

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

**ΣΧΕΔΙΟ ΠΑΡΟΥΣΙΑΣΗΣ
ΑΤΟΜΙΚΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ**

Μάιος 2022

Ατομική Διπλωματική Εργασία

Επιχειρηματολογία στην Μηχανική Μάθηση στην Ιατρική

Εφαρμογή στην αναγνώριση ενδομήτριου καρκίνου

Αθηνά Γαβριελίδου

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Μάιος 2022

ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΥΠΡΟΥ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Επιχειρηματολογία στην Μηχανική Μάθηση στην Ιατρική
Εφαρμογή στην αναγνώριση ενδομήτριου καρκίνου
Αθηνά Γαβριελίδου

Επιβλέπων Καθηγητής
Αντώνης Κάκας

Η Ατομική Διπλωματική Εργασία υποβλήθηκε προς μερική εκπλήρωση των απαιτήσεων απόκτησης του πτυχίου Πληροφορικής του Τμήματος Πληροφορικής του Πανεπιστημίου Κύπρου

Μάιος 2022

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον καθηγητή του τμήματος Πληροφορικής και επιβλέπων καθηγητή μου κ. Αντώνη Κάκα για την συνεχή βοήθεια και υποστήριξη που μου παρείχε, καθ' όλη την διάρκεια των δύο εξαμήνων όπου χρειάστηκαν για να ολοκληρώσω την πτυχιακή μου εργασία. Θα ήθελα να τον ευχαριστήσω για την πάντα άμεση και ακριβή ανταπόκριση σε κάθε δυσκολία που αντιμετώπισα, καθώς και τις χρήσιμες συμβουλές που μου παρείχε.

Ακόμη, δεν θα μπορούσα να μην ευχαριστήσω τον κ. Μάριο Νεοφύτου για τις γνώσεις που μου παρείχε όσο αφορά το κομμάτι του καρκίνου και τις υποδείξεις που μου έκανε όσο αφορά τα δεδομένα που έπρεπε να χρησιμοποιήσω καθώς και την πολύτιμη καθοδήγηση που μου παρείχε όσον αφορά την διαδικασία κατανόησης και ανάλυσης του απαιτούμενου πλαισίου περιεχομένου.

Πίνακας Περιεχομένων

Ευχαριστίες.....	4
Πίνακας Περιεχομένων.....	5
Κεφάλαιο 1 – Εισαγωγή	7
1.1 Γενική Εισαγωγή.....	7
1.2 Διατύπωση προβλήματος.....	7
1.3 Στόχος πτυχιακής έρευνας	8
1.4 Δομή πτυχιακής έρευνας.....	8
Κεφάλαιο 2 – Ανασκόπηση τεχνολογιών	10
2.1 Μηχανική Μάθηση	10
2.2 Επιχειρηματολογία.....	11
2.3 Επεξηγήσεις στην Επιχειρηματολογία	13
Κεφάλαιο 3 – Ανασκόπηση του προβλήματος.....	15
3.1 Καρκίνος	15
3.1.1 Γυναικολογικός καρκίνος.....	15
3.2 Προηγούμενη εργασία στο πρόβλημα.....	16
Κεφάλαιο 4 – Μέθοδος ανάλυσης των δεδομένων.....	18
4.1 Δεδομένα.....	18
4.2 Διαδικασία	19
4.3 Επεξηγήσεις επιχειρηματολογίας	25
4.4 Ομαδοποιήσεις επεξηγήσεων επιχειρηματολογίας	31
Κεφάλαιο 5 – Αποτελέσματα	38
5.1 Αξιολόγηση δεδομένων εκπαίδευσης.....	38
5.2 Αξιολόγηση δεδομένων ελέγχου	40
5.3 Επεξήγηση	42
5.4 Συμπεράσματα	44
Κεφάλαιο 6 – Συζήτηση	45
6.1 Σύντομη περίληψη	45

6.2 Σύντομη κριτική.....	46
6.3 Σύντομη συζήτηση.....	46
Βιβλιογραφία	48
Παράρτημα	49

Κεφάλαιο 1

1. Εισαγωγή

1.1 Γενική Εισαγωγή

Η ασθένεια του καρκίνου από καταβολής της αποτελούσε ένα μείζον ζήτημα το οποίο απασχολούσε την επιστημονική κοινότητα. Στο σύγχρονο επιστημονικό περιβάλλον παρατηρείται η συνεργασία πολλαπλών επιστημών με στόχο την εύρεση μεθόδων πρόγνωσης και θεραπείας της συγκεκριμένης ασθένειας. Στα πλαίσια της εν λόγω διεπιστημονικής έρευνας ιδιαίτερα ενεργητικό ρόλο έχουν τεχνολογίες της Πληροφορικής όπως η Τεχνική Νοημοσύνη, Μηχανική Μάθηση και επιχειρηματολογία, πάντα σε συνδυασμό με Ιατρικές πρακτικές και μεθοδολογίες.

Εστιάζοντας στην παρούσα έρευνα, γίνεται ενασχόληση με τον γυναικολογικό καρκίνο και τις μεθοδολογίες έγκαιρης διάγνωσης του. Εξερευνάται, δηλαδή, το κατά πόσο διάφορα δεδομένα τα οποία συλλέγονται με της παραδοσιακές ιατρικές μεθόδους, μπορούν να χρησιμοποιηθούν από προγράμματα Τεχνικής Νοημοσύνης και επιχειρηματολογίας ούτως ώστε να βοηθήσουν στην έγκαιρη και έγκυρη διάγνωση των καρκινωμάτων στα γυναικεία γεννητικά όργανα.

1.2 Διατύπωση προβλήματος

Τα τελευταία χρόνια παρατηρείται μεγάλη έξαρση της ασθένειας του καρκίνου, ειδικότερα του γυναικολογικού. Λαμβάνοντας υπόψη το γεγονός ότι η εν λόγω ασθένεια έχει βαρυσήμαντες και επίπονες επιπτώσεις στην υγεία των ασθενών, καθίσταται απαραίτητη η ανάγκη για εύρεση αποτελεσματικού τρόπου αντιμετώπισης του προβλήματος.

Η υφιστάμενη βιβλιογραφία έχει αποδείξει την σημαντικότητα της γρήγορης πρόγνωσης του γυναικολογικού καρκίνου, καθώς επιτρέπει την πιο αποδοτική αντιμετώπιση του, προτού επεκταθεί και προσβάλει πολλαπλά όργανα.

Το μεγάλο πρόβλημα είναι το γεγονός ότι δεν μπορεί εύκολα να εντοπιστούν οι όγκοι που προκαλούν καρκινώματα καθώς και το γεγονός ότι δεν μπορείς εύκολα να ξεχωρίσεις έναν καλοήγη από έναν κακοήγη όγκο. Το γεγονός αυτό δεν επιτρέπει την εύκολη αναγνώριση των κακοήθεις όγκων σε πρώιμα στάδια με τις σύγχρονες ιατρικές πρακτικές. Για τον λόγο αυτό είναι επιτακτική η ανάγκη για δημιουργία συστημάτων που να επιτρέπουν την ανάλυση των εικόνων των όγκων που συλλέγονται μέσω ενδοσκόπησης, υστεροσκόπησης και η λαπαροσκόπησης ούτως ώστε να τους κατατάσσουν σε καλοήθεις και κακοήθεις.

1.3 Στόχος πτυχιακής έρευνας

Η εν λόγω έρευνα έχει ως κύριο στόχο την δημιουργία ενός συστήματος επιχειρηματολογίας το οποίο θα μπορεί να βοηθήσει στην πρώιμη πρόγνωση του γυναικολογικού καρκίνου. Δηλαδή, ενός συστήματος το οποίο θα είναι υποβοηθητικό για τους ιατρούς, ούτως ώστε να μπορούν πιο εύκολα να διαχωρίζουν τις κατηγορίες όγκων. Για να γίνει αυτό εφικτό υπάρχει η ανάγκη για:

- a) Δημιουργία κανόνων οι οποίοι θα αποφασίζουν
- b) Δημιουργία κανόνων που καλύπτουν τις πλείστες περιπτώσεις όγκων

Αποδοτικοί κανόνες

Υπάρχει η ανάγκη μέσω της ανάλυσης των υπάρχοντων δεδομένων να εντοπιστούν κανόνες για τις τιμές των χαρακτηριστικών των όγκων, οι οποίοι θα κατατάσσουν έξυπνα τους όγκους σε καλοήθεις και κακοήθεις.

Πληρότητα κανόνων

Πέρα από την αποδοτικότητα των κανόνων υπάρχει η ανάγκη για εντοπισμό κανόνων οι οποίοι θα καλύπτουν την όλες τις περιπτώσεις, δηλαδή, κανόνες οι οποίοι θα μπορούν να κατατάξουν όλες τις περιπτώσεις καρκίνων σε μια κατηγορία. Πρέπει δηλαδή να δημιουργηθεί ένα σύνολο από κανόνες, το οποίο σε κάθε περίπτωση να μπορεί να αποφασίσει και να αιτιολογήσει αν κάποιος καρκίνος είναι καλοήθης ή κακοήθης.

1.4 Δομή πτυχιακής έρευνας

- Το κεφάλαιο 2 παρουσιάζει τις διάφορες τεχνολογίες που αναμένεται να χρησιμοποιηθούν κατά την εκπόνηση της συγκεκριμένης έρευνας. Παρουσιάζουν τις βασικές τεχνικές και χαρακτηριστικά τόσο των διαφόρων αλγορίθμων Μηχανικής Μάθησης όσο και των μοντέλων επιχειρηματολογίας.
- Το κεφάλαιο 3 παρουσιάζει το επιστημονικό υπόβαθρο στο οποίο στηρίζεται η έρευνα. Γίνεται αναφορά στα διάφορα χαρακτηριστικά του γυναικολογικού καρκίνου, που αποτελεί αντικείμενο μελέτης, και παρουσιάζει τα ευρήματα της υφιστάμενης βιβλιογραφίας στον τομέα της πρόγνωσης με ενδοσκοπικές μεθόδους.
- Το κεφάλαιο 4 παρουσιάζει την μορφή της συλλογής δεδομένων που χρησιμοποιήθηκε στα πλαίσια της έρευνας και τις μεθοδολογίες που ακολουθήθηκαν καθ' όλη τη διάρκεια εκπόνησης της μελέτης και ενδεικτικές επεξηγήσεις που επιστρέφει το Gorgias Cloud.
- Το κεφάλαιο 5 παρουσιάζει τα αποτελέσματα που προέκυψαν από την έρευνα μαζί με διάφορες επεξηγήσεις.

- Το κεφάλαιο 6 κάνει μια μικρή περίληψη διαδικασίας που ακολουθήθηκε καθ' όλη την διάρκεια της εκπόνησης της εν λόγω έρευνας, παρουσιάζει επιγραμματικά τα διάφορα αποτελέσματα και συμπεράσματα που προέκυψαν και κλείνει με μια σύντομη συζήτηση.
- Το παράρτημα παρουσιάζει ενδεικτικά παραδείγματα της μορφής των δεδομένων καθώς και του κώδικα επιχειρηματολογίας που δημιουργήθηκε στο Cloud Gorgias.

Κεφάλαιο 2

2. Ανασκόπηση τεχνολογιών

2.1 Μηχανική Μάθηση

Μηχανική Μάθηση [4] [5]

Αρχικά, αξίζει να αναφερθεί η σημασία και η ιδιότητες της μάθησης. Σε ένα γνωστικό σύστημα -το οποίο μπορεί να το αντιληφθεί κάποιος στην καθημερινή ζωή, η μάθηση συνδέεται με δύο βασικές ιδιότητες. Η πρώτη από αυτές τις ιδιότητες, είναι η ικανότητα της αύξησης των ήδη υπάρχον γνώσεων την στιγμή της αλληλεπίδρασης με το περιβάλλον, και η δεύτερη βασική ιδιότητα είναι η ικανότητα που έχει να βελτιώνει και να διορθώνει μέσα από την επανάληψη τον τρόπο με τον οποίο εκτελείτε μία ενέργεια [5].

Ο Άρθουρ Σάμουελ το 1959 όρισε τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί" [4].

Η Μηχανική Μάθηση είναι ένα από τα πολλά πεδία της επιστήμης των υπολογιστών, το οποίο έχει αναπτυχθεί από την μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Διερευνά την μελέτη και την κατασκευή διαφόρων αλγορίθμων, οι οποίοι θα μπορούν να μαθαίνουν από τα δεδομένα που λαμβάνουν και βάσει αυτών να κάνουν τις διάφορες προβλέψεις ή να καταλήγουν σε κάποια συμπεράσματα που θα εκφράζονται και ως το αποτέλεσμα.

Η εφαρμογή της μηχανικής μάθησης γίνεται σε μια σειρά από διάφορες υπολογιστικές εργασίες, στις οποίες υπάρχει ανέφικτος σχεδιασμός και ρητός προγραμματισμός αλγορίθμων. Διάφορα παραδείγματα αυτών των εφαρμογών είναι τα spam filtering, η οπτική αναγνώριση χαρακτήρων (OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση.

Ανάλογα με το πρόβλημα που υπάρχει και την φύση αυτού, χρησιμοποιείται και η ανάλογη τεχνική μηχανικής μάθησης η οποία ανήκει σε ένα από τα δύο είδη μηχανικής μάθησης.

Το πρώτο από τα δύο αυτά είδη είναι η Μάθηση με επίβλεψη (Supervised Learning), ή αλλιώς Μάθηση με Παραδείγματα (Learning from Examples). Όπου το σύστημα θα πρέπει επαγωγικά να 'μάθει' μια συνάρτηση ονομαζόμενη ως συνάρτηση στόχος (target function) η οποία αποτελεί την έκφραση του μοντέλου το οποίο περιγράφει τα δεδομένα. Η συνάρτηση αυτή, χρησιμοποιείται για να προβλέπει βάση των τιμών που υπάρχουν σε ένα σύνολο μεταβλητών -ανεξάρτητων μεταβλητών ή μεταβλητών εισόδου, την τιμή κάποιας μεταβλητής ονομαζόμενη ως μεταβλητή στόχος ή μεταβλητή εξόδου.

Το δεύτερο είδος μάθησης είναι η Μάθηση χωρίς Επίβλεψη (Unsupervised Learning) ή Μάθηση από Παρατήρηση (Learning from Observation). Σε αυτό το είδος, στόχος του

συστήματος, είναι να ανακαλύπτει συσχετίσεις και ομάδες μέσα από τα δεδομένα, βασισμένο στις ιδιότητες τους. Αποτέλεσμα αυτού του είδους είναι τα διάφορα πρότυπα που προκύπτουν όπου κάθε ένα από αυτά περιγράφει ένα μέρος από τα δεδομένα [5].

2.2 Επιχειρηματολογία

Η επιχειρηματολογία ή αλλιώς η θεωρία επιχειρημάτων, είναι μια διεπιστημονική μελέτη για τον τρόπο που μπορούν να εξαχθούν τα διάφορα συμπεράσματα από τις προϋποθέσεις μέσω της λογικής συλλογιστικής και περικλείει τις τέχνες και επιστήμες της πολιτικής συζήτησης, της πειθούς, της συνομιλίας και του διαλόγου τόσο του εριστικού όσο και του διδακτικού. Αυτές η τέχνες και επιστήμες, χρησιμοποιούνται από τους ανθρώπους για να προστατεύσουν τα προσωπικά τους συμφέροντα και τις πεποιθήσεις τους. Επίσης, μελετά τους κανόνες των συμπερασμάτων, της λογικής και τους διαδικαστικούς κανόνες όχι μόνο σε τεχνητό αλλά και σε πραγματικό κόσμο.

Γενικά η επιχειρηματολογία σύμφωνα με την βιβλιογραφία [3] χρησιμοποιείται για την παροχή υποστήριξης για λήψη αποφάσεων. Ιδιαίτερα χρήσιμη είναι όταν υπάρχουν προβλήματα τα οποία στηρίζονται σε σύνθετες πολιτικές προτιμήσεων. Μέσω λοιπόν της επιχειρηματολογίας επιτρέπεται η δημιουργία συστημάτων που επιτρέπουν την ύπαρξη πολλαπλών επιχειρημάτων που πιθανόν να έρχονται σε αντιπαράθεση, παρέχοντας διάφορους κανόνες προτεραιότητας που επιτρέπουν με βάση το εκάστοτε σενάριο να ληφθούν οι κατάλληλες αποφάσεις. Αναλυτικότερα, η επιχειρηματολογία επιτρέπει ανά πάσα χρονική στιγμή να υπάρχουν ενεργά διλήμματα ως προς την απόφαση που πρέπει να ληφθεί και τα εν λόγω διλήμματα να καταλήγουν σε διαφορετικές αποφάσεις σε ενδεχόμενα διαφορετικά σενάρια.

Ένα ενδεικτικό παράδειγμα όπως της βιβλιογραφίας [3] που δείχνει την ιεραρχική δομή που επιτρέπει η επιχειρηματολογία είναι το εξής:

```
complement(malignant(Tumor), benign(Tumor)).  
complement(benign(Tumor), malignant(Tumor)).
```

```
rule(r1(Tumor), benign(Tumor),[]).
```

```
rule(r2(Tumor), benign(Tumor),[]):-big_tumor(Tumor).  
rule(pr1(Tumor), prefer(r2(Tumor), r1(Tumor)),[]).
```

```
rule(pr2(Tumor),prefer(r1(Tumor), r2(Tumor)),[]):-light_colour(Tumore).  
rule(c1(Tumor),prefer(pr2(Tumor),pr1(Tumor)),[]).
```

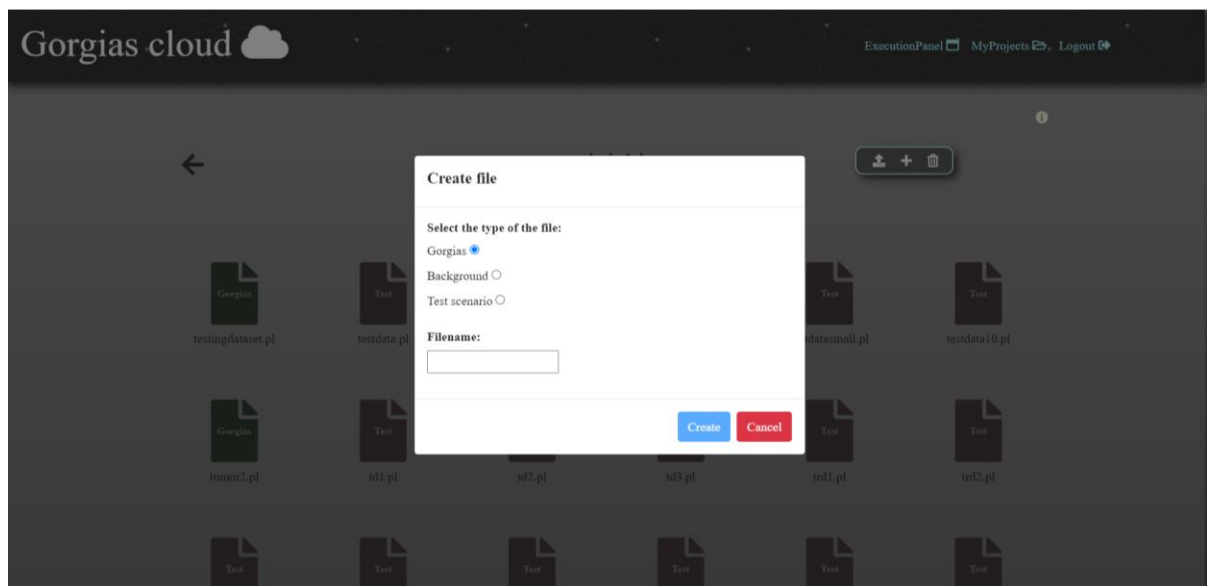
Στο πιο πάνω παράδειγμα στο πρώτο επίπεδο υπάρχει ο κανόνας 1 ο οποίος δηλώνει ότι γενικά κάποιος όγκος είναι καλοήθης. Σε δεύτερο επίπεδο υπάρχει ο κανόνας 2 ο οποίος λέει ότι αν ο όγκος είναι μεγάλος τότε είναι κακοήθης. Ο κανόνας του δεύτερου επιπέδου

υπερέχει του πρώτου ένεκα των προτεραιοτήτων και επομένως αν ισχύει ο κανόνας 2 αναιρείται ο κανόνας 1. Σε τρίτο επίπεδο υπάρχει ο κανόνας 3 ο οποίος λέει ότι αν το χρώμα του είναι ανοιχτό τότε είναι καλοήθης. Ο εν λόγω κανόνας υπερέχει της προτεραιότητας του δεύτερου επιπέδου και την αναιρεί.

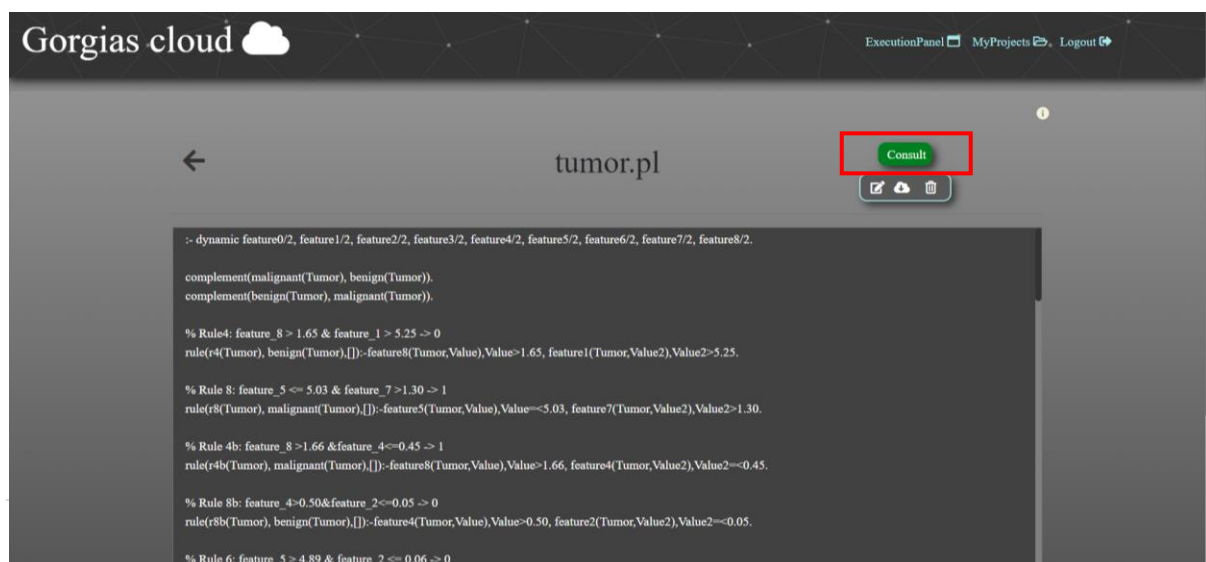
Μια τεχνολογία που χρησιμοποιείται ως πλαίσιο για την ανάπτυξη του προγράμματος επιχειρηματολογίας είναι το πλαίσιο Gorgias. Στο εν λόγω πλαίσιο επιτρέπεται ανάπτυξη προγραμμάτων επιχειρηματολογίας, η δημιουργία ανεξάρτητων σεναρίων εκτέλεσης και ο έλεγχος των αποφάσεων του προγράμματος επιχειρηματολογίας με τις αντίστοιχες επεξηγήσεις.

Αναλυτικότερα στο Execution Panel ο χρήστης μπορεί επιλέγοντας το «+» να δημιουργήσει ένα νέο αρχείο το οποίο μπορεί να είναι:

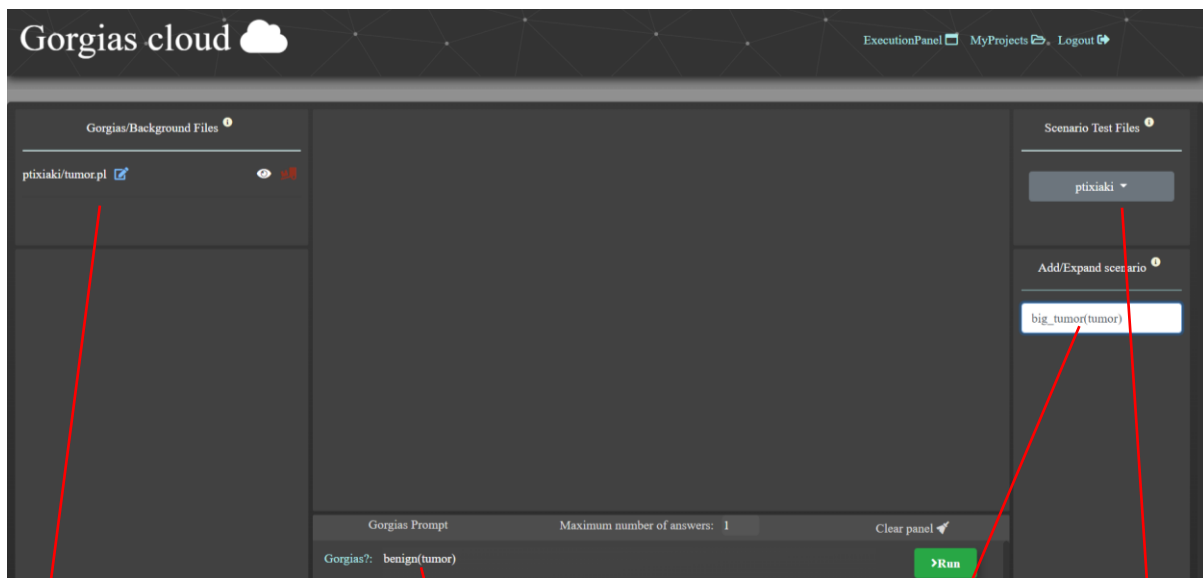
- a) Κώδικας επιχειρηματολογίας – επέλεξε Gorgias
- b) Σενάριο εκτέλεσης – επέλεξε Test Scenario



Όταν ολοκληρωθεί ο κώδικας, για να μπορέσει ο χρήστης να ελέγξει τις αποφάσεις που λαμβάνει το μοντέλο επιχειρηματολογίας επιλέγει να κάνει το αντίστοιχο αρχείο “Gorgias” Consult.



Εν συνεχεία ο χρήστης μπορεί να ελέγξει τις αποφάσεις που λαμβάνει το πρόγραμμα επιχειρηματολογίας σε κάθε σενάριο.



Το πρόγραμμα
επιχειρηματολογίας το
οποίο έχει γίνει Consult

Προσθήκη της απόφασης
που θα ελεγχθεί αν ισχύει
ή όχι

Προσθήκη αρχείου με
συγκεκριμένο σενάριο

Δημιουργία νέου
σεναρίου χειροποίητα

2.3 Επεξηγήσεις στην Επιχειρηματολογία

Δεδομένης της υφιστάμενης βιβλιογραφίας, πλέον δημιουργούνται πολλαπλά συστήματα Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης τα οποία παρέχουν συμβουλευτική υποστήριξη στη λήψη αποφάσεων. Ιδιαίτερη εφαρμογή των συγκεκριμένων συστημάτων εντοπίζεται στον τομέα της Ιατρικής. Παρόλα αυτά, στα εν λόγω συστήματα διαπιστώνεται και μια μεγάλη ανάγκη όσον αφορά την κατάλληλη αιτιολόγηση. Αναλυτικότερα, ένα σύστημα όπως αυτά που περιγράφονται πιο πάνω εκτός από την αναγκαιότητα για ακρίβεια στην πρόβλεψη των αποτελεσμάτων, απαιτεί αντίστοιχη βαρύτητα στον επικοινωνία της πρόβλεψης στον χρήστη. Δηλαδή, η πρόβλεψη του συστήματος θα πρέπει να παρουσιαστεί με τέτοιο τρόπο στον χρήστη ούτως ώστε να είναι κατανοητή και αξιόπιστη. Για να είναι κατανοητή θα πρέπει να χρησιμοποιείται κατάλληλο συντακτικό και σημασιολογικό περιεχόμενο στις αποκρίσεις. Για να είναι αξιόπιστη, θα πρέπει να παρέχονται οι κατάλληλες αιτιολογήσεις, ώστε να μπορεί ο χρήστης να δει τους λόγους που οδήγησαν το σύστημα να κάνει την εν λόγω πρόβλεψη. Ιδιαίτερα στα Ιατρικά συστήματα είναι απαραίτητο εχέγγυο η

παροχή μιας ολοκληρωμένης αιτιολόγησης, ώστε να μπορεί ο Ιατρός να κρίνει αν οι διάφορες αιτιολογήσεις που οδήγησαν στην πρόβλεψη είναι έμπιστες ή όχι.

Για τον λόγο αυτό πλέον γίνεται μια έρευνα στον τομέα του Explainable AI (XAI)**Error! Reference source not found.**, δηλαδή στην προσπάθεια για δημιουργία αλγορίθμων που να παράγουν πλήρεις, κατανοητές και αξιόπιστες αιτιολογήσεις για κάθε πρόβλεψη που παράγουν τα συστήματα Τεχνητής Νοημοσύνης. Επομένως, το XAI δημιούργησε μια συλλογή από τεχνικές που χρησιμοποιούνται ούτως ώστε να είναι δυνατή η παραγωγή των κατάλληλων επεξηγήσεων για τα συστήματα Τεχνητής Νοημοσύνης.

Τεχνικές XAI για παραγωγή επεξηγήσεων με την χρήση Μηχανικής Μάθησης [6]**Error!**

Error! Reference source not found.:

1. Model Agnostic: Πρόκειται για μια τεχνική η οποία χειρίζεται το μοντέλου που δημιουργήθηκε σαν «μαύρο κουτί», το οποίο και εκπαιδεύει είτε με την παροχή παραδειγμάτων, είτε με την ιεραρχία σχετικότητας για τα διάφορα χαρακτηριστικά, είτε με την δημιουργία υπό-μοντέλων που επεξηγούν της προβλέψεις που παράγονται.
2. Interpretable Models: Πρόκειται για μια τεχνική η οποία δημιουργεί αυτό-επεξηγηματικά μοντέλα τα οποία ενσωματώνουν τις διάφορες ερμηνείες στην δομή τους και με βάση τα χαρακτηριστικά εισόδου παράγουν τις αντίστοιχες επεξηγήσεις.
3. Explaining DNN: Πρόκειται για μια τεχνική στην οποία διάφορα Βαθιά Νευρωνικά Δίκτυα δεχόμενα μια σειρά από δεδομένα προσπαθούν να ανακαλύψουν διάφορα κοινά μοτίβα ώστε να δημιουργήσουν αναπαραστάσεις.

Κεφάλαιο 3

3. Ανασκόπηση του προβλήματος

3.1 Καρκίνος

Ο καρκίνος ή κακοήθης όγκος είναι μια σοβαρή ασθένεια η οποία προκύπτει από την αφύσικη (μη φυσιολογική) κυτταρική ανάπτυξη και διαίρεση. Δηλαδή, τα κύτταρα αναπτύσσονται και συνεχίζουν να διαιρούνται ανεξέλεγκτα, σε αντίθεση με τα φυσιολογικά κύτταρα, τα οποία αυξάνονται, διαιρούνται και πεθαίνουν με αυστηρά ελεγχόμενο τρόπο. Με την ανεξέλεγκτη αυτή ανάπτυξη και διαίρεση, δημιουργείται μία μάζα κυττάρων, ονομαζόμενη ως όγκος. Αν ο όγκος αυτός είναι κακοήθης και όχι καλοήθης, μπορεί να εισβάλει σε γειτονικούς ιστούς ή να διαδοθεί σε άλλα μέρη του σώματος, προκαλώντας έτσι μετάσταση.

Ανά το παγκόσμιο, περίπου 10 εκατομμύρια άνθρωποι ετησίως, χάνουν την μάχη με τον καρκίνο. Ο καρκίνος, μπορεί να προσβάλει από παιδιά μέχρι ηλικιωμένους, γυναίκες και άντρες, ανεξαρτήτως φυλής. Μπορεί στο κάθε μέρος του σώματος να προκληθεί καρκίνος, ενώ υπάρχουν περισσότερα από 100 διαφορετικά γνωστά είδη καρκίνου. Κάποια από τα πιο γνωστά είδη καρκίνου είναι ο καρκίνος στους πνεύμονες, ο καρκίνος του παχέος εντέρου, ο γυναικολογικός καρκίνος, ο καρκίνος του στομάχου, ο καρκίνος του προστάτη, ο καρκίνος του μαστού, ο καρκίνος των οστών κ.α.

Όπως προαναφέρθηκε, ο κάθε άνθρωπος μπορεί να προσβληθεί από τον καρκίνο, ωστόσο κάποιοι διατρέχουν υψηλότερο κίνδυνο λόγω κληρονομικών παραγόντων (5%-10% των καρκίνων), ή λόγω της συμπεριφοράς του ή κάποιων περιβαλλοντικών παραγόντων (έκθεση στις ρυπαντικές ουσίες και την ακτινοβολία). Αρχικά, το κάπνισμα, είναι η πρώτη αιτία καρκίνου, αφού 12% των θανάτων οφείλονται σε αυτό. Ακόμη, η καθιστική ζωή (η έλλειψη της φυσικής δραστηριότητας), η παχυσαρκία, η κατανάλωση αλκοόλ και η ανθυγιεινή διατροφή προκαλούν ένα άλλο 10% των θανάτων από καρκίνο. Τέλος, στις αναπτυσσόμενες χώρες, παρατηρείται μια αύξηση του καρκίνου στο 20% που προκαλούνται από σεξουαλικά μεταδιδόμενα νοσήματα (όπως η ηπατίτιδα).

3.1.1 Γυναικολογικός καρκίνος

Ο γυναικολογικός καρκίνος ονομάζεται ο καρκίνος ο οποίος εμφανίζεται στο γυναικείο γενετικό σύστημα. Όπως, ο ενδομήτριος καρκίνος, ο καρκίνος της μήτρας, των ωοθηκών, του αιδοίου, του κόλπου και του τράχηλου της μήτρας.

Στην συγκεκριμένη πτυχιακή έρευνα θα ασχοληθούμε με τον γυναικολογικό καρκίνο και πιο συγκεκριμένα με τον ενδομήτριο καρκίνο.

3.2 Προηγούμενη εργασία στο πρόβλημα

Ένεκα της εξέλιξης της Μηχανικής Μάθησης και της τεχνολογίας γενικότερα, πλέον η τεχνική της επεξεργασίας και ανάλυσης εικόνων αυτόματα από συστήματα είναι ιδιαίτερα διαδεδομένη. Η συγκεκριμένη τεχνολογία βρίσκει εφαρμογή και στον τομέα της Ιατρικής, σε διάφορους τομείς όπως η ογκολογία, καρδιολογία, νευρολογία. Σύγχρονα συστήματα μπορούν να χρησιμοποιήσουν τις εικόνες οι οποίες συλλέγηκαν X-ray, MRI, ultrasounds και να τις αναλύσουν αυτόματα ούτως ώστε να εκτελέσουν διάφορες εργασίες οι οποίες θα ήταν ιδιαίτερα χρονοβόρες ή για να παρέχουν διάφορες συμβουλές όσον αφορά την διάγνωση.

Ειδικότερα, όσον αφορά την ογκολογία πλέον μέσω της ανάλυσης των εικόνων των όγκων μπορούν συστήματα Μηχανικής Μάθησης και Τεχνητής Νοημοσύνης να παρέχουν μια συμβουλευτική πρόβλεψη κατά πόσο ο όγκος είναι καλοήθης ή κακοήθης. Στον τομέα του γυναικολογικού καρκίνου εφαρμόζεται επεξεργασία και ανάλυση των εικόνων των όγκων που συλλέγονται μέσω της ενδοσκόπησης, είτε με λαπαροσκόπηση είτε με υστεροσκόπηση [7].

Γενικά ενδοσκόπηση είναι η ιατρική διαδικασία κατά την οποία γίνεται εισαγωγή ενός σωλήνα στην μήτρα ούτως ώστε να εξεταστεί η περιοχή για τυχόν λοιμώξεις. Η εν λόγω ιατρική πράξη διακρίνεται σε υστεροσκόπηση και λαπαροσκόπηση. Η υστεροσκόπηση είναι η διαδικασία κατά την οποία εξετάζεται η περιοχή της κοιλότητας της μήτρας, περνώντας τον σωλήνα στον τράχηλο της μήτρας [8].

Στο άρθρο [1] παρουσιάζεται η μελέτη που έγινε για την δημιουργία ενός υπολογιστικού διαγνωστικού συστήματος για διάγνωση ενδομητρίου καρκίνου σε πρόωρα στάδια. Αναλυτικότερα, το εν λόγω σύστημα χρησιμοποιεί εικόνες που συλλέχθηκαν μέσω ενδοσκόπησης – υστεροσκόπησης σε διάφορα χρώματα, YCrCb, HSV και RGB, ούτως ώστε να κατατάξει τα ενδομητριακά καρκινώματα σε καλοήθη ή κακοήθη. Για την κατηγοριοποίηση των όγκων χρησιμοποιήθηκαν οι εξής αλγορίθμοι:

1. Statistical features (SF) – κάνει κατανομή του gray level ιστογράμματος χωρίς να λαμβάνεται υπόψη η χωρική εξάρτηση [2].
2. Spatial gray-level dependence matrices (SGLDM) - βασίζεται στην εκτίμηση των συναρτήσεων πυκνότητας πιθανοτήτων δεύτερης τάξης ότι δύο εικονοστοιχεία (k, l) και (m, n) με απόσταση d στην κατεύθυνση που καθορίζεται από τη γωνία θ , έχουν εντάσεις gray level [2].

3. Gray-level difference statistics (GLDS) – βασίζεται στην υπόθεση ότι χρήσιμες πληροφορίες υφής μπορούν να εξαχθούν χρησιμοποιώντας στατιστικά πρώτης τάξης μιας εικόνας [2].

Αναλυτικότερα, η έρευνα καταλήγει στο ότι τα χαρακτηριστικά στα οποία διακρίνεται μεγάλη διαφορά ανάμεσα στους καλοήθεις και κακοήθεις όγκους, και επομένως τα πιο σημαντικά για την κατηγοριοποίηση των όγκων είναι τα εξής:

SF	SGLDM	GLDS
Energy (feature 2)	Homogeneity (feature 0)	Homogeneity (feature 4)
Entropy (feature 3)	Entropy (feature 1)	Contrast (feature 5)
		Energy (feature 6)
		Entropy (feature 7)
		Mean (feature 8)

Πίνακας 1: Τα 10 πιο σημαντικά χαρακτηριστικά υφής των όγκων[1]

Παράλληλα η έρευνα καταλήγει στο συμπέρασμα ότι ο αλγόριθμος Support Vector Machines (SVM) επιτυγχάνει μεγαλύτερη ακρίβεια ως προς την κατάταξη των όγκων σε καλοήθεις ή κακοήθεις σε σχέση με τον Probabilistic Neural Networks (PNN).

Κεφάλαιο 4

4. Μέθοδοι ανάλυσης των δεδομένων

4.1 Δεδομένα

Για την εκπόνηση της εν λόγω έρευνας χρησιμοποιήθηκαν τα δεδομένα στην μορφή που παρουσιάζονται στα άρθρα [1] [2]. Αναλυτικότερα, για κάθε όγκο συλλέχθηκαν τα 26 χαρακτηριστικά υφής που παρουσιάζονται στον Πίνακα 2: Τα 26 χαρακτηριστικά υφής των όγκων[1]. Επιπλέον κάθε όγκος άνηκε στην κατηγορία καλοήθους (1) ή κακοήθους (0).

Στην εν λόγω συλλογή δεδομένων, έγινε τυχαίος καταμερισμός των δεδομένων σε δεδομένα εκπαίδευσης και δεδομένα ελέγχου. Αναλυτικότερα, το 90% του συνόλου δεδομένων χρησιμοποιήθηκε σαν δεδομένα εκπαίδευσης και το υπόλοιπο 10% σαν δεδομένα ελέγχου. Ο λόγος του διαχωρισμού αυτού, ήταν η ανάγκη για μελέτη, ανάλυση των δεδομένων εκπαίδευσης και δημιουργίας αλγορίθμων επιχειρηματολογίας που να είναι αποδοτικοί στο συγκεκριμένο σύνολο, και η επαλήθευση ή διάψευση των ευρημάτων στα δεδομένα ελέγχου.

SF	SGLDM	GLDS
Mean	angular second moment (ASM)	Homogeneity
Variance	contrast	Contrast
Median	correlation	Energy
Energy	homogeneity	Entropy
Skewness	variance	Mean
Kurtosis	entropy	
Mode	sum entropy	
Entropy	sum average	
	sum variance	
	difference entropy	
	difference variance	
	information measurement of correlation 1	
	information measurement of correlation 2	

Πίνακας 2: Τα 26 χαρακτηριστικά υφής των όγκων[1]

4.2 Διαδικασία

Προετοιμασία δεδομένων

Στα πρώτα στάδια της διαδικασίας έγινε η προσαρμογή των δεδομένων στις ανάγκες της εν λόγω έρευνας. Αναλυτικότερα, με βάση τα ευρήματα των ερευνών [1] [2] δεν έχουν όλα τα χαρακτηριστικά την ίδια συνεισφορά στην κατηγοριοποίηση των όγκων σε καλοήθη ή κακοήθη. Επομένως, λαμβάνοντας υπόψη τα αποτελέσματα της έρευνας [1] έγινε επιλογή των 10 χαρακτηριστικών υφής του Πίνακα 2: Τα 26 χαρακτηριστικά υφής των όγκων[1], ως τα χαρακτηριστικά στα οποία εντοπίζεται η μεγάλη διαφορά στις τιμές τους όταν ο όγκος είναι καλοήθης από όταν είναι κακοήθης. Παράλληλα, για την επιλογή των πιο κατάλληλων χαρακτηριστικών δημιουργήθηκε ένας αλγόριθμος Random Forest ο οποίος δεχόταν τις τιμές των χαρακτηριστικών και την κατάταξη του όγκου σε καλοήθη ή κακοήθη, και παρουσίαζε το ποσοστό συνεισφοράς που έχει το κάθε χαρακτηριστικό στην κατηγοριοποίηση του όγκου όπως φαίνεται στον Πίνακα 3: Σημαντικότητα χαρακτηριστικών με βάση τον αλγόριθμο Random Forest.

Χαρακτηριστικά	Σημαντικότητα
sgldm_con	0.064438
sgldm_diff_ent	0.063552
sgldm_homog	0.05969
gldm_eng	0.05855
gldm_hom	0.056461
fos_mean	0.051617
fos_med	0.049913
gldm_mean	0.048916
gldm_con	0.048628
sgldm_sumav	0.046912
fos_mode	0.043663
gldm_ent	0.040057
sgldm_difvar	0.036667
sgldm_sumvar	0.031368
fos_kurt	0.031063
fos_skew	0.027394
sgldm_asm	0.027125
sgldm_infcor2	0.026722
sgldm_var	0.026693
fos_ent	0.025724
fos_ener	0.024186
sgldm_cor	0.023787
sgldm_sument	0.021991
fos_var	0.021886
sgldm_entr	0.021867
sgldm_infcor1	0.021129

Πίνακας 3: Σημαντικότητα χαρακτηριστικών με βάση τον αλγόριθμο Random Forest

Παρατηρήθηκε ότι τα αποτελέσματα του αλγορίθμου δεν συμφωνούν απόλυτα με τα αποτελέσματα της προαναφερθέντος έρευνας, επομένως θεωρήθηκε πιο δόκιμο στα πλαίσια της εν λόγω μελέτης να χρησιμοποιηθούν τα ευρήματα της έρευνας και επομένως να χρησιμοποιηθούν τα χαρακτηριστικά του Πίνακα 1: Τα 10 πιο σημαντικά χαρακτηριστικά υφής των όγκων[1]

Μετά την επιλογή των σημαντικών χαρακτηριστικών, ακολούθησε ο καταμερισμός της συλλογής δεδομένων σε δεδομένα εκπαίδευσης και δεδομένων ελέγχου, όπως ορίστηκε πιο πάνω.

Δημιουργία βασικών κανόνων

Δημιουργήθηκε ένας αλγόριθμος κατηγοριοποίησης Random Forest, ο οποίος δεχόμενος ως είσοδο τα δεδομένα εκπαίδευσης δημιουργούσε δεντρικές δομές με πιθανούς κανόνες κατηγοριοποίησης των όγκων. Ενδεικτικό παράδειγμα η εικόνα 1. Ο αλγόριθμος επιτρέπει την επιλογή μεταβλητού βάθους δεντρικής δομής, ούτως ώστε να επιλεγεί από τον προγραμματιστή ο αριθμός των χαρακτηριστικών που μετέχουν σε κάθε κανόνα.

Με την βοήθεια, λοιπόν, του συγκεκριμένου αλγορίθμου έγινε η δημιουργία των βασικών κανόνων που χρησιμοποιήθηκαν μεταγενέστερα στην εφαρμογή επιχειρηματολογίας. Οι βασικοί κανόνες είναι οι εξής:

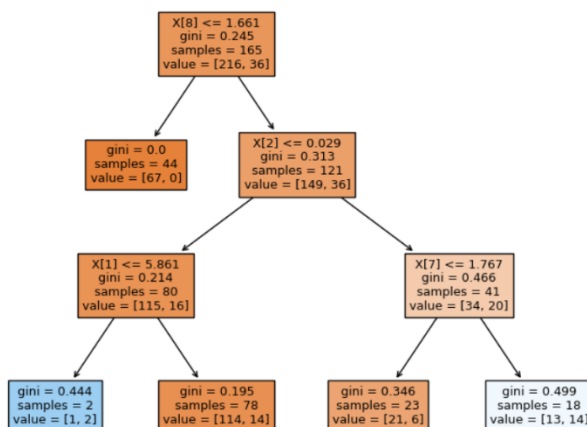
Κανόνας 4: $\text{feature_8} > 1.65 \ \& \ \text{feature_1} > 5.25 \rightarrow 0$

Κανόνας 6: $\text{feature_5} > 4.89 \ \& \ \text{feature_2} \leq 0.06 \rightarrow 0$

Κανόνας 8: $\text{feature_5} \leq 5.03 \ \& \ \text{feature_7} > 1.30 \rightarrow 1$

Κανόνας 10: $\text{feature_8} \leq 1.65 \ \& \ \text{feature_0} > 0.45 \rightarrow 1$

Η επιλογή των βασικών κανόνων έγινε με πολύ προσεκτική ανάλυση των δεδομένων καθώς υπήρχε η ανάγκη για επιλογή κανόνων οι οποίοι θα «κάλυπταν» σχεδόν όλο το σύνολο των δεδομένων εκπαίδευσης όπως παρουσιάζεται στον πίνακα 3. Συνολικά μόλις 1 από τα 400 δεδομένα του συνόλου δεδομένων εκπαίδευσης δεν καλύπτεται από τους συγκεκριμένους κανόνες.



```
--- feature_8 <= 1.66
|--- class: 0.0
--- feature_8 > 1.66
|--- feature_2 <= 0.03
|   |--- feature_1 <= 5.86
|   |   |--- class: 1.0
|   |   |--- feature_1 > 5.86
|   |       |--- class: 0.0
|   |--- feature_2 > 0.03
|   |   |--- feature_7 <= 1.77
|   |   |   |--- class: 0.0
|   |   |   |--- feature_7 > 1.77
|   |       |--- class: 1.0
```

Εικόνα 1: Δεντρική δομή κανόνων σε δύο μορφές

	Περιπτώσεις που ισχύει	Περιπτώσεις που δεν ισχύει	Αριθμός δεδομένων που επηρεάζει
Κανόνας 4	152	40	192
Κανόνας 6	156	43	199
Κανόνας 8	145	56	201
Κανόνας 10	144	57	201

Πίνακας 4: Ποσότητα επηρεασμού του κάθε κανόνα στα δεδομένα εκπαίδευσης (ο συνολικός αριθμός των δεδομένων εκπαίδευσης είναι 400)

Καθορισμός προτεραιοτήτων

Αφού εξάχθηκαν οι βασικοί κανόνες, δημιουργήθηκαν δυο ζεύγη «αντίπαλων» κανόνων. Στο κάθε ζεύγος ο ένας κανόνας αποφάσιζε ότι ο όγκος είναι καλοήθης (1) και ο δεύτερος ότι είναι κακοήθης (0). Πιο συγκεκριμένα, τα ζεύξης επιλέγηκαν ως εξής:

Κανόνας 4 – Κανόνας 8

Κανόνας 6 – Κανόνας 10

Κάθε ζεύγος διαπιστώθηκε ότι έχει έναν αριθμό διλημάτων, περιπτώσεις δηλαδή στις οποίες ισχύουν και οι δύο κανόνες και επομένως ο ένας κανόνας αποφασίζει ότι ο όγκος είναι καλοήθης ενώ ο άλλος ότι ο όγκος είναι κακοήθης. Για τον λόγο αυτό, μετά από μελέτη και ανάλυση των περιπτώσεων που προκύπταν διλήμματα, δημιουργήθηκαν κανόνες προτεραιότητας για το κάθε ζεύγος κανόνων. Συγκεκριμένα οι προτεραιότητες που τέθηκαν για τα δύο ζεύγη είναι οι εξής:

Κανόνας 4 – Κανόνας 8:

Αν $feature_0 > 0.445$ και $feature_4 > 0.445$ τότε ο κανόνας 8 έχει προτεραιότητα έναντι του κανόνα 4, αλλιώς ο κανόνας 4 έχει προτεραιότητα έναντι του κανόνα 8.

Κανόνας 6 – Κανόνας 10

Αν $feature_0 > 0.454$ και $feature_0 < 0.46$ τότε ο κανόνας 10 έχει προτεραιότητα έναντι του κανόνα 6 αλλιώς ο κανόνας 6 έχει προτεραιότητα έναντι του κανόνα 10.

Δημιουργία επιπρόσθετων κανόνων

Έχοντας διασφαλίσει τον ελεγχόμενο χειρισμό των διλημάτων μεταξύ των ζευγών βασικών κανόνων, υπήρχε η ανάγκη για επιδιόρθωση των περιπτώσεων στις οποίες οι βασικοί κανόνες έκαναν λάθος. Για τον λόγο αυτό, χρησιμοποιώντας και πάλι τον αλγόριθμο που είδαμε στην δημιουργία των βασικών κανόνων, έγινε ανάλυση των δεδομένων για δημιουργία κανόνων οι οποίοι διόρθωναν τις περιπτώσεις στις οποίες κάθε βασικός κανόνας έκανε λάθος. Οι βελτιωτικοί κανόνες που δημιουργήθηκαν είναι οι εξής:

Για τον κανόνα 4:

$feature_8 > 1.66 \ \& \ feature_4 \leq 0.45 \rightarrow 1$

Για τον κανόνα 6:

$\text{feature_7} > 1.67 \ \& \ \text{feature_1} \leq 5.93 \ \& \ \text{feature_6} \leq 0.19 \rightarrow 0$

Για τον κανόνα 8:

$\text{feature_4} > 0.50 \ \& \ \text{feature_2} \leq 0.05 \rightarrow 0$

Για τον κανόνα 10:

$\text{feature_4} > 0.50 \ \& \ \text{feature_0} > 0.50 \ \& \ \text{feature_3} > 3.31 \rightarrow 0$

Οι συγκεκριμένοι κανόνες αναμένεται να έχουν προτεραιότητα σε σχέση με τους βασικούς τους κανόνες ούτως ώστε να μπορούν να διορθώνουν τα σφάλματα των βασικών κανόνων. Για να διασφαλιστεί, ωστόσο, το γεγονός ότι οι βελτιωτικοί κανόνες δεν θα δημιουργήσουν νέα λάθη, γενικά οι υπόλοιποι τρεις βασικοί κανόνες θα υπερέχουν του βελτιωτικού κανόνα του τέταρτου βασικού.

Προσομοίωση αλγορίθμου επιχειρηματολογίας

Πριν την δημιουργία του μοντέλου επιχειρηματολογίας, δημιουργήθηκε ένας πιλοτικός αλγόριθμος ο οποίος προσομοιάζει την λειτουργία του. Ο αλγόριθμος σε ψευδοκώδικα είναι ο εξής:

#####

Έστω ότι 4β ο βελτιωτικός κανόνας του βασικού κανόνα 4 και αντίστοιχα για τους υπόλοιπους
Βασικός αλγόριθμος

Αν υπάρχει δίλημμα μεταξύ του κανόνα 4-8 τότε

Αν έχει ο 4 προτεραιότητα τότε

Εκτέλεσε πρώτα το πακέτο κανόνα 4 και αν δεν αποφασίσει εκτέλεσε το πακέτο κανόνα 8.

Αλλιώς

Εκτέλεσε πρώτα το πακέτο κανόνα 8 και αν δεν αποφασίσει εκτέλεσε το πακέτο κανόνα 4.

Αλλιώς

Εκτέλεσε τα πακέτα κανόνα 8 και κανόνα 4 με οποιαδήποτε σειρά αφού είτε θα ισχύει μόνο ένας κανόνας είτε κανένας.

Πακέτο κανόνα 4

Αν ισχύει ο 4β τότε

Αν ισχύει μόνο ένας εκ των 6 ή 10 τότε εκτέλεσε αυτούς αλλιώς, εκτέλεσε τον κανόνα 4β

Αλλιώς

Εκτέλεσε τον κανόνα 4

Πακέτο κανόνα 8

Αν ισχύει ο 8β τότε

Αν ισχύει μόνο ένας εκ των 6 ή 10 τότε εκτέλεσε αυτούς αλλιώς, εκτέλεσε τον κανόνα 8β

Αλλιώς

Εκτέλεσε τον κανόνα 8

Αντίστοιχα και για το ζεύγος κανόνων 6-10

#####

Σημαντικά σημεία ευρήματα αλγορίθμου:

- Όταν ισχύει κάποιος βελτιωτικός κανόνας, τότε δώσε προτεραιότητα στους υπόλοιπους βασικούς κανόνες ούτως ώστε ο βελτιωτικός κανόνας να διορθώνει τα σφάλματα του βασικού του κανόνα χωρίς να δημιουργεί καινούρια σφάλματα.
- Αν ισχύει κάποιος βελτιωτικός κανόνας και στους υπόλοιπους βασικούς κανόνες υπάρχουν διλήμματα, τότε είναι η μόνη περίπτωση στην οποία δίνεται προτεραιότητα και έναντι των υπολοίπων βασικών κανόνων για να αποφευχθούν διλήμματα.

* Ο εν λόγω αλγόριθμος δεν έχει ακριβώς τα ίδια αποτελέσματα με το μοντέλο επιχειρηματολογίας καθώς οι δυνατότητες και η λογική ενός μοντέλου επιχειρηματολογίας και ενός αλγορίθμου με συνθήκες ifs δεν είναι οι ίδιες. Η δημιουργία του εν λόγω αλγορίθμου αποσκοπούσε στην μελέτη, κατανόηση και ανάλυση της λογικής.

Δημιουργία μοντέλου επιχειρηματολογίας

Με βάση τους κανόνες που δημιουργήθηκαν, τις προτεραιότητες που τέθηκαν και τα σημαντικά σημεία που εντοπίστηκαν μέσω των αλγορίθμων δημιουργείται το μοντέλο επιχειρηματολογίας.

Αφορά τους κανόνες 4 & 8

Κανόνας 4 > Κανόνας 8 AN { }

Κανόνας 8 > Κανόνας 4 AN { feature0 > 0.445 and feature4 > 0.445 }

Κανόνας 4β > Κανόνας 4 AN { }

Κανόνας 8β > Κανόνας 8 AN { }

Κανόνας 6 > Κανόνας 4β

Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών

Κανονας 6 > Κανόνας 8β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών
Κανονας 10 > Κανόνας 4β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών
Κανονας 10 > Κανόνας 8β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών

Αφορά τους κανόνες 6 & 10

Κανόνας 6 > Κανόνας 10 AN { }

Κανόνας 10 > Κανόνας 6 AN { feature0>0.454 και feature0<0.46 }

Κανονας 6β > Κανόνας 6 AN { }

Κανονας 10β > Κανόνας 10 AN { }

Κανονας 4 > Κανόνας 6β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών
Κανονας 4 > Κανόνας 10β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών
Κανονας 8 > Κανόνας 6β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών
Κανονας 8 > Κανόνας 10β	Προτεραιότητα βασικών κανόνων έναντι βελτιωτικών

Περιγραφή κώδικα επιχειρηματολογίας

Για κάθε στιγμιότυπο – όγκος δόθηκε ένας μοναδικός αριθμός, ο οποίος αποτελεί την ταυτότητα του συγκεκριμένου όγκου, μέσω της οποίας θα δοθούν οι τιμές στα χαρακτηριστικά του και θα χρησιμοποιηθεί για να επερωτηθεί το μοντέλο επιχειρηματολογίας αν ο συγκεκριμένος όγκος είναι καλοήθης ή κακοήθης.

Οι ανάθεση των τιμών των χαρακτηριστικών του κάθε όγκου γίνεται ως εξής:

feature0(1,0.495107).

feature1(1,5.054545).

feature2(1,0.048805).

feature3(1,3.159896).

feature4(1,0.495371).

feature5(1,3.049931).

feature6(1,0.275601).

feature7(1,1.445991).

feature8(1,1.326772).

Στο συγκεκριμένο παράδειγμα γίνεται η ανάθεση των τιμών των χαρακτηριστικών του όγκου με ταυτότητα 1. Για παράδειγμα, με τη χρήση του **feature0(1,0.495107)**, δηλώνεται ότι για τον όγκο 1 η τιμή του feature 0 είναι 0.495107. Αντίστοιχα και για τα υπόλοιπα χαρακτηριστικά.

Το μοντέλο επιχειρηματολογίας επιτρέπει την δυαδική κατηγοριοποίηση του κάθε στιγμιότυπου – όγκου. Οι κατηγορίες που επιτρέπει το μοντέλο να κατηγοριοποιηθούν οι όγκοι είναι benign(ID) ή malignant(ID). Επομένως αν ο χρήστης εισάγει για παράδειγμα το benign(1) τότε αν το μοντέλο επιχειρηματολογίας βρει κάποια έγκυρη δικαιολογία τότε θα επιστρέψει «αληθές» μαζί με την αντίστοιχη επιχειρηματολογία, αλλιώς θα επιστρέψει «λάθος».

4.3 Επεξηγήσεις επιχειρηματολογίας

Περιπτώσεις όπου η απόφαση που παίρνει το Gorgias Cloud είναι ορθές:

Παράδειγμα 1:

Η επεξήγηση η οποία επιστρέφει το Gorgias Cloud είναι η εξής:

```
prove([benign(21)], Explanation).
false
```

```
prove([malignant(21)], Explanation).
Solution 1
```

```
Explanation=[c6(21),pr13(21),r10(21)],
```

```
'=====Application Level Explanation=====
The statement "malignant(21)" is supported by:
- feature8(21,_5554),_5554=<1.65,feature0(21,_5572),_5572>0.45
This reason is strengthened:
- against "benign(21)" by: feature0(21,_6634),_6634>0.454,feature0(21,_6652),_6652<0.46
```

Στην συνέχεια όταν αντικατασταθούν τα feature με την πραγματική τους ονομασία δίνεται η πιο κάτω αιτιολόγηση:

```
prove([benign(21)], Explanation).
false
```

```
prove([malignant(21)], Explanation).
Solution 1
```

```
Explanation=[c6(21),pr13(21),r10(21)],
```

```
'=====Application Level Explanation=====
The statement "malignant(21)" is supported by:
- gldm_mean(21,_5554),_5554=<1.65, sgldm_homog (21,_5572),_5572>0.45
This reason is strengthened:
- against "benign(21)" by: sgldm_homog (21,_6634),_6634>0.454, sgldm_homog
(21,_6652),_6652<0.46
```

Τέλος η πιο απλοποιημένη μορφή της αιτιολόγησης είναι η εξής:

Ο ασθενής με ταυτότητα 21 έχει κακοήγη όγκο.

Το $gldm_mean$ είναι ≤ 1.65 και το $sgldm_homog > 0.45$ και η θεωρία που ισχυρίζεται ότι ο όγκος που έχει ο ασθενής 21 είναι κακοήθης, ενδυναμώνεται με τις τιμές $sgldm_homog > 0.454$ και $sgldm_homog < 0.46$.

Παράδειγμα 2:

Η επεξήγηση η οποία επιστρέφει το Gorgias Cloud είναι η εξής:

*prove([benign(286)], Explanation).
false*

*prove([malignant(286)], Explanation).
Solution 1*

Explanation=[c1(286),pr2(286),r8(286)],

'=====Application Level Explanation=====

The statement "malignant(286)" is supported by:

- feature5(286,_5554),_5554=<5.03,feature7(286,_5572),_5572>1.3

This reason is strengthened:

- against "benign(286)" by:

feature0(286,_6634),_6634>0.445,feature4(286,_6652),_6652>0.445

Στην συνέχεια όταν αντικατασταθούν τα feature με την πραγματική τους ονομασία δίνεται η πιο κάτω αιτιολόγηση:

*prove([benign(286)], Explanation).
false*

*prove([malignant(286)], Explanation).
Solution 1*

Explanation=[c1(286),pr2(286),r8(286)],

'=====Application Level Explanation=====

The statement "malignant(286)" is supported by:

*- **gldm_con** (286,_5554),_5554=<5.03, **gldm_ent** (286,_5572),_5572>1.3*

This reason is strengthened:

*- against "benign(286)" by: **sgldm_homog**(286,_6634),_6634>0.445,*

***gldm_hom**(286,_6652),_6652>0.445*

Τέλος η πιο απλοποιημένη μορφή της αιτιολόγησης είναι η εξής:

Ο ασθενής με ταυτότητα 286 έχει κακοήγη όγκο.

Το $gldm_con \leq 5.03$ και το $gldm_ent > 1.3$ και η θεωρία που ισχυρίζεται ότι ο όγκος που έχει ο ασθενής 286 είναι κακοήθης, ενδυναμώνεται με τις τιμές $sgldm_homog > 0.445$ και το $gldm_hom > 0.445$.

Περιπτώσεις όπου η απόφαση που παίρνει το Gorgias Cloud είναι λάθος:

Παράδειγμα 3:

Η επεξήγηση η οποία επιστρέφει το Gorgias Cloud είναι η εξής:

```
prove([benign(12)], Explanation).  
false
```

```
prove([malignant(12)], Explanation).  
Solution 1
```

```
Explanation=[r8(12)],
```

```
'=====Application Level Explanation=====  
The statement "malignant(12)" is supported by:  
- feature5(12,_4552),_4552=<5.03,feature7(12,_4570),_4570>1.3
```

Στην συνέχεια όταν αντικατασταθούν τα feature με την πραγματική τους ονομασία δίνεται η πιο κάτω αιτιολόγηση:

```
prove([benign(12)], Explanation).  
false
```

```
prove([malignant(12)], Explanation).  
Solution 1
```

```
Explanation=[r8(12)],
```

```
'=====Application Level Explanation=====  
The statement "malignant(12)" is supported by:  
- gldm_con (12,_4552),_4552=<5.03, gldm_ent (12,_4570),_4570>1.3
```

Τέλος η πιο απλοποιημένη μορφή της αιτιολόγησης είναι η εξής:

Ο ασθενής με ταυτότητα 12 έχει κακοήγη όγκο, με την αιτιολόγηση ότι ισχύουν το $gldm_con \leq 5.03$ και το $gldm_ent > 1.3$.

Παράδειγμα 4:

Η επεξήγηση η οποία επιστρέφει το Gorgias Cloud είναι η εξής:

```
prove([benign(101)], Explanation).  
Solution 1
```

```
Explanation=[pr4(101),r4(101),r6(101)],
```

```
'=====Application Level Explanation=====  
The statement "benign(101)" is supported by:
```

```
-
feature8(101,_5562),_5562>1.65,feature1(101,_5580),_5580>5.25,feature5(101,_5694),_5694>4.89,feature2(101,_5712),_5712=<0.06
This reason is strengthened:
- against "malignant(101)"
=====

prove([malignant(101)], Explanation).
false
```

Στην συνέχεια όταν αντικατασταθούν τα feature με την πραγματική τους ονομασία δίνεται η πιο κάτω αιτιολόγηση:

```
prove([benign(101)], Explanation).
Solution 1

Explanation=[pr4(101),r4(101),r6(101)],

'=====Application Level Explanation=====
The statement "benign(101)" is supported by:
- gldm_mean(101,_5562),_5562>1.65, sgldm_entr(101,_5580),_5580>5.25, gldm_con
(101,_5694),_5694>4.89, fos_ener(101,_5712),_5712=<0.06
This reason is strengthened:
- against "malignant(101)"
=====

prove([malignant(101)], Explanation).
false
```

Τέλος η πιο απλοποιημένη μορφή της αιτιολόγησης είναι η εξής:

Ο ασθενής με ταυτότητα 101 έχει καλοήγη όγκο, με την αιτιολόγηση ότι ισχύουν το $gldm_mean > 1.65$, το $sgldm_entr > 5.25$, το $gldm_con > 4.89$, και το $fos_ener \leq 0.06$

Περιπτώσεις όπου το Gorgias Cloud δεν παίρνει κάποια απόφαση (υπάρχει δίλημμα):

Παράδειγμα 5:

Η επεξήγηση η οποία επιστρέφει το Gorgias Cloud είναι η εξής:

```
prove([malignant(2)], Explanation).
Solution 1

Explanation=[r8(2)],

'=====Application Level Explanation=====
The statement "malignant(2)" is supported by:
- feature5(2,_4552),_4552=<5.03,feature7(2,_4570),_4570>1.3
=====
```

prove([benign(2)], Explanation).

Solution 1

Explanation=[r6(2)],

'=====Application Level Explanation=====

The statement "benign(2)" is supported by:

- feature5(2,_4552),_4552>4.89,feature2(2,_4570),_4570=<0.06

Στην συνέχεια όταν αντικατασταθούν τα feature με την πραγματική τους ονομασία δίνεται η πιο κάτω αιτιολόγηση:

prove([malignant(2)], Explanation).

Solution 1

Explanation=[r8(2)],

'=====Application Level Explanation=====

The statement "malignant(2)" is supported by:

- gldm_con (2,_4552),_4552=<5.03, gldm_ent (2,_4570),_4570>1.3

'=====

prove([benign(2)], Explanation).

Solution 1

Explanation=[r6(2)],

'=====Application Level Explanation=====

The statement "benign(2)" is supported by:

- gldm_con (2,_4552),_4552>4.89, fos_ener(2,_4570),_4570=<0.06

Τέλος η πιο απλοποιημένη μορφή της αιτιολόγησης είναι η εξής:

Δεν μπορεί να αποφασιστεί αν ο όγκος που έχει ο ασθενής με ταυτότητα 2 είναι κακοήθης ή καλοήθης.

Αυτό συμβαίνει γιατί ισχύουν τα $gldm_con \leq 5.03$ και $gldm_ent > 1.3$, τα οποία ισχυρίζονται ότι ο όγκος είναι κακοήθης και τα $gldm_con > 4.89$, και $fos_ener \leq 0.06$, που ισχυρίζονται ότι ο όγκος είναι καλοήθης.

Παράδειγμα 6:

Η επεξήγηση η οποία επιστρέφει το Gorgias Cloud είναι η εξής:

prove([benign(134)], Explanation).

Solution 1

Explanation=[pr15(134),pr3(134),r10b(134),r8b(134)],

'=====Application Level Explanation=====

The statement "benign(134)" is supported by:

-

feature4(134,_6474),_6474>0.5,feature2(134,_6492),_6492=<0.05,feature4(134,_6620),_6620>0.5,feature0(134,_6638),_6638>0.5,feature3(134,_6656),_6656>3.31

This reason is strengthened:

- against "malignant(134)"
- against "malignant(134)"

=====

prove([malignant(134)], Explanation).

Solution 1

Explanation=[pr19(134),pr7(134),r10(134),r8(134)],

'=====Application Level Explanation=====

The statement "malignant(134)" is supported by:

-

feature5(134,_6430),_6430=<5.03,feature7(134,_6448),_6448>1.3,feature8(134,_6562),_6562=<1.65,feature0(134,_6580),_6580>0.45

This reason is strengthened:

- against "benign(134)"
- against "benign(134)"

Στην συνέχεια όταν αντικατασταθούν τα feature με την πραγματική τους ονομασία δίνεται η πιο κάτω αιτιολόγηση:

prove([benign(134)], Explanation).

Solution 1

Explanation=[pr15(134),pr3(134),r10b(134),r8b(134)],

'=====Application Level Explanation=====

The statement "benign(134)" is supported by:

- **gldm_hom** (134,_6474),_6474>0.5, **fos_ener** (134,_6492),_6492=<0.05, **gldm_hom** (134,_6620),_6620>0.5, **sgldm_homog** (134,_6638),_6638>0.5, **fos_ent** (134,_6656),_6656>3.31

This reason is strengthened:

- against "malignant(134)"
- against "malignant(134)"

=====

prove([malignant(134)], Explanation).

Solution 1

Explanation=[pr19(134),pr7(134),r10(134),r8(134)],

'=====Application Level Explanation=====

The statement "malignant(134)" is supported by:

- **gldm_con** (134,_6430),_6430=<5.03, **gldm_ent** (134,_6448),_6448>1.3, **gldm_mean** (134,_6562),_6562=<1.65, **sgldm_homog** (134,_6580),_6580>0.45

This reason is strengthened:

- against "benign(134)"
- against "benign(134)"

Τέλος η πιο απλοποιημένη μορφή της αιτιολόγησης είναι η εξής:

Δεν μπορεί να αποφασιστεί αν ο όγκος που έχει ο ασθενής με ταυτότητα 2 είναι κακοήθης ή καλοήθης.

Αυτό συμβαίνει γιατί ισχύουν τα $gldm_hom > 0.5$, $fos_ener \leq 0.05$, $sgldm_homog > 0.5$ και $fos_ent > 3.31$ τα οποία ισχυρίζονται ότι ο όγκος είναι καλοήθης και τα $gldm_con \leq 5.03$, $gldm_ent > 1.3$, $gldm_mean \leq 1.65$ και $sgldm_homog > 0.45$ που ισχυρίζονται ότι ο όγκος είναι κακοήθης.

4.4 Ομαδοποιήσεις επεξηγήσεων επιχειρηματολογίας

Μέσα από τα αποτελέσματα που παράχθηκαν κατά την εκτέλεση των δεδομένων εκπαίδευσης στο πρόγραμμα επιχειρηματολογίας στο Gorgias Cloud, εντοπίστηκαν δεκατέσσερις διαφορετικές κατηγορίες επεξηγήσεων. Δηλαδή, κάθε όγκος που βρίσκεται στα δεδομένα εκπαίδευσης με την αντίστοιχη πρόγνωση του είχε μια επεξήγηση ίδια με μια εκ των πιο κάτω κατηγοριών.

Γενική εικόνα της θεωρίας κανόνων

A4: {r4(Tumor), pr1(Tumor), pr16(Tumor)}

A6: {r6(Tumor), pr12(Tumor), pr4(Tumor)}

A8: {r8(Tumor), pr19(Tumor)}

A10: {r10(Tumor), pr7(Tumor)}

B4b: {r4b(Tumor), pr3(Tumor)}

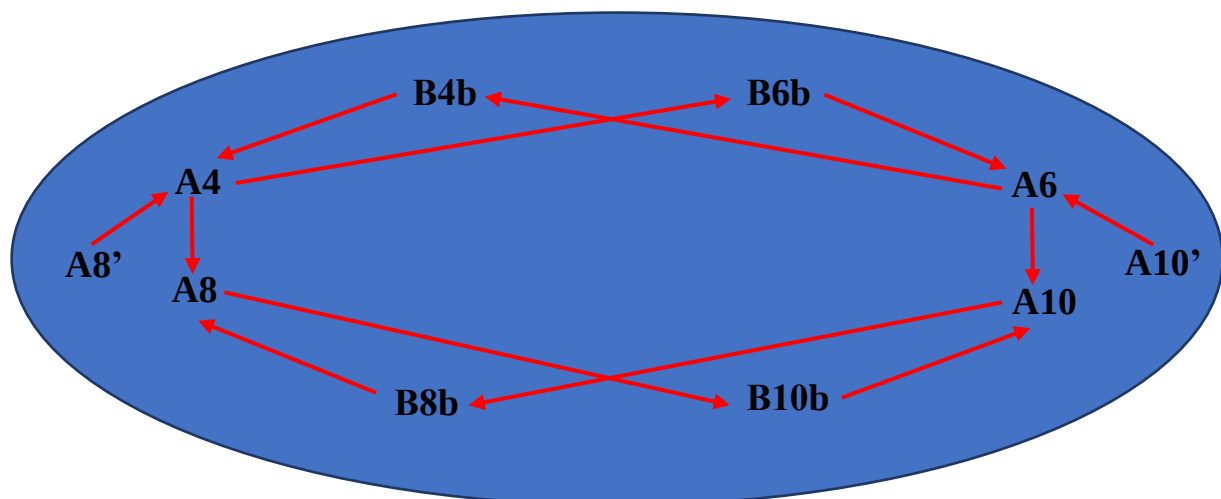
B6b: {r6b(Tumor), pr14(Tumor)}

B8b: {r8b(Tumor), r6b(Tumor), pr5(Tumor)}

B10b: {r10b(Tumor), pr15(Tumor)}

A8': {r8(Tumor), pr2(Tumor), c1(Tumor)}

A10': {r10(Tumor), pr13(Tumor), c6(Tumor)}



Παρόλα αυτά στην πιο πάνω γενική θεωρία γίνεται η υπόθεση ότι ισχύουν όλες οι πιθανές συνθήκες, γεγονός το οποίο δεν είναι εφικτό αφού υπάρχουν αλληλοσυγκρουόμενα χαρακτηριστικά. Για τον λόγο αυτό δημιουργήθηκαν τα σενάρια για τις δεκατέσσερις κατηγορίες επεξήγησης.

Ομάδα1: Explanation=[pr7(_),r10(_),r8(_)]

1, 37, 49, 63, 110, 133, 196, 267, 271, 308, 319, 323, 342, 372

Σενάριο 1: {f5 < 5.03, f7>1.3, f8<1.65, f0>0.45, f4>0.5, f2<0.05}

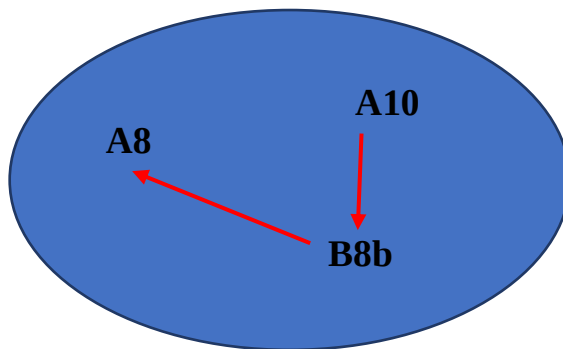
Σενάριο 1 σε κανονική μορφή: {gldm_con < 5.03, gldm_ent >1.3, gldm_mean<1.65, sgldm_homog >0.45, gldm_hom >0.5, fos_ener <0.05}

Ισχύουν: r10, r8, r8b, r8b > r8 , r10 > r8b

A8: {r8(Tumor)}

A10: {r10(Tumor), pr7(Tumor): r10>r8b}

B8b: {r8b(Tumor), pr5(Tumor): r8b > r8}



Ομάδα2: Explanation=[r6(_)]

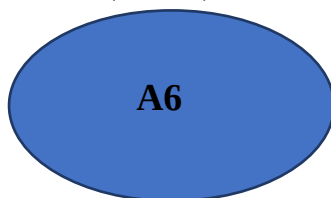
2, 71, 81, 130, 334, 349

Σενάριο 2: {f5>4.89, f2<=0.06}, τα επιχειρήματα που υπερτερούν του r6 είναι τα r6b και r10 των οποίων δεν ικανοποιούνται όλα τα conditions τους.

Σενάριο 2 σε κανονική μορφή: {gldm_con>4.89, fos_ener<=0.06}

Ισχύει: r6

A6: {r6(Tumor)}



Ομάδα3: Explanation=[r8(_)]

2, 3, 8, 9, 11, 12, 13, 16, 17, 22, 24, 26, 27, 29,32, 34, 35, 36, 40, 41, 42, 43, 44, 52, 55, 58, 67, 68, 69, 70, 73, 75, 80, 81, 82, 89, 91, 92, 96, 98, 100, 104, 105, 107, 113, 117, 118, 122, 125, 126, 128, 129, 130, 132, 136, 139, 140, 142, 143, 149, 153, 155, 158, 168, 169, 171, 173, 176, 177, 180, 186, 187, 189, 190, 192, 193, 194, 200, 208, 211, 212, 215, 216, 218, 219, 221, 223, 224, 232, 233, 234, 236, 240, 242, 244, 246, 253, 254, 255, 259, 260, 261, 263, 264, 266, 268, 269, 278, 280, 281, 283, 291, 299, 300, 302, 303, 307, 309, 310, 311, 312, 315, 316, 317, 318, 320, 322, 326, 328, 329, 330, 333, 338, 339, 345, 346,

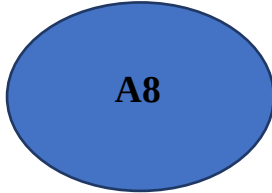
348, 350, 351, 354, 356, 358, 363, 365, 374, 376, 378, 380, 381, 382, 383, 387, 391, 392, 397, 398, 400

Σενάριο 3: { $f5 \leq 5.03$, $f7 > 1.30$ }, τα επιχειρήματα που υπερτερούν του r8 είναι τα r8b και r4 των οποίων δεν ικανοποιούνται όλα τα conditions τους.

Σενάριο 3 σε κανονική μορφή: { $gldm_con \leq 5.03$, $gldm_ent > 1.30$ }

Ισχύει: r8

A8: {r8(Tumor)}



Ομάδα4: Explanation=[pr4(_),r4(_),r6(_)]

4, 5, 6, 7, 10, 14, 15, 18, 19, 20, 23, 25, 31, 33, 48, 50, 51, 54, 56, 57, 59, 60, 61, 62, 64, 65, 66, 72, 74, 76, 77, 78, 79, 83, 84, 85, 86, 87, 88, 93, 94, 95, 99, 101, 102, 108, 109, 111, 112, 114, 116, 117, 119, 121, 131, 137, 141, 144, 145, 146, 147, 151, 152, 156, 160, 162, 164, 165, 166, 167, 172, 174, 175, 178, 179, 182, 183, 184, 185, 188, 195, 197, 201, 203, 205, 206, 207, 209, 210, 213, 217, 220, 222, 225, 226, 229, 230, 231, 235, 237, 238, 239, 245, 248, 249, 250, 251, 252, 256, 257, 258, 262, 272, 273, 274, 276, 277, 282, 287, 289, 292, 294, 296, 297, 301, 304, 305, 324, 325, 327, 331, 332, 335, 337, 340, 341, 343, 344, 347, 352, 353, 355, 360, 362, 364, 366, 367, 368, 369, 370, 371, 375, 377, 384, 385, 386, 388, 389, 393, 394, 395, 396, 399

Σενάριο 4: { $f2 \leq 0.06$, $f5 > 4.89$, $f8 > 1.65$, $f1 > 5.25$, $f8 > 1.66$, $f4 \leq 0.45$ }

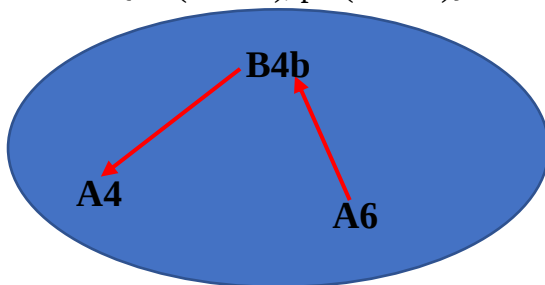
Σενάριο 4 σε κανονική μορφή: { $fos_ener \leq 0.06$, $gldm_con > 4.89$, $gldm_mean > 1.65$, $sgldm_entr > 5.25$, $gldm_mean > 1.66$, $gldm_hom \leq 0.45$ }

Ισχύουν: r4, r4b, r6, r4b>r4, r6 > r4b

A4: {r4(Tumor)}

A2: {r6(Tumor), pr4(Tumor)}

B4b: {r4b(Tumor), pr3(Tumor)}



Ομάδα5: Explanation=[c6(_),pr13(_),r10(_)]

21

Σενάριο 5: { $f8 \leq 1.65$, $f0 > 0.45$, $f0 > 0.454$, $f0 < 0.46$, $f5 > 4.89$, $f2 \leq 0.06$ }

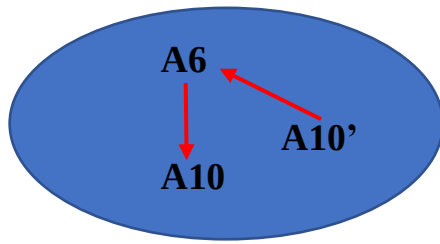
Σενάριο 5 σε κανονική μορφή: { $gldm_mean \leq 1.65$, $sgldm_homog > 0.45$, $sgldm_homog > 0.454$, $sgldm_homog < 0.46$, $gldm_con > 4.89$ & $fos_ener \leq 0.06$ }

Ισχύουν: r10, r6, r10 > r6, r6 > r10

A6: {r6(Tumor), pr12(Tumor)}

A10: {r10(Tumor)}

A10': {r10(Tumor), pr13(Tumor), c6(Tumor)}



Ομάδα6: Explanation=[r10(_)]

28, 157, 227

Σενάριο 6: {f8 ≤ 1.65, f0 > 0.45} τα επιχειρήματα που υπερτερούν του r10 είναι τα r10b και r6 των οποίων δεν ικανοποιούνται όλα τα conditions τους.

Σενάριο 6 σε κανονική μορφή: {gldm_mean ≤ 1.65, sgldm_homog > 0.45}

Ισχύουν: r10

A10: {r10(Tumor)}



Ομάδα7: Explanation=[r4(_)]

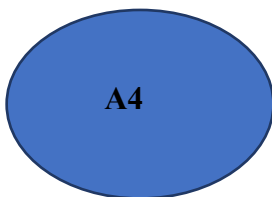
30, 45, 115, 124, 154, 159, 198, 202, 204, 284, 293, 295, 298

Σενάριο 7: {f8 > 1.65, f1 > 5.25} τα επιχειρήματα που υπερτερούν του r4 είναι τα r6 των οποίων δεν ικανοποιούνται όλα τα conditions του.

Σενάριο 7 σε κανονική μορφή: {gldm_mean > 1.65, sgldm_entr > 5.25}

Ισχύουν: r4

A4: {r4(Tumor)}



Ομάδα 8: Explanation=[pr15(_),pr3(_),r10b(_),r8b(_)]

38, 46, 47, 53, 106, 120, 127, 134, 135, 161, 170, 181, 228, 243, 265, 270, 285, 290, 306, 313, 336, 357, 359, 361, 373, 379

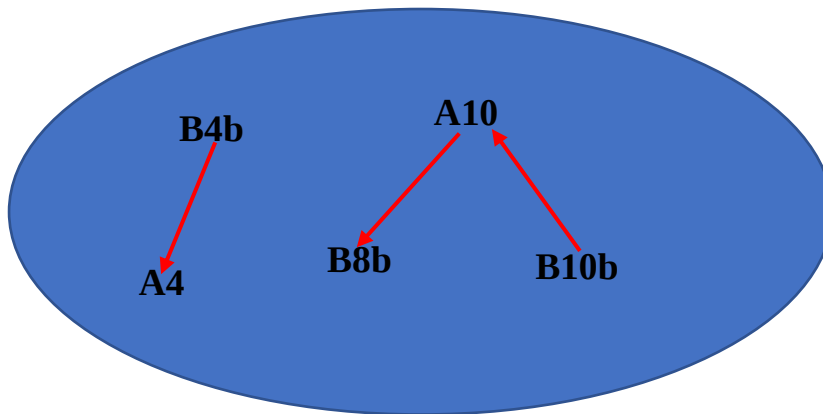
Σενάριο 8: {f4 > 0.50, f8 ≤ 1.65, f0 > 0.50, f3 > 3.31, f0 > 0.45, f4 > 0.50, f2 ≤ 0.05, f8 > 1.65, f1 > 5.25, f8 > 1.66, f4 ≤ 0.45}

Σενάριο 8 σε κανονική μορφή: {gldm_hom > 0.50, gldm_mean ≤ 1.65, sgldm_homog > 0.50, fos_entr > 3.31, sgldm_homog > 0.45, gldm_hom > 0.50, fos_ener ≤ 0.05, gldm_mean > 1.65, sgldm_entr > 5.25, gldm_mean > 1.66, gldm_hom ≤ 0.45}

Ισχύουν: r4, r4b, r8b, r10, r10b, r4b > r4, r10b > r10, r10 > r8b

A4: {r4(Tumor)}

A10: {r10(Tumor), pr7(Tumor)}
 B4b: {r4b(Tumor), pr3(Tumor)}
 B8b: {r8b(Tumor)}
 B10b: {r10b(Tumor), pr15(Tumor)}



Ομάδα9: Explanation=[pr19(_),pr7(_),r10(_),r8(_)]

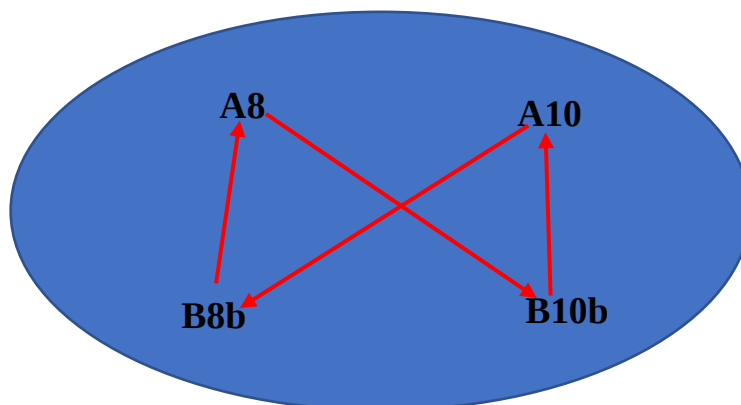
38, 46, 47, 53, 106, 120, 127, 134, 135, 161, 170, 181, 228, 243, 265, 270, 285, 290, 306, 313, 336, 357, 359, 361, 373, 379,

Σενάριο 9: { f8<=1.65, f0 > 0.45, f4 >0.50, f0 > 0.50, f3 > 3.31, f4>0.50, f2<=0.05, f5 <= 5.03, f7 >1.30 }

Σενάριο 9 σε κανονική μορφή: {gldm_mean <=1.65, sgldm_homog> 0.45, gldm_hom>0.50, sgldm_homog> 0.50, fos_ent> 3.31, gldm_hom>0.50, fos_ener<=0.05, gldm_con <= 5.03, gldm_ent >1.30 }

Ισχύουν: r8, r8b, r10, r10b, r8b>r8, r10b>r10, r10 > r8b, r8>r10b

A8: {r8(Tumor), pr19(Tumor)}
 A10: {r10(Tumor), pr7(Tumor)}
 B8b: {r8b(Tumor), pr5(Tumor)}
 B10b: {r10b(Tumor), pr15(Tumor)}



Ομάδα10: Explanation=[r4b(_)]

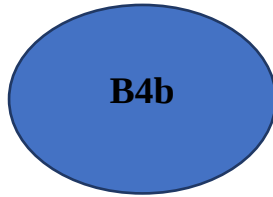
39, 90, 123, 150

Σενάριο 10: {f8 >1.66, f4<=0.45}

Σενάριο 10 σε κανονική μορφή: { gldm_mean >1.66, gldm_hom <=0.45}

Ισχύουν: r4b

B4b: {r4b(Tumor)}



Ομάδα11: Explanation=[pr16(_),pr4(_),r4(_),r6(_)]

103, 148, 163, 191, 199, 214, 241, 247, 275, 288, 314, 321, 390

Σενάριο 11: { f8 >1.65, f1 >5.25, f8 >1.66, f4<=0.45, f5 > 4.89, f2 <=0.06, f7 >1.67, f1 <=5.93 , f6 <= 0.19 }

Σενάριο 11 σε κανονική μορφή: { gldm_mean >1.65, sgldm_entr >5.25, gldm_mean >1.66, gldm_hom <=0.45, gldm_con > 4.89, fos_ener<=0.06, gldm_ent >1.67, sgldm_entr <=5.93 , gldm_eng <= 0.19 }

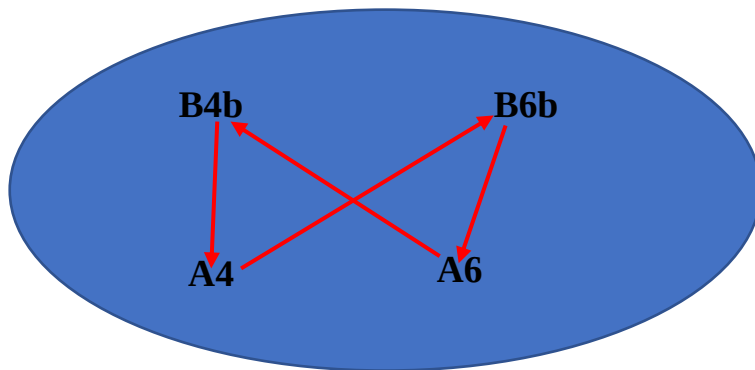
Ισχύουν: r4, r6, r4b, r6b, r6 > r4b, r4 > r6b

A4: {r4(Tumor), pr16(Tumor)}

A6: {r6(Tumor), pr4(Tumor)}

B4b: {r4b(Tumor), pr3(Tumor)}

B6b: {r6b(Tumor), pr14(Tumor)}



Ομάδα12: Explanation=[pr3(_),r4b(_),r8(_)]

138

Σενάριο 12: { f5 <= 5.03, f7 >1.30, f8 >1.66, f4<=0.45, f8>1.65, f1 >5.25}

Σενάριο 12 σε κανονική μορφή: { gldm_con <= 5.03, gldm_ent >1.30, gldm_mean >1.66, gldm_hom <=0.45, gldm_mean >1.65, sgldm_entr >5.25}

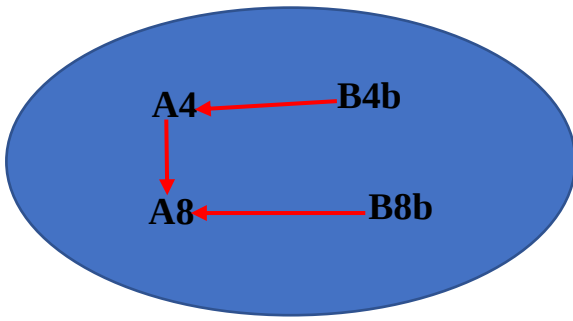
Ισχύουν: r8, r4b, r4, r4b>r4, r4>r8

A4: {r4(Tumor), pr1(Tumor)}

A8: {r8(Tumor)}

B4b: {r4b(Tumor), pr3(Tumor)}

B8b: {r8b(Tumor), pr5(Tumor)}



Ομάδα13: Explanation=[pr12(_),r6(_),r8b(_)]
279

Σενάριο 13: {f8 <= 1.65, f0 > 0.45, f5 > 4.89, f2 <= 0.06, f4>0.50, f2<=0.05}

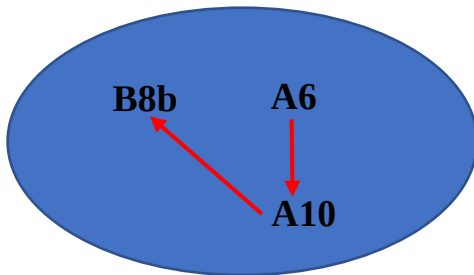
Σενάριο 13 σε κανονική μορφή: {gldm_mean <= 1.65, sgldm_homog > 0.45, gldm_con>4.89, fos_ener <= 0.06, gldm_hom >0.50, fos_ener <=0.05}

Ισχύουν: r8b, r6, r10, r6 >r10, r10 >r8b

A6: {r6(Tumor), pr12(Tumor)}

A10: {r10(Tumor), pr7(Tumor)}

B8b: {r8b(Tumor)}



Ομάδα14: Explanation=[c1(_),pr2(_),r8(_)]
286

Σενάριο 14: { f5 <= 5.03, f7 >1.30, f8 > 1.65, f1 > 5.25 ,f0>0.445, f4>0.445}

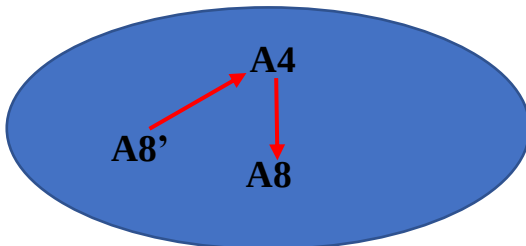
Σενάριο 14 σε κανονική μορφή: { gldm_con<= 5.03, gldm_ent>1.30, gldm_mean > 1.65, sgldm_entr > 5.25 , sgldm_homog >0.445, gldm_hom >0.445}

Ισχύουν: r8, r4, r8> r4, r4>r8

A4: {r4(Tumor), pr1(Tumor)}

A8: {r8(Tumor)}

A8': {r8(Tumor), pr2(Tumor), c1(Tumor)}



Κεφάλαιο 5

5. Αποτελέσματα

5.1 Αξιολόγηση δεδομένων εκπαίδευσης

	Ταυτότητες όγκων
Σωστή πρόβλεψη	1 3 4 5 6 8 9 10 11 13 14 15 17 19 20 21 22 23 24 27 29 30 31 32 33 34 35 36 39 40 41 42 44 45 48 49 50 51 52 54 55 56 61 62 63 64 65 66 67 68 69 72 73 74 77 78 79 82 83 84 85 86 87 88 89 92 93 94 95 96 98 99 100 102 104 105 108 109 110 111 112 113 114 116 117 118 121 122 123 124 125 126 129 131 132 133 136 136 139 140 141 142 143 144 145 146 147 149 151 152 153 154 155 158 159 162 164 166 167 169 171 172 173 174 175 176 177 179 180 182 183 184 185 186 188 190 192 193 194 196 197 198 200 201 202 203 205 206 208 212 215 217 218 219 221 224 225 226 227 229 230 231 233 234 235 236 237 239 240 242 244 245 246 248 249 250 251 252 254 255 256 257 258 259 260 261 262 264 266 267 268 271 272 273 276 277 278 279 282 283 286 287 289 291 292 294 295 296 297 298 299 300 301 302 304 205 307 309 310 311 312 315 316 317 318 319 320 322 324 325 327 328 329 331 332 333 334 335 337 338 339 340 342 343 345 346 347 348 349 350 351 352 354 355 356 358 360 362 364 365 366 367 368 369 371 372 374 375 376 377 378 380 381 382 383 384 385 386 387 388 391 393 394 395 396 400
Λανθασμένη πρόβλεψη	7 12 16 18 25 26 28 37 43 57 58 59 60 70 71 75 76 80 90 91 101 107 115 119 128 138 150 156 157 160 165 168 178 187 189 195 204 207 209 210 211 213 216 220 222 223 232 238 253 263 269 274 280 281 284 293 303 308 323 326 330 341 344 353 363 370 389 392 397 398 399
Δίλημμα	2 38 46 47 53 81 103 106 120 127 130 134 135 148 161 163 170 181 191 199 214 228 241 243 247 265 270 275 285 288 290 306 313 314 321 336 357 359 361 373 379 390

Πίνακας 5: Κατάταξη κάθε όγκου από το πρόγραμμα Επιχειρηματολογίας Γοργία στα δεδομένα εκπαίδευσης

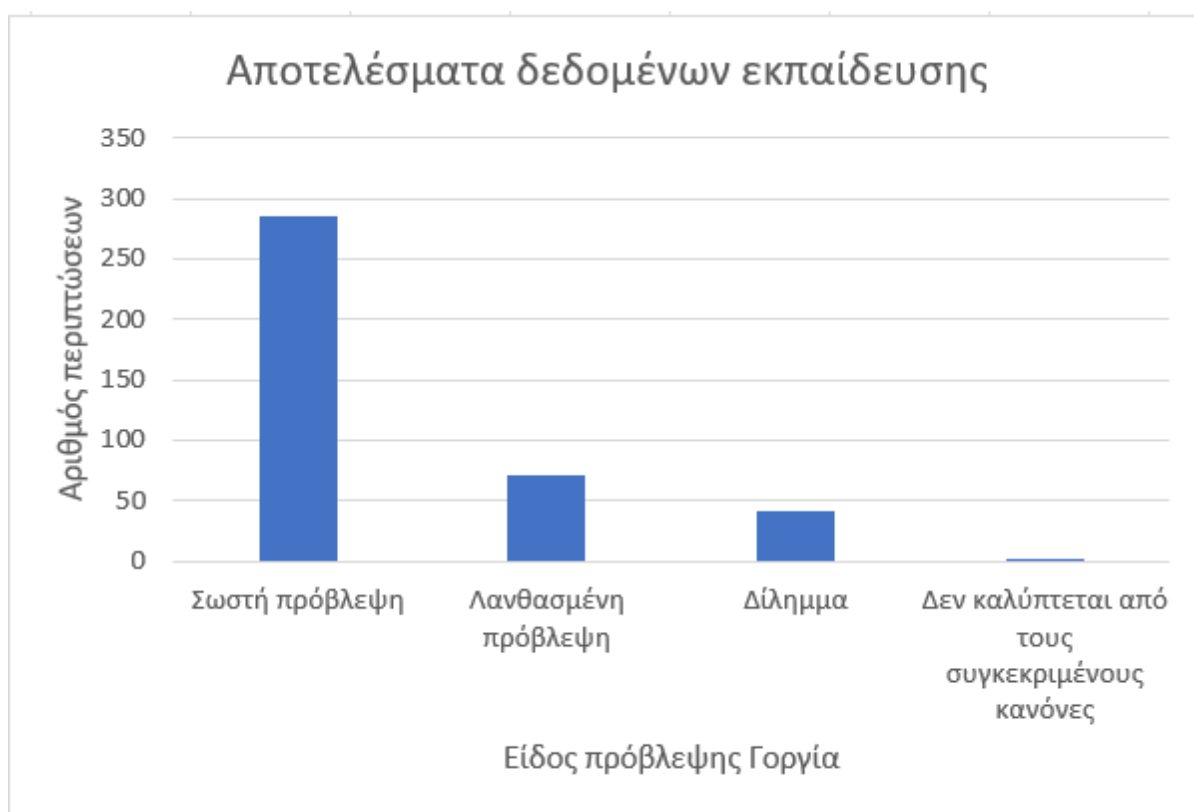
Στον εν λόγω πίνακα παρουσιάζονται τα αποτελέσματα που επέστρεψε το μοντέλο επιχειρηματολογίας για τα δεδομένα εκπαίδευσης. Οι αριθμοί που αναγράφονται αφορούν

την ταυτότητα του κάθε στιγμιότυπου – όγκου. Ο πίνακας δείχνει για κάθε στιγμιότυπο αν το μοντέλο επιχειρηματολογίας αποφάσισε σωστά, λανθασμένα ή δημιουργήθηκε δίλημμα (βλέπε Πίνακας 5: Κατάταξη κάθε όγκου από το πρόγραμμα Επιχειρηματολογίας Γοργία στα δεδομένα εκπαίδευσης).

Ο συνολικός αριθμός σωστών, λανθασμένων προβλέψεων και διλημμάτων παρουσιάζεται στον πιο κάτω πίνακα (βλέπε Πίνακας 6: Αποτελέσματα Επιχειρηματολογίας Γοργία στα δεδομένα εκπαίδευσης).

	Συνολικός αριθμός
Σωστή πρόβλεψη	286
Λανθασμένη πρόβλεψη	71
Δίλημμα	42
Δεν καλύπτεται από τους συγκεκριμένους κανόνες	1

Πίνακας 6: Αποτελέσματα Επιχειρηματολογίας Γοργία στα δεδομένα εκπαίδευσης



Εικόνα 2: Αποτελέσματα Επιχειρηματολογίας Γοργία στα δεδομένα εκπαίδευσης

Στην εν λόγω γραφική παρουσιάζονται γραφικά τα αποτελέσματα του προγράμματος επιχειρηματολογίας Gorgias, ούτως ώστε να μπορεί πιο εύκολα να γίνει αντιληπτή πιο εύκολα η τάση της ορθότητας των αποτελεσμάτων. Διαφαίνεται ότι γενικά το πρόγραμμα του Gorgia προβλέπει σωστά την κατάσταση του κάθε όγκου. Επίσης διαφαίνεται ότι οι

κανόνες που επιλέγηκαν καλύπτουν σχεδόν όλες τις περιπτώσεις (399/400). Τέλος, παρατηρείται ότι είναι μικρό το ποσοστό των περιπτώσεων στις οποίες ο αλγόριθμος δεν αποφασίζει (διλήμματα) (βλέπε Εικόνα 2: Αποτελέσματα Επιχειρηματολογίας Γοργία στα δεδομένα εκπαίδευσης).



Εικόνα 3: Διάκριση περιπτώσεων στις οποίες το πρόγραμμα Επιχειρηματολογίας Gorgia αποφασίζει ή όχι στα δεδομένα εκπαίδευσης

Εφόσον στην περίπτωση του διλήμματος το πρόγραμμα επιχειρηματολογίας αποφασίζει ότι ισχύουν και οι δύο κατηγορίες τότε θεωρείται πως το πρόγραμμα δεν αποφασίζει για τις συγκεκριμένες περιπτώσεις. Διαφαίνεται ότι το πρόγραμμα επιχειρηματολογίας Gorgia στο μεγαλύτερο ποσοστό των περιπτώσεων αποφασίζει να κατατάξει τον όγκο σε μια από τις δύο κατηγορίες (βλέπε Εικόνα 3: Διάκριση περιπτώσεων στις οποίες το πρόγραμμα Επιχειρηματολογίας Gorgia αποφασίζει ή όχι).

5.2 Αξιολόγηση δεδομένων ελέγχου

Στον εν λόγω πίνακα παρουσιάζονται τα αποτελέσματα που επέστρεψε το μοντέλο επιχειρηματολογίας για τα δεδομένα ελέγχου. Οι αριθμοί που αναγράφονται αφορούν την ταυτότητα του κάθε στιγμιότυπου – όγκου. Ο πίνακας δείχνει για κάθε στιγμιότυπο αν το μοντέλο επιχειρηματολογίας αποφάσισε σωστά, λανθασμένα ή δημιουργήθηκε δίλημμα.

	Ταυτότητες όγκων
Σωστή πρόβλεψη	1 2 3 4 6 7 8 9 11 13 14 15 16 17 18 19 20 22 24 25 28 30 31 32 33 34 36 37 38 42 44 45
Λανθασμένη πρόβλεψη	10 12 23 26 35 39 41 43
Δίλημμα	5 21 27 29 40

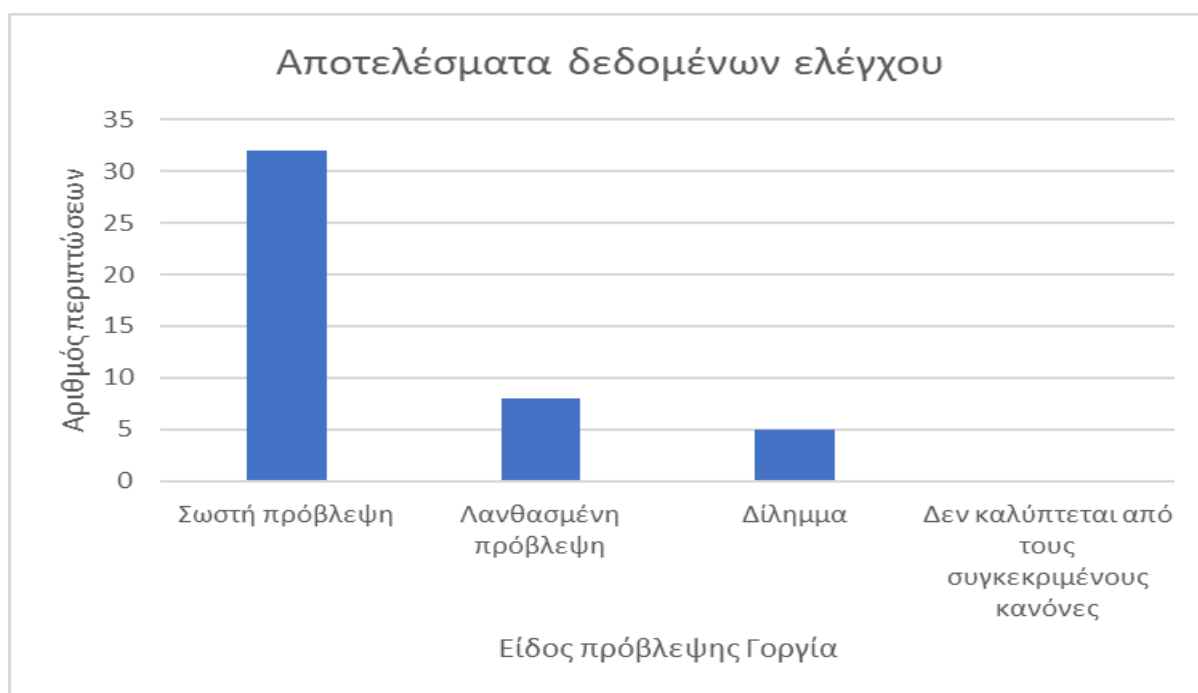
Πίνακας 7: Κατάταξη κάθε όγκου από το πρόγραμμα Επιχειρηματολογίας Gorgias στα δεδομένα ελέγχου

Στον εν λόγω πίνακα παρουσιάζονται τα αποτελέσματα που επέστρεψε το μοντέλο επιχειρηματολογίας για τα δεδομένα ελέγχου. Οι αριθμοί που αναγράφονται αφορούν την ταυτότητα του κάθε στιγμιότυπου – όγκου. Ο πίνακας δείχνει για κάθε στιγμιότυπο αν το μοντέλο επιχειρηματολογίας αποφάσισε σωστά, λανθασμένα ή δημιουργήθηκε δίλημμα (βλέπε Πίνακας 7: Κατάταξη κάθε όγκου από το πρόγραμμα Επιχειρηματολογίας Gorgias στα δεδομένα ελέγχου).

Ο συνολικός αριθμός σωστών, λανθασμένων προβλέψεων και διλημάτων παρουσιάζεται στον πιο κάτω πίνακα (βλέπε Πίνακας 8: Αποτελέσματα Επιχειρηματολογίας Gorgia στα δεδομένα ελέγχου).

	Συνολικός αριθμός
Σωστή πρόβλεψη	32
Λανθασμένη πρόβλεψη	8
Δίλημμα	5
Δεν καλύπτεται από τους συγκεκριμένους κανόνες	0

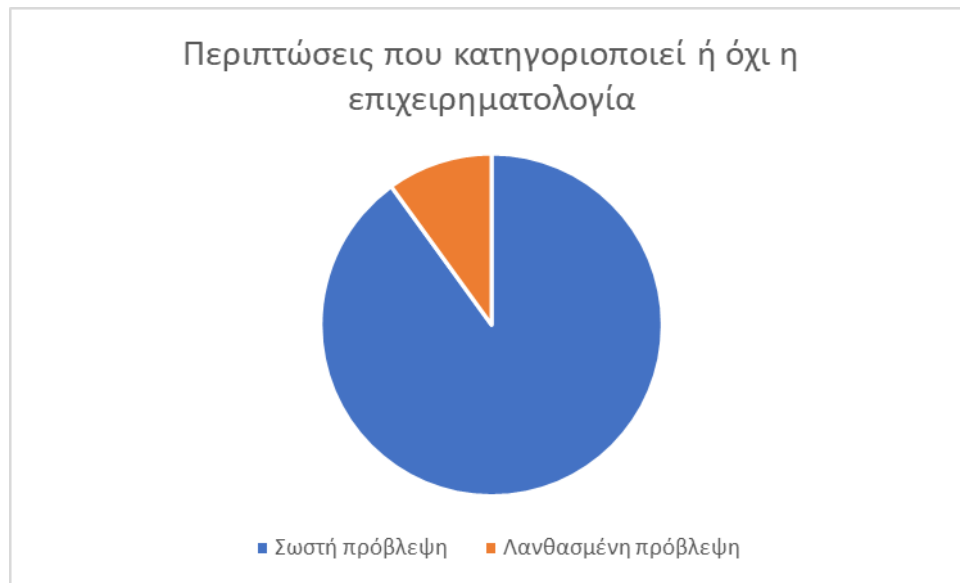
Πίνακας 8: Αποτελέσματα Επιχειρηματολογίας Gorgia στα δεδομένα ελέγχου



Εικόνα 4: Αποτελέσματα Επιχειρηματολογίας Gorgia στα δεδομένα ελέγχου

Στην εν λόγω γραφική παρουσιάζονται γραφικά τα αποτελέσματα του προγράμματος επιχειρηματολογίας Γοργία, ούτως ώστε να μπορεί πιο εύκολα να γίνει αντιληπτή πιο εύκολα η τάση της ορθότητας των αποτελεσμάτων. Διαφαίνεται ότι γενικά το πρόγραμμα του Gorgia, προβλέπει σωστά την κατάσταση του κάθε όγκου. Επίσης διαφαίνεται ότι οι κανόνες

που επιλέγηκαν καλύπτουν όλες τις περιπτώσεις (45/45). Τέλος, παρατηρείται ότι είναι μικρό το ποσοστό των περιπτώσεων στις οποίες ο αλγόριθμος δεν αποφασίζει (διλήμματα) (βλέπε Εικόνα 4: Αποτελέσματα Επιχειρηματολογίας Gorgia στα δεδομένα ελέγχου).



Εικόνα 5: Διάκριση περιπτώσεων στις οποίες το πρόγραμμα Επιχειρηματολογίας Gorgia αποφασίζει ή όχι στα δεδομένα ελέγχου

Εφόσον στην περίπτωση του διλήμματος το πρόγραμμα επιχειρηματολογίας αποφασίζει ότι ισχύουν και οι δύο κατηγορίες τότε θεωρείται πως το πρόγραμμα δεν αποφασίζει για τις συγκεκριμένες περιπτώσεις. Διαφαίνεται ότι το πρόγραμμα επιχειρηματολογίας Gorgia στο μεγαλύτερο ποσοστό των περιπτώσεων αποφασίζει να κατατάξει τον όγκο σε μια από τις δύο κατηγορίες (βλέπε Εικόνα 5: Διάκριση περιπτώσεων στις οποίες το πρόγραμμα Επιχειρηματολογίας Gorgia αποφασίζει ή όχι στα δεδομένα ελέγχου).

5.3 Επεξήγηση

Ακρίβεια προβλέψεων:

Όπως ορίστηκε και στην παρουσίαση των αποτελεσμάτων, στις περιπτώσεις στις οποίες το πρόγραμμα επιχειρηματολογίας περιήλθε σε δίλημμα θεωρείται ότι ουσιαστικά δεν αποφάσισε στο αν ο όγκος είναι καλοήθης ή κακοήθης. Επομένως υπάρχουν δύο πυλώνες αξιολόγησης της ακρίβειας του συστήματος:

- Ο ένας αφορά το ποσοστό των περιπτώσεων στις οποίες ο αλγόριθμος κατατάσσει τον όγκο σε μια εκ των δύο κατηγοριών.
- Ο δεύτερος αφορά το ποσοστό των σωστών προβλέψεων σε συνάρτηση των περιπτώσεων που ο αλγόριθμος αποφάσισε.

Με βάση τα παραπάνω προκύπτουν τα εξής αποτελέσματα:

Δεδομένα εκπαίδευσης

	Ποσοστό
Περιπτώσεις που αποφασίζει	89,23 %
Περιπτώσεις που δεν αποφασίζει	10,77 %

Δεδομένα ελέγχου

	Ποσοστό
Περιπτώσεις που αποφασίζει	88.89 %
Περιπτώσεις που δεν αποφασίζει	11,11 %

Εκ των περιπτώσεων που ο αλγόριθμος αποφασίζει να κατατάξει τον όγκο σε μια εκ των δύο περιπτώσεων η ακρίβεια επιτυχημένων/αποτυχημένων προβλέψεων είναι ως εξής:

Δεδομένα εκπαίδευσης

	Ποσοστό
Επιτυχημένες προβλέψεις	80.11 %
Αποτυχημένες προβλέψεις	19.99 %

Δεδομένα ελέγχου

	Ποσοστό
Επιτυχημένες προβλέψεις	80.00 %
Αποτυχημένες προβλέψεις	20.00 %

5.4 Συμπεράσματα

Ομοιομορφία αποτελεσμάτων

	Δεδομένα εκπαίδευσης	Δεδομένα ελέγχου
Ποσοστό περιπτώσεων που ο αλγόριθμος δεν αποφασίζει	10,77 %	11,11 %
Ποσοστό Ακρίβειας	80.11 %	80.00 %
Αριθμός περιπτώσεων που δεν καλύπτεται από τους κανόνες	1 στις 400	0 στις 45

Πίνακας 9: Ομοιομορφία δεδομένων μεταξύ δεδομένων εκπαίδευσης και ελέγχου

Καταρχάς, παρατηρείται το γεγονός ότι τα αποτελέσματα που συλλέχθηκαν για τα δεδομένα εκπαίδευσης είναι απόλυτα ανάλογα με τα αποτελέσματα που συλλέχθηκαν για τα δεδομένα ελέγχου. Διαφαίνεται, ότι σε όλους τους άξονες αξιολόγησης, δηλαδή ακρίβεια, κάλυψη κανόνων και περιπτώσεις που αποφασίζει ο αλγόριθμος, τα αποτελέσματα των δύο συλλογών δεδομένου είναι σχεδόν τα ίδια, αποδεικνύοντας έτσι, ότι όσα πειράματα ελέγχθηκαν και αξιολογήθηκαν στα δεδομένα εκπαίδευσης επαληθεύονται από τα δεδομένα ελέγχου (βλέπε Πίνακας 9: Ομοιομορφία δεδομένων μεταξύ δεδομένων εκπαίδευσης και ελέγχου).

Ακρίβεια – Ποσοστό Διλημμάτων

Παρατηρείται επίσης το γεγονός ότι ο αλγόριθμος σε ένα ποσοστό κοντά στο 90% αποφασίζει για το που θα κατατάξει τον κάθε όγκο, ποσοστό πολύ ικανοποιητικό αν αναλογιστεί κανείς ότι η εν λόγω μελέτη ασχολείται με την δημιουργία ενός αλγορίθμου που να λειτουργεί ως ένα υποβοηθητικό εργαλείο για τους γιατρούς στην πρόγνωση διαγνώσεων όγκων, παρέχοντας συμβουλές με τις αντίστοιχες επεξηγήσεις, και όχι σαν ένα αυτόνομο σύστημα πρόβλεψης.

Για τον ίδιο λόγο, το ποσοστό επιτυχίας το οποίο φτάνει στο 80% θεωρείται ως ιδιαίτερα ικανοποιητικό και ενθαρρυντικό, λαμβάνοντας υπόψη ότι κάθε πρόβλεψη συνοδεύεται από την αντίστοιχη επεξήγηση και ο γιατρός μπορεί να κρίνει αν είναι ικανοποιητική ή όχι.

Πληρότητα Κανόνων

Πρόκειται για έναν βασικό άξονα της ποιότητας της έρευνας και έναν άξονα στον οποίο τα αποτελέσματα είναι εξαιρετικά ικανοποιητικά. Αναλυτικότερα, δεδομένου του ότι στα δεδομένα εκπαίδευσης μόλις μια περίπτωση στις 400 βρέθηκε που κανένας κανόνας δεν κάλυπτε και καμία περίπτωση για τα δεδομένα ελέγχου, δεικνύει το γεγονός ότι οι κανόνες που επιλέχθηκαν καλύπτουν πλήρως, ολόκληρη την συλλογή των δεδομένων.

Κεφάλαιο 6

6. Συζήτηση

6.1 Σύντομη περίληψη

Ένεκα της μεγάλης έξαρσης των καρκινογόνων όγκων που εντοπίζεται στην σύγχρονη κοινωνία, διαπιστώθηκε η ανάγκη για δημιουργία συστημάτων που να μπορούν να διευκολύνουν το έργο των γιατρών όσον αφορά την πρόγνωση τους. Για τον λόγο αυτό, η εν λόγω έρευνα ασχολείται με την δημιουργία ενός συστήματος επιχειρηματολογίας το οποίο θα παρέχει συμβουλευτική υποστήριξη στους γιατρούς κατά την εξέταση των γυναικολογικών όγκων. Αναλυτικότερα, δεδομένο κάποιων χαρακτηριστικών των όγκων τα οποία συλλέγονται με τις υφιστάμενες ιατρικές μεθόδους, δημιουργήθηκε ένας αλγόριθμος επιχειρηματολογίας ο οποίος θα κατατάσσει τον κάθε όγκο σε καλοήγη ή κακοήγη, παρέχοντας την αντίστοιχη επεξήγηση. Με αυτό τον τρόπο ο γιατρός θα μπορεί να κρίνει αν τα αίτια που οδήγησαν το πρόγραμμα στην κατάταξη του όγκου είναι ικανοποιητικά και άρα μπορεί να τα εμπιστευθεί στην πρόγνωση του ή όχι.

Η διαδικασία που ακολουθήθηκε για την δημιουργία του συστήματος επιχειρηματολογίας στον Gorgia περιλάμβανε αρκετή ανάλυση των δεδομένων, δημιουργία κανόνων και αξιολόγηση. Αναλυτικότερα, πρώτα – πρώτα έγινε η επιλογή των πιο σημαντικών χαρακτηριστικών τα οποία θα λαμβάνονταν υπόψη κατά την κατηγοριοποίηση των όγκων. Η επιλογή έγινε με βάση προηγούμενες έρευνες και με εκτέλεση Feature Importance αλγορίθμων με την χρήση Random Forest μοντέλων. Αφού ολοκληρώθηκε η επιλογή των χαρακτηριστικών, έγινε ο διαχωρισμός των δεδομένων σε δεδομένα εκπαίδευσης, που θα χρησιμοποιούνταν για την πειραματική δόμηση του αλγορίθμου, και δεδομένα ελέγχου που θα χρησιμοποιούνταν για την επαλήθευση των ευρημάτων και συμπερασμάτων που διαπιστώθηκαν από την πρώτη συλλογή δεδομένων.

Με την ολοκλήρωση της προεργασίας των δεδομένων, ακολούθησε η δημιουργία των βασικών κανόνων για το πρόγραμμα επιχειρηματολογίας. Η δημιουργία των κανόνων έγινε με την χρήση ενός αλγορίθμου Random Forest που δημιουργούσε δεντρικές δομές με τους κανόνες κατάταξης. Η επιλογή των τεσσάρων πιο κατάλληλων κανόνων έγινε με βάση τους άξονες της ακρίβειας και της πληρότητας κάλυψης όλων των δεδομένων. Στην συνέχεια, δημιουργήθηκαν δυο ζεύγη κανόνων και στο κάθε ζεύγος τέθηκαν κανόνες προτεραιότητας για να επιλυθούν τα διλήμματα που προέκυψαν σε κάθε ζεύγος. Εν συνεχεία για κάθε ένα από τους τέσσερις βασικούς κανόνες επιλέγθηκε ένας κανόνας προτεραιότητας, ο οποίος υπερείχε του βασικού του κανόνα ούτως ώστε να διορθώνει τις περιπτώσεις λανθασμένων προβλέψεων που υπέπιπτε ο βασικός κανόνας. Για να διασφαλιστεί, ωστόσο, το γεγονός ότι οι βελτιωτικοί κανόνες δεν θα δημιουργήσουν νέα λάθη, γενικά οι υπόλοιποι τρεις βασικοί κανόνες θα υπερέχουν του βελτιωτικού κανόνα του τέταρτου βασικού.

Μετά τον καθορισμό ολόκληρης της συλλογής κανόνων και των προτεραιοτήτων τους, έγινε η δημιουργία του προγράμματος επιχειρηματολογίας στο Gorgias Cloud. Ακολούθησε η αξιολόγηση των αποτελεσμάτων, τόσο για τα δεδομένα εκπαίδευσης όσο για τα δεδομένα ελέγχου. Οι βασικοί πυλώνες της αξιολόγησης αφορούσαν το ποσοστό των περιπτώσεων στις οποίες ο αλγόριθμος αποφάσιζε (δεν υπήρχε δίλημμα), η ακρίβεια των προβλέψεων και η πληρότητα κάλυψης των συλλογών δεδομένων.

6.2 Σύντομη κριτική

Τα αποτελέσματα που προέκυψαν από την εκτέλεση του προγράμματος επιχειρηματολογίας είναι ιδιαίτερα ενθαρρυντικά. Καταρχάς, όσον αφορά την ακρίβεια των προβλέψεων προκύπτει ένα ποσοστό κοντά στο 80% το οποίο είναι ικανοποιητικό για ένα σύστημα το οποίο θα παρέχει συμβουλευτική υποστήριξη για πρόγνωση με την παροχή των απαραίτητων επεξηγήσεων. Επίσης, θετικό είναι το γεγονός ότι διατηρείται αρκετά χαμηλά το ποσοστό των διλημμάτων, κοντά στο 11%. Αυτό σημαίνει ότι ο αλγόριθμος, στην πλειοψηφία των περιπτώσεων αποφασίζει να κατατάξει τον όγκο είτε ως καλοήθη είτε ως κακοήθη. Τον τρίτο άξονα αξιολόγησης αποτελεί το ποσοστό κάλυψης των δεδομένων από τους κανόνες, δηλαδή το κατά πόσο υπάρχει τουλάχιστον ένας κανόνας που να μπορεί να κατατάξει τον κάθε όγκο σε μια κατηγορία. Η επίδοση στην συγκεκριμένη πτυχή αξιολόγησης είναι εξαιρετική καθώς η κάλυψη στα δεδομένα ελέγχου είναι στο 100% ενώ στα δεδομένα εκπαίδευσης μόλις 1 στις 400 περιπτώσεις δεν καλύπτεται από τους κανόνες. Τέλος, ένα γεγονός που αποδεικνύει την αξιοπιστία των αποτελεσμάτων είναι το γεγονός ότι όλες οι μετρήσεις που συλλέχθηκαν είναι ομοιόμορφες για τις δύο συλλογές δεδομένων, δηλαδή για τα δεδομένα εκπαίδευσης και ελέγχου. Αυτό δείχνει ότι όλα τα αποτελέσματα και συμπεράσματα προκύπτουν από τα δεδομένα εκπαίδευσης, επαληθεύονται από τα δεδομένα ελέγχου.

6.3 Σύντομη συζήτηση

Η σημαντικότητα και η συνεισφορά της εν λόγω πτυχιακής έρευνας διακρίνεται σε τρεις διαφορετικούς άξονες:

1. Αξία αποτελεσμάτων
2. Αξία μεθοδολογίας
3. Αξία επιχειρηματολογίας

Αξία αποτελεσμάτων

Από την παρουσίαση των αποτελεσμάτων πέρα από την σημαντικότητα της επίτευξης μιας αρκετά καλής ακρίβειας, ιδιαίτερα ενθαρρυντικό είναι το γεγονός ότι εντοπίζεται άμεση

αντιστοιχία των αποτελεσμάτων που προέκυψαν στα αποτελέσματα των δεδομένων εκπαίδευσης και των δεδομένων ελέγχου. Η εν λόγω ομοιομορφία δείχνει ότι όλα τα ευρήματα που εντοπίστηκαν στην πρώτη συλλογή δεδομένων επαληθεύονται και ενισχύονται από την δεύτερη.

Αξία μεθοδολογίας

Πρωτεύον στόχος της έρευνας ήταν ο εντοπισμός μιας απλής διαδικασίας ανάλυσης των δεδομένων και δημιουργίας συστήματος επιχειρηματολογίας καθώς και η κατάλληλη περιγραφή όλης της διαδικασίας. Επομένως, επιτεύχθηκε η εκτέλεση και περιγραφή μιας απλής διαδικασίας η οποία μπορεί εύκολα να επαναληφθεί ακολουθώντας τα βήματα που περιγράφονται στην έρευνα. Η σημαντικότητα της μεθοδολογίας αυξάνεται από το γεγονός ότι η συγκεκριμένη μέθοδος είναι ανεξάρτητη του πλαισίου εφαρμογής, δηλαδή μπορεί πολύ εύκολα να εφαρμοστεί σε οποιοδήποτε άλλο τομέα πέρα από τον καρκίνο, που μελετά η συγκεκριμένη έρευνα.

Αξία επιχειρηματολογίας

Ακόμα μια σημαντική πτυχή είναι η παρουσίαση των αποτελεσμάτων συνοδευόμενη με την κατάλληλη επεξήγηση. Αναλυτικότερα, στόχος της έρευνας ήταν η δημιουργία ενός συστήματος το οποίο θα παρέχει συμβουλευτική υποστήριξη στους γιατρούς για την πρόγνωση του καρκίνου. Διαφαίνεται, λοιπόν, η μεγάλη σημασία της παροχής κατάλληλων επεξηγήσεων μαζί με την κάθε πρόβλεψη, ούτως ώστε ο εκάστοτε γιατρός να μπορεί να κρίνει αν τα ευρήματα που έκαναν το σύστημα να πάρει την κάθε πρόγνωση, είναι αξιόπιστα ή όχι.

Βιβλιογραφία

- [1] Neofytou, M. S., Tanos, V., Constantinou, I., Kyriacou, E. C., Pattichis, M. S., & Pattichis, C. S. (2014). Computer-aided diagnosis in hysteroscopic imaging. *IEEE Journal of Biomedical and Health Informatics*, 19(3), 1129-1136.
- [2] Neofytou, M. S., Tanos, V., Pattichis, M. S., Pattichis, C. S., Kyriacou, E. C., & Koutsouris, D. D. (2007). A standardised protocol for texture feature analysis of endoscopic images in gynaecological cancer. *BioMedical Engineering OnLine*, 6(1), 1-21.
- [3] Vassiliades, Alexandros, Ilias Papadimitriou, Nick Bassiliades and Theodore Patkos. “Visual Gorgias: A Mechanism for the Visualization of an Argumentation Dialogue.” 25th Pan-Hellenic Conference on Informatics (2021): n. pag.
- [4] Wikimedia Foundation. (2021, August 9). Μηχανική μάθηση. Wikipedia. Retrieved May 25, 2022, from https://el.wikipedia.org/wiki/%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE_%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7
- [5] Κεφάλαιο 18 - aibook.csd.auth.gr. (n.d.). Retrieved May 24, 2022, from <http://aibook.csd.auth.gr/include/slides/Chap18.pdf>
- [6] Vassiliades, A., Bassiliades, N., & Patkos, T. (2021). Argumentation and explainable artificial intelligence: a survey. *The Knowledge Engineering Review*, 36.
- [7] Imielski, R. (2020, November 4). How medical image analysis can impact healthcare: Objectivity blog. Objectivity. Retrieved May 25, 2022, from <https://www.objectivity.co.uk/blog/medical-image-analysis/>
- [8] Endoscopy. NOVA IVF. (n.d.). Retrieved May 25, 2022, from <https://www.novaivffertility.com/infertility-treatments/endoscopy>
- [9] Prentzas, N., Kakas, A., & Pattichis, C. S. (n.d.). Technical Report, Department of Computer Science, University of Cyprus, 2021.

Παράρτημα

Ενδεικτικό παράδειγμα δεδομένων

sgldm_asm	0.008576
sgldm_con	2.795231
sgldm_cor	0.964766
sgldm_var	39.76377
sgldm_homog	0.516985
sgldm_sumav	237.8087
sgldm_sumvar	156.2598
sgldm_sument	3.885115
sgldm_entr	5.055711
sgldm_difvar	1.166222
sgldm_diff_ent	1.380824
sgldm_infcor1	-0.42942
sgldm_infcor2	0.965615
fos_mean	117.8972
fos_var	40.42327
fos_med	118.1537
fos_mode	123
fos_skew	0.031574
fos_kurt	2.408296
fos_ener	0.044393
fos_ent	3.225842
gldm_hom	0.517302
gldm_con	2.790776
gldm_eng	0.284793
gldm_ent	1.413936
gldm_mean	1.250719
Y	1

Όπου Y η κατηγοριοποίηση του όγκου σε καλοήγη (0) ή κακοήγη (1) και τα υπόλοιπα αποτελούν τα χαρακτηριστικά του όγκου. Το πρώτο μισό των χαρακτηριστικών είναι η μέθοδος (sf (fos), glds (gldm), sgldm (sgldm)).

sgldm_homog	sgldm_entr	fos_ener	fos_ent	gldm_hom	gldm_con	gldm_eng	gldm_ent	gldm_mean	Y
0.516985	5.055711	0.044393	3.225842	0.517302	2.790776	0.284793	1.413936	1.250719	1
0.380123	6.011866	0.029813	3.659241	0.380428	9.595113	0.184369	1.926381	2.227534	0
0.413026	5.902123	0.029083	3.685005	0.413379	7.541495	0.204071	1.81595	1.958318	0
0.362175	6.381823	0.020749	4.018208	0.362502	11.518373	0.177567	1.97324	2.374016	0
0.40969	5.845777	0.028882	3.650776	0.410155	6.865142	0.202373	1.79834	1.934383	1
0.497419	5.688058	0.024228	3.80836	0.497723	3.018935	0.277394	1.441226	1.317398	1
0.470936	5.928159	0.022988	3.934273	0.47128	4.627234	0.247715	1.600173	1.53687	1

Κώδικας επιχειρηματολογίας στο Cloud Gorgias

:- dynamic feature0/2, feature1/2, feature2/2, feature3/2, feature4/2, feature5/2, feature6/2, feature7/2, feature8/2.

complement(malignant(Tumor), benign(Tumor)).
complement(benign(Tumor), malignant(Tumor)).

% Rule4: feature_8 > 1.65 & feature_1 > 5.25 -> 0
rule(r4(Tumor), benign(Tumor),[]):-feature8(Tumor,Value),Value>1.65,
feature1(Tumor,Value2),Value2>5.25.

% Rule 4b: feature_8 >1.66 &feature_4<=0.45 -> 1
rule(r4b(Tumor), malignant(Tumor),[]):-feature8(Tumor,Value),Value>1.66,
feature4(Tumor,Value2),Value2=<0.45.

% R4b > R4
rule(pr3(Tumor), prefer(r4b(Tumor), r4(Tumor)),[]).

% Rule 8: feature_5 <= 5.03 & feature_7 >1.30 -> 1
rule(r8(Tumor), malignant(Tumor),[]):-feature5(Tumor,Value),Value=<5.03,
feature7(Tumor,Value2),Value2>1.30.

% Rule 8b: feature_4>0.50&feature_2<=0.05 -> 0
rule(r8b(Tumor), benign(Tumor),[]):-feature4(Tumor,Value),Value>0.50,
feature2(Tumor,Value2),Value2=<0.05.

% R8b > R8
rule(pr3(Tumor), prefer(r8b(Tumor), r8(Tumor)),[]).

% R4 > R8
rule(pr1(Tumor), prefer(r4(Tumor), r8(Tumor)),[]).

% R8 > R4
rule(pr2(Tumor),prefer(r8(Tumor), r4(Tumor)),[]):-feature0(Tumor,Value),Value>0.445,
feature4(Tumor,Value2),Value2>0.445.
rule(c1(Tumor),prefer(pr2(Tumor),pr1(Tumor)),[]).

% Rule 6: feature_5 > 4.89 & feature_2 <= 0.06 -> 0
rule(r6(Tumor), benign(Tumor),[]):-feature5(Tumor,Value),Value>4.89,
feature2(Tumor,Value2),Value2=<0.06.

% Rule 6b: feature_7 > 1.67 & feature_1 <= 5.93 & feature_6 <= 0.19
rule(r6b(Tumor), malignant(Tumor),[]):-feature7(Tumor,Value),Value>1.67,
feature1(Tumor,Value2),Value2=<5.93, feature6(Tumor,Value3),Value3=<0.19.

% R6b > R6
rule(pr14(Tumor), prefer(r6b(Tumor), r6(Tumor)),[]).

% R6 > R4b
rule(pr4(Tumor), prefer(r6(Tumor), r4b(Tumor)),[]).

% R4 > R6b
rule(pr16(Tumor), prefer(r4(Tumor), r6b(Tumor)),[]).

% Rule 10: feature_8 <= 1.65 & feature_0 > 0.45 -> 1
 rule(r10(Tumor), malignant(Tumor),[]):-feature8(Tumor,Value),Value=<1.65,
 feature0(Tumor,Value2),Value2>0.45.

% Rule 10b: feature_4 > 0.50 & feature_0 > 0.50 & feature_3 > 3.31
 rule(r10b(Tumor), benign(Tumor),[]):-feature4(Tumor,Value),Value>0.50,
 feature0(Tumor,Value2),Value2>0.50, feature3(Tumor,Value3),Value3>3.31.

% R10b > R10
 rule(pr15(Tumor), prefer(r10b(Tumor), r10(Tumor)),[]).

% R6 > R10
 rule(pr12(Tumor), prefer(r6(Tumor), r10(Tumor)),[]).

% R10 > R6 AN feature0>0.454 και feature0<0.46
 rule(pr13(Tumor),prefer(r10(Tumor), r6(Tumor)),[]):-feature0(Tumor,Value),Value>0.454,
 feature0(Tumor,Value2),Value2<0.46.
 rule(c6(Tumor),prefer(pr13(Tumor),pr12(Tumor)),[]).

% R8 > R10b
 rule(pr19(Tumor), prefer(r8(Tumor), r10b(Tumor)),[]).

% R10 > R8b
 rule(pr7(Tumor), prefer(r10(Tumor), r8b(Tumor)),[]).

Δείγμα δεδομένων που χρησιμοποιήθηκαν από τον Gorgia

Όγκος με ταυτότητα 1	Όγκος με ταυτότητα 2	Όγκος με ταυτότητα 3	Όγκος με ταυτότητα 4
feature0(1,0.495107).	feature0(2,0.482821).	feature0(3,0.315535).	feature0(4,0.481576).
feature1(1,5.054545).	feature1(2,5.754925).	feature1(3,6.658286).	feature1(4,4.902363).
feature2(1,0.048805).	feature2(2,0.025003).	feature2(3,0.019756).	feature2(4,0.059395).
feature3(1,3.159896).	feature3(2,3.790581).	feature3(3,4.114404).	feature3(4,2.967904).
feature4(1,0.495371).	feature4(2,0.483115).	feature4(3,0.31579).	feature4(4,0.48196).
feature5(1,3.049931).	feature5(2,4.065492).	feature5(3,19.72166).	feature5(4,3.659105).
feature6(1,0.275601).	feature6(2,0.252686).	feature6(3,0.141263).	feature6(4,0.254847).
feature7(1,1.445991).	feature7(2,1.567833).	feature7(3,2.224246).	feature7(4,1.531406).
feature8(1,1.326772).	feature8(2,1.468941).	feature8(3,3.082115).	feature8(4,1.432883).

Στον πιο πάνω πίνακα φαίνονται κάποια παραδείγματα με το πως γίνεται η ανάθεση των τιμών των χαρακτηριστικών των όγκων που χρησιμοποιούνται από τον Gorgia. Για παράδειγμα, με τη χρήση του **feature0(1,0.495107)**, δηλώνεται ότι για τον όγκο 1 η τιμή του feature 0 είναι 0.495107. Αντίστοιχα και για τα υπόλοιπα χαρακτηριστικά.