



ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

«Υλοποίηση του Wart Immunotherapy Prediction»

Για το μάθημα

Συστήματα Υποστήριξης Αποφάσεων

Ονοματεπώνυμο: Γρίβας Πασχάλης

Εξάμηνο Φοίτησης: ΣΤ΄

A.M.: Π2017082

Περίληψη

Η εφαρμογή “Wart Immunotherapy Prediction” είναι ένα ιατρικό σύστημα υποστήριξης αποφάσεων το οποίο είναι ικανό να βοηθήσει τους ιατρούς να αποφασίσουν αν η Ανοσοθεραπεία είναι ο κατάλληλος τύπος θεραπείας για την αντιμετώπιση των κονδυλωμάτων των ασθενών τους. Η βασική λειτουργία της εφαρμογής είναι η συλλογή πληροφοριών σχετικά με τον ασθενή όπως, η ηλικία, το φύλο, το χρονικό διάστημα πριν από τη θεραπεία, ο αριθμός κονδυλωμάτων, ο τύπος κονδυλωμάτων, το μέγεθος της επιφάνειας του μεγαλύτερου κονδυλώματος και η διάμετρος της σκλήρυνσης. Στη συνέχεια, η εφαρμογή βασισμένη σε ένα μοντέλο πρόβλεψης, ταξινομεί το δείγμα το οποίο ο χρήστης εισήγαγε, σε μία από τις δύο κατηγορίες, η θεραπεία πρόκειται να είναι επιτυχής ή όχι. Με αυτόν τον τρόπο ο ειδήμων ιατρός θα μπορεί να λάβει την απόφαση αν θα εφαρμόσει την συγκεκριμένη μέθοδο ή αν θα ακολουθήσει κάποια άλλη.

Στόχος αυτής της εργασίας είναι η υλοποίηση του μοντέλου πρόβλεψης δεδομένου ενός σχετικού συνόλου δεδομένων, χρησιμοποιώντας τον κατάλληλο αλγόριθμο και η δημιουργία της εφαρμογής η οποία θα κάνει χρήση αυτού του μοντέλου για να προβλέψει νέα δείγματα.

Λέξεις Κλειδιά

Ανοσοθεραπεία, Κονδυλώματα, Σύστημα υποστήριξης αποφάσεων, Μοντέλο πρόβλεψης, Weka, Android, Java

Περιεχόμενα

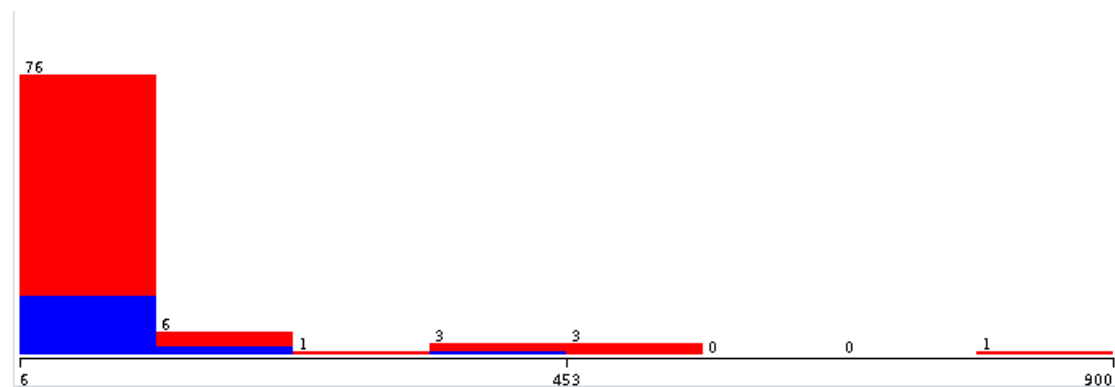
Περίληψη	2
Περιεχόμενα	3
1. Δημιουργία του μοντέλου	4
1.1 Περιγραφή των βημάτων εξαγωγής του μοντέλου	4
1.2 Σύγκριση αλγορίθμων	5
2. Υλοποίηση εφαρμογής	9
2.1 Εισαγωγή νέου δείγματος	9
2.2 Πρόβλεψη του αποτελέσματος	10
2.3 Διεπαφή χρήστη	11
Βιβλιογραφία - Παραπομπές	12

Κεφάλαιο 1

Δημιουργία του μοντέλου

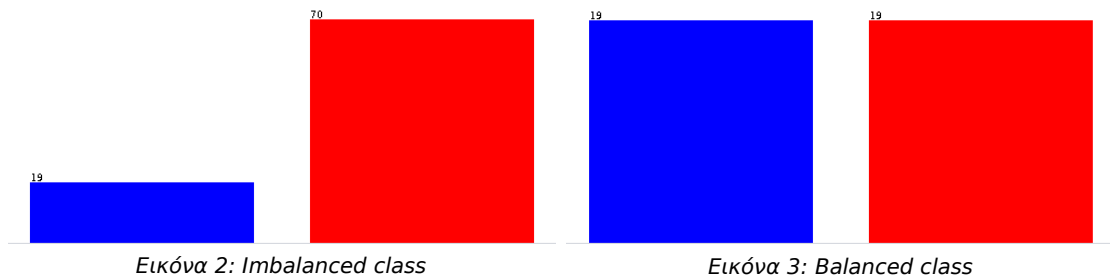
1.1 Περιγραφή των βημάτων εξαγωγής του μοντέλου

Για την παραγωγή του μοντέλου χρησιμοποιήθηκε το λογισμικό μηχανικής μάθησης “Weka 3.8.4”. Αρχικά δημιουργήθηκε ένα αρχείο .arff, με τα δεδομένα από το dataset, με κατάλληλη δομή ώστε να αναγνωρίζεται από το Weka. Το αρχείο αυτό περιλαμβάνει 90 δείγματα (instances) ασθενών με κονδυλώματα, στους οποίους εφαρμόστηκε η μέθοδος της ανοσοθεραπείας και κάθε ένα από αυτά περιλαμβάνει 8 χαρακτηριστικά. Τα χαρακτηριστικά αυτά είναι: η ηλικία (age), το φύλο (sex), το χρονικό διάστημα πριν από την θεραπεία (time), ο αριθμός των κονδυλωμάτων (number_of_warts), ο τύπος κονδυλωμάτων (type), το μέγεθος της επιφάνειας του μεγαλύτερου κονδυλώματος (area), η διάμετρος της σκλήρυνσης (induration_diameter) και το αποτέλεσμα της θεραπείας (result_of_treatment) (κλάση). Παράλληλα, από το αρχείο αφαιρέθηκε ένα δείγμα το οποίο έπαιρνε μία ακραία τιμή στο χαρακτηριστικό area όπως φαίνεται και στην εικόνα 1. παρακάτω.



Εικόνα 1: Ακραία τιμή του χαρακτηριστικού area

Στη συνέχεια, ένα άλλο πρόβλημα το οποίο υπήρχε στα δεδομένα ήταν το γεγονός ότι υπήρχε μεγάλη απόκλιση στον αριθμό των δειγμάτων που ταξινομούνται σε κάθε κατηγορία (class imbalance), όπως φαίνεται στην εικόνα 2. Προκειμένου να αντιμετωπιστεί αυτό το πρόβλημα, εφαρμόστηκε το φίλτρο του Weka “SpreadSubsample” το οποίο διατήρησε 19 δείγματα σε κάθε μία από τις 2 κατηγορίες, όπως φαίνεται και στην εικόνα 3.



1.2 Σύγκριση αλγορίθμων

Προκειμένου να γίνει η εξαγωγή ενός μοντέλου κανόνων το οποίο θα μπορεί να προβλέπει νέα δείγματα, είναι απαραίτητη η χρήση αλγορίθμων μηχανικής μάθησης. Μετά τις αλλαγές στα δεδομένα, οι οποίες περιγράφηκαν στην ενότητα 1.1, εφαρμόστηκαν διάφοροι αλγόριθμοι ώστε να επιτευχθούν τα καλύτερα δυνατά αποτελέσματα. Παρακάτω, στις εικόνες 4 έως 9 φαίνονται τα αποτελέσματα των αλγορίθμων Naive Bayes, ενός κοντινότερου γείτονα (ibk k=1), τριών κοντινότερων γειτόνων (ibk k=3), J48, Adaboost με J48 και Bagging με J48 αντίστοιχα. Για καθένα από τα αποτελέσματα συγκρίνονται οι τιμές “TP Rate” (Sensitivity) και “Precision” της κάθε κατηγορίας. Στη συνέχεια, επιλέγεται ο αλγόριθμος εκείνος ο οποίος δίνει τις καλύτερες τιμές “TP Rate” στο χαρακτηριστικό το οποίο είναι πιο σημαντικό για την εφαρμογή να προβλέπεται σωστά, δεδομένου ότι σε όλες τις υπόλοιπες κατηγορίες υπάρχει πιθανότητα σωστής πρόβλεψης μεγαλύτερη από την θεωρητικά τυχαία (π.χ. 0.5 στην περίπτωση των δύο κατηγοριών).

```
Correctly Classified Instances      23          60.5263 %
Incorrectly Classified Instances   15          39.4737 %
Kappa statistic                   0.2105
Mean absolute error                0.447
Root mean squared error            0.5396
Relative absolute error            89.2749 %
Root relative squared error       107.7682 %
Total Number of Instances         38

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.789	0.579	0.577	0.789	0.667	0.226	0.587	0.547	0
	0.421	0.211	0.667	0.421	0.516	0.226	0.587	0.606	1
Weighted Avg.	0.605	0.395	0.622	0.605	0.591	0.226	0.587	0.576	

```

=== Confusion Matrix ===
  a  b   <-- classified as
15  4 | a = 0
11  8 | b = 1

```

Εικόνα 4: Αποτελέσματα αλγορίθμου Naive Bayes

```

Correctly Classified Instances      20          52.6316 %
Incorrectly Classified Instances    18          47.3684 %
Kappa statistic                    0.0526
Mean absolute error                0.4751
Root mean squared error            0.6695
Relative absolute error            94.8943 %
Root relative squared error        133.7055 %
Total Number of Instances          38

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.632   0.579   0.522      0.632   0.571      0.054    0.522    0.511     0
                0.421   0.368   0.533      0.421   0.471      0.054    0.522    0.512     1
Weighted Avg.   0.526   0.474   0.528      0.526   0.521      0.054    0.522    0.512

=== Confusion Matrix ===

  a  b  <-- classified as
12  7  |  a = 0
11  8  |  b = 1

```

Εικόνα 5: Αποτελέσματα αλγορίθμου *ibk* για $k=1$

```

Correctly Classified Instances      21          55.2632 %
Incorrectly Classified Instances    17          44.7368 %
Kappa statistic                    0.1053
Mean absolute error                0.4656
Root mean squared error            0.5464
Relative absolute error            92.9842 %
Root relative squared error        109.1208 %
Total Number of Instances          38

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.579   0.474   0.550      0.579   0.564      0.105    0.551    0.553     0
                0.526   0.421   0.556      0.526   0.541      0.105    0.551    0.583     1
Weighted Avg.   0.553   0.447   0.553      0.553   0.552      0.105    0.551    0.568

=== Confusion Matrix ===

  a  b  <-- classified as
11  8  |  a = 0
 9 10  |  b = 1

```

Εικόνα 6: Αποτελέσματα αλγορίθμου *ibk* για $k=3$

```

Correctly Classified Instances      29          76.3158 %
Incorrectly Classified Instances     9          23.6842 %
Kappa statistic                    0.5263
Mean absolute error                0.299
Root mean squared error            0.4577
Relative absolute error            59.7237 %
Root relative squared error        91.4082 %
Total Number of Instances          38

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.684   0.158   0.813      0.684   0.743      0.533    0.737    0.697     0
                0.842   0.316   0.727      0.842   0.780      0.533    0.737    0.678     1
Weighted Avg.   0.763   0.237   0.770      0.763   0.762      0.533    0.737    0.687

=== Confusion Matrix ===

  a  b  <-- classified as
13  6  |  a = 0
 3 16  |  b = 1

```

Εικόνα 7: Αποτελέσματα αλγορίθμου *J48*

```

Correctly Classified Instances      26          68.4211 %
Incorrectly Classified Instances   12          31.5789 %
Kappa statistic                    0.3684
Mean absolute error                0.3076
Root mean squared error           0.5176
Relative absolute error            61.4364 %
Root relative squared error       103.37 %
Total Number of Instances         38

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.842    0.474    0.640      0.842    0.727      0.388    0.765    0.716     0
                0.526    0.158    0.769      0.526    0.625      0.388    0.765    0.793     1
Weighted Avg.    0.684    0.316    0.705      0.684    0.676      0.388    0.765    0.755

=== Confusion Matrix ===

  a  b  <-- classified as
16  3  |  a = 0
 9 10  |  b = 1

```

Εικόνα 8: Αποτελέσματα αλγορίθμου Adaboost με J48

```

Correctly Classified Instances      28          73.6842 %
Incorrectly Classified Instances   10          26.3158 %
Kappa statistic                    0.4737
Mean absolute error                0.3568
Root mean squared error           0.4682
Relative absolute error            71.2502 %
Root relative squared error       93.4934 %
Total Number of Instances         38

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0.737    0.263    0.737      0.737    0.737      0.474    0.730    0.739     0
                0.737    0.263    0.737      0.737    0.737      0.474    0.730    0.679     1
Weighted Avg.    0.737    0.263    0.737      0.737    0.737      0.474    0.730    0.709

=== Confusion Matrix ===

  a  b  <-- classified as
14  5  |  a = 0
 5 14  |  b = 1

```

Εικόνα 9: Αποτελέσματα αλγορίθμου Bagging με J48

Όπως φαίνεται στα αποτελέσματα παραπάνω αλλά και στον συγκεντρωτικό πίνακα 1. ο οποίος ακολουθεί, ο αλγόριθμος J48 δίνει την υψηλότερη ακρίβεια σωστής πρόβλεψης των δειγμάτων αλλά και την υψηλότερη ακρίβεια σωστής πρόβλεψης της κατηγορίας 1, κάτι το οποίο είναι επιθυμητό για την συγκεκριμένη εφαρμογή. Πιο συγκεκριμένα, είναι περισσότερο σημαντικό να υπάρχει μεγαλύτερη επιτυχία σωστής πρόβλεψης των ασθενών οι οποίοι δεν πρόκειται να θεραπευτούν χρησιμοποιώντας την μέθοδο της ανοσοθεραπείας καθώς έτσι δεν θα υποβάλλονται άδικα σε αυτήν την θεραπεία με κίνδυνο να επιδεινωθεί η κατάστασή τους.

Αλγόριθμος	Ακρίβεια	Class/TP Rate/Precision		
Naive Bayes	60.5%	0	0.78	0.57
		1	0.42	0.66
ibk (k=1)	52.6%	0	0.63	0.52
		1	0.42	0.53
ibk (k=3)	55.2%	0	0.57	0.55
		1	0.52	0.55
J48	76.3%	0	0.68	0.81
		1	0.84	0.72
Adaboost – J48	68.4%	0	0.84	0.64
		1	0.52	0.76
Bagging – J48	73.6%	0	0.73	0.73
		1	0.73	0.73

Πίνακας 1: Αποτελέσματα των αλγορίθμων

Κεφάλαιο 2

Υλοποίηση εφαρμογής

2.1 Εισαγωγή νέου δείγματος

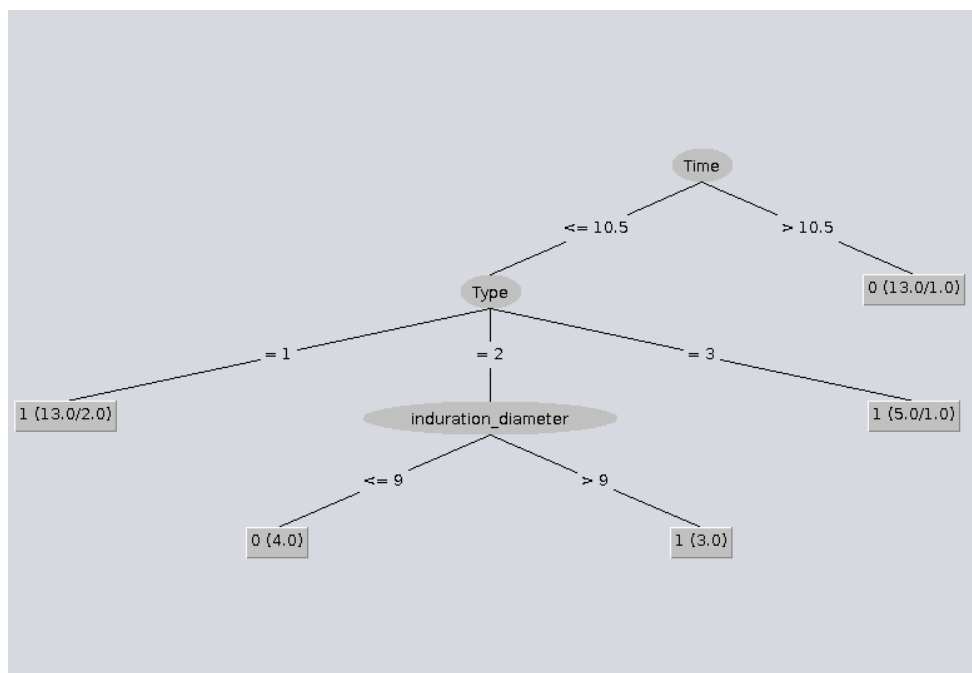
Η υλοποίηση της εφαρμογής έγινε χρησιμοποιώντας το περιβάλλον ανάπτυξης Android Studio 3.6.3. σε γλώσσα προγραμματισμού Java. Προκειμένου να γίνει χρήση του μοντέλου προβλέψεων, η δημιουργία του οποίου αναλύθηκε στο κεφάλαιο 1, είναι απαραίτητο η εφαρμογή να ζητάει από τον χρήστη να εισάγει ένα νέο δείγμα ασθενή ώστε να προβλεφθεί αν η ανοσοθεραπεία είναι η κατάλληλη μέθοδος θεραπείας. Στον κώδικα παρακάτω, φαίνονται κάποιοι από τους τρόπους με τους οποίους συλλέγονται από τον χρήστη όλα τα χαρακτηριστικά του νέου δείγματος (π.χ. input boxes, radio buttons, seekbars κ.α.).

```
public void UIHandler()
{...
    if(currentQuery == 1) //Get gender
    {...
        RadioButton male = (RadioButton) findViewById(R.id.maleRadioButton);
        if(male.isChecked())
            gender = "male";
        else
            gender = "female";
    }...
    else if(currentQuery == 3) //Get time
    {...
        query25SeekBar.setOnSeekBarChangeListener(new SeekBar.OnSeekBarChangeListener() {
            @Override
            public void onProgressChanged(SeekBar seekBar, int progress, boolean fromUser) {
                if(progress <= 7)
                {
                    seekBarTextView2.setText("<1");
                    time = 0;
                }
                else if(progress > 7 && progress <= 14)
                {
                    seekBarTextView2.setText("1");
                    time = 1;
                }
            }...
        });
    }...
}
```

2.2 Πρόβλεψη του αποτελέσματος

Αφού ο χρήστης έχει εισάγει όλες τις τιμές του νέου δείγματος, η εφαρμογή σύμφωνα με τους κανόνες του μοντέλου το οποίο είχε δημιουργηθεί, εμφανίζει αν η ανοσοθεραπεία θα θεραπεύσει τον ασθενή ή όχι. Στον κώδικα παρακάτω γίνεται ο έλεγχος των δεδομένων εισόδου με βάση των κανόνων του μοντέλου το οποίο φαίνεται στην εικόνα 10.

```
public String makePrediction()
{
    if(time > 10)
        return "Immunotherapy will cure warts!";
    else
    {
        if(typeOfWart.equals("common"))
            return "Immunotherapy\n is not the right method to cure warts!";
        else if(typeOfWart.equals("plantar"))
        {
            if(indurationDiameterOfWart <= 9)
                return "Immunotherapy will cure warts!";
            else
                return "Immunotherapy\n is not the right method to cure warts!";
        }
        else
            return "Immunotherapy\n is not the right method to cure warts!";
    }
}
```



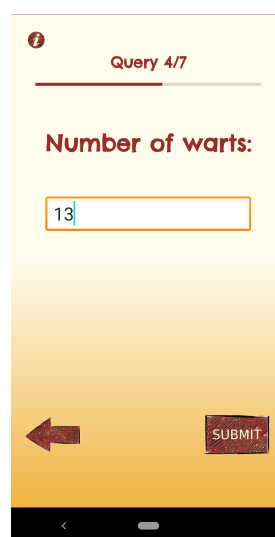
Εικόνα 10: Μοντέλο πρόβλεψης

2.3 Διεπαφή χρήστη

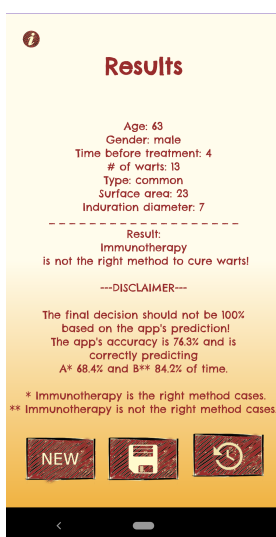
Στην ενότητα αυτή παρουσιάζεται η διεπαφή χρήστη (user interface) της εφαρμογής. Στις εικόνες 11. και 12. ενδεικτικά, φαίνονται δύο οθόνες στις οποίες ο χρήστης - ιατρός καλείται να καταχωρήσει πληροφορίες για νέο δείγμα – ασθενή. Έπειτα, στην εικόνα 13. παρουσιάζεται η οθόνη η οποία ενημερώνει τον χρήστη για το αποτέλεσμα της πρόβλεψης όπως επίσης και για την ακρίβεια των αποτελεσμάτων. Τέλος, από αυτήν την οθόνη ο χρήστης έχει την δυνατότητα, μέσω των κουμπιών στο κάτω μέρος, να καταχωρήσει ένα νέο δείγμα, να αποθηκεύσει τα αποτελέσματα της τρέχουσας πρόβλεψης ή να προσπελάσει όλες τις αποθηκευμένες προβλέψεις, όπως φαίνεται και στην εικόνα 14.



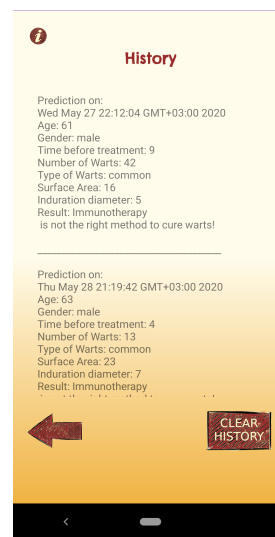
Εικόνα 11: Καταχώρηση δείγματος 3/7



Εικόνα 12: Καταχώρηση δείγματος 4/7



Εικόνα 14: Αποτέλεσμα πρόβλεψης



Εικόνα 13: Αποθηκευμένες προβλέψεις

Βιβλιογραφία - Παραπομπές

- Fahime Khozeimeh, Roohallah Alizadehsani, Mohamad Roshanzamir, Abbas Khosravi, Poursan Layegh, Saeid Nahavandi. (2017). *An expert system for selecting wart treatment method*. Volume 81.

<https://www.sciencedirect.com/science/article/abs/pii/S001048251730001X?via%3Dihub>

- Το αποθετήριο με τον πλήρη κώδικα της εφαρμογής και το εκτελέσιμο αρχείο:

<https://github.com/p17griv/wart-immunotherapy-prediction-app>