εναλλαγή προσωπικού εννοούμε την αποχώρηση ενός εργαζομένου, οικειοθελή ή ακούσια, με στόχο την πλήρωση μιας ανοιχτής θέσης. Η αποχώρηση και η εναλλαγή μπορεί να προκύψουν και οι δύο από τους ίδιους παράγοντες, ωστόσο η αποχώρηση θεωρείται συνήθως αρνητική και ως κόστος για τους εργοδότες.

Ένας υπάλληλος που εγκαταλείπει οικειοθελώς την εταιρεία μπορεί να υποκινήθηκε από μια καλύτερη προσφορά εργασίας, έλλειψη περιθωρίου για πρόοδο ή εξέλιξη στην υπάρχουσα θέση, κακή αποζημίωση, εχθρικό περιβάλλον εργασίας, την αίσθηση ότι δεν ταίριαζε με την εταιρική κουλτούρα ή ποικιλία άλλων παραγόντων.

Η λέξη "ακούσια εναλλαγή" αναφέρεται στην απόλυση ή την απόλυση εργαζομένου (με σκοπό την αντικατάστασή του) για λόγους όπως η υποβαθμισμένη εργασία, η προβληματική συμπεριφορά και άλλα.



Εικόνα 3. Αποχώρηση Προσωπικού VS Εναλλαγή προσωπικού

2.4 Μείωση του ποσοστού αποχώρησης εργαζομένων



Εικόνα 4. Τρόποι για την μείωση της αποχώρησης των εργαζομένων

Η πρόληψη για την μείωση του ποσοστού απώλειας των εργαζομένων πρέπει να γίνεται εξαρχής κατά την διαδικασία την πρόσληψης τους. Είναι πρόπων να γίνονται κατανοητά και από τις δύο πλευρές τι ζητάνε και τι μπορούν να καλύψουν.

Πολλές επιχειρήσεις σήμερα είναι επιτυχημένες επειδή κατανοούν πως να διατηρήσουν το προσωπικό τους. Για να πάρουν έμπνευση οι υπάλληλοι και να μείνουν στην εταιρεία όσο το δυνατό περισσότερο χρονικό διάστημα, έχει βγεί το συμπέρασμα ότι λειτουργούν καλύτερα οι άτυπες συνδέσεις από ότι οι ιεραρχίες στους οργανισμούς για την επίλυση ζητημάτων. Παρακάτω παρουσιάζεται μια σειρά από βέλτιστες πρακτικές:

- ο Αρχικά να προσληφθεί το κατάλληλο προσωπικό.
- ο Να δίνετε στους εργαζόμενους η ευκαιρία να ολοκληρώνουν εργασίες με σκοπό την ενδυνάμωσή τους.
- ο Να τους παρέχεται η βοήθεια προκειμένου να καταλάβουν ότι είναι το πολυτιμότερο περιουσιακό στοιχείο της εταιρείας.
- ο Να υπάρχει πίστη στους εργαζόμενους της εταιρείας, εμπιστοσύνη και ο απαραίτητος σεβασμός.
- ο Να μοιραστείτε μαζί τους πληροφορίες και γνώσεις πάνω στον τομέα τους με σκοπό την άμεση βελτίωσή τους.
- ο Να δίνετε συνεχή σχόλια για την απόδοσή τους είτε θετικά είτε αρνητικά.
- ο Εκτιμήστε και αναγνωρίστε τα επιτεύγματά τους και τις επιτυχίες τους.
- ο Δημιουργήστε ένα περιβάλλον όπου οι εργαζόμενοι θέλουν να εργαστούν και να διασκεδάσουν.

Προκειμένου να διατηρηθούν οι εργαζόμενοι ,το τμήμα ανθρωπίνου δυναμικού μια εταιρείας θα πρέπει να δώσει προτεραιότητα στην ανάπτυξη ταλέντων, θα πρέπει να προσφέρει ευκαιρίες ανάπτυξης στους εργαζόμενους της ώστε ο υπάλληλος να ξέρει ότι υπάρχει το περιθώριο εξέλιξης του στην ίδια την εταιρεία. Το θετικό εργασιακό περιβάλλον, οι ευκαιρίες σταδιοδρομίας και η οικοδόμηση της αυτοεκτίμησης των εργαζομένων είναι από τα σημαντικότερα κίνητρα για να παραμείνει ένας εργαζόμενος. Η προσφορά μπόνους με βάση την απόδοση μπορεί να προσφέρει ένα επιπλέον κίνητρο.

Η παροχή προγραμμάτων επαγγελματικής κατάρτισης είναι ένα επιπλέον όφελος του εργαζομένου και της εταιρείας αντίστοιχα καθώς υπάρχει προσωπική ανάπτυξη , παραπάνω αποτελεσματικότητα και αυξάνεται ο χρόνος παραμονής στην εταιρεία. Ο εργαζόμενος οφείλει να μπορεί να εκφράσει την γνώμη του σχετικά με τις συνθήκες εργασίας και αυτές να λαμβάνονται υπόψιν από την διοίκηση. Να υπάρχει ανταγωνιστικό πακέτο γενικότερα σε σχέση με τις άλλες εταιρείες και ως προς την επαγγελματική ανάπτυξη αλλά και τις οικονομικές απολαβές που θα λαμβάνουν οι εργαζόμενοι.

Κεφάλαιο 3: Παρουσίαση προηγούμενων μελετών employee attrition (αποχώρησης εργαζομένων)

Η έρευνα Predicting employee attrition using machine learning (Alduayj & Rajpoot,2018), είχε σαν κύριο στόχο την πρόβλεψη της αποχώρηση των εργαζομένων με την χρήση μοντέλων μηχανών μάθησης (Machine Learning). Για την υλοποίηση της έρευνας πραγματοποιήθηκαν τρία διαφορετικά πειράματα. Αρχικά, χρησιμοποιήθηκαν τρεις διαφορετικοί αλγόριθμοι (SVM, kNN, Random Forest) στο πρωτότυπο, μη ισορροπημένο dataset. Στη συνέχεια εφαρμόστηκε ο αλγόριθμος ADASYN (προσεγγιστική συνθετική δειγματοληψία) για την επίλυση της ανισορροπίας των κλάσεων, και τέλος εφαρμόστηκε τυχαία δειγματοληψία των δεδομένων όπου αποτελούνται από ίσο υποσύνολο της κάθε τάξης. Και τα τρία πειράματα αξιολογήθηκαν με γνωστά μέτρα απόδοσης (Accuracy, Precision, Recall, F1-score). Το μεγαλύτερο F1-score επετεύχθη στο δεύτερο πείραμα, που έγινε χρήση του αλγορίθμου ADASYN.

Η έρευνα Analyzing Employee Attrition using decision tree algorithms (Alao D. και Adeyemo A. B., 2013) επιχείρησε να προβλέψει την αποχώρηση των εργαζομένων σύμφωνα με τα χαρακτηριστικά τους με την χρήση των μοντέλων των δέντρων αποφάσεων και των συνόλων κανόνων. Χρησιμοποιήθηκε αρχείο 309 υπαλλήλων ενός από τα ανώτατα ιδρύματα στην Νιγηρία που εργάστηκαν και αποχώρησαν από το ίδρυμα μεταξύ 1978 έως το 2006. Η αποχώρηση των εργαζομένων προβλέφθηκε με βάση των δημογραφικών τους χαρακτηριστικών και μεταβλητών που σχετίζονται με την εργασία του καθενός. Πιο συγκεκριμένα το φύλο του κάθε εργαζομένου, την καταγωγή του, την διάρκεια εργασίας στην εταιρεία, ο τίτλος της κατηγορίας όπου εργάζεται, ο μισθός του για έναν χρόνο και οι αιτίες που ο εργαζόμενος αποχώρησε.

Η μέθοδος που χρησιμοποιήθηκε ήταν η ταξινόμηση (classification), μια κλασική τεχνική εξόρυξης δεδομένων, η οποία βασίζεται στην μηχανική μάθηση και ταξινομεί κάθε στοιχείο σε μια κλάση δεδομένων. Επιλέχθηκαν οι αλγόριθμοι J48 και δέντρα αποφάσεων REPTree και SimpleCart. Ο ταξινομητής J48 είχε καλύτερη απόδοση από τους άλλους, όμως το Boost SeeTree είχαν την μεγαλύτερη. Τα βασικά συμπεράσματα ήταν ότι οι εργαζόμενοι που εργάζονταν για λιγότερο από είκοσι χρόνια στην εταιρεία και λάμβαναν μικρότερους μισθούς, αλλά και αυτοί που ήταν παραπάνω στην εταιρεία αλλά εξακολουθούσαν να βρίσκονται σε χαμηλές θέσεις έτειναν να αποχωρήσουν ευκολότερα.

Η έρευνα με τίτλο Employee Attrition prediction using neural network cross validation method (Dutta & Kumar, 2020) είχε σαν στόχο της να προβλέψει τις πιθανότητες αποχώρησης των εργαζομένων εκ των προτέρων, με την εφαρμογή ενός αυτοματοποιημένου εργαλείου βασισμένο σε νευρωνικά δίκτυα.

Και σε αυτήν την μεθοδολογία χρησιμοποιήθηκαν 1470 δείγματα που αποτελούν το σύνολο των δεδομένων και τα χαρακτηριστικά που θα χρησιμοποιηθούν σαν κριτήρια για την αποχώρηση των εργαζομένων είναι ποικίλα όπου κάποια από αυτά είναι η ηλικία, ο μισθός, η εκπαίδευση, ικανοποίηση, ποσοστό αύξησης μισθού, επαγγελματικό επίπεδο, ώρες εργασίας κλπ. Το νευρωνικό δίκτυο δημιουργείται λαμβάνοντας υπόψιν τρία επίπεδα που αποτελούνται από 32,16,1 κόμβους αντίστοιχα. Για κάθε επίπεδο χρησιμοποιήθηκαν 30 epochs και το μέγεθος της παρτίδας είναι 10. Στην συνέχεια πραγματοποιείται με 10πλάσια μέθοδο διασταυρούμενης επικύρωσης για τον υπολογισμό

των δεξιοτήτων του μοντέλου. Στην συγκεκριμένη μέθοδο σε 10 ομάδες χωρίζεται το σύνολο των δεδομένων, όπου σε κάθε επανάληψη 1 ομάδα θεωρείται δεδομένο δοκιμής και οι άλλες εννιά ως δεδομένα προπόνησης.

Η αξιολόγηση των μοντέλων γίνεται με βάση την απόδοση (accuracy) όπου προσδιορίζει την αναλογία αληθών προβλέψεων και το μέσο τετράγωνο του σφάλματος(MSE) όπου υπολογίζει την απόλυτη διαφορά μεταξύ της πρόβλεψης και της πραγματικής παρατήρησης των δειγμάτων.

Τα αποτελέσματα της έρευνας είναι τα παρακάτω:

Performance Measure Metrics	SVM	K-NN	Naïve Bayes	Decision Tree	Adaboost	Random Forest
Accuracy	85.6%	83.74%	83.74%	78.4%	85.8%	85.8%
MSE	0.144	0.1626	0.16	0.216	0.14	0.142

Performance Measure Metrics	Accuracy	MSE
Neural Network with Cross-Validation	87.01%	0.1299

Το νευρωνικό δίκτυο με 10πλάσια διασταυρούμενη επικύρωση επιτυγχάνει μέγιστη απόδοση 87,01%. Η δεκαπλάσια διασταυρούμενη επικύρωση και ένα νευρωνικό δίκτυο είναι μία ενιαία πλατφόρμα που μπορεί να καθορίσει εκ των προτέρων την αποχώρηση των εργαζομένων. Η προτεινόμενη μέθοδος επιτυγχάνει πολλά υποσχόμενα αποτελέσματα με ακρίβεια 87,01% και βελτιστοποιημένη τιμή MSE 0,1299.

Κεφάλαιο 4: Ποιοτική συγκριτική ανάλυση της QCA (Qualitative Comparative Analysis)

4.1 Ποιοτική προσέγγιση της QCA

Ο Αμερικάνος κοινωνικός επιστήμονας Charles Ragin πρότεινε την ποιοτική συγκριτική ανάλυση (QCA) στα τέλη της δεκαετίας του 1980. Η τεχνική αυτή αποτελεί έναν τρόπο αντιπαράθεσης μεταξύ θεωρίας και δεδομένων, μία “συνθετική στρατηγική” όπου λειτουργεί με τα καλύτερα στοιχεία μίας ποιοτικής, δηλαδή στην σε βάθος εξέταση συγκεκριμένων περιπτώσεων, και μίας ποσοτικής προσέγγισης, όπου εστιάζουν στις σχέσεις μεταξύ των μεταβλητών σε πολλές περιπτώσεις, σύμφωνα με τον Ragin.

Η QCA βασίζεται σε μια λογική η οποία προέρχεται από την θεωρία συνόλων και όχι την λογική της συνδιακύμανσης που έχουν οι στατιστικές μέθοδοι, και γι’ αυτό το λόγο εφαρμόζει άλγεβρα Boole και ασαφή άλγεβρα. Σε γενικές γραμμές, ο κύριος στόχος της QCA είναι να εξηγήσει πώς παράγεται ένα συγκεκριμένο αποτέλεσμα.

Η QCA μπορεί να χρησιμοποιηθεί για πέντε διαφορετικούς σκοπούς:

1. Με την δημιουργία ενός πίνακα αλήθειας όπου με αυτόν τον τρόπο συνοψίζει τα δεδομένα και τα παρουσιάζει συνδυαστικά.
2. Με την αναγνώριση αντιφάσεων μεταξύ των δεδομένων (contradiction) ελέγχει την συνοχή τους.
3. Μπορεί να κάνει τον απαραίτητο έλεγχο προκειμένου να επιβεβαιώσει ή να διαψεύσει μία θεωρία.
4. Μπορεί να χρησιμοποιηθεί για τον έλεγχο νέων ιδεών και προτάσεων.
5. Με την εφαρμογή της μεθόδου μπορούμε να συγκρίνουμε νέες και παλιές υποθέσεις και να βγάλουμε νέα συμπεράσματα.

Χαρακτηριστικό της QCA είναι ότι αναπτύσσει την έννοια της σύνθετης αιτιότητας. Η έννοια αυτή περιλαμβάνει: 1) την πολλαπλή συνδυαστική αιτιότητα (multiple Conjunctural Causation), 2) την ισοδυναμία των λύσεων (Equifinality) και 3) τη μη συμμετρική αιτιότητα (Asymmetric Causation) (Ragin 2000, 2008, Rihoux, 2003, 2006, Wageman, 2009).

1. Conjunctural Causation (Συνδυαστική Αιτιότητα): αναφέρεται στην υπόθεση ότι είναι συνήθως ένας συνδυασμός παραγόντων και όχι ένας παράγοντας που οδηγεί σε αποτέλεσμα. Σε αυτήν τη διαμόρφωση, όχι μόνο η παρουσία αλλά και η απουσία ενός συγκεκριμένου παράγοντα εκτιμάται ότι επηρεάζει το αποτέλεσμα και επομένως μετριέται.

2. Equifinality (Ισοδυναμία): αναφέρεται στην υπόθεση ότι πολλοί συνδυασμοί παραγόντων παράγουν το ίδιο αποτέλεσμα.

3. Asymmetric Causation (Ασύμμετρη αιτιότητα): Η βασική ιδέα είναι ότι το σύνολο των αιτιωδών συνθηκών που οδηγούν στην παρουσία του αποτελέσματος μπορεί συχνά να είναι διαφορετικό από το σύνολο των συνθηκών που οδηγούν στην απουσία του αποτελέσματος. Έτσι, θα πρέπει να γίνεται χωριστός έλεγχος των αιτιωδών συνθηκών για την παρουσία και την απουσία του αποτελέσματος που εξετάζεται κάθε φορά. Η παρουσία ή η απουσία λοιπόν των συνθηκών μπορεί να έχει σημαντικά διαφορετικούς ρόλους στην εμφάνιση ή όχι ενός αποτελέσματος.

4.2 Ποσοτική προσέγγιση της QCA

Η μέθοδος χαρακτηρίζεται και από ποσοτικές προσεγγίσεις όπου σύμφωνα με τον Rihoux το 2006 κάποια από τα πλεονεκτήματα είναι τα εξής:

- i. Η ικανότητα να παράγει γενικεύσεις. Υπάρχει η δυνατότητα ο αναλυτής να εξάγει γενικούς κανόνες και συμπεράσματα λόγω ότι αναλύει μεγάλο όγκο δεδομένων. Με αυτόν τον τρόπο διευκολύνεται για μελλοντική χρήση.
- ii. Η επανάληψη των αποτελεσμάτων. Η κάθε περίπτωση πρέπει να μειώνεται σε μία σειρά μεταβλητών, σύμφωνα με την άλγεβρα Boolean. Η επανάληψη επιτρέπεται λόγω της αναλυτικής αυτής προσέγγισης έχοντας την δυνατότητα αλλού αναλυτές να διασταυρώσουν τα τελικά αποτελέσματα.
- iii. Ο αναλυτής έχει την δυνατότητα να εντοπίζει φειδωλούς συνδυασμούς, οι οποίοι μπορούν να ερμηνευτούν με τις λιγότερες αιτιώδεις συνθήκες.

4.3 Κατηγορίες της QCA

Crisp-set QCA (Cs-QCA)

Η αρχική έκδοση του QCA ονομάζεται Crisp-set QCA ((Ragin, 1987). Η τεχνική αυτή χρησιμοποιεί την δυαδική λογική, δηλαδή δηλώνει την παρουσία ή την απουσία ενός χαρακτηριστικού ή ενός αποτελέσματος με τις τιμές 0 (απουσία) ή 1 (παρουσία). Όταν το αποτέλεσμα είναι παρόν αναζητούνται οι παράγοντες που εμφανίζονται σε όλα τα στοιχεία και αποτελούν τις αναγκαίες συνθήκες (necessary conditions). Ο σκοπός της Cs-QCA είναι η ανάλυση του αποτελέσματος σε όρους αναγκαίων και ικανών συνθηκών.

Multi-Value QCA (mv-QCA)

Η multi-value QCA (mvQCA) στοχεύει να αντιμετωπίσει τον ίδιο βασικό περιορισμό του csQCA με το fsQCA, δηλαδή την υποχρέωση χρήσης μόνο διχοτόμων συνθηκών παρουσίας / απουσίας (Cronqvist και Berg-Schlusser2009). Η multi-value QCA (mvQCA) επιτρέπει τη συμπερίληψη πολυκατηγορικών συνθηκών στην ανάλυση. Στην ιδανική περίπτωση θα πρέπει να χρησιμοποιούνται τρεις ή τέσσερις εκτιμημένες συνθήκες. Αυτή η προσέγγιση επιτρέπει περισσότερη διαφοροποίηση και περισσότερες ομοιογενείς ομαδοποιήσεις συνθηκών από το crispset QCA, αλλά το πιο σημαντικό είναι ότι επιτρέπει την καλύτερη κατανόηση των πολυκατηγορικών παραγόντων.

Fuzzy-set QCA (fs-QCA)

Μια σημαντική κριτική σχετικά με την csQCA είναι η δυαδική της μορφή, δηλαδή η απαίτηση από τους ερευνητές να αξιολογήσουν τα αποτελέσματα ως αληθής ή ψευδής. Για τον λόγο αυτό ο Ragin δημιούργησε την fsQCA επιτρέποντας τους να καθορίσουν την αξία των συνθηκών όχι μόνο σε διχοτομικό πεδίο αλλά σε διάστημα μεταξύ 0 και 1. Αυτή η προσέγγιση επιτρέπει περισσότερη διαφοροποίηση και ακριβέστερη περιγραφή. Στα επόμενα κεφάλαια, γίνεται μια αναλυτική παρουσίαση τόσο των ασαφών συνόλων όσο και της ποιοτικής συγκριτικής ανάλυσης με τη χρήση ασαφών συνόλων (fs-QCA) .

4.4 Διαδικασία της QCA

Και για τις τρεις τεχνικές που προαναφέρθηκαν η διαδικασία που γίνεται στην QCA είναι παρόμοια. Αρχικά θα δημιουργηθεί ένας πίνακας δεδομένων, όπου η κάθε περίπτωση αποτελείται από ένα συγκεκριμένο συνδυασμό συνθηκών και ένα αποτέλεσμα, και το λογισμικό στην συνέχεια θα παράγει έναν πίνακα αλήθειας όπου συνδυάζονται τα δεδομένα με το αποτέλεσμα. Στην συνέχεια με την Boolean ελαχιστοποίηση θα μειωθούν οι περιγραφές του πίνακα αλήθειας ώστε να φτάσουμε στην ελάχιστη αιτιώδη συνταγή όπου θα ερμηνευτεί θεωρητικά και θα σχολιάσουμε τα αποτελέσματα.

Ύστερα από την περιγραφή των τριών στρατηγικών καταλήξαμε ότι αποτελούνται από δυο κοινά σημεία:

- Αντιμετωπίζουν τις περιπτώσεις των συνόλων σαν να ήταν συνδυασμοί των διαφόρων συνθηκών.
- Προσπαθούν να εξηγήσουν το αποτέλεσμα σε όρους ικανών και αναγκαίων συνθηκών.

Συμπερασματικά θα μπορούσε κανείς να ισχυριστεί ότι το διχοτομικό QCA (csQCA) είναι κατασκευασμένο για περιστάσεις με μικρό μέγεθος δείγματος (30-40 περιπτώσεις). Αντίθετα τα ασαφή σύνολα (fuzzy sets) μπορούν να πραγματοποιήσουν μεγαλύτερο αριθμό δειγμάτων. Η mnQCA είναι μεταξύ της csQCA και των ασαφών συνόλων και είναι πιο αξιόπιστη σε πλήθος μεσαίου αριθμού δείγματος.

Κεφάλαιο 5: Η διαδικασία της fs/QCA

5.1 Εισαγωγή

Για να εξακριβωθεί ο βαθμός στον οποίο υπάρχουν ή απουσιάζουν ορισμένοι παράγοντες ή συνδυασμοί παραγόντων η fs/QCA χρησιμοποιεί άλγεβρα Boole. Αυτή η συγκεκριμένη μεθοδολογία επιδιώκει να εντοπίσει τις περίπλοκες αιτιώδεις σχέσεις μεταξύ των μεταβλητών που έχουν σαν αποτέλεσμα συγκεκριμένο αποτέλεσμα. Οι αιτίες του φαινομένου αναφέρονται ως συνθήκες (conditions) ενώ το ίδιο φαινόμενο αναφέρεται ως αποτέλεσμα (outcome).

5.2 Calibration – Βαθμονόμηση

Τα δεδομένα για τη μελέτη των κοινωνικών επιστημών συχνά χωρίζονται σε τέσσερις κατηγορίες: ονομαστικά, τακτικά (τάξεως), μεσοδιαστήματα και αναλογίες. Μπορούν σχεδόν όλα να βαθμονομηθούν σε δυαδικά ευκρινή σύνολα, σχεδόν όλα σε ευκρινή σύνολα πολλαπλών τιμών, και μόνο τα επίπεδα μέτρησης "διαστήματος" και "αναλογίας" μπορούν να χρησιμοποιηθούν για ασαφή σύνολα. Οι όροι "βαθμονόμηση" και "μέτρηση", ωστόσο, δεν σημαίνουν απαραίτητα το ίδιο πράγμα.

Η βιβλιογραφία χρησιμοποιεί συνήθως δύο τεχνικές βαθμονόμησης. Τα τρία ποιοτικά σημεία αποκοπής που συνθέτουν ασαφή σύνολα στη άμεση μέθοδο (direct method) είναι τα κατώφλια για την πλήρη συμμετοχή, το κατώφλι για τη συνολική μη συμμετοχή και το καθορισμένο σημείο διέλευσης (Woodside & Zhang, 2013). Η έμμεση μέθοδος, αντίθετα, χρησιμοποιεί τεχνικές παλινδρόμησης για την αξιολόγηση του βαθμού συμμετοχής και βασίζεται σε ένα σχήμα αποκοπής έξι σημείων. Οι δύο τεχνικές παράγουν επίσης ακριβείς βαθμονομήσεις των επιπέδων συμμετοχής με βάση είτε ομάδες ποιότητας (έμμεση προσέγγιση) είτε σημεία ποιότητας (άμεση μέθοδος).

Το όριο για πλήρη συμμετοχή – full membership (ασαφής βαθμολογία = 0,95), το όριο για πλήρη μη συμμετοχή – full nonmembership (ασαφής βαθμολογία = 0,05) και το σημείο διασταύρωσης / σημείο μέγιστης ασάφειας – crossover point (βαθμολογία ασαφής = 0,5) είναι τα τρία ποιοτικά σημεία αποκοπής (Qualitative Anchors) που δομούν ένα ασαφές σύνολο. Και τα τρία σημεία αναφοράς έχουν χρησιμοποιηθεί για τη μετατροπή της μεταβλητής αναλογίας ή διαστήματος σε ασαφείς βαθμολογίες με βάση τις αποδόσεις στην πλήρη συμμετοχή (Ragin, 2007; Fiss, 2011; Woodside, 2013).

Σε αντίθεση με την άμεση μέθοδο, η έμμεση μέθοδος βασίζεται σε μεγάλες ομάδες περιπτώσεων ανάλογα με το επίπεδο εμπλοκής τους στο σύνολο που πραγματεύεται η έρευνα. Ουσιαστικά, ο ερευνητής κάνει μια προκαταρκτική ταξινόμηση περιπτώσεων σε διάφορα επίπεδα συμμετοχής, εκχωρεί τα προκαταρκτικά αποτελέσματα συμμετοχής σε αυτά τα διάφορα επίπεδα και στη συνέχεια ενισχύει αυτές τις βαθμολογίες μελών χρησιμοποιώντας την κλίμακα διαβαθμισμένων δεδομένων. Το πρώτο και πιο κρίσιμο βήμα στην έμμεση μέθοδο είναι η ποιοτική ταξινόμηση των εργαζομένων με βάση το υποτιθέμενο επίπεδο εμπλοκής τους στην προβλεπόμενη ομάδα. Οι ποιοτικές κατηγορίες μπορεί να είναι ad hoc και υπόκεινται σε αλλαγές.

Ωστόσο, θα πρέπει να βασίζονται όσο το δυνατόν περισσότερο στη θεμελιώδη γνώση και την υπάρχουσα θεωρία.

Υπάρχουν πολλές τεχνικές για τη διαίρεση των δεδομένων σε ασαφή σύνολα, όπως ο

ασαφής διαχωρισμός συνόλων με συνεχείς τιμές, σε τέσσερα, έξι και οκτώ σημεία στο [0,1]. Οι έξι θεμελιώδεις κατηγορίες ποιότητας που χρησιμοποιούνται σε έναν διαχωρισμό των 6 είναι οι ακόλουθες:

Οι ενδείξεις συχνά περιλαμβάνουν:

- α) πλήρης συμμετοχή (βαθμός συμμετοχής=1)·
- (β) ως επί το πλείστον εντός (βαθμός εμπλοκής = 0,8), αλλά όχι εξ ολοκλήρου·
- (γ) περισσότερο εντός παρά όχι (επίπεδο συμμετοχής = 0,6).
- δ) περισσότερο έξω από το εσωτερικό (ποσοστό συμμετοχής = 0,4).
- (ε) κυρίως εκτός αλλά όχι εξ ολοκλήρου (επίπεδο συμμετοχής 0,2%) και
- (στ) συνολική μη συμμετοχή (επίπεδο συμμετοχής 0).

Συνάρτηση συμμετοχής (membership function)

Η λογιστική συνάρτηση, μια παραδοσιακή σιγμοειδής συνάρτηση, χρησιμοποιείται από την τεχνική fs/QCA για τον προσδιορισμό του επιπέδου συμμετοχής κάθε περίπτωσης.

$$\mu_A(x) = \frac{\exp(z_A(x))}{1 + \exp(z_A(x))}$$

$$z_A(x) = \begin{cases} 3(x - x_2)/(x_2 - x_1) & x < x_2 \\ 3(x - x_2)/(x_3 - x_2) & x \geq x_2 \end{cases}$$

Τα τρία σημεία αποκοπής στα οποία αναφέρεται η άμεση τεχνική είναι εύκολα ορατά κατά τη σχεδίαση της σιγμοειδούς λογιστικής συνάρτησης. Η μέγιστη ασάφεια της καμπύλης (σημείο διασταύρωσης), όπου αλλάζει η κλίση, βρίσκεται στον αριθμό 0,5.

5.3 Πίνακας Αλήθειας

Οι πίνακες αλήθειας, οι οποίοι βρίσκονται στον πυρήνα όλων των QCA και, κατά συνέπεια, της ανάλυσης fs/QCA, βοηθούν στη λογική ταξινόμηση των δεδομένων που συλλέγονται για τις περιπτώσεις.

Οι πίνακες αλήθειας, σύμφωνα με τους Schneider και Grofman (2006), παρουσιάζουν σε βάθος τις ομοιότητες και τις διαφορές μεταξύ των παραδειγμάτων που περιλαμβάνονται στη μελέτη. Υποδηλώνουν επίσης αντικρουόμενες γραμμές ή καταστάσεις με τα ίδια σύνολα συνθηκών αλλά

διαφορετικά αποτελέσματα. Εμφανίζουν τον βαθμό ποικιλομορφίας δεδομένων, δηλαδή ποια λογικά δεδομένα υπάρχουν.

Όταν αναλυθούν σωστά, αυτά τα δεδομένα μπορούν να βοηθήσουν τον ερευνητή να επαναπροσδιορίσει τις περιπτώσεις και τις συνθήκες που ελήφθησαν υπόψη για την ανάλυση καθώς και τις πολυάριθμες συνδέσεις μεταξύ των συνθηκών και του αποτελέσματος.

Οι πίνακες αλήθειας δεν πρέπει να συγχέονται με κανονικούς πίνακες που χρησιμοποιούνται για τη μετάδοση δεδομένων. Ενώ μια σειρά ενός πίνακα αλήθειας έχει πληροφορίες για έναν από τους λογικά εφικτούς συνδυασμούς αιτιακών συνθηκών, μια σειρά ενός κανονικού πίνακα δεδομένων περιέχει πληροφορίες για μια περίπτωση σε κάθε σειρά.

Τα δύο στάδια της ανάλυσης πίνακα αλήθειας, το κύριο συστατικό της ανάλυσης δεδομένων με την fsQCA, είναι:

1. Η δημιουργία πίνακα αλήθειας από ασαφή σύνολα
2. Η μείωση του αριθμού των πιθανών διαμορφώσεων του πίνακα αιτιολογικής συνταγής.

Οι περιπτώσεις σε ένα σύνολο δεδομένων, που παρέχονται ως διάφορες διαμορφώσεις περιστάσεων, περιγράφονται με διακριτικό τρόπο με αυτή τη μέθοδο. Ολόκληρος ο πίνακας εμφανίζει όλους τους πιθανούς συνδυασμούς των αιτιακών συνθηκών, με κάθε γραμμή να αντιπροσωπεύει έναν συγκεκριμένο συνδυασμό ιδιοτήτων, τις αιτιακές συνθήκες. Σύμφωνα με τις τιμές αυτών των χαρακτηριστικών (αιτιώδεις συνθήκες), οι περιπτώσεις στο σύνολο δεδομένων ταξινομούνται σε σειρές του πίνακα αλήθειας, με ορισμένες γραμμές να περιέχουν πολλαπλές περιπτώσεις, άλλες μόνο λίγες και άλλες να μην περιέχουν καθόλου περιπτώσεις, εάν δεν υπάρχει εμπειρικό παράδειγμα του συγκεκριμένου συνόλου χαρακτηριστικών που συνδέεται με τη δεδομένη γραμμή (Ragin & Rihoux, 2004; Fiss, 2011).

Οι περιπτώσεις ταξινομούνται στις γραμμές του πίνακα αλήθειας σε αναλύσεις που χρησιμοποιούν ευκρινείς πίνακες αλήθειας (crisp sets) σύμφωνα με ορισμένους συνδυασμούς των βαθμολογιών τους για παρουσία ή απουσία στις διάφορες συνθήκες. Κάθε περίπτωση εισάγεται σε ξεχωριστή γραμμή και κάθε γραμμή περιέχει διαφορετικό υποσύνολο των περιπτώσεων της μελέτης. Ένας πίνακας αλήθειας περιλαμβάνει συνολικά 2^k γραμμές, όπου k είναι ο αριθμός των αιτιακών περιστάσεων που περιλαμβάνει το μοντέλο. (Grofman & Schneider, 2009; Ragin, 2009).

Πριν την ελαχιστοποίηση των γραμμών του πίνακα αλήθειας πρέπει να καθοριστούν δύο κατώφλια για τον προσδιορισμό της συχνότητας για τον ελάχιστο αριθμό περιπτώσεων που απαιτούνται για να εξεταστεί μια γραμμή του πίνακα αλήθειας και το ελάχιστο επίπεδο συνέπειας που πρέπει να έχει ένας αιτιώδης συνδυασμός. (Ragin, 2009; Fiss, 2011).

5.4 Κατώφλι Συχνότητας

Το πρώτο βήμα είναι να επιλεχθεί ένα κατώφλι συχνότητας, το οποίο καθορίζει πόσες περιπτώσεις πρέπει να περιλαμβάνει μια σειρά του πίνακα προκειμένου να ληφθούν υπόψη στην ανάλυση για την αξιολόγηση των ασαφειών στις ασαφείς συνδέσεις υποσυνόλου. Με βάση μια ιδιότητα συνδυασμών ασαφών συνόλων, η οποία δηλώνει ότι κάθε περίπτωση μπορεί να έχει μόνο στο ένα μέλος βαθμολογία μεγαλύτερη από 0,5 εντός των ορίων εύλογων δυνατών συνδυασμών που δημιουργούνται

από ένα συγκεκριμένο σύνολο συνθηκών, η κατανομή των περιπτώσεων στις σειρές του πίνακα βασίζεται σε αυτήν την ιδιότητα (Ragin, 2009).

Είναι δυνατό να προσδιοριστεί ποιο σύνολο συνθηκών απεικονίζει καλύτερα κάθε περίπτωση στο σύνολο δεδομένων μετά την ανάθεση βαθμολογιών συμμετοχής-μέλους στις περιπτώσεις για κάθε ασαφές σύνολο. Θα υπάρξει μόνο μια συγκεκριμένη κατασκευή αιτιακών συνθηκών στις οποίες εμπίπτει κάθε περίπτωση. Μια περίπτωση βρίσκεται περισσότερο εντός του εν λόγω αιτιολογικού συνδυασμού παρά εκτός αυτής, εάν έχει βαθμολογία συμμετοχής άνω του 0,5. Ο Ragin (2008) ισχυρίζεται ότι είναι σημαντικό να βεβαιωθείτε ότι η ανάλυση του πίνακα αλήθειας θα καλύπτει τουλάχιστον το 75% έως 80% των περιπτώσεων από ολόκληρο το σύνολο δεδομένων κατά τον υπολογισμό του ορίου συχνότητας. Ο αλγόριθμος fs/QCA αντιμετωπίζει τους συνδυασμούς συνθηκών κάτω από το ελάχιστο όριο συχνότητας που έχει καθορίσει ως λογικά υπόλοιπα. Τα λογικά υπόλοιπα είναι πιθανές διαμορφώσεις αιτιακών συνθηκών που απουσιάζουν από τη συλλογή δεδομένων υπό εξέταση των εμπειρικών παραδειγμάτων. Γενικά, η επιλογή του ορίου του αριθμού των περιπτώσεων από τον ερευνητή θα πρέπει να λαμβάνει υπόψη τόσο τους στόχους της μελέτης όσο και τη φύση των δεδομένων.

Ο συνολικός αριθμός των περιπτώσεων που περιλαμβάνονται στη μελέτη, ο αριθμός των συνθηκών, ο βαθμός ακρίβειας βαθμονόμησης του ασαφούς συνόλου, το μέγεθος της μέτρησης και του σφάλματος εκχώρησης, κ.λπ. είναι σημαντικές εκτιμήσεις (Ragin, 2005, 2009).

5.5 Κατώφλι Συνέπειας

Σε μια ανάλυση που αναζητά ικανές συνθήκες, μας ενδιαφέρει ο βαθμός όπου συγκεκριμένοι αιτιακοί παράγοντες (συνθήκες) ή διαμορφώσεις (configurations) παραγόντων, οι οποίοι είναι υποσύνολα των αποτελεσμάτων. Βαθμολογία συμφωνίας για μια αιτιακή συνθήκη είναι ένα μέτρο αυτής της σχέσης υποσυνόλου. Ο βαθμός στον οποίο η βαθμολογία συμμετοχής στα αποτελέσματα είναι μεγαλύτερη ή ίση με Βαθμολογίες συμμετοχής σε αιτιολογικούς συνδυασμούς. Οι βαθμολογίες συνέπειας κυμαίνονται από 0 έως 1, με το 0 να είναι καμία σχέση υποσυνόλου, η βαθμολογία 1 υποδηλώνει μια τέλεια σχέση υποσυνόλου.

Η σημαντική απόφαση που πρέπει να ληφθεί είναι ποια βαθμολογία συνέπειας χρησιμοποιείται ως διαχωρισμοί για τον προσδιορισμό των αιτιολογικών συνδυασμών για να θεωρείται ένα λογικό υποσύνολο του αποτελέσματος. Συνδυασμός με βαθμολογία συνέπειας ακριβώς ή πάνω από την αποκοπή ορίζεται ως ένα ασαφές υποσύνολο των αποτελεσμάτων, και μπαίνει 1 στη στήλη του αποτελέσματος. Αντίθετα, αυτά κάτω από την τιμή αποκοπής δεν είναι ένα ασαφές υποσύνολο του αποτελέσματος, και κωδικοποιείται ως 0 (Schneider, 2010). Το ελάχιστο επίπεδο σύνολο-θεωρητικής συνέπειας μπορεί να ληφθεί με τον ορισμό του ορίου συνέπειας σε τουλάχιστον 0,75 ή περισσότερο (Ragin, 2005; Ragin et al, 2008).

5.6 Ελαχιστοποίηση Πίνακα

Προκειμένου να απλοποιηθούν αυτές οι βασικές δηλώσεις και να παρέχονται πιο κατανοητές λύσεις, η fsQCA χρησιμοποιεί Boolean ελαχιστοποίηση. Για να απλοποιηθούν οι σειρές του πίνακα αλήθειας, χρησιμοποιείται μια τεχνική που χρησιμοποιεί άλγεβρα Boole για να τις αναλύσει σε μικρότερους, πιο διαχειρίσιμους συνδυασμούς.

Μια προσέγγιση για την ελαχιστοποίηση του πίνακα αλήθειας που χρησιμοποιείται συχνά βασίζεται σε μια αντίθετη ανάλυση των αιτιακών καταστάσεων, την Ανάλυση Αντιπαραδειγμάτων (Ragin, 2005, 2008). Στην πραγματικότητα, τα πακέτα λογισμικού εκτελούν αυτήν την ελαχιστοποίηση αυτόματα (για παράδειγμα, το εργαλείο ανάλυσης πίνακα αλήθειας στο fs/QCA).

5.7 Είδη λύσεων

Τρεις εναλλακτικές απαντήσεις μπορούν να ληφθούν μέσω η ανάλυσης του πίνακα αληθείας ανάλογα με τον τρόπο προσέγγισης των απλουστευτικών παραδοχών στο fs/QCA:

Το περίπλοκο, το συνετό και το ενδιάμεσο

Οι αιτιώδεις συνταγές αυτών των λύσεων μπορεί να διαφέρουν κάπως η μία από την άλλη, αλλά είναι πάντα ισοδύναμες ως προς τη λογική και την αλήθεια και δεν περιέχουν ποτέ πληροφορίες που είναι αντιφατικές (Ragin & Sonnett, 2005)

Complex

Η ανάλυση δεν μπορεί να κάνει απλοποιητικές υποθέσεις κατά τη χρήση της σύνθετης λύσης. Όταν υπάρχουν πολλοί αιτιολογικοί παράγοντες, καθίσταται δύσκολο να μειωθεί η πολυπλοκότητα της λύσης, η οποία συνεισφέρει το λιγότερο στην ερμηνεία των δεδομένων. Φυσικά, όταν δεν υπάρχουν πολλές αιτιολογικές συνθήκες, συνιστάται αυτή η λύση.

Parsimonious

Η Parsimonious Solution μειώνει τους όρους της λύσης (αιτιακές συνταγές) ώστε να περιέχει όσο το δυνατόν λιγότερες συνθήκες και ενσωματώνει όλες τις υποθέσεις απλοποίησης, ανεξάρτητα από το αν βασίζονται σε απλά ή σύνθετα αντιπαραδείγματα (εύκολα / δύσκολα αντιπαραδείγματα). Εκτός από οποιαδήποτε άλλη λύση στον πίνακα αλήθειας, οι όροι που περιλαμβάνονται σε αυτήν τη λύση δεν μπορούν να αφαιρεθούν.

Οι αυτόματες αποφάσεις σχετικά με εύλογα υπόλοιπα λαμβάνονται χωρίς να λαμβάνονται υπόψη θεωρητικές ή εμπειρικές γνώσεις σχετικά με την εγκυρότητα μιας απλουστευτικής υπόθεσης.

Ωστόσο, δεδομένης αυτής της ισχυρής υπόθεσης, η φειδωλή προσέγγιση θα πρέπει να εφαρμόζεται μόνο εάν υποστηρίζονται πλήρως οι υποθέσεις σχετικά με τις κατάλληλες ισορροπίες.

Intermediate

Η Intermediate solution για την μείωση της πολυπλοκότητας αποτελείται μόνο από τις απλουστευτικές υποθέσεις. Θα μπορούσαμε να πούμε ότι είναι η σύνθετη λύση ύστερα από αφαίρεση των συνθηκών που δεν συμφωνούν με τις θεμελιώδεις θεωρητικές και εμπειρικές γνώσεις. Η ενδιάμεση λύση συνιστάται ως το κύριο μέσο για την επεξήγηση των αποτελεσμάτων εφόσον γίνεται ορθή χρήση των απλουστευτικών υποθέσεων.

Οι τύποι λύσεων χρησιμοποιούνται για την παρουσίαση των αποτελεσμάτων από το fs/QCA (τύποι διαλύματος). Μια μορφή λύσης χρησιμοποιεί ονόματα μεταβλητών που βασίζονται σε γράμματα για να υποδείξουν την επίδραση και τις αντίστοιχες αιτιολογικές συνθήκες.

Boolean H (+), KAI (*), and OXI () τελεστές.

Ενώ οι διακριτές αιτιακές διαδρομές που προσδιορίζονται λόγω της ύπαρξης ισοδύναμων λύσεων συνδέονται μεταξύ τους με τη λογική H (+), οι πολυάριθμες αιτιακές συνθήκες που περιλαμβάνουν αιτιακές διαδρομές συνδέονται με το λογικό KAI (*). Η λογική σύνδεση μεταξύ αιτιολογικών συνθηκών (ή συνδυασμών συνθηκών) και του αποτελέσματος συμβολίζεται με ένα βέλος που δείχνει προς τα δεξιά (Schneider & Grofman, 2006; Grofman & Schneider, 2009).

Ο στόχος της παρουσίασης των ευρημάτων με αυτόν τον τρόπο είναι να επιστήσουμε την προσοχή στους αστερισμούς των περιστάσεων που συνέβαλαν στο συμπέρασμα.

Θα πρέπει να σημειωθεί ότι καμία από τις συνθήκες A, B ή Γ δεν είναι κατάλληλη σε αυτή τη λύση. Αν το A ήταν επαρκής συνθήκη, τότε η υπονοούμενη ύπαρξη του Y δεν θα απαιτούσε την προσθήκη του B. Ομοίως, εφόσον κανένα από τα συστατικά του στοιχείου δεν είναι απαραίτητη προϋπόθεση, δεν υπάρχει καθόλου προϋπόθεση.

5.8 Απλουστευτικές Υποθέσεις (Simplifying Assumptions)

Ο Ragin προτείνει την χρήση απλουστευτικών υποθέσεων καθώς είναι δύσκολο να βρεθούν ζεύγη διαμορφώσεων που διαφέρουν σε μια μόνο συνθήκη και ταυτόχρονα συμφωνούν. Οι απλουστευτικές υποθέσεις βοηθούν στην ανάλυση των αντιπαραδειγμάτων.

Ανάλογα με την κλίμακα της χρήσης της εξωτερικής θεωρητικής γνώσης από τον ερευνητή, η χρήση απλοποιητικών υποθέσεων στη διαδικασία ελαχιστοποίησης μπορεί να είναι περισσότερο ή λιγότερο ενοχλητική. Οι έννοιες των εύκολων και δύσκολων αντιπαραδειγμάτων αναπτύχθηκαν από τους Ragin και Sonnett (2005) για να βοηθήσουν στον προσδιορισμό του πότε μια απλουστευτική υπόθεση είναι λογική.

Εύκολα Αντιπαραδείγματα (Easy Counterfactuals)

Τα εύκολα αντιπαραδείγματα είναι σενάρια όταν μια πλεονάζουσα αιτιολογική συνθήκη εισάγεται σε μια συλλογή αιτιακών συνθηκών που, αν ήταν μόνες, θα παρήγαγαν ήδη το επιθυμητό αποτέλεσμα.

Για παράδειγμα, εάν κάποιος έχει στοιχεία που δείχνουν ότι το αποτέλεσμα του συνδυασμού ($A*B*\sim C$) είναι πιθανό να είναι παρόν. Παρόλο που ο Πίνακας Αλήθειας δεν περιέχει στοιχεία που να αποδεικνύουν ότι ο συνδυασμός ($A*B*C$) θα παράγει επίσης το αποτέλεσμα, θεωρητικές ή εμπειρικές πληροφορίες συνδέουν την ύπαρξη του C - αντί της απουσίας του - με το αποτέλεσμα. Μια ανάλυση απλών αντιπαραδειγμάτων δείχνει ότι σε αυτήν την περίπτωση, και οι δύο σχέσεις ($A*B*C$) και ($A*B*\sim C$) θα οδηγήσουν στο αποτέλεσμα. Ως αποτέλεσμα, η φράση μπορεί να μειωθεί σε ($A*B$) επειδή το C μπορεί να υπάρχει ή να απουσιάζει χωρίς να αλλάξει το αποτέλεσμα.

Δύσκολα Αντιπαραδείγματα (Difficult Counterfactuals)

Αντίθετα, τα δύσκολα αντιπαραδείγματα είναι σενάρια στα οποία μια συνθήκη εξαλείφεται από μια αλυσίδα γεγονότων που προκαλεί αποτέλεσμα με την αιτιολογία ότι είναι περιττή.

Για παράδειγμα, εάν κάποιος έχει στοιχεία που δείχνουν ο συνδυασμός ($A*B*C$), θα φτάσει στο αποτέλεσμα αλλά δεν υπάρχει ένδειξη για τον συνδυασμό ($A*B*\sim C$). Αυτή η συνθήκη είναι αντίθετη από αυτήν που αναφέρθηκε παραπάνω.

Ο ερευνητής ανησυχεί εάν η αφαίρεση μιας αιτιολογικής συνθήκης θα οδηγήσει σε κάποια διαφοροποίηση σε μια ανάλυση προκλητικών αντιπαραδειγμάτων. Λόγω της έλλειψης επαρκών εμπειρικών παραδειγμάτων ($A*B*\sim C$), είναι πολύ πιο δύσκολο να αξιολογηθεί εάν το C είναι πραγματικά μια περιττή συνθήκη που μπορεί να παραλειφθεί, απλοποιώντας τη λύση μόνο στο ($A*B$). Οι εξωτερικές θεωρητικές πληροφορίες συσχετίζουν την ύπαρξη, όχι την απουσία, του C με το αποτέλεσμα.

5.9 Μέτρα προσαρμογής

Οι συνθήκες (ή οι συνδυασμοί συνθηκών) που μοιράζονται όλες οι περιπτώσεις σε ένα σύνολο δεδομένων είναι εξαιρετικά ασυνήθιστες. Θα υπάρξουν τουλάχιστον μερικές περιπτώσεις που διαφέρουν από τον κανόνα. Εξαιτίας αυτού, είναι σημαντικό για τον ερευνητή να μπορεί να κρίνει εάν τα παραδείγματα σε ένα σύνολο δεδομένων ταιριάζουν με μια σχέση που θεωρείται ότι είναι ικανή ή απαραίτητη για το αποτέλεσμα. Επιπλέον, λόγω της αιτιακής πολυπλοκότητας και της παρουσίας ισοδύναμων λύσεων, ένα αποτέλεσμα μπορεί συχνά να εξηγηθεί από πολλούς συνδυασμούς συνθηκών (αιτιώδη μονοπάτια).

Με βάση τα παρακάτω γίνεται η αξιολόγηση των λύσεων που προκύπτουν καθώς και κάθε διαφορετικό αιτιώδες μονοπάτι:

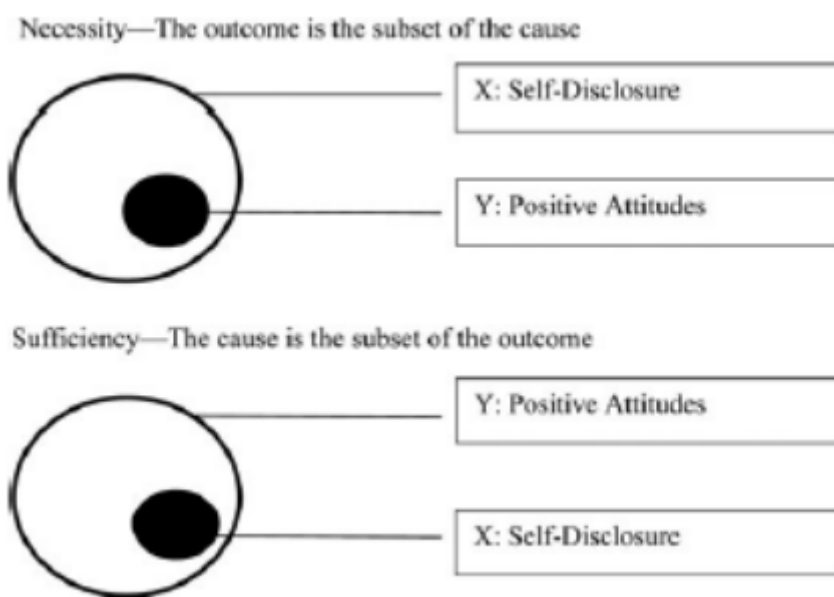
- Συνολοθεωρητική Συνέπεια (Consistency), η οποία δείχνει τον βαθμό στον οποίο έχει προσεγγιστεί μία σχέση υποσυνόλου.
- Συνολοθεωρητική Κάλυψη (Coverage), η οποία δείχνει την εμπειρική σημασία μιας συνεπούς σχέσης υποσυνόλου.

5.10 Αναγκαιότητα και Επάρκεια

Οι ασύμμετρες σχέσεις επάρκειας και αναγκαιότητας είναι δύο περαιτέρω ιδέες που μπορούν να εξεταστούν χρησιμοποιώντας fs/QCA.

Αν και η ύπαρξη της αιτίας δεν εγγυάται πάντα την ύπαρξη της συνέπειας, Αναγκαιότητα είναι όταν η αιτία (X) πρέπει να υπάρχει για να υπάρχει το αποτέλεσμα (Y). Αυτό το είδος ασύμμετρης σχέσης, σύμφωνα με τον Ragin (2000, 2008), αποκαλύπτει τότε το αποτέλεσμα είναι ένα υποσύνολο της αιτίας.

Όταν η αιτία (X) μπορεί να δημιουργήσει ένα αποτέλεσμα (Y) λέγεται ότι είναι Επαρκές, ενώ άλλες αιτίες θα μπορούσαν επίσης να παράγουν το ίδιο αποτέλεσμα. Σύμφωνα με τον Ragin (2000, 2008), η αιτία σε αυτό το είδος ασύμμετρης σχέσης είναι ένα υποσύνολο του αποτελέσματος. Αν υπάρχει X, πρέπει να υπάρχει και το Y.



Εικόνα 5:Αναγκαιότητα και Επάρκεια

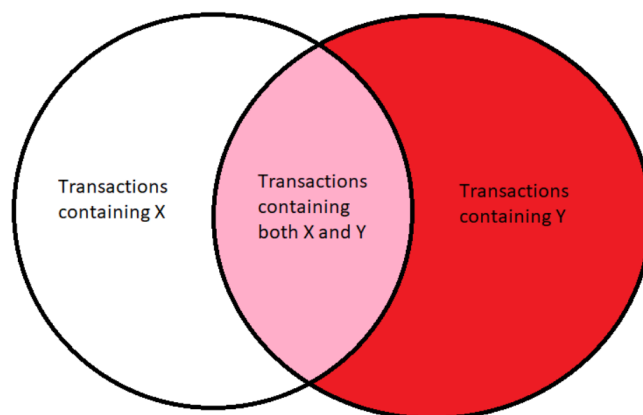
Κεφάλαιο 6: Association Rules

6.1 Θεωρία Κανόνων Συσχέτισης (Association Rules)

Στο πιο βασικό της επίπεδο, η εξόρυξη κανόνων συσχέτισης (association rules) χρησιμοποιεί μοντέλα μηχανικής μάθησης για να αναζητήσει βάσεις δεδομένων για μοτίβα ή συν-εμφανίσεις σε δεδομένα. Ανιχνεύει κοινές σχέσεις, οι οποίες είναι οι κανόνες συσχέτισης από μόνες τους. Ένα επακόλουθο και ένα προηγούμενο (αν) συνθέτουν έναν κανόνα συσχέτισης (τότε).

Ένας κανόνας συσχέτισης (Association Rule) έχει την μορφή $X \rightarrow Y$ όπου $X \subset M$, $Y \subset M$ και $X \cap Y = \emptyset$. Σύμφωνα με τα δικά μας δεδομένα το X είναι ένα υποσύνολο των αποτελεσμάτων της έρευνας και το Y είναι ένα άλλο υποσύνολο, το οποίο δεν έχει κανένα κοινό μέλος με το X . Ο κανόνας $X \rightarrow Y$ διατυπώνεται ως εξής: αν για κάποιον εργαζόμενο ισχύουν τα δεδομένα του X , τότε θα ισχύουν και τα δεδομένα του Y .

Ο κανόνας $X \rightarrow Y$ καθορίζεται για το πόσο ισχυρός είναι από δύο ποσοτικά μεγέθη, την υποστήριξη (support) και την εμπιστοσύνη (confidence). Η υποστήριξη είναι το ποσοστό των περιπτώσεων (επί του συνόλου των περιπτώσεων) που περιέχουν και το X και το Y , $\text{supp}(X \rightarrow Y) = P(X \cup Y)$. Η εμπιστοσύνη μας δείχνει την πιθανότητα εμφάνισης του Y όταν εμφανίζεται το X , $\text{conf}(X \rightarrow Y) = P(Y|X)$.



Εικόνα 6: Κανόνες συσχέτισης

Έστω ένα στοιχείο ή γεγονός, το $X = \{i_1, i_2, \dots, i_n\}$ αντιπροσωπεύει το σύνολο όλων των συμβάντων που σχετίζονται με αυτό το πρόβλημα. Το $T = \{t_1, t_2, \dots, t_n\}$ είναι ένα σύνολο συναλλαγών όπου κάθε συναλλαγή είναι ένα σύνολο στοιχείων. Ο κανόνας συσχέτισης $X \Rightarrow Y$ υποδεικνύει μια ορισμένη σχέση μεταξύ δύο συνόλων.

Η υποστήριξη διασφαλίζει ότι ένας επαρκής αριθμός περιπτώσεων εμπίπτει στην κάλυψη του κανόνα. Οι κανόνες που έχουν ελάχιστα στοιχεία πίσω τους μπορούν να θεωρηθούν ότι εκφράζουν ένα τυχαίο γεγονός. Επιπλέον, η εμπιστοσύνη είναι ένας μετρητής όπου μας δείχνει το πόσο η εμφάνιση του αριστερού μέρους του κανόνα ευνοεί την εμφάνιση του δεξιού μέρους. Επίσης παρόλο τα μεγάλα ποσοστά που μπορεί να έχουν οι δείκτες της υποστήριξης και της εμπιστοσύνης οι κανόνες που προκύπτουν δεν δείχνουν πάντα μία πραγματική σχέση.

Κατά τη σύγκριση της εμπιστοσύνης με την προβλεπόμενη εμπιστοσύνη ή τον αριθμό των φορών που αναμένεται να είναι αληθής μια δήλωση αν-τότε, μπορεί να χρησιμοποιηθεί ένα τρίτο στατιστικό που είναι γνωστό ως ανύψωση(Lift) . Σύμφωνα με το ποσοτικό μέτρο Lift βλέπουμε πόσο καλύτερη ή χειρότερη είναι η επίδοση ενός κανόνα σε σχέση με έναν τυχαίο κανόνα επιλογής. Επιπλέον για τον υπολογισμό του βαθμού ενδιαφέροντος ενός κανόνα χρησιμοποιούμε το στατιστικό μέτρο Lift:

$$Lift(X \rightarrow Y) = \frac{supp(X \rightarrow Y)}{supp(X) * supp(Y)}$$

Όπου αν η τιμή του lift=1 ,τότε δεν υπάρχει συσχέτιση μεταξύ του X και του Y και ο κανόνας εκφράζει ένα τυχαίο γεγονός. Αν η τιμή του lift>1 τότε υπάρχει θετική συσχέτιση μεταξύ του X και του Y και η εμφάνιση του X ενισχύει την εμφάνιση του Y. Κλείνοντας αν το lift<1 τότε τα X και Y είναι αρνητικά συσχετισμένα και η εμφάνιση του X μειώνει την εμφάνιση του Y.

Στο δικό μας παράδειγμα το μέτρο Lift τον παράγοντα αύξησης ή μείωσης της πιθανότητας αποχώρησης ενός εργαζομένου για τον λόγο Y, με δεδομένο ότι έχει αποχωρήσει για τον λόγο X.

Τα σύνολα δύο ή περισσότερων στοιχείων - χρησιμοποιούνται για τον υπολογισμό κανόνων συσχέτισης. Εάν αναλυθούν όλα τα πιθανά σύνολα στοιχείων για τη δημιουργία κανόνων, μπορεί να υπάρχουν τόσοι πολλοί κανόνες που να μην έχουν νόημα. Υπό το πρίσμα αυτό, οι κανόνες συσχέτισης αναπτύσσονται συχνά χρησιμοποιώντας κανόνες που αντιπροσωπεύονται καλά στα δεδομένα.

6.2 Αλγόριθμος Apriori

Ο αλγόριθμος Apriori (Agrawal & Srikant, 1994) είναι ο πρώτος αλγόριθμος για την εύρεση κανόνων συσχέτισης όπου έχουν υποστήριξη μεγαλύτερη ή ίση από ένα ορισμένο κατώφλι και ικανοποιούν την συνθήκη της εμπιστοσύνης. Ο Apriori χρησιμοποιεί προηγούμενη γνώση σχετικά με την αντιμονότονη ιδιότητα της υποστήριξης, όπου ορίζει ότι η υποστήριξη ενός στοιχειοσυνόλου είναι ίση ή μικρότερη από την υποστήριξη κάθε δυνατού υποσυνόλου του. $\forall X, Y: (X \subseteq Y) \Rightarrow supp(X) \geq supp(Y)$.

Ο αλγόριθμος Apriori περιλαμβάνει τα παρακάτω βήματα:

1. Ελέγχει την βάση δεδομένων και γίνεται καταμέτρηση του κάθε στοιχείου
2. Δημιουργεί ένα σύνολο (L_k όπου $k=1$) με τα στοιχεία(k) όπου η συχνότητα τους ισούται ή υπερβαίνει ένα ορισμένο κατώφλι.
3. Δημιουργεί ένα νέο σύνολο C_{k+1} όπου αποτελείται από την συνένωση του L_k με τον εαυτό του.
4. Το σύνολο C_{k+1} σύμφωνα με την αντιμονότονη ιδιότητα της υποστήριξης ελέγχεται για μη συχνά στοιχειοσύνολα, όπου όσα στοιχειοσύνολα βρεθούν διαγράφονται από το σύνολο C_{k+1} .
5. Πραγματοποιεί έλεγχο στην βάση δεδομένων και υπολογίζεται η υποστήριξη για τα μέλη του C_{k+1} . όσα έχουν υποστήριξη μικρότερη από την τιμή του κατωφλιού αφαιρούνται. Άρα προκύπτει το σύνολο των συχνών στοιχειοσυνόλων L_{k+1} .

Ύστερα από την εύρεση των συχνών στοιχειοσυνόλων δημιουργούμε τους κανόνες συσχέτισης σύμφωνα με την παρακάτω διαδικασία.

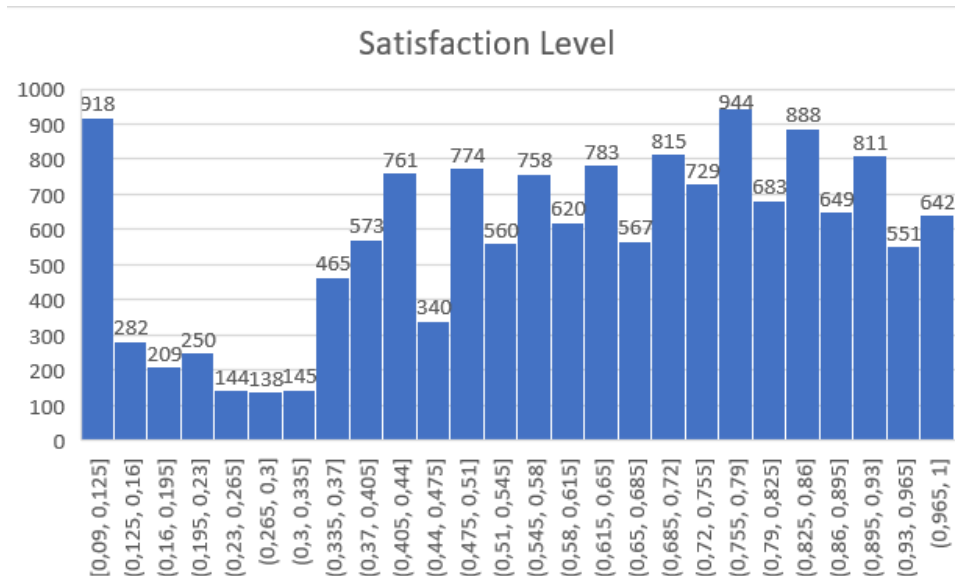
- Για κάθε συχνό στοιχειοσύνολο Z δημιουργούνται όλα τα μη κενά του υποσύνολα.
- Για κάθε μη κενό υποσύνολο X δημιουργείται κανόνας της μορφής $X \rightarrow Z-X$
- Κάθε κανόνας που δημιουργήθηκε ελέγχεται ως προς την τιμή της εμπιστοσύνης καθώς πρέπει να υπερβαίνει την τιμή του κατωφλίου, τότε ο κανόνας θεωρείται ισχυρός. Η εμπιστοσύνη υπολογίζεται ως εξής:

$$conf(X \rightarrow Y) = P(Y|X) = \frac{sup_count(X \cup Y)}{sup_count(X)} = \frac{sup_count(Z)}{sup_count(X)}$$

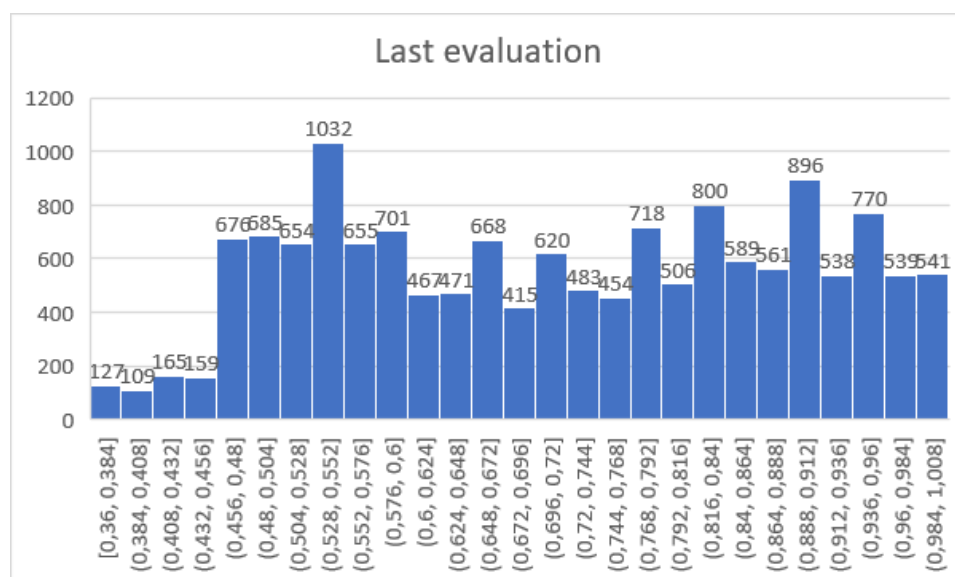
Κεφάλαιο 7: Στατιστική Ανάλυση Δεδομένων

7.1 Γραφήματα Κριτηρίων

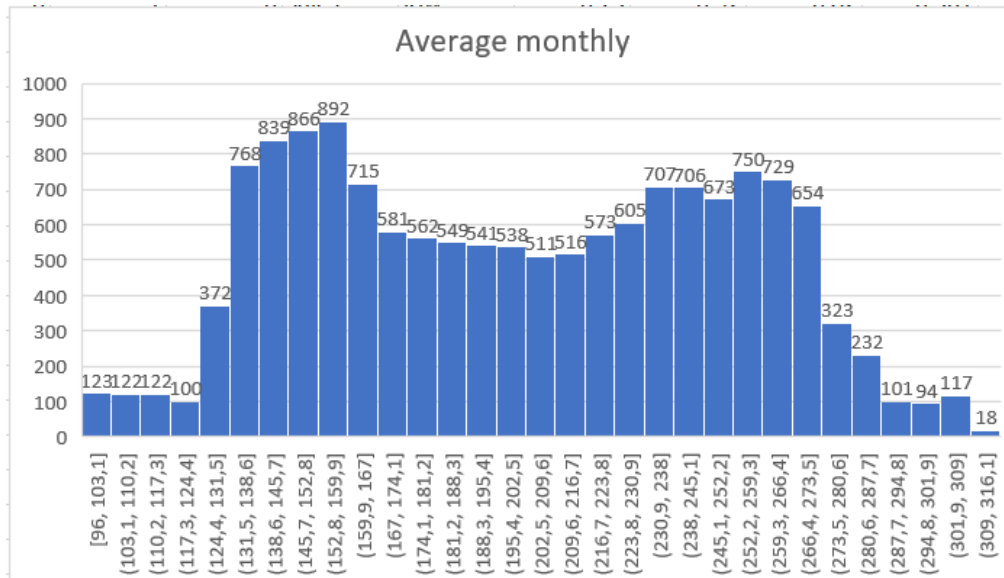
Στο dataset που αναλύουμε συμμετείχαν 14.999 εργαζόμενοι. Παρακάτω παρουσιάζονται όλα τα δεδομένα που προέκυψαν.



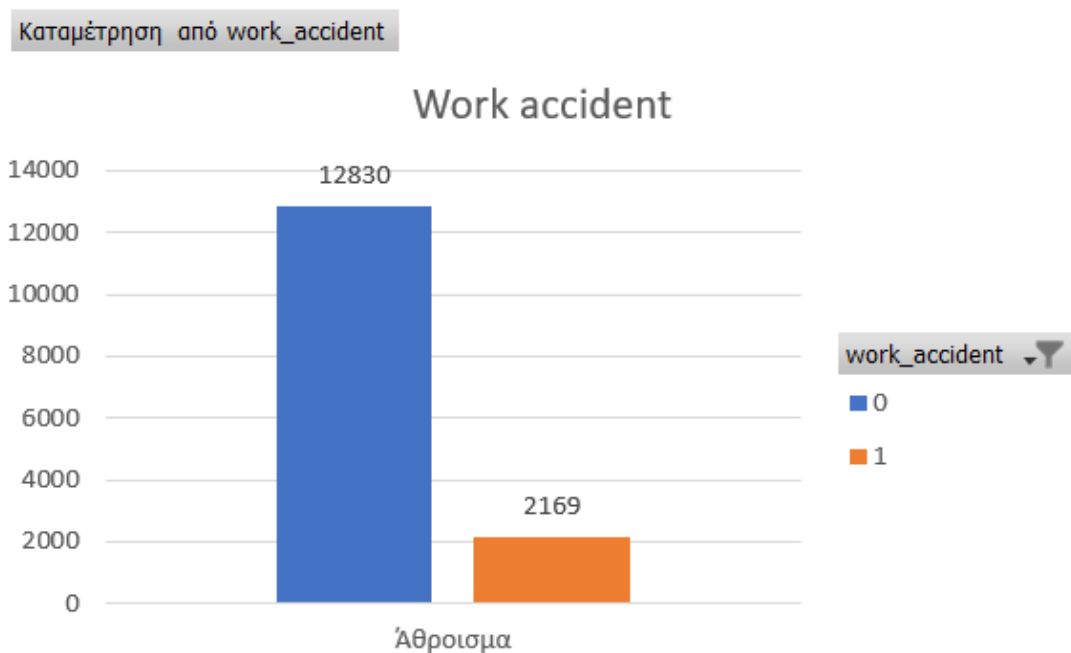
Το επίπεδο ικανοποίησης (satisfaction level) ,δείχνει πόσο ικανοποιημένοι είναι οι εργαζόμενοι στην εταιρεία. Πάνω από το 50% του δείγματος έχουν βαθμολογήσει πάνω από 0,5 άρα ανήκουν στους ικανοποιημένους εργαζόμενους.



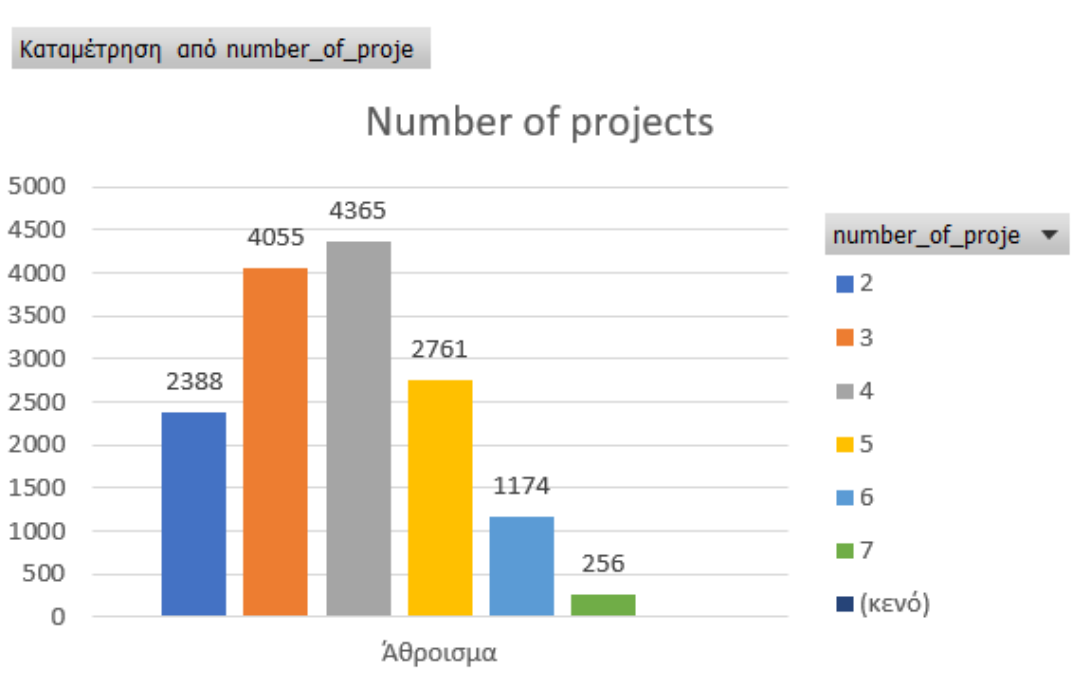
Το δεύτερο κριτήριο είναι το σκορ που επιτεύχθηκε στην τελευταία αξιολόγηση(last evaluation). Οι πλειοψηφία έχει υψηλό σκορ.



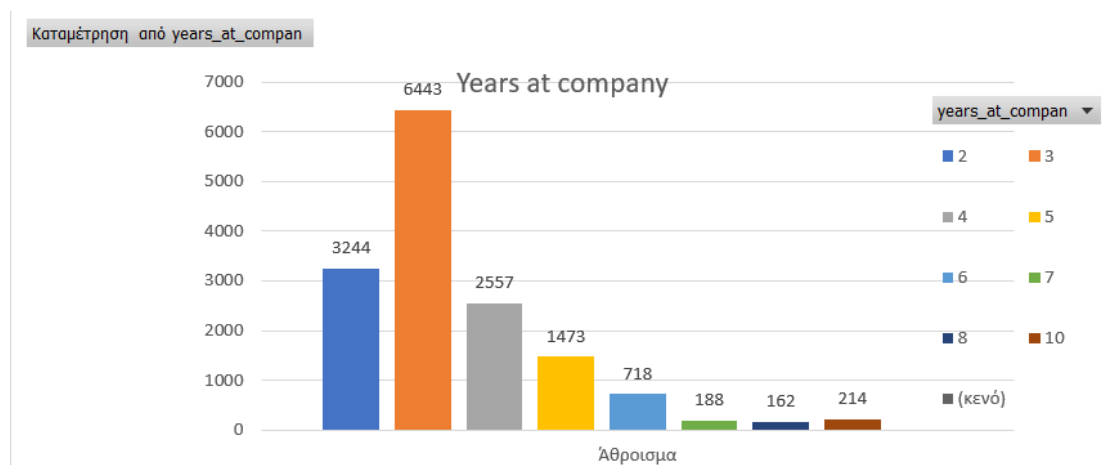
Ο μέσος όρος ωρών που έχουν εργαστεί μηνιαίως φαίνεται παρακάτω, η μικρότερη τιμή είναι 96 ώρες και η μεγαλύτερη φτάνει έως και 316 ώρες μηνιαίως.



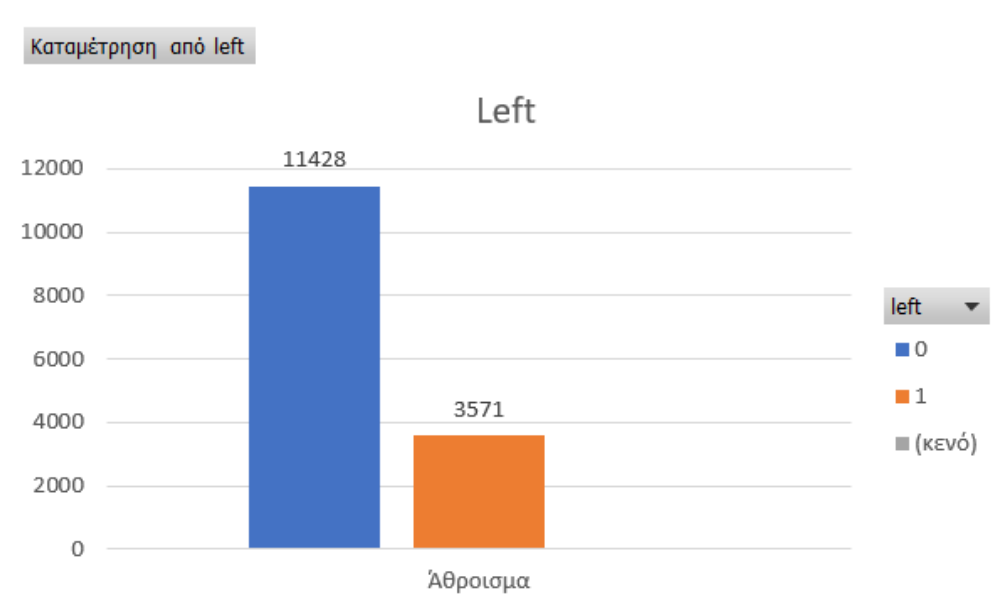
Παραπάνω βλέπουμε ότι από τους 14.999 εργαζόμενους οι 12.830 δεν είχαν εργατικό ατύχημα και οι 2.169 είχαν.



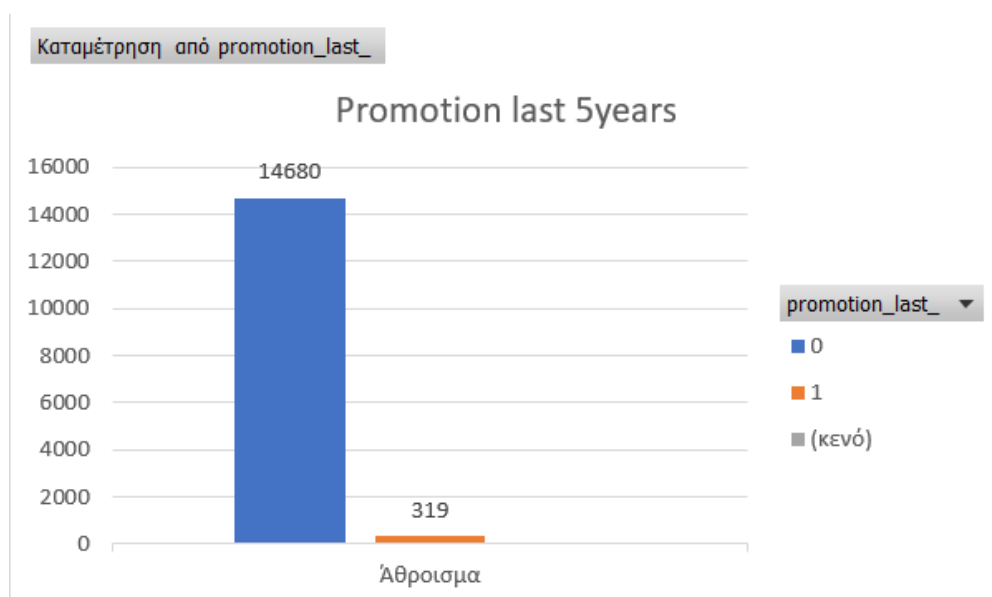
Επόμενο κριτήριο είναι ο αριθμός των προτζεκτ που είχαν ανατεθεί στον εργαζόμενο. Οι περισσότεροι εργαζόμενοι είχαν αναλάβει 3-4 προτζεκτ ενώ οι λιγότεροι είχαν αναλάβει επτά.



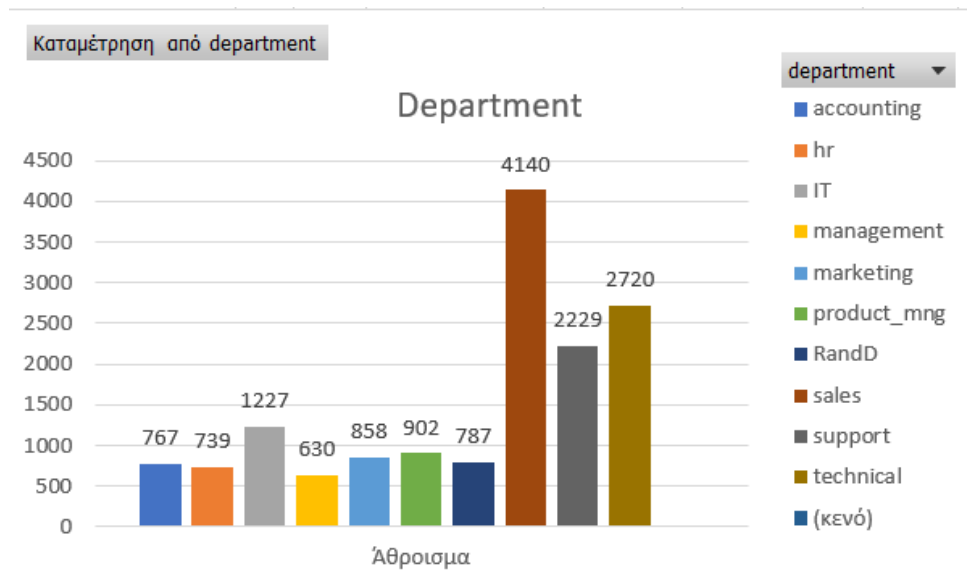
Αναλύθηκαν τα χρόνια που εργάζεται ο κάθε εργαζόμενος στην εταιρεία και βλέπουμε ότι όλοι οι εργαζόμενοι είναι τουλάχιστον 2 χρόνια στην εταιρεία και φτάνουν να παραμένουν έως 10 χρόνια.



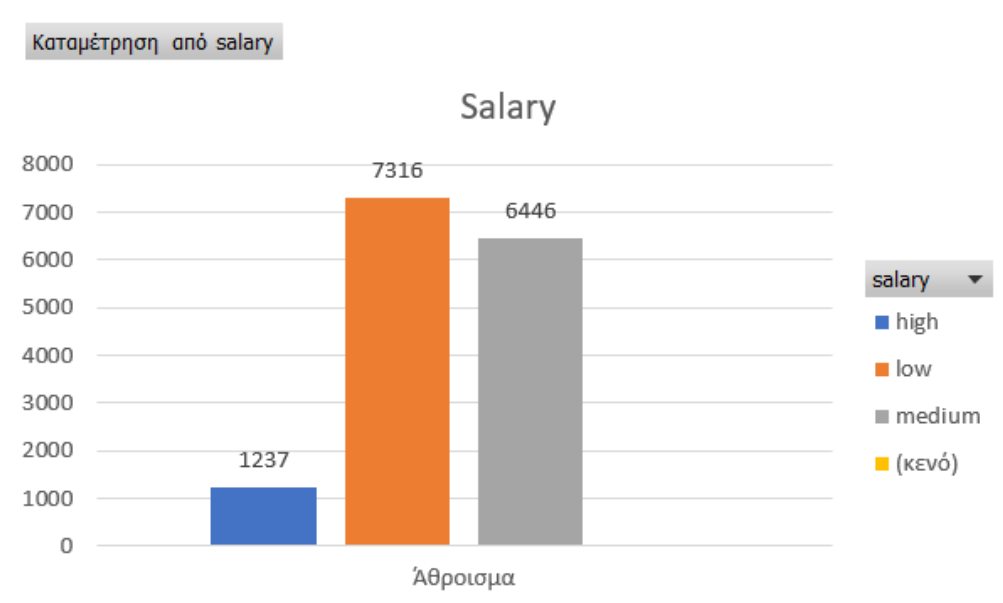
Αναλύθηκε αν έφυγαν ή έμειναν οι εργαζόμενοι από την εταιρεία. Βλέπουμε ότι 11.428 εργαζόμενοι παρέμειναν στην εταιρεία ενώ μόλις 3.571 είναι αυτοί που έφυγαν.



Επόμενο κριτήριο είναι αν είχαν πάρει προαγωγή τα τελευταία πέντε χρόνια. Οι πλειοψηφία των εργαζομένων 14.680 δεν είχε λάβει προαγωγή τα τελευταία 5 χρόνια ενώ μόνο 319 εργαζόμενοι είχαν λάβει.



Αναλύθηκε σε τι τμήμα ανήκει ο κάθε εργαζόμενος, οι περισσότεροι ανήκουν στο τμήμα των πωλήσεων της εταιρείας.



Αναλύθηκε ο μισθός των εργαζομένων σε 3 υποκατηγορίες ως υψηλός, χαμηλός και μέτριος. Οι περισσότεροι (7316) λαμβάνουν χαμηλό μισθό και 6446 λαμβάνουν μέτριο.

7.2 Πίνακας Συσχετίσεων

Το πόσο στενά συνδέονται δύο μεταβλητές η μία με την άλλη εκφράζεται μέσω της συσχέτισης. Ο συντελεστής συσχέτισης κυμαίνεται μεταξύ του -1 και του 1. Η συσχέτιση είναι θετική εάν οι δύο μεταβλητές κινούνται προς την ίδια κατεύθυνση, αρνητική εάν κινούνται προς αντίθετες κατευθύνσεις και εάν είναι μηδέν τότε δεν υπάρχει σχέση μεταξύ τους.

	satisfaction_le	last_evaluation	umber_of_proj	average_monthly	years_at_compa	work_accident	left	romotion_last
satisfaction_le	1							
last_evaluation	0,113298304	1						
number_of_proj	-0,15650969	0,484702602	1					
average_monthly	-0,066625465	0,461889192	0,574517992	1				
years_at_compa	-0,120573472	0,304127375	0,34754086	0,319791583	1			
work_accident	0,085297849	0,004417823	-0,01290934	-0,03308581	-0,0599938	1		
left	-0,43150036	0,00264396	0,027922023	0,087645438	0,35824065	-0,18751165	1	
promotion_last	0,030488809	-0,01141301	-0,00042195	-0,02432729	-0,0337739	0,03147585	-0,056936	1

Πίνακας 1: Πίνακας Συσχετίσεων

Σύμφωνα με τον πίνακα συσχετίσεων καταλήγουμε στο συμπέρασμα ότι οι μεταβλητές με την μεγαλύτερη συσχέτιση και που επηρεάζουν η μία την άλλη είναι:

- Ο αριθμός των προτζεκτ με τον βαθμό στην τελευταία αξιολόγηση
- Ο μέσος όρος ωρών που εργάζεται μηνιαίως με τον βαθμό στην τελευταία αξιολόγηση
- Ο μέσος όρος ωρών που εργάζεται μηνιαίως με τον αριθμό των προτζεκτ
- Τα χρόνια που εργάζεται στην εταιρεία με τον βαθμό στην τελευταία αξιολόγηση
- Τα χρόνια που εργάζεται στην εταιρεία με τον αριθμό των προτζεκτ
- Τα χρόνια που εργάζεται στην εταιρεία με τις ώρες που εργάζεται μηνιαίως
- Αν έφυγε από την εταιρεία με το πόσα χρόνια εργάζεται στην εταιρεία

Κεφάλαιο 8: Ανάλυση Αποτελεσμάτων fs/QCA

8.1 Εισαγωγή

Ακολουθεί μια μελέτη που στοχεύει στον εντοπισμό των αιτιολογικών παραγόντων (συνδυασμοί των πτυχών της Ικανοποίησης) που μπορούν να οδηγήσουν σε υψηλή Ολική Ικανοποίηση. Η μέθοδος της Ποιοτικής Συγκριτικής Ανάλυσης με χρήση ασαφών συνόλων (fs/QCA) χρησιμοποιείται για τη διερεύνηση των συνδέσεων μεταξύ των επιμέρους κριτηρίων ικανοποίησης και της Ολικής Ικανοποίησης. Η συνολική ικανοποίηση είναι το αποτέλεσμα που λαμβάνεται υπόψη στην fs/QCA, ενώ τα ατομικά κριτήρια ικανοποίησης είναι οι αιτιώδεις παράγοντες που συμβάλλουν στο αποτέλεσμα.

Πρώτα απ' όλα, η fs/QCA επιτρέπει την ανάλυση των επιπτώσεων των διαμορφώσεων αιτιώδους συνθήκης και όχι απλώς των ανεξάρτητων επιδράσεων κάθε αιτιολογικής συνθήκης σε κάθε αποτέλεσμα που μελετάται ξεχωριστά. Αυτό δείχνει ποια κριτήρια, καθώς και ποιοι συνδυασμοί τους, μπορούν να οδηγήσουν σε υψηλή συνολική ικανοποίηση.

Ωστόσο, επιτρέπει στην Ολική Ικανοποίηση να είναι το αποτέλεσμα της μελέτης ανεξάρτητα από το αν υπάρχει ή όχι κάποιο από τα πολυάριθμα μεμονωμένα κριτήρια (αιτιώδεις συνθήκες) (Whittington & Bell, 2016).

Επιπλέον, επιτρέπει τον εντοπισμό πολυάριθμων πιθανών συνδυασμών των διαφόρων κριτηρίων (αιτιακές διαδρομές ή διαμορφώσεις) που έχουν ως αποτέλεσμα το ίδιο αποτέλεσμα ή ίσες λύσεις (equifinality).

8.2 Δεδομένα

Όπως αναφέρθηκε στην εισαγωγή, στόχος της παρούσας ανάλυσης είναι να εντοπίσει τυχόν αιτιολογικούς παράγοντες —μεμονωμένα ή σε συνδυασμό— που συμβάλλουν στο υψηλό επίπεδο συνολικής ικανοποίησης του προσωπικού της εταιρείας.

Η αποχώρηση των εργαζομένων (left) καταγράφεται ως αποτέλεσμα, με τα κριτήρια ή άλλες πτυχές της ικανοποίησης (ικανοποίηση εργαζομένων, αριθμός προτζεκτ, χρόνια που εργάζονται στην εταιρεία, αν έχουν πάρει αύξηση ή όχι κλπ.) να αναφέρονται ως αιτιώδεις συνθήκες (causal conditions).

Η ανάλυση περιλαμβάνει συνολικά 14.999 υπαλλήλους όπου κάθε άτομο αντιμετωπίζεται ως διαφορετική περίπτωση (case) και αναλύθηκαν σύμφωνα με τα 7 κριτήρια ικανοποίησης συν το αποτέλεσμα (την εξαρτημένη μεταβλητή).

8.3 Βαθμονόμηση (calibration)

Όπως αναφέρθηκε σε προηγούμενο κεφάλαιο, το πρώτο βήμα κατά την διαδικασία της fs/QCA είναι η βαθμονόμηση των δεδομένων. Υπάρχει ένα ειδικό σημείο στα ασαφή σύνολα μεταξύ της πλήρους ένταξης και της πλήρους μη ένταξης: το crossover point ή σημείο μέγιστης ασάφειας.

Σύμφωνα με την άμεση μέθοδο του Ragin(2008) όπου εστιάζει σε τρία σημεία αποκοπής:

- Πλήρης συμμετοχή ($\mu = 0.95$)
- Πλήρης μη συμμετοχή ($\mu=0.05$)
- Σημείο Διασταύρωσης ($\mu = 0.50$)

πραγματοποιήθηκε η βαθμονόμηση των επιμέρους κριτηρίων.

Για την βαθμονόμηση των κριτηρίων ικανοποίησης πραγματοποιήθηκαν δύο διαδικασίες ανάλογα την μορφή τους: Descriptive statistics όπου βασιστήκαμε στα περιγραφικά στατιστικά κάθε κριτηρίου για να καθορίσουμε τα κατώφλια και Frequencies statistics όπου βασιστήκαμε στην ανάλυση συχνοτήτων κάθε κριτηρίου για να καθορίσουμε τα κατώφλια.

8.4 Descriptive Statistics

Για τα κριτήρια **average monthly hours** (ο μέσος όρος των ωρών που έχει δουλέψει μηνιαίως) , **last evaluation** (το σκορ που πέτυχε στην τελευταία αξιολόγηση) και **satisfaction level** (επίπεδο ικανοποίησης), η βαθμονόμηση έγινε χρησιμοποιώντας τα αντίστοιχα μέτρα θέσης και διασποράς (μέσος και τυπική απόκλιση).

Human_Resources_Employee_Attrition.csv: Descriptive Statistics				
Variable	Mean	Std. Dev.	Minimum	Maximum
N Cases Missing				
average_monthly	201.0503	49.94143	96	310
14999 0				
last_evaluation	0.7161017	0.1711634	0.36	1
14999 0				
satisfaction_le	0.6128335	0.2486224	0.09	1
14999 0				

Πίνακας 2. Descriptive Statistics

Επομένως για το κριτήριο Average Monthly το σημείο πλήρους συμμετοχής ($\mu=0,95$) θα πάρει την τιμή 250, όπου αυτό σημαίνει ότι όσοι εργαζόμενοι έχουν βαθμολογίες από 250 και πάνω ανήκουν πλήρως στο σύνολο των εργαζομένων της εταιρείας με μεγάλο φόρτο εργασίας. Επιπλέον, το σημείο για τη μη συμμετοχή ($\mu=0,05$) θα είναι 150, ενώ το σημείο διασταύρωσής crossover ($\mu=0,5$) θα πάρει την τιμή 201, που αντιστοιχεί στο μέσο όρο του κριτηρίου. Η ίδια διαδικασία ακολουθείται και για τα υπόλοιπα κριτήρια, όπως φαίνεται στον παρακάτω πίνακα.

Κριτήριο	Πλήρης συμμετοχή ($\mu = 0.95$)	Σημείο Διασταύρωσης ($\mu = 0.50$)	Πλήρης μη συμμετοχή ($\mu=0.05$)
Average Monthly Hours	250	201	150
Last Evaluation	0.9	0.7161	0.5
Satisfaction Level	0.86	0.6128	0.36

8.5 Frequencies

Για τα κριτήρια **Number of projects** (αριθμός project που έχουν ανατεθεί στον εργαζόμενο) και **years at company** (πόσα χρόνια εργάζεται στην εταιρεία) πραγματοποιήθηκε ανάλυση συχνοτήτων για τον καθορισμό των σημείων αποκοπής.

Human_Resources_Employee_Attrition.csv: Frequencies

number_of_proje	f	Pct
2	2388	15.9
3	4055	27.0
4	4365	29.1
5	2761	18.4
6	1174	7.8
7	256	1.7

Total	14999
Missing	0

years_at_compan	f	Pct
2	3244	21.6
3	6443	43.0
4	2557	17.0
5	1473	9.8
6	718	4.8
7	188	1.3
8	162	1.1
10	214	1.4

Total	14999
Missing	0

Πίνακας 3. Πίνακας Συχνοτήτων

Για το κριτήριο Number of projects το σημείο πλήρους συμμετοχής ($\mu=0,95$) θα πάρει την τιμή 6, όπου αυτό σημαίνει ότι όσοι εργαζόμενοι έχουν βαθμολογίες από 6 και πάνω ανήκουν πλήρως στο σύνολο των εργαζομένων με τα περισσότερα project. Το σημείο για την πλήρη μη συμμετοχή ($\mu=0,05$) θα είναι 3, ενώ το σημείο διασταύρωσής ($\mu=0,5$) θα πάρει την τιμή 4.

Αντίστοιχα, για το κριτήριο Years at company το σημείο πλήρους συμμετοχής ($\mu=0,95$) θα πάρει την τιμή 8, το σημείο για την μη πλήρη συμμετοχή ($\mu=0,05$) θα είναι 3 και τέλος το σημείο διασταύρωσής ($\mu=0,5$) θα πάρει την τιμή 5.

8.6 Αναγκαίες συνθήκες

Η αναζήτηση για τυχόν αναγκαίες συνθήκες αποτελεί το επόμενο βήμα στην εφαρμογή της fs/QCA. Όπως αναφέρθηκε προηγουμένως, μια συνθήκη μπορεί να χαρακτηριστεί ως αναγκαία εάν πρέπει να υπάρχει για να συμβεί το αποτέλεσμα που αναλύουμε, αλλά η απλή παρουσία της δεν εγγυάται την εμφάνισή του. Οποιοσδήποτε προϋποθέσεις κρίνονται αναγκαίες μπορούν στη συνέχεια να αφαιρεθούν από την διαδικασία του πίνακα αλήθειας αλλά για την πραγματοποίηση του αποτελέσματος θα πρέπει να συμπεριλαμβάνονται.

Σύμφωνα με την θεωρία των Schneider & Wagemann (2010) για να είναι μια αιτιώδης συνθήκη αναγκαία θα πρέπει η συνέπεια να είναι αρκετά υψηλή (Consistency > 0.9) και η κάλυψη θα πρέπει να είναι μεγαλύτερη του 0,5 (Coverage > 0.5).

Analysis of Necessary Conditions		
Outcome variable: left		
Conditions tested:		
	Consistency	Coverage
promotion_last_	0.005321	0.119497
~promotion_last_	0.994679	0.390587
faverage	0.528421	0.402560
~faverage	0.471579	0.368853
flast	0.521509	0.391088
~flast	0.478491	0.380459
fsatisfaction	0.287163	0.228531
~fsatisfaction	0.712837	0.534124
fnumber	0.444144	0.420734
~fnumber	0.555856	0.362001
fyears	0.226808	0.475431
~fyears	0.773192	0.365732
work_accident	0.047326	0.144568
~work_accident	0.952674	0.420831

Πίνακας 4. Πίνακας Αναγκαίων Συνθηκών

Στον παραπάνω πίνακα παρουσιάζονται τα αποτελέσματα, όπου φαίνεται ότι καμία συνθήκη (ή η απουσία τους) δεν είναι αναγκαία για την αποχώρηση των εργαζομένων.

8.7 Ικανές Συνθήκες

Το 3^ο βήμα της διαδικασίας της fs/QCA αποτελεί την κατασκευή του πίνακα αλήθειας. Ο πίνακας που δημιουργείται περιέχει 128 γραμμές ($=2^7$, όπου 7 είναι ο αριθμός των αιτιωδών συνθηκών) όπου παρουσιάζει όλους τους δυνατούς συνδυασμούς των επιμέρους κριτηρίων, το αποτέλεσμα (left), την

συνέπεια του κάθε συνδυασμού και την στήλη number όπου δείχνει πόσοι εργαζόμενοι ανήκουν στον κάθε συνδυασμό.

Ένα μέρος του πίνακα αλήθειας παρουσιάζεται παρακάτω:

fsatisfaction	flastevaluation	fnumber	fyears	fanerage	promotion_la...	work_accident	number ▾	left	raw consist.	PRI consist.	SYM consist
0	0	0	0	0	0	0	1976 (13%)		0.496388	0.496388	0.496395
0	1	1	0	1	0	0	1160 (20%)		0.451340	0.451340	0.451358
1	1	1	1	1	0	0	895 (26%)		0.535272	0.535272	0.535260
1	1	1	0	1	0	0	831 (32%)		0.250679	0.250679	0.250685
1	1	1	0	0	0	0	767 (37%)		0.101785	0.101785	0.101786
1	0	1	0	1	0	0	729 (42%)		0.053207	0.053207	0.053208
1	0	1	0	0	0	0	712 (47%)		0.044338	0.044338	0.044339
1	1	0	0	1	0	0	590 (51%)		0.159890	0.159890	0.159894
1	1	0	0	0	0	0	572 (54%)		0.103257	0.103257	0.103258
1	0	0	0	1	0	0	511 (58%)		0.074595	0.074595	0.074596
1	0	0	0	0	0	0	503 (61%)		0.078251	0.078251	0.078252
0	0	1	0	0	0	0	422 (64%)		0.090936	0.090936	0.090937
0	0	1	0	1	0	0	389 (67%)		0.158049	0.158049	0.158050
0	1	1	0	0	0	0	374 (69%)		0.133423	0.133423	0.133426
0	1	0	0	1	0	0	271 (71%)		0.156467	0.156467	0.156473
0	1	0	0	0	0	0	254 (73%)		0.147342	0.147342	0.147346
0	0	0	0	1	0	0	240 (74%)		0.131429	0.131429	0.131434
0	1	1	1	1	0	0	239 (76%)		0.393318	0.393318	0.393316
1	1	1	0	1	0	1	199 (77%)		0.078608	0.078608	0.078608
0	0	0	0	0	0	1	183 (78%)		0.192076	0.192076	0.192076
1	0	1	0	0	0	1	149 (79%)		0.015990	0.015990	0.015990
0	0	1	1	1	0	0	139 (80%)		0.237535	0.237535	0.237535
0	0	1	1	0	0	0	134 (81%)		0.151671	0.151671	0.151672
1	1	1	0	0	0	1	132 (82%)		0.036336	0.036336	0.036336
0	1	1	1	0	0	0	131 (83%)		0.219093	0.219093	0.219092
1	0	0	0	0	0	1	126 (84%)		0.020814	0.020814	0.020814
1	0	1	0	1	0	1	121 (85%)		0.017760	0.017760	0.017760
1	1	0	0	1	0	1	119 (85%)		0.044553	0.044553	0.044553
1	1	0	0	0	0	1	111 (86%)		0.029868	0.029868	0.029868
0	1	1	0	1	0	1	111 (87%)		0.165088	0.165088	0.165088
1	0	0	0	1	0	1	100 (87%)		0.019525	0.019525	0.019525
1	0	1	1	0	0	0	92 (88%)		0.103970	0.103970	0.103970

Πίνακας 5. Πίνακας Αλήθειας

Θα επιλεχτούν κατώφλια συχνότητας και συνέπειας με σκοπό να μειώσουμε τον αριθμό των γραμμών του πίνακα αλήθειας και κατ' επέκταση και τον αριθμό των συνδυασμών. Για το όριο συχνότητας γνωρίζουμε ότι πρέπει να συμπεριληφθούν τουλάχιστον το 75% - 85% των περιπτώσεων του συνόλου (Ragin 2008). Έτσι στην δική μας περίπτωση ορίσαμε σαν κατώφλι συχνότητας το 119 (85%) λόγω πολλών δεδομένων.

Στην συνέχεια κατώφλι συνέπειας (consistency) ορίζουμε το επίπεδο συνέπειας ώστε να καταλήξουμε στο συμπέρασμα ότι τις συνδυασμός είναι συνεπές υποσύνολο του αποτελέσματος. Σαν κατώφλι συνέπειας προτείνεται το 0,8, στην δική μας περίπτωση, επειδή η συνέπεια είναι σε χαμηλά επίπεδα, το κατώφλι θα το ορίσουμε στην τιμή του 0,2.

Τέλος, θα πραγματοποιηθεί η λογική μείωση του Πίνακα Αλήθειας, μέσω του αλγορίθμου Quine-McCluskey, , και θα προκύψουν τα τρία είδη λύσεων, η Σύνθετη (Complex), η Φειδωλή (Parsimonious) και η Ενδιάμεση (Intermediate).

```

--- COMPLEX SOLUTION ---
frequency cutoff: 119.000000
consistency cutoff: 0.219093

raw      unique

coverage  coverage  consistency
-----  -
~work_accident*~promotion_last_*flastevaluation*fnumber*fanera
ge          0.375432    0.292839
0.464417
~work_accident*~promotion_last_*~fsatisfaction*flastevaluation
*fnumber*fyears          0.084006    0.001905
0.351215
~work_accident*~promotion_last_*~fsatisfaction*fnumber*fyears*
fanerage          0.083700    0.001698
0.359167
~work_accident*~promotion_last_*~fsatisfaction*~flastevaluatio
n*~fnumber*~fyears*~fanerage    0.394359    0.376164
0.496388
solution coverage: 0.755757
solution consistency: 0.507460

```

Πίνακας 7. Σύνθετη Λύση

Σύμφωνα με την σύνθετη λύση βλέπουμε ότι υπάρχουν τέσσερα μονοπάτια που οδηγούν τους εργαζόμενους στην αποχώρηση από την εταιρεία.

Αρχικά το πρώτο μονοπάτι αναφέρει ότι αν δεν υπάρχει εργατικό ατύχημα, δεν έχει πάρει αύξηση ο εργαζόμενος τα τελευταία πέντε χρόνια, έχει υψηλή βαθμολογία στην τελευταία αξιολόγηση, υψηλό αριθμό προτζεκτ και υψηλό μέσο όρο ωρών εργασίας μηνιαίως τότε θα αποχωρήσει.

Το δεύτερο μονοπάτι αναφέρει ότι αν δεν υπάρχει εργατικό ατύχημα δεν έχει πάρει αύξηση ο εργαζόμενος, δεν είναι ικανοποιημένος, έχει υψηλό βαθμό στην τελευταία αξιολόγηση, υψηλό αριθμό προτζεκτ και πολλά χρόνια στην εταιρεία, τότε θα αποχωρήσει από αυτήν.

```

--- PARSIMONIOUS SOLUTION ---
frequency cutoff: 119.000000
consistency cutoff: 0.219093

raw      unique

coverage  coverage  consistency

-----
flastevaluation*fyears
0.217765    0.004413    0.390761
fyears*fanerage
0.212386    0.003332    0.397661
~work_accident*flastevaluation*fnumber*fanerage
0.376104    0.198750    0.457477
~work_accident*~fsatisfaction*~flastevaluation*~fnumber*~faner
age    0.397589    0.359356    0.478443
solution coverage: 0.781373
solution consistency: 0.448966

```

Πίνακας 8. Φειδωλή Λύση

Σύμφωνα με την φειδωλή λύση βλέπουμε ότι υπάρχουν και εδώ τέσσερα μονοπάτια που οδηγούν τους εργαζόμενους στην αποχώρηση από την εταιρεία.

Αρχικά το πρώτο μονοπάτι αναφέρει ότι αν έχει υψηλή βαθμολογία στην τελευταία αξιολόγηση και υψηλό μέσο όρο ωρών εργασίας μηνιαίως τότε θα αποχωρήσει.

Το δεύτερο μονοπάτι αναφέρει ότι αν εργάζεται πολλά χρόνια στην εταιρεία και έχει υψηλό μέσο όρο ωρών μηνιαίως εργασίας , τότε θα αποχωρήσει από αυτήν.

```

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 119.000000
consistency cutoff: 0.219093
Assumptions:

raw      unique

coverage  coverage  consistency

-----
fanerage*fnumber*flastevaluation*~promotion_last_*~work_accide
nt    0.375432    0.292839
0.464417
fanerage*fyears*fnumber*~fsatisfaction*~promotion_last_*~work_
accident    0.083700    0.001698
0.359167
fyears*fnumber*flastevaluation*~fsatisfaction*~promotion_last_
*~work_accident    0.084006    0.001905
0.351215
~fanerage*~fyears*~fnumber*~flastevaluation*~fsatisfaction*~pr
omotion_last_*~work_accident    0.394359    0.376164
0.496388
solution coverage: 0.755757
solution consistency: 0.507460

```

Πίνακας 9. Ενδιάμεση Λύση

Από τις τρεις παραπάνω λύσεις η επικρατέστερη και αυτή που θα λάβω υπόψιν όπως προαναφέρθηκε και παραπάνω είναι η **Intermediate** (ενδιάμεση λύση) η οποία περιέχει όσες απλουστευτικές υποθέσεις χρειάζεται καθώς αποτελεί ένα υπερσύνολο της σύνθετης λύσης, όπου δεν περιέχει καμία απλουστευτική υπόθεση, και ένα υποσύνολο της φειδωλής όπου ανεξάρτητα από την εγκυρότητα τους περιέχει όλες τις απλουστευτικές υποθέσεις.

Σύμφωνα με την ενδιάμεση λύση προκύπτουν τέσσερα μονοπάτια που οδηγούν τους εργαζόμενους να αποχωρήσουν από την εταιρεία. Αρχικά το πρώτο μονοπάτι ορίζει ότι αν ο εργαζόμενος έχει υψηλό μέσο όρο ωρών εργασίας στην εταιρεία, υψηλό αριθμό προτζεκτ, υψηλή βαθμολογία στην τελευταία αξιολόγηση αλλά δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια και δεν έχει εργατικό ατύχημα τότε θα αποχωρήσει.

Το δεύτερο μονοπάτι αναφέρει ότι αν ο εργαζόμενος έχει υψηλό μέσο όρο ωρών εργασίας, εργάζεται πολλά χρόνια στην εταιρεία, έχει μεγάλο αριθμό προτζεκτ, αν και δεν είναι ικανοποιημένος, δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια και δεν έχει εργατικό ατύχημα τότε θα οδηγηθεί στο να φύγει από την εταιρεία.

Το τρίτο μονοπάτι θεωρεί ότι αν εργάζεται πολλά χρόνια ,έχει υψηλό αριθμό προτζεκτ, υψηλή βαθμολογία στην τελευταία αξιολόγηση, δεν είναι ικανοποιημένος, δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια και δεν έχει εργατικό ατύχημα τότε θα αποχωρήσει.

Το τέταρτο μονοπάτι ορίζει ότι αν δεν έχει υψηλό μέσο όρο ωρών εργασίας, δεν εργάζεται πολλά χρόνια στην εταιρεία, δεν έχει υψηλό αριθμό προτζεκτ, δεν έχει υψηλή βαθμολογία στην αξιολόγηση, δεν είναι ικανοποιημένος , δεν έχει λάβει αύξηση και δεν έχει εργατικό ατύχημα πάλι ο εργαζόμενος θα οδηγηθεί στην αποχώρησή του.

Λόγω του χαμηλού solution consistency που έχει 0,507 παρόλο της ικανοποιητικής κάλυψης solution coverage (0,755) η λύση αυτή δεν είναι πολύ αξιόπιστη και τα αποτελέσματά της δεν είναι ικανοποιητικά.

Κεφάλαιο 9: fsQCA με χρήση Undersampling

Διατηρώντας όλα τα δεδομένα στην κατηγορία μειοψηφίας και μειώνοντας το μέγεθος της κλάσης πλειοψηφίας, το Undersampling είναι μια μέθοδος για την εξισορρόπηση μη ισορροπημένων συνόλων δεδομένων. Είναι μία από τις πολλές μεθόδους που μπορούν να χρησιμοποιήσουν οι επιστήμονες δεδομένων για να αντλήσουν ακριβέστερες πληροφορίες από σύνολα δεδομένων που αρχικά δεν ήταν ισορροπημένα. Παρόλο που έχει μειονεκτήματα, συμπεριλαμβανομένης της απώλειας δυνητικά κρίσιμων πληροφοριών, είναι ωστόσο μια συχνή και κρίσιμη δεξιότητα για τους επιστήμονες δεδομένων.

Από τον πίνακα δεδομένων επιλέγουμε τυχαία τα μισά δεδομένα.

9.1 Παρουσία αποτελέσματος (left)

Η βαθμονόμηση των κριτηρίων ικανοποίησης (calibration) που πραγματοποιήθηκε με δύο διαδικασίες ανάλογα την μορφή τους: με περιγραφικά στατιστικά(descriptive) και με συχνότητες (frequencies) παραμένει ίδια όπως πριν. Οπότε το επόμενο βήμα είναι η αναζήτηση για τυχόν αναγκαίες συνθήκες που όπως αναφέρθηκε για να είναι μια αιτιώδης συνθήκη αναγκαία θα πρέπει η συνέπεια να είναι αρκετά υψηλή (Consistency > 0.9) και η κάλυψη θα πρέπει να είναι μεγαλύτερη του 0,5 (Coverage > 0.5). Σύμφωνα με τα παρακάτω, καταλήγουμε ότι η απουσία αύξησης και η απουσία εργατικού ατυχήματος είναι πιθανόν αναγκαίες συνθήκες.

Conditions tested:		
	Consistency	Coverage
faverage	0.528421	0.402560
~faverage	0.471579	0.368853
flast	0.521509	0.391088
~flast	0.478491	0.380459
fsatisfaction	0.287163	0.228531
~fsatisfaction	0.712837	0.534124
fnumber	0.444144	0.420734
~fnumber	0.555856	0.362001
fyears	0.226808	0.475431
~fyears	0.773192	0.365732
promotion_last_	0.005321	0.119497
~promotion_last_	0.994679	0.390587
work_accident	0.047326	0.144568
~work_accident	0.952674	0.420831

Πίνακας 10. Αναγκαίες συνθήκες

Πίνακας Αλήθειας:

Το επόμενο βήμα είναι η κατασκευή του πίνακα αλήθειας με outcome μεταβλητή την left , όπου θα επιλεγτούν κατώφλια συχνότητας και συνέπειας με σκοπό να μειώσουμε τον αριθμό των γραμμών του πίνακα αλήθειας και κατ' επέκταση και τον αριθμό των συνδυασμών.

Για το όριο συχνότητας γνωρίζουμε ότι πρέπει να συμπεριληφθούν τουλάχιστον το 75% - 85% των περιπτώσεων του συνόλου (Ragin 2008). Έτσι στην δική μας περίπτωση ορίσαμε σαν κατώφλι συχνότητας το 100 (82%) λόγω πολλών δεδομένων. Στην συνέχεια κατώφλι συνέπειας(consistency) ορίζουμε το επίπεδο συνέπειας ώστε να καταλήξουμε στο συμπέρασμα ότι ένας συνδυασμός είναι συνεπές υποσύνολο του αποτελέσματος. Σαν κατώφλι συνέπειας προτείνετε 0,8 κατά Ragin και σύμφωνα με τον Crilly μπορεί να οριστεί σαν την διαφορά μεταξύ των τιμών της συνέπειας δυο κοντινών γραμμών.

Στην δική μας περίπτωση επειδή η συνέπεια είναι σε χαμηλά επίπεδα το κατώφλι θα το ορίσουμε στην τιμή του 0,56.

promotion_la...	faverage	flast	fsatisfaction	fnumber	fyears	work_accident	number	left	raw consist. ▾	PRI consist.	SYM consist
0	1	1	1	1	1	0	846	1	0.699048	0.699048	0.699032
0	0	0	0	0	0	0	1705	1	0.686419	0.686419	0.686437
0	1	1	0	1	0	0	975	1	0.626425	0.626425	0.626425
0	1	1	0	1	1	0	166	1	0.564113	0.564113	0.564105
0	1	1	1	1	0	0	421	0	0.395848	0.395848	0.395851
0	0	0	0	0	0	1	127	0	0.341628	0.341628	0.341627
0	1	0	0	1	0	0	223	0	0.278925	0.278925	0.278926
0	1	1	1	0	0	0	291	0	0.274545	0.274545	0.274553
0	1	1	0	0	0	0	141	0	0.261598	0.261598	0.261601
0	0	1	0	0	0	0	135	0	0.239948	0.239948	0.239951
0	1	0	0	0	0	0	134	0	0.237442	0.237442	0.237447
0	0	0	1	0	0	0	253	0	0.179729	0.179729	0.179731
0	0	1	0	1	0	0	196	0	0.179123	0.179123	0.179122
0	0	1	1	0	0	0	295	0	0.174313	0.174313	0.174314
0	1	1	1	1	0	1	100	0	0.145164	0.145164	0.145164
0	1	0	1	0	0	0	258	0	0.143982	0.143982	0.143983
0	0	1	1	1	0	0	385	0	0.139072	0.139072	0.139072
0	0	0	0	1	0	0	212	0	0.136550	0.136550	0.136550
0	1	0	1	1	0	0	374	0	0.101937	0.101937	0.101937
0	0	0	1	1	0	0	350	0	0.074921	0.074921	0.074922

Πίνακας 11. Πίνακας Αλήθειας

Θα προκύψουν τρία είδη λύσεων, η Σύνθετη (Complex), η Φειδωλή (Parsimonious) και η Ενδιάμεση (Intermediate).

```

--- COMPLEX SOLUTION ---
frequency cutoff: 100.000000
consistency cutoff: 0.564113

```

	raw coverage	unique coverage	consistency
~promotion_last*faverage*flast*~fsatisfaction*fnumber*~work_accident	0.238863	0.162916	0.641078
~promotion_last*faverage*flast*fnumber*fyears*~work_accident	0.176518	0.101113	0.715766
~promotion_last*~faverage*~flast*~fsatisfaction*~fnumber*~fyears*~work_accident	0.381573	0.369794	0.686419
solution coverage: 0.709775			
solution consistency: 0.728191			

Πίνακας 12. Σύνθετη Λύση

Σύμφωνα με την σύνθετη λύση προκύπτουν τρία μονοπάτια όπου οδηγούν τους εργαζόμενους να αποχωρήσουν από την εταιρεία. Το μονοπάτι με το μεγαλύτερο Unique coverage (0.369794) είναι το τρίτο όπου αναφέρει, ότι αν ο εργαζόμενος δεν είχε λάβει προαγωγή, είχε χαμηλό μέσο όρο ωρών μηνιαίως, χαμηλή βαθμολογία στην τελευταία αξιολόγηση, δεν ήταν ικανοποιημένος, είχε αναλάβει χαμηλό αριθμό προτζεκτ, δεν ήταν πολλά χρόνια στην εταιρεία και δεν είχε εργατικό ατύχημα τότε θα αποχωρήσει από αυτήν.

```

--- PARSIMONIOUS SOLUTION ---
frequency cutoff: 100.000000
consistency cutoff: 0.564113

```

	raw coverage	unique coverage	consistency
fyears	0.226802	0.124749	0.475410
faverage*flast*~fsatisfaction*fnumber	0.251105	0.171166	0.607690
~faverage*~flast*~fsatisfaction*~fnumber*~work_accident	0.384296	0.349824	0.669387
solution coverage: 0.748349			
solution consistency: 0.629119			

Πίνακας 13. Φειδωλή Λύση

Σύμφωνα με την φειδωλή λύση προκύπτουν τρία μονοπάτια όπου οδηγούν τους εργαζόμενους να αποχωρήσουν από την εταιρεία. Το μονοπάτι με το μεγαλύτερο Unique coverage (0.349824) είναι το τρίτο όπου αναφέρει, ότι αν ο εργαζόμενος είχε χαμηλό μέσο όρο ωρών μηνιαίως, χαμηλή βαθμολογία στην τελευταία αξιολόγηση, δεν ήταν ικανοποιημένος, είχε αναλάβει χαμηλό αριθμό προτζεκτ και δεν είχε εργατικό ατύχημα τότε θα αποχωρήσει από αυτήν.


```

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 100.000000
consistency cutoff: 0.564113
Assumptions:

```

	raw coverage	unique coverage	consistency
~work_accident*fnumber*~fsatisfaction*flast*faverage*~promotion_last_	0.238863	0.162916	0.641078
~work_accident*fyears*fnumber*flast*faverage*~promotion_last_	0.176518	0.101113	0.715766
~work_accident*~fyears*~fnumber*~fsatisfaction*~flast*~faverage*~promotion_last_	0.381573	0.369794	0.686419
solution coverage:	0.709775		
solution consistency:	0.728191		

Πίνακας 14. Ενδιάμεση Λύση

Από τις τρεις παραπάνω λύσεις η επικρατέστερη και αυτή που θα λάβω υπόψιν όπως προαναφέρθηκε και παραπάνω είναι η Intermediate (ενδιάμεση λύση) η οποία περιέχει όσες απλουστευτικές υποθέσεις χρειάζεται καθώς αποτελεί ένα υπερσύνολο της σύνθετης λύσης, όπου δεν περιέχει καμία απλουστευτική υπόθεση, και ένα υποσύνολο της φειδωλής όπου ανεξάρτητα από την εγκυρότητα τους περιέχει όλες τις απλουστευτικές υποθέσεις.

Σύμφωνα με την ενδιάμεση λύση προκύπτουν τρία μονοπάτια που οδηγούν τους εργαζόμενους να αποχωρήσουν από την εταιρεία. Αρχικά το πρώτο μονοπάτι ορίζει ότι αν ο εργαζόμενος δεν έχει εργατικό ατύχημα, έχει πραγματοποιήσει πολλά προτζεκτ, δεν είναι ικανοποιημένος, έχει υψηλή βαθμολογία στην αξιολόγηση, υψηλό μέσο όρο ωρών εργασίας και δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια τότε θα αποχωρήσει.

Το δεύτερο μονοπάτι θεωρεί ότι αν δεν υπάρχει εργατικό ατύχημα, έχει δουλέψει πολλά χρόνια, έχει υψηλό αριθμό προτζεκτ, υψηλή βαθμολογία στην αξιολόγηση, υψηλό μέσο όρο ωρών εργασίας μηνιαίως και δεν έχει πάρει αύξηση τότε θα αποχωρήσει.

Το τρίτο και τελευταίο μονοπάτι μας δείχνει ότι αν δεν υπάρχει εργατικό ατύχημα, δεν έχει δουλέψει πολλά χρόνια στην εταιρεία, δεν έχει υψηλό αριθμό προτζεκτ, δεν είναι ικανοποιημένος, δεν είχε υψηλή βαθμολογία στην τελευταία αξιολόγηση, δεν έχει υψηλό μέσο όρο ωρών εργασίας μηνιαίως στην εταιρεία και δεν έχει λάβει και αύξηση τα τελευταία 5 χρόνια τότε θα οδηγηθεί στην αποχώρηση του.

Στην μέθοδο αυτή βλέπουμε ότι το solution coverage και solution consistency είναι σε πολύ υψηλές τιμές 0,709775 και 0,728191 αντίστοιχα. Πολύ πιο ικανοποιητικές τιμές σε σχέση με πριν γι' αυτό και η μέθοδος Undersampling όπως και τα αποτελέσματά της θεωρούνται πιο αξιόπιστα.

Από αυτά τα τρία μονοπάτια αυτό που θα επιλέξω σαν πιο αξιόπιστο είναι αυτό με το μεγαλύτερο unique coverage, όπου το έχει το τρίτο μονοπάτι 0,39794 όπου μας δείχνει ότι περίπου το 39% των περιπτώσεων καλύπτεται-επεξηγούνται από αυτό το μονοπάτι.

9.2 Στατιστική μέθοδος χ^2

Χρησιμοποιούμε την στατιστική ανάλυση χ^2 για την περαιτέρω ανάλυση των αναγκαίων συνθηκών. Παρατηρούμε ότι σε όλα τα αιτιώδη μονοπάτια που προκύπτουν παρουσιάζεται η απουσία του εργατικού ατυχήματος και η απουσία της αύξησης των τελευταίων 5 χρόνων. Γι' αυτό τον λόγο πραγματοποιούμε την στατιστική ανάλυση χ^2 ώστε να δούμε αν οι δύο αυτές μεταβλητές είναι ανεξάρτητες ή όχι με την αποχώρηση των εργαζομένων.

Η μηδενική υπόθεση για ονομαστικά δεδομένα που αναφέρονται ως συχνότητες ελέγχεται χρησιμοποιώντας την ανάλυση χ^2 . Η μηδενική υπόθεση δηλώνει ότι η κατανομή του πληθυσμού συμμορφώνεται με ένα σύνολο υποθετικών αναλογιών για την δοκιμή μιας μεταβλητής. Με βάση την μηδενική υπόθεση καταλήγουμε στο αν οι δύο μεταβλητές είναι ανεξάρτητες μεταξύ τους. Δημιουργούνται ένα σύνολο προβλεπόμενων συχνοτήτων με βάση την μηδενική υπόθεση και ένα παρατηρούμενων συχνοτήτων όπου συγκρίνονται.

Η μηδενική υπόθεση θα είναι πιο αμφίβολη όπως θα δείξουμε καθώς όσο μεγαλύτερες είναι οι διαφορές μεταξύ παρατηρούμενων και αναμενόμενων συχνοτήτων, τόσο μεγαλύτερη είναι και η τιμή του χ^2 . Οι θετικές αποκλίσεις προκύπτουν από τον τετραγωνισμό κάθε απόκλισης επομένως η τιμή του χ^2 δεν μπορεί να είναι ποτέ αρνητική. Επιπλέον διαιρούμε με την αναμενόμενη συχνότητα (f_e), καθώς οι αποκλίσεις πρέπει να αξιολογούνται σε σχέση με τις αναμενόμενες συχνότητες.

DEGREES OF FREEDOM (TWO-VARIABLE χ^2 TEST)

$$df = (c - 1)(r - 1)$$

Όπου c ισούται με τον αριθμό των κατηγοριών για την μεταβλητή στήλης και r ισούται με τον αριθμό των κατηγοριών για την μεταβλητή της γραμμής. Γνωρίζουμε ότι όλες οι παρατηρούμενες συχνότητες έχουν ίσες καταστάσεις και όλοι οι συνδυασμοί των συχνοτήτων αυτών είναι πιθανοί. Ωστόσο μόνο χ συχνότητες είναι ελεύθερες να ποικίλλουν όπου το χ προκύπτει από τον βαθμό ελευθερίας που βρίσκουμε ($df = \chi$).

9.2.1 Squared Cramer's Phi Coefficient (ϕ^2_c)

Υπολογίζει το ποσοστό της εξηγούμενης διακύμανσης- προβλεψιμότητας μεταξύ δύο ποιοτικών μεταβλητών.

PROPORTION OF EXPLAINED VARIANCE (TWO-VARIABLE χ^2)

$$\phi^2_c = \frac{\chi^2}{n(k - 1)}$$

Όπου χ^2 είναι το αποτέλεσμα της στατιστικής ανάλυσης της μεθόδου χ^2 , n είναι η συνολική παρατηρούμενη συχνότητα (μέγεθος δείγματος) και k είναι το μικρότερο από τις στήλες c ή σειρές r .

Με βάση τον υπολογισμό του ϕ^2_c καθορίζεται η ισχύς της σχέσης μεταξύ των δύο μεταβλητών. Η ισχύς της σχέσης είναι μικρή αν το ϕ^2_c προσεγγίζει την τιμή 0,01, μεσαία αν προσεγγίζει το 0,09 και μεγάλη αν προσεγγίζει ή υπερβαίνει την τιμή του 0,25.

GUIDELINES FOR ϕ_c^2

ϕ_c^2	EFFECT
.01	Small
.09	Medium
.25	Large

9.2.2 Έλεγχος στατιστικής σχέσης μεταξύ αποχώρησης των εργαζομένων και αν είχαν εργατικό ατύχημα

left * work_accident Crosstabulation

			work_accident		Total
			0	1	
left	0	Count	4682	1000	5682
		Expected Count	4964,2	717,8	5682,0
	1	Count	3402	169	3571
		Expected Count	3119,8	451,2	3571,0
Total		Count	8084	1169	9253
		Expected Count	8084,0	1169,0	9253,0

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	328,911 ^a	1	,000		
Continuity Correction ^b	327,746	1	,000		
Likelihood Ratio	372,239	1	,000		
Fisher's Exact Test				,000	,000
N of Valid Cases	9253				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 451,15.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-,189	,000
	Cramer's V	,189	,000
N of Valid Cases		9253	

H_0 : Η αποχώρηση των εργαζομένων και αν είχαν εργατικό ατύχημα είναι ανεξάρτητα

H_1 : Να μην ισχύει το H_0

Βλέπουμε ότι χ^2 είναι 328,911, επιπλέον έχει βαθμό ελευθερίας=1 όπου γνωρίζουμε ότι για βαθμό ελευθερίας=1 και βαθμό σημαντικότητας $P < 0,05$ έχουμε =3,84. Θέλουμε το χ^2 να είναι μικρότερο από 3,84 όπου σε εμάς αυτό δεν ισχύει καθώς όπως αναφέρθηκε το χ^2 είναι 328,911, άρα το σενάριο H_0 απορρίπτεται, άρα η αποχώρηση των εργαζομένων και το αν είχαν εργατικό ατύχημα δεν είναι ανεξάρτητα.

Επιπλέον βλέπουμε ότι η τιμή της ϕ^2 είναι 0,189 η οποία σύμφωνα με όσα προαναφέρθηκαν προσεγγίζει την τιμή 0,25 άρα η ισχύς μεταξύ των μεταβλητών είναι υψηλή.

9.2.3 Έλεγχος στατιστικής σχέσης μεταξύ αποχώρησης των εργαζομένων και αν είχαν πάρει προαγωγή τα τελευταία 5 χρόνια

left * promotion Crosstabulation

			promotion		Total
			0	1	
left	0	Count	5542	140	5682
		Expected Count	5584,4	97,6	5682,0
	1	Count	3552	19	3571
		Expected Count	3509,6	61,4	3571,0
Total		Count	9094	159	9253
		Expected Count	9094,0	159,0	9253,0

Βλέπουμε ότι η παρατηρούμενη συχνότητα (count) στο να μην φύγει και να μην έχει πάρει προαγωγή είναι 5542 ενώ η αναμενόμενη είναι 5584,4. Αντίστοιχα στο να μην φύγει και να έχει πάρει προαγωγή η παρατηρούμενη είναι 140 και η αναμενόμενη 97,6 κλπ.

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	48,459 ^a	1	,000		
Continuity Correction ^b	47,322	1	,000		
Likelihood Ratio	57,189	1	,000		
Fisher's Exact Test				,000	,000
N of Valid Cases	9253				

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 61,36.

b. Computed only for a 2x2 table

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	-,072	,000
	Cramer's V	,072	,000
N of Valid Cases		9253	

H_0 : Η αποχώρηση των εργαζομένων και αν είχαν εργατικό ατύχημα είναι ανεξάρτητα

H_1 : Να μην ισχύει το H_0

Βλέπουμε ότι χ^2 είναι 48,459 ,με βαθμό ελευθερίας=1 όπου γνωρίζουμε ότι για βαθμό ελευθερίας=1 και βαθμό σημαντικότητας $P < 0,05$ έχουμε =3,84. Θέλουμε το χ^2 να είναι μικρότερο από 3,84 όπου σε

εμάς αυτό δεν ισχύει, άρα το σενάριο H_0 απορρίπτεται ,άρα η αποχώρηση των εργαζομένων και αν είχαν πάρει προαγωγή δεν είναι ανεξάρτητα, είναι εξαρτημένα.

Επιπλέον βλέπουμε ότι η τιμή της ϕ^2_c είναι 0,072 η οποία σύμφωνα με όσα προαναφέρθηκαν προσεγγίζει την τιμή 0,09 άρα η ισχύς μεταξύ των μεταβλητών είναι μεσαία.

9.3 Απουσία αποτελέσματος (~ left)

Το επόμενο βήμα είναι η αναζήτηση για τυχόν αναγκαίες συνθήκες με outcome μεταβλητή την απουσία του Left (παραμονή των εργαζομένων), που όπως αναφέρθηκε για να είναι μια αιτιώδης συνθήκη αναγκαία θα πρέπει η συνέπεια να είναι αρκετά υψηλή (Consistency > 0.9) και η κάλυψη θα πρέπει να είναι μεγαλύτερη του 0,5 (Coverage > 0.5).

Analysis of Necessary Conditions

Outcome variable: ~left

Conditions tested:

	Consistency	Coverage
promotion_last_	0.024639	0.880503
~promotion_last_	0.975361	0.609413
faverage	0.492869	0.597440
~faverage	0.507131	0.631147
flast	0.510306	0.608912
~flast	0.489694	0.619541
fsatisfaction	0.609243	0.771469
~fsatisfaction	0.390757	0.465876
fnumber	0.384312	0.579266
~fnumber	0.615688	0.637999
fyears	0.157276	0.524569
~fyears	0.842724	0.634268
work_accident	0.175994	0.855432
~work_accident	0.824006	0.579169

Πίνακας 15. Πίνακας Αναγκαίων Συνθηκών

Θα κατασκευαστεί ο πίνακας αλήθειας με Outcome μεταβλητή την **απουσία της left (~left)**. Για το όριο συχνότητας γνωρίζουμε ότι πρέπει να συμπεριληφθούν τουλάχιστον το 75% - 85% των περιπτώσεων του συνόλου (Ragin 2008). Έτσι στην δική ένας περίπτωση ορίσαμε σαν κατώφλι συχνότητας το 100 (82%) λόγω πολλών δεδομένων. Στην συνέχεια κατώφλι συνέπειας(consistency) ορίζουμε το επίπεδο συνέπειας ώστε να καταλήξουμε στο συμπέρασμα ότι ένας συνδυασμός είναι συνεπές υποσύνολο του αποτελέσματος. Σαν κατώφλι συνέπειας προτείνετε 0,8 κατά Ragin.

work_accident	promotion_la...	faverage	flast	fsatisfaction	fnumber	fyears	number	~left	raw consist.	PRI consist.	SYM consist
0	0	0	0	1	1	0	350	1	0.925070	0.925070	0.925078
0	0	1	0	1	1	0	374	1	0.898057	0.898057	0.898063
0	0	0	0	0	1	0	212	1	0.863449	0.863449	0.863450
0	0	0	1	1	1	0	385	1	0.860926	0.860926	0.860928
0	0	1	0	1	0	0	258	1	0.856007	0.856007	0.856017
1	0	1	1	1	1	0	100	1	0.854836	0.854836	0.854836
0	0	0	1	1	0	0	295	1	0.825682	0.825682	0.825686
0	0	0	1	0	1	0	196	1	0.820879	0.820879	0.820878
0	0	0	0	1	0	0	253	1	0.820264	0.820264	0.820269
0	0	1	0	0	0	0	134	0	0.762538	0.762538	0.762553
0	0	0	1	0	0	0	135	0	0.760040	0.760040	0.760049
0	0	1	1	0	0	0	141	0	0.738392	0.738392	0.738399
0	0	1	1	1	0	0	291	0	0.725427	0.725427	0.725447
0	0	1	0	0	1	0	223	0	0.721071	0.721071	0.721074
1	0	0	0	0	0	0	127	0	0.658375	0.658375	0.658373
0	0	1	1	1	1	0	421	0	0.604145	0.604145	0.604149
0	0	1	1	0	1	1	166	0	0.435901	0.435901	0.435895
0	0	1	1	0	1	0	975	0	0.373574	0.373574	0.373575
0	0	0	0	0	0	0	1705	0	0.313555	0.313555	0.313563
0	0	1	1	1	1	1	846	0	0.300975	0.300975	0.300968

Πίνακας 16. Πίνακας Αλήθειας

Θα προκύψουν τρία είδη λύσεων, η Σύνθετη (Complex), η Φειδωλή (Parsimonious) και η Ενδιάμεση (Intermediate).

```

--- COMPLEX SOLUTION ---
frequency cutoff: 100.000000
consistency cutoff: 0.820264

raw    unique
coverage coverage consistency
-----
~work_accident*~promotion_last_*~flast*fsatisfaction*~fyears 0.247452 0.089795 0.869738
~work_accident*~promotion_last_*~faverage*fsatisfaction*~fyears 0.256313 0.056406 0.859211
~work_accident*~promotion_last_*~faverage*fnumber*~fyears 0.171563 0.038183 0.871764
work_accident*~promotion_last_*faverage*flast*fsatisfaction*fnumber*~fyears 0.021695 0.021698 0.854836
solution coverage: 0.405986
solution consistency: 0.868957

```

Πίνακας 17. Σύνθετη Λύση

Προκύπτουν τέσσερα μονοπάτια ,το μονοπάτι με το μεγαλύτερο Unique coverage είναι το πρώτο μονοπάτι που προκύπτει , όπου μας αναφέρει ότι αν ο εργαζόμενος δεν είχε εργατικό ατύχημα, δεν πήρε προαγωγή τα τελευταία 5 χρόνια, είχε χαμηλή βαθμολογία στην τελευταία αξιολόγηση, είναι ικανοποιημένος και δεν εργάζεται πολλά χρόνια στην εταιρεία τότε θα παραμείνει σε αυτήν.

```

--- PARSIMONIOUS SOLUTION ---
frequency cutoff: 100.000000
consistency cutoff: 0.820264

raw coverage unique consistency
-----
~flast*fsatisfaction 0.332442 0.101527 0.891566
~faverage*fsatisfaction 0.338042 0.063045 0.880216
~faverage*fnumber 0.228389 0.056421 0.885734
work_accident*fsatisfaction 0.108583 0.027477 0.924826
solution coverage: 0.546281
solution consistency: 0.885068

```

Πίνακας 18. Φειδωλή Λύση

Προκύπτουν τέσσερα μονοπάτια, το μονοπάτι με το μεγαλύτερο unique coverage είναι το δεύτερο, όπου αναφέρει ότι αν ο εργαζόμενος είχε χαμηλό μέσο όρο ωρών εργασίας τον μήνα και είναι ικανοποιημένος από την εταιρεία τότε θα παραμείνει σε αυτήν.

```

--- INTERMEDIATE SOLUTION ---
frequency cutoff: 100.000000
consistency cutoff: 0.820264
Assumptions:

raw coverage unique consistency
-----
~fyears*fnumber*~faverage*~promotion_last_*~work_accident 0.171563 0.038183 0.871764
~fyears*fsatisfaction*~faverage*~promotion_last_*~work_accident 0.256313 0.056406 0.859211
~fyears*fsatisfaction*~flast*~promotion_last_*~work_accident 0.247452 0.089795 0.869738
~fyears*fnumber*fsatisfaction*flast*faverage*~promotion_last_*~work_accident 0.021695 0.021698 0.854836
solution coverage: 0.405986
solution consistency: 0.868957

```

Πίνακας 19. Ενδιάμεση Λύση

Από στους τρεις παραπάνω λύσεις η επικρατέστερη και αυτή που θα λάβω υπόψιν στους προαναφέρθηκε και παραπάνω είναι η Intermediate (ενδιάμεση λύση) η οποία περιέχει όσες απλουστευτικές υποθέσεις χρειάζεται καθώς αποτελεί ένα υπερσύνολο στους σύνθετης λύσης, όπου δεν περιέχει καμία απλουστευτική υπόθεση, και ένα υποσύνολο στους φειδωλής όπου ανεξάρτητα από την εγκυρότητα τους περιέχει όλες τις απλουστευτικές υποθέσεις.

Σύμφωνα με την ενδιάμεση λύση προκύπτουν τέσσερα μονοπάτια που οδηγούν στους εργαζόμενους να μην αποχωρήσουν από την εταιρεία άρα να μείνει. Αρχικά το πρώτο μονοπάτι ορίζει ότι αν ο εργαζόμενος δεν εργάζεται πολλά χρόνια στην εταιρεία, έχει υψηλό αριθμό προτζεκτ, δεν έχει υψηλό μέσο όρο ωρών μηνιαίως, δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια και δεν έχει εργατικό ατύχημα τότε θα παραμείνει στην εταιρεία.

Το δεύτερο μονοπάτι αναφέρει ότι αν ο εργαζόμενος δεν εργάζεται πολλά χρόνια στην εταιρεία, είναι ικανοποιημένος, δεν εργάζεται στους ώρες κατά μέσο όρο μηνιαίως, δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια και δεν έχει εργατικό ατύχημα τότε δεν θα παραμείνει στην εταιρεία.

Το τρίτο μονοπάτι στους δείχνει ότι αν ο εργαζόμενος δεν εργάζεται πολλά χρόνια στην εταιρεία, είναι ικανοποιημένος, δεν έχει υψηλή βαθμολογία στην τελευταία αξιολόγηση, δεν έχει πάρει αύξηση και δεν έχει εργατικό ατύχημα, τότε θα παραμείνει στην εταιρεία.

Το τέταρτο και τελευταίο μονοπάτι ορίζει ότι αν ο εργαζόμενος δεν εργάζεται πολλά χρόνια στην εταιρεία, έχει υψηλό αριθμό προτζεκτ, είναι ικανοποιημένος, έχει υψηλή βαθμολογία στην αξιολόγηση, υψηλό μέσο όρο ωρών μηνιαίως, δεν έχει πάρει αύξηση και είχε εργατικό ατύχημα τότε θα παραμείνει.

Στην μέθοδο αυτή βλέπουμε ότι το solution coverage και solution consistency είναι σε πολύ υψηλές τιμές 0,405986 και 0,868957 αντίστοιχα. Αρκετά υψηλές τιμές, ειδικά το consistency πολύ πιο βελτιωμένο σε σχέση με τις άλλες μεθόδους.

Από αυτά τα τέσσερα μονοπάτια αυτό που θα επιλέξω σαν πιο αξιόπιστο είναι αυτό με το μεγαλύτερο unique coverage, όπου το έχει το τρίτο μονοπάτι 0,089795 όπου στους δείχνει ότι περίπου το 8% των συνολικών περιπτώσεων καλύπτεται-επεξηγούνται από αυτό το μονοπάτι.

Κεφάλαιο 10: Αλγόριθμος Apriori

10.1 Εφαρμογή του Αλγορίθμου Apriori

Τα δεδομένα μας αποτελούνται από επτά κριτήρια αξιολόγησης (satisfaction level, last evaluation, number of project, average monthly, years at company, work accident και promotion last 5 years) και την εξαρτημένη μεταβλητή που είναι το left.

Αρχικά εφαρμόσαμε το φίλτρο Discretize ώστε να χωρίσουμε το κάθε δεδομένο σε 5 ομάδες (bins=5). Στην συνέχεια εφαρμόσαμε το φίλτρο NumerictoNominal ώστε να μετατρέψουμε το σύνολο του dataset σε ονομαστικές μεταβλητές.

Εφαρμόζουμε τον αλγόριθμο Apriori χωρίς να λάβουμε υπόψιν την μεταβλητή left. Στόχος στους είναι να βρούμε στους κανόνες με την υψηλότερη τιμή support για μία ορισμένη ελάχιστη τιμή confidence (εμπιστοσύνης). Η τιμή της εμπιστοσύνης θα πρέπει να είναι ιδανικά κοντά στην μονάδα. Το Lift μας δείχνει εάν τα χαρακτηριστικά είναι εξαρτημένα μεταξύ τους, για τιμές μεγαλύτερες της μονάδας τα χαρακτηριστικά έχουν σημαντική συσχέτιση ενώ για τιμές μικρότερες της μονάδας τα χαρακτηριστικά είναι μάλλον ανεξάρτητα.

Εφαρμόζουμε τον αλγόριθμο Apriori στο σύνολο των δεδομένων στο πρόγραμμα Weka ώστε να καταλήξουμε στους παρακάτω κανόνες συσχέτισης.

Best rules found:

```
1. satisfaction_le='(0.818-inf)' average_monthly='(181.6-224.4]' years_at_compan='(-inf-3.6]' work_accident='(-inf-0.2]' promotion_last_='(-inf-0.2]' 295 ==> left='(-inf-0.2]' 295
2. number_of_proje='(6-inf)' 256 ==> left='(0.8-inf)' 256 acc:(0.99499)
3. satisfaction_le='(0.636-0.818]' years_at_compan='(-inf-3.6]' work_accident='(0.8-inf)' 214 ==> left='(-inf-0.2]' 214 acc:(0.99499)
4. satisfaction_le='(0.818-inf)' last_evaluation='(0.744-0.872]' years_at_compan='(-inf-3.6]' promotion_last_='(-inf-0.2]' 293 ==> left='(-inf-0.2]' 292 acc:(0.99499)
5. years_at_compan='(6.8-8.4]' 208 ==> left='(-inf-0.2]' 208 acc:(0.99498)
6. satisfaction_le='(0.818-inf)' last_evaluation='(0.744-0.872]' years_at_compan='(-inf-3.6]' work_accident='(-inf-0.2]' 249 ==> left='(-inf-0.2]' 248 acc:(0.99497)
7. last_evaluation='(0.616-0.744]' years_at_compan='(-inf-3.6]' work_accident='(0.8-inf)' 163 ==> left='(-inf-0.2]' 163 acc:(0.99496)
8. satisfaction_le='(0.818-inf)' number_of_proje='(3-4]' average_monthly='(138.8-181.6]' 162 ==> left='(-inf-0.2]' 162 acc:(0.99496)
9. satisfaction_le='(0.454-0.636]' last_evaluation='(0.616-0.744]' number_of_proje='(-inf-3]' years_at_compan='(-inf-3.6]' 154 ==> left='(-inf-0.2]' 154 acc:(0.99495)
10. satisfaction_le='(0.818-inf)' last_evaluation='(0.616-0.744]' number_of_proje='(3-4]' 151 ==> left='(-inf-0.2]' 151 acc:(0.99495)
11. satisfaction_le='(0.818-inf)' last_evaluation='(0.488-0.616]' number_of_proje='(3-4]' 147 ==> left='(-inf-0.2]' 147 acc:(0.99494)
12. satisfaction_le='(0.454-0.636]' last_evaluation='(0.872-inf)' number_of_proje='(3-4]' 141 ==> left='(-inf-0.2]' 141 acc:(0.99494)
13. last_evaluation='(0.488-0.616]' number_of_proje='(3-4]' average_monthly='(181.6-224.4]' 140 ==> left='(-inf-0.2]' 140 acc:(0.99494)
14. last_evaluation='(0.616-0.744]' average_monthly='(181.6-224.4]' years_at_compan='(-inf-3.6]' 293 ==> left='(-inf-0.2]' 291 acc:(0.99493)
15. satisfaction_le='(0.818-inf)' years_at_compan='(-inf-3.6]' work_accident='(0.8-inf)' 210 ==> left='(-inf-0.2]' 209 acc:(0.99492)
16. satisfaction_le='(0.636-0.818]' number_of_proje='(-inf-3]' work_accident='(0.8-inf)' 133 ==> left='(-inf-0.2]' 133 acc:(0.99492)
17. satisfaction_le='(0.636-0.818]' last_evaluation='(0.488-0.616]' number_of_proje='(-inf-3]' years_at_compan='(-inf-3.6]' promotion_last_='(-inf-0.2]' 130 ==> left='(-inf-0.2]' 130
18. satisfaction_le='(0.818-inf)' last_evaluation='(0.488-0.616]' average_monthly='(138.8-181.6]' 123 ==> left='(-inf-0.2]' 123 acc:(0.9949)
19. satisfaction_le='(0.818-inf)' last_evaluation='(0.488-0.616]' average_monthly='(181.6-224.4]' 120 ==> left='(-inf-0.2]' 120 acc:(0.9949)
20. average_monthly='(181.6-224.4]' years_at_compan='(-inf-3.6]' work_accident='(0.8-inf)' 187 ==> left='(-inf-0.2]' 186 acc:(0.99488)
```

Πίνακας 20. Κανόνες Συσχέτισης

Θα εμφανίσουμε τους παραπάνω κανόνες συσχέτισης με διαφορετική κωδικοποίηση σύμφωνα με τις 5 ομάδες που δημιουργήθηκαν σε κάθε κριτήριο για την καλύτερη κατανόησή τους.

1	Satisfaction=5	Average=3	Years=4	Accident=0	Promotion=0	Left=0	295	0.995
2	Projects=7	Left=1					256	0.99499
3	Satisfaction=4	Years=4	Accident=1	Left=0			214	0.99499
4	Satisfaction=5	Evaluation=4	Years=4	Promotion=0	Left=0		292	0.99499
5	Years=10	Left=0					208	0.99498
6	Satisfaction=5	Evaluation=4	Years=4	Accident=0	Left=0		248	0.99497
7	Evaluation=3	Years=4	Accident=1	Left=0			163	0.99496
8	Satisfaction=5	Projects=4	Average=2	Left=0			162	0.99496
9	Satisfaction=3	Evaluation=3	Projects=3	Years=4	Left=0		154	0.99495
10	Satisfaction=5	Evaluation=3	Projects=4	Left=0			151	0.99495
11	Satisfaction=5	Evaluation=2	Projects=4	Left=0			147	0.99494
12	Satisfaction=3	Evaluation=5	Projects=4	Left=0			141	0.99494
13	Evaluation=2	Projects=4	Average=3	Left=0			140	0.99494
14	Evaluation=3	Average=3	Years=4	Left=0			291	0.99493
15	Satisfaction=5	Years=4	Accident=1	Left=0			209	0.99492
16	Satisfaction=4	Projects=3	Accident=1	Left=0			133	0.99492
17	Satisfaction=4	Evaluation=2	Projects=3	Years=4	Promotion=0	Left=0	130	0.99492
18	Satisfaction=5	Evaluation=2	Average=2	Left=0			123	0.9949
19	Satisfaction=5	Evaluation=2	Average=3	Left=0			120	0.9949
20	Average=3	Years=4	Accident=1	Left=0			186	0.99488

Θα αναλύσουμε τους κανόνες με την μεγαλύτερη τιμή support (υποστήριξη):

Ο 1^{ος} κανόνας αναφέρει ότι αν ο εργαζόμενος είναι ικανοποιημένος από την εταιρεία, έχει υψηλό μέσο όρο ωρών εργασίας, εργάζεται για τέσσερα χρόνια, δεν είχε εργατικό ατύχημα και δεν πήρε προαγωγή τα τελευταία πέντε χρόνια τότε παρέμεινε στην εταιρεία.

Ο 4^{ος} κανόνας με support=292 αναφέρει ότι αν ο εργαζόμενος είναι ικανοποιημένος, πήρε υψηλό σκορ στην τελευταία αξιολόγηση, εργάζεται τέσσερα χρόνια και δεν πήρε προαγωγή τα τελευταία πέντε χρόνια, θα παραμείνει στην εταιρεία.

Ο 2^{ος} κανόνας με support=256 μας δείχνει ότι αν στον εργαζόμενο είχαν ανατεθεί εφτά προτζεκτ τότε αυτός αποχώρησε από την εταιρεία.

Ο 6^{ος} κανόνας μας δείχνει ότι αν ο εργαζόμενος ήταν ικανοποιημένος από την εταιρεία, είχε υψηλή βαθμολογία στην τελευταία αξιολόγηση, εργάζεται για τέσσερα χρόνια και δεν είχε εργατικό ατύχημα τότε αυτός παρέμεινε στην εταιρεία.

Ο 3^{ος} κανόνας αναφέρει ότι αν ο εργαζόμενος ήταν ικανοποιημένος, εργαζόταν για τέσσερα χρόνια και είχε εργατικό ατύχημα τότε παρέμεινε στην εταιρεία.

Κεφάλαιο 11: Σύγκριση αποτελεσμάτων της μεθόδου fsQCA και της Apriori

Σύμφωνα με την μέθοδο fsQCA με χρήση Undersampling τα αποτελέσματα στα οποία καταλήξαμε είναι τα παρακάτω αιτιώδη μονοπάτια:

Σύμφωνα με την ενδιάμεση λύση προκύπτουν τρία μονοπάτια που οδηγούν τους εργαζόμενους να αποχωρήσουν από την εταιρεία. Αρχικά το πρώτο μονοπάτι ορίζει ότι αν ο εργαζόμενος δεν έχει εργατικό ατύχημα, έχει πραγματοποιήσει πολλά προτζεκτ, δεν είναι ικανοποιημένος, έχει υψηλή βαθμολογία στην αξιολόγηση, υψηλό μέσο όρο ωρών εργασίας και δεν έχει πάρει αύξηση τα τελευταία 5 χρόνια τότε θα αποχωρήσει.

Το δεύτερο μονοπάτι θεωρεί ότι αν δεν υπάρχει εργατικό ατύχημα, έχει δουλέψει πολλά χρόνια, έχει υψηλό αριθμό προτζεκτ, υψηλή βαθμολογία στην αξιολόγηση, υψηλό μέσο όρο ωρών εργασίας μηνιαίως και δεν έχει πάρει αύξηση τότε θα αποχωρήσει.

Το τρίτο και τελευταίο μονοπάτι μας δείχνει ότι αν δεν υπάρχει εργατικό ατύχημα, δεν έχει δουλέψει πολλά χρόνια στην εταιρεία, δεν έχει υψηλό αριθμό προτζεκτ, δεν είναι ικανοποιημένος, δεν είχε υψηλή βαθμολογία στην τελευταία αξιολόγηση, δεν έχει υψηλό μέσο όρο ωρών εργασίας μηνιαίως στην εταιρεία και δεν έχει λάβει και αύξηση τα τελευταία 5 χρόνια τότε θα οδηγηθεί στην αποχώρηση του.

Στην μέθοδο αυτή βλέπουμε ότι το solution coverage και solution consistency είναι σε πολύ υψηλές τιμές 0,709775 και 0,728191 αντίστοιχα. Πολύ πιο ικανοποιητικές τιμές σε σχέση με πριν γι' αυτό και η μέθοδος Undersampling όπως και τα αποτελέσματά της θεωρούνται πιο αξιόπιστα.

Από αυτά τα τρία μονοπάτια αυτό που θα επιλέξω σαν πιο αξιόπιστο είναι αυτό με το μεγαλύτερο unique coverage, όπου το έχει το τρίτο μονοπάτι 0,39794 όπου μας δείχνει ότι περίπου το 39% των περιπτώσεων καλύπτεται-επεξηγούνται από αυτό το μονοπάτι

Σύμφωνα με την μέθοδο Apriori καταλήξαμε στους παρακάτω 5 πιο επικρατέστερους κανόνες συσχέτισης σύμφωνα με την μεγαλύτερη τιμή support.

Ο 1^{ος} κανόνας με το μεγαλύτερο support αναφέρει ότι αν ο εργαζόμενος είναι ικανοποιημένος από την εταιρεία, έχει υψηλό μέσο όρο ωρών εργασίας, εργάζεται για τέσσερα χρόνια, δεν είχε εργατικό ατύχημα και δεν πήρε προαγωγή τα τελευταία πέντε χρόνια τότε παρέμεινε στην εταιρεία.

Ο 4^{ος} κανόνας με support=292 αναφέρει ότι αν ο εργαζόμενος είναι ικανοποιημένος, πήρε υψηλό σκορ στην τελευταία αξιολόγηση, εργάζεται τέσσερα χρόνια και δεν πήρε προαγωγή τα τελευταία πέντε χρόνια, θα παραμείνει στην εταιρεία.

Ο 2^{ος} κανόνας με support=256 μας δείχνει ότι αν στον εργαζόμενο είχαν ανατεθεί εφτά προτζεκτ τότε αυτός αποχώρησε από την εταιρεία.

Ο 6^{ος} κανόνας μας δείχνει ότι αν ο εργαζόμενος ήταν ικανοποιημένος από την εταιρεία, είχε υψηλή βαθμολογία στην τελευταία αξιολόγηση, εργάζεται για τέσσερα χρόνια και δεν είχε εργατικό ατύχημα τότε αυτός παρέμεινε στην εταιρεία.

Ο 3^{ος} κανόνας αναφέρει ότι αν ο εργαζόμενος ήταν ικανοποιημένος, εργαζόταν για τέσσερα χρόνια και είχε εργατικό ατύχημα τότε παρέμεινε στην εταιρεία.

Σύμφωνα με τις δύο παραπάνω μεθόδους και τα επικρατέστερα αποτελέσματά τους καταλήγουμε στο συμπέρασμα ότι και στις δύο το κύριο χαρακτηριστικό που τους οδηγεί στην αποχώρηση ή την παραμονή στην εταιρεία είναι αν δεν είναι ή είναι αντίστοιχα ικανοποιημένοι από αυτήν.

Κεφάλαιο 12: Συμπεράσματα

Ο σκοπός της παρούσας εργασίας ήταν η μελέτη της σχέσης μιας ομάδας κριτηρίων με την αποχώρηση των εργαζομένων από μια υποθετική εταιρεία. Τα κριτήρια με βάση τα οποία έγινε η αξιολόγηση είχαν να κάνουν με την εμπειρία του εργαζόμενου στην εταιρεία.

Η μέθοδος ποιοτικής συγκριτικής ανάλυσης με ασαφή σύνολα (fs/QCA) χρησιμοποιήθηκε για την διερεύνηση της σχέσης μεταξύ των κριτηρίων και της αποχώρησης του εργαζόμενου. Με αυτόν τον τρόπο ήταν δυνατό να προσδιοριστεί ποιοι αιτιώδεις συνδυασμοί απαιτούνται ώστε ο εργαζόμενος να αποχωρήσει.

Η διαδικασία της fs/QCA όπως έχει ήδη αναφερθεί αποτελείται αρχικά από την διαδικασία της βαθμονόμησης και στην συνέχεια σειρά έχει η αναζήτηση για τυχόν αναγκαίες συνθήκες. Δημιουργείται ο πίνακας αλήθειας και εξάγονται από αυτόν οι τρεις λύσεις που προκύπτουν, η σύνθετη (Complex), η φειδωλή (Parsimonious), και η ενδιάμεση (Intermediate). Τα αιτιώδη μονοπάτια που προέκυψαν μας αναλύουν τους λόγους που οδήγησαν τους εργαζόμενους στην αποχώρηση. Στην συγκεκριμένη περίπτωση προκύπτουν τέσσερα μονοπάτια αλλά λόγω του χαμηλού solution consistency που έχει 0,507460 παρόλο της ικανοποιητικής κάλυψης solution coverage (0,755757) η λύση αυτή δεν είναι πολύ αξιόπιστη και τα αποτελέσματά της δεν είναι ικανοποιητικά.

Πραγματοποιήθηκε επιπλέον η μέθοδος Apriori όπου στόχος μας είναι να βρούμε τους κανόνες συσχέτισης με την υψηλότερη τιμή support για μία ορισμένη ελάχιστη τιμή confidence (εμπιστοσύνης) έτσι καταλήγουμε στους 5 πιο επικρατέστερους κανόνες συσχέτισης. Ο κανόνας με την μεγαλύτερη τιμή confidence αναφέρει ότι αν ο εργαζόμενος είναι ικανοποιημένος από την εταιρεία, έχει υψηλό μέσο όρο ωρών εργασίας, εργάζεται για τέσσερα χρόνια, δεν είχε εργατικό ατύχημα και δεν πήρε προαγωγή τα τελευταία πέντε χρόνια τότε παρέμεινε στην εταιρεία.

Στην συνέχεια χρησιμοποιήθηκε η διαδικασία του Undersampling, η οποία είναι μια μέθοδος για την εξισορρόπηση μη ισορροπημένων συνόλων δεδομένων. Με την μείωση των δεδομένων περίπου κατά 50%, και πραγματοποιώντας την ίδια διαδικασία με πριν, καταλήγουμε στην εύρεση 3 αιτιωδών μονοπατιών. Στην μέθοδο αυτή βλέπουμε ότι το solution coverage και solution consistency είναι σε πολύ υψηλές τιμές 0,71 και 0,73 αντίστοιχα, πολύ πιο ικανοποιητικές τιμές σε σχέση με πριν, γι' αυτό και η μέθοδος Undersampling, όπως και τα αποτελέσματά της θεωρούνται πιο αξιόπιστα.

Από αυτά τα τρία μονοπάτια επιλέχθηκε αυτό με το μεγαλύτερο unique coverage, το 3^ο μονοπάτι το οποίο μας δείχνει ότι αν δεν υπάρχει εργατικό ατύχημα, ο εργαζόμενος δεν έχει δουλέψει πολλά χρόνια στην εταιρεία, δεν έχει υψηλό αριθμό προτζεκτ, δεν είναι ικανοποιημένος, δεν είχε υψηλή βαθμολογία στην τελευταία αξιολόγηση, δεν έχει υψηλό μέσο όρο ωρών εργασίας μηνιαίως στην εταιρεία και δεν έχει λάβει και αύξηση τα τελευταία 5 χρόνια τότε θα οδηγηθεί στην αποχώρηση του.

Πραγματοποιήσαμε στην συγκεκριμένη περίπτωση έναν ακόμα έλεγχο, όπου με την στατιστική ανάλυση χ^2 ελέγχθηκαν αν οι μεταβλητές work accident και promotion, είναι ανεξάρτητες ή όχι με την αποχώρηση των εργαζομένων. Από τα αποτελέσματα προέκυψε ότι και οι δύο αυτές μεταβλητές δεν είναι ανεξάρτητες σε σχέση με το αποτέλεσμα.

Στην συνέχεια ορίσαμε σαν Outcome Μεταβλητή την απουσία της αποχώρησης των εργαζομένων άρα ότι παρέμειναν στην εταιρεία προκύπτουν τέσσερα αιτιώδη μονοπάτια. Στην μέθοδο αυτή βλέπουμε ότι το solution coverage και solution consistency είναι σε πολύ υψηλές τιμές 0,405986 και 0,868957 αντίστοιχα

Από αυτά τα τέσσερα μονοπάτια αυτό που θα επιλέξω σαν πιο αξιόπιστο είναι αυτό με το μεγαλύτερο unique coverage, όπου το έχει το τρίτο μονοπάτι 0,089795 όπου μας δείχνει ότι περίπου το 8% των συνολικών περιπτώσεων καλύπτεται-επεξηγούνται από αυτό το μονοπάτι. Πιο αναλυτικά μας δείχνει ότι αν ο εργαζόμενος δεν εργάζεται πολλά χρόνια στην εταιρεία, είναι ικανοποιημένος, δεν έχει υψηλή βαθμολογία στην τελευταία αξιολόγηση, δεν έχει πάρει αύξηση και δεν έχει εργατικό ατύχημα, τότε θα παραμείνει στην εταιρεία.

ΒΙΒΛΙΟΓΡΑΦΙΑ

- Alao, D. A. B. A., & Adeyemo, A. B. (2013). Analyzing employee attrition using decision tree algorithms. *Computing, Information Systems, Development Informatics and Allied Research Journal*, 4(1), 17-28.
- Alduayj, S. S., & Rajpoot, K. (2018, November). Predicting employee attrition using machine learning. In *2018 international conference on innovations in information technology (iit)* (pp. 93-98). IEEE.
- Doss, S. (2022). Analytic Approach of Predicting Employee Attrition Using Data Science Techniques.
- Dutta, Shawni, Samir Kumar Bandyopadhyay, and S. Kumar Bandyopadhyay. "Employee attrition prediction using neural network cross validation method." *International Journal of Commerce and Management Research* 6.3 (2020): 80-85.
- Jadhav, A. (2021). Churn Prediction of Employees Using Machine Learning Techniques. *Tehnički glasnik*, 15(1), 51-59.
- Krassadaki, E., & Grigoroudis, E. (2018). Analyzing perceived quality of health care services: A multicriteria decision analysis approach based on the theory of attractive quality. In *Preference Disaggregation in Multiple Criteria Decision Analysis* (pp. 211-235). Springer, Cham.
- Krassadaki, E., & Grigoroudis, E. (2018). Analyzing perceived quality of health care services: A multicriteria decision analysis approach based on the theory of attractive quality
- Legewie, N. (2013, September). An introduction to applied data analysis with qualitative comparative analysis. In *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research* (Vol. 14, No. 3).
- Lin, W. C., Tsai, C. F., Hu, Y. H., & Jhang, J. S. (2017). Clustering-based undersampling in class-imbalanced data. *Information Sciences*, 409, 17-26.
- Witte, R. S., & Witte, J. S. (2017). *Statistics*. John Wiley & Sons.v
- Yuan, X. (2017, March). An improved Apriori algorithm for mining association rules. In *AIP conference proceedings* (Vol. 1820, No. 1, p. 080005). AIP Publishing LLC.
- Κύρκος, Ε. (2016). Επιχειρηματική ευφυΐα και εξόρυξη δεδομένων.
- Ragin, C. C. (2013). New directions in the logic of social inquiry. *Political Research*
- Ragin, C. C., & Rihoux, B. (2004). Qualitative comparative analysis (QCA): State of the art and prospects. *Qualitative Methods*, 2(2), 3-13.
- Schneider, M. R., Schulze-Bentrop, C., & Paunescu, M. (2010). Mapping the institutional capital of high-tech firms: A fuzzy-set analysis of capitalist variety and export performance. *Journal of International Business Studies*, 41(2), 246-266.