



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΟΛΟΓΩΝ ΚΑΙ ΑΕΡΟΝΑΥΠΗΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ
ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΙΟΥΛΙΟΣ 2020

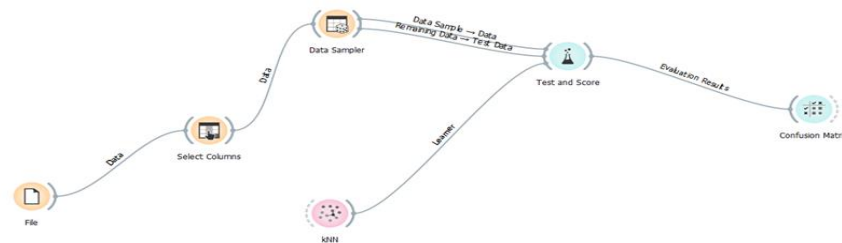
Ανάπτυξη Υβριδικού Συστήματος Συστάσεων με αξιοποίηση τεχνικών Εξόρυξης Δεδομένων και Επιχειρησιακής Έρευνας
ΠΑΝΑΓΙΩΤΗΣ Γ. ΓΙΑΝΝΟΠΟΥΛΟΣ, AM 1051351
ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΝΙΚΟΛΑΟΣ ΚΑΡΑΚΑΠΙΛΙΔΗΣ

Περίληψη: Στόχος της παρούσας διπλωματικής εργασίας είναι η μελέτη και ανάπτυξη συστημάτων συστάσεων με αξιοποίηση τεχνικών Εξόρυξης Δεδομένων και Επιχειρησιακής Έρευνας. Στην εργασία πραγματεύονται συστήματα, τα οποία μπορούν να ενταχθούν είτε στα κλασσικά είτε στα υβριδικά μοντέλα. Τα κλασσικά συστήματα συστάσεων υλοποιούνται κάνοντας χρήση τεχνικών εξόρυξης δεδομένων, και συγκεκριμένα των αλγορίθμων $k - NN$ και δέντρων αποφάσεων. Τα υβριδικά μοντέλα προκύπτουν από τον συνδυασμό των αλγορίθμων αυτών με μοντέλα της πολυκριτηριακής μεθόδου PROMETHEE.

Στην εργασία γίνεται αναλυτική παρουσίαση και αξιολόγηση των τεχνικών και των μοντέλων που χρησιμοποιούνται στις προτεινόμενες λύσεις. Έμφαση δίνεται στην αξιοποίηση των δυνατοτήτων τους κατά την ανάπτυξη ενός πρωτότυπου υβριδικού συστήματος συστάσεων ξενοδοχείων.

Σύστημα συστάσεων με χρήση του αλγορίθμου $k - NN$:

-Υλοποίηση μέσω του λογισμικού γραφικού προγραμματισμού **Orange**.



-Διερεύνηση για την τιμή της μεταβλητής k και για τον διαμοιρασμό σε training και test data set, μέσω των δεικτών AUC και CA αντίστοιχα. Επιλογή $k = 18$ και 90 – 10 αναλογία στην βάση.

Ποσοστό training data set					
AC	70%	75%	80%	85%	90%
	0,722	0,722	0,75	0,72	0,76

Ποσοστό training data set					
	70%	75%	80%	85%	90%
k	AUC				
17	0,519	0,487	0,501	0,52	0,551
18	0,524	0,495	0,503	0,531	0,561
19	0,525	0,49	0,494	0,533	0,555
20	0,528	0,489	0,492	0,514	0,549
21	0,524	0,481	0,488	0,5	0,511
22	0,513	0,474	0,486	0,49	0,507
23	0,507	0,468	0,482	0,498	0,514
24	0,509	0,476	0,472	0,499	0,493
25	0,515	0,476	0,468	0,492	0,482
26	0,513	0,481	0,468	0,486	0,478
27	0,51	0,48	0,465	0,481	0,474

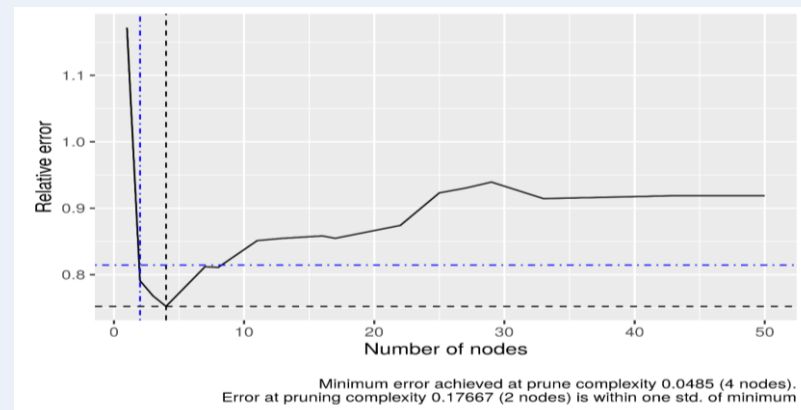
-Τα αποτελέσματα των επιτυχών και ανεπιτυχών προβλέψεων και συστάσεων για το μοντέλο παρουσιάζονται στον **Confusion Matrix**.

Actual \ Predicted	Predicted		Σ
	NO	YES	
NO	66.7 %	23.4 %	13
YES	33.3 %	76.6 %	37
Σ	3	47	50

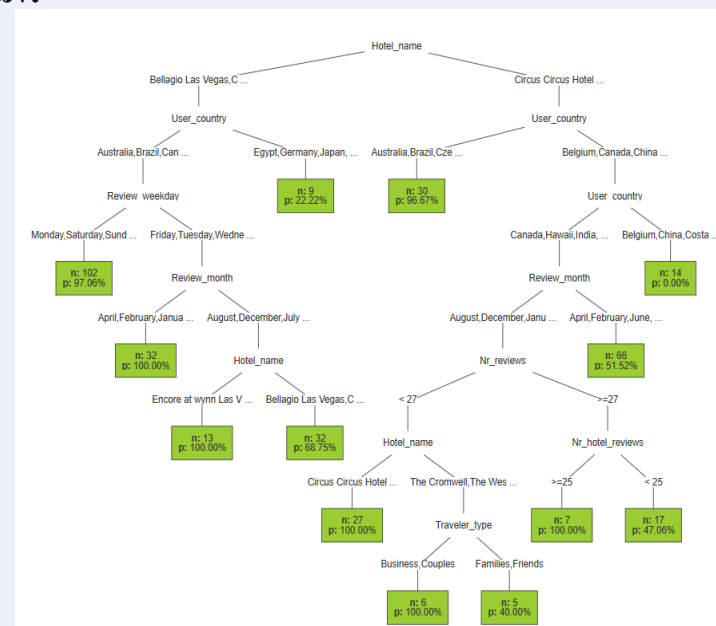
Σύστημα συστάσεων με χρήση δέντρου αποφάσεων:

-Υλοποίηση μέσω του λογισμικού **Radiant**.

-Διερεύνηση έκτασης δέντρου με στόχο την μείωση της υπέρ – εστίασης στην συγκεκριμένη βάση.



-Υλοποίηση δέντρου αποφάσεων με επιλογή 15 κόμβων.

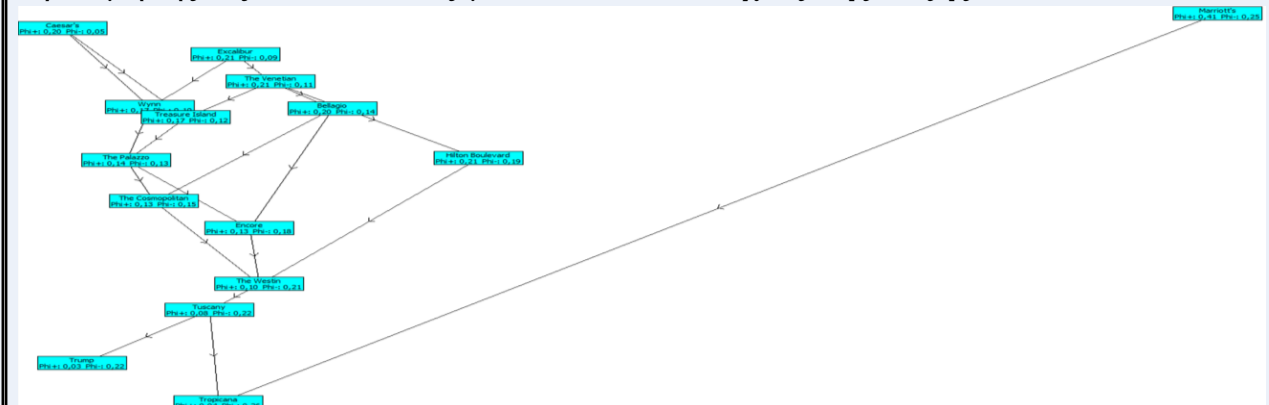


-Ποσοστά επιτυχών προβλέψεων της τάξης του 65% στο YES και του 55% στο NO.

Υβριδικά συστήματα συστάσεων:

-Δημιουργία μοντέλων δύο επιπέδων ανάλυσης.

-Στο πρώτο επίπεδο αξιοποιούνται τα αποτελέσματα από τα προηγούμενα συστήματα και κατόπιν ανάλυσης, τροφοδοτείται ένα πολυκριτηριακό μοντέλο. Το υβριδικό πολυκριτηριακό μοντέλο κατατάσσει σε φθίνουσα σειρά προτίμησης τις εναλλακτικές με ποσοστό επιτυχίας της τάξης του 70%.



-Στο δεύτερο επίπεδο χρησιμοποιούνται οι αλγόριθμοι εξόρυξης δεδομένων για κάθε μια από τις τρεις ζώνες προτίμησης.

-Συγκριτική θεώρηση για όλα τα μοντέλα.

	Πεδίο εφαρμογής (2 - βάθμια κλίμακα)	Χρόνος και κόστος υλοποίησης (3 - βάθμια κλίμακα)	Μέση αποτελεσματικότητα (Ποσοτικό κριτήριο)	Επιπλέον γνώση (5 - βάθμια κλίμακα)
.				
Απλά μοντέλα εξόρυξης δεδομένων	Σύνολο βάσης	Μέτρια	64%	Αρκετά
Υβριδικό πολυκριτηριακό μοντέλο	Σύνολο βάσης	Πολύ	70%	Πολύ
Υβριδικά μοντέλα ανάλυσης δεδομένων	Τμήμα βάσης	Πολύ	Πρώτη ζώνη προτίμησης: 80% Δεύτερη ζώνη προτίμησης: 62% Τρίτη ζώνη προτίμησης: 47%	Πάρα πολύ

-Τα μοντέλα δεύτερου επιπέδου είναι χρήσιμα γιατί παρέχουν την δυνατότητα εξαγωγής συγκεκριμένης γνώσης για κάθε ζώνη προτίμησης.