



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ**

**Πρόγραμμα Μεταπτυχιακών σπουδών:
«Επιστήμη και Τεχνολογία της Πληροφορικής και των Υπολογιστών»**

Μεταπτυχιακή Διπλωματική Εργασία

**«Εξόρυξη Δεδομένων και Σεμινάρια Επιμόρφωσης Φορέων
του Δημόσιου Τομέα»**

**Παναγιώτης Παπαϊωάννου
Α.Μ. 19027**

**Επιβλέποντες καθηγητές:
Χρήστος Σκουρλάς, Αικατερίνη Μαρινάγη**

Αθήνα, Δεκέμβριος 2021

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος ΠΑΝΑΓΙΩΤΗΣ ΠΑΠΑΙΩΑΝΝΟΥ του ΙΩΑΝΝΗ, με αριθμό μητρώου 19027, φοιτητής του Προγράμματος Μεταπτυχιακών Σπουδών ΕΠΙΣΤΗΜΗ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΩΝ ΥΠΟΛΟΓΙΣΤΩΝ του Τμήματος ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ της Σχολής ΜΗΧΑΝΙΚΩΝ του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



Παναγιώτης Παπαϊωάννου

Τριμελής Εξεταστική Επιτροπή

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

Χρήστος Σκουρλάς
Καθηγητής

Αικατερίνη Μαρινάγη
Καθηγήτρια

Βασίλειος Μάμαλης
Καθηγητής

Περίληψη

Σε έναν κόσμο όπου πλέον η κάθε είδους πληροφορία αποθηκεύεται σε ψηφιακή μορφή, ο ρυθμός παραγωγής δεδομένων αυξάνεται εκθετικά χρόνο με το χρόνο. Οι πάσης φύσεως οργανισμοί διαθέτουν πλέον στις βάσεις δεδομένων τους ή στις ψηφιακές τους «αποθήκες» τεράστιο όγκο δεδομένων.

Τα τελευταία χρόνια, η Εξόρυξη Δεδομένων βρίσκει εφαρμογή σε πολλούς οργανισμούς που δραστηριοποιούνται σε τομείς όπως η Οικονομία, οι Τηλεπικοινωνίες, η Ιατρική, η Ασφάλεια Δικτύων, η Ενέργεια και φυσικά η Εκπαίδευση. Έρχεται να καλύψει την ανάγκη για την αξιοποίηση του τεράστιου αυτού όγκου δεδομένων, με τρόπους που διαφέρουν από την «απλή» στατιστική ανάλυση και που επικεντρώνονται στην αναγνώριση προτύπων μέσα στα δεδομένα και στην εύρεση συσχετίσεων μεταξύ τους, με στόχο την εξαγωγή πληροφορίας και γνώσης τέτοιας, που να μπορεί να οδηγήσει σε ανάλυση δεδομένων και λήψη αποφάσεων.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η διερεύνηση τρόπων με τους οποίους τεχνικές εξόρυξης μπορούν να εφαρμοστούν σε δεδομένα που διατηρούνται στο Ολοκληρωμένο Πληροφοριακό Σύστημα ενός δημόσιου οργανισμού ο οποίος διοργανώνει επιμορφωτικά σεμινάρια για δημοσίους υπαλλήλους και τα οποία περιλαμβάνουν, μεταξύ άλλων, στοιχεία που αφορούν στις εκπαιδευτικές ανάγκες φορέων του δημόσιου τομέα, στην οργάνωση των σεμιναρίων επιμόρφωσης, καθώς και στην αξιολόγησή τους από εκπαιδευτές και εκπαιδευόμενους.

Στόχος της εργασίας είναι η εξαγωγή χρήσιμων συμπερασμάτων τα οποία θα μπορούσαν να οδηγήσουν τον φορέα σε λήψη αποφάσεων για τη βελτίωση της εκπαιδευτικής διαδικασίας.

Abstract

In a world where all kinds of information is eventually stored in digital form, the rate of data production is increasing exponentially year by year. Organizations of all kinds find at some point themselves with a huge amount of data stored in their databases or in their digital "warehouses".

In recent years, Data Mining is being applied by many organizations in sectors such as Economy, Telecommunications, Medicine, Network Security, Energy and of course Education. It tries to meet the need to utilize this huge amount of data, in ways that differ from the "simple" statistical analysis and focus more on identifying patterns in the data and finding correlations between them, in order to extract information and knowledge of such nature that can lead to data analysis and decision making.

The aim of this thesis is to explore ways in which data mining techniques can be applied to data stored in the Integrated Information System of a public Organization that implements training seminars for civil servants and which include, inter alia, information on educational public sector body's needs, information about the organizational process of the training seminars, as well as data concerning the evaluation of the seminars by trainers and trainees.

The aim of this work is to draw useful conclusions that could lead the Organization to make decisions that could improve the educational process as a whole.

Περιεχόμενα

1	ΕΙΣΑΓΩΓΗ	10
1.1	ΑΝΤΙΚΕΙΜΕΝΟ ΚΑΙ ΣΤΟΧΟΙ ΤΗΣ ΕΡΓΑΣΙΑΣ	10
2	ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ	11
2.1	ΓΕΝΙΚΑ ΓΙΑ ΤΗΝ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ	11
2.2	ΦΑΣΕΙΣ ΚΑΙ ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΜΕΓΑΛΗΣ ΚΛΙΜΑΚΑΣ	12
2.3	ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ	14
2.3.1	<i>Association Analysis / Frequent Pattern Mining</i>	14
2.3.2	<i>Clustering Analysis</i>	14
2.3.3	<i>Correlation Analysis</i>	14
2.3.4	<i>Classification</i>	14
2.3.5	<i>Decision Tree Induction</i>	14
2.3.6	<i>Regression Analysis</i>	15
2.4	ΔΕΔΟΜΕΝΑ ΜΕΓΑΛΗΣ ΚΛΙΜΑΚΑΣ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ.....	15
3	ΤΑ ΕΠΙΜΟΡΦΩΤΙΚΑ ΣΕΜΙΝΑΡΙΑ ΤΟΥ ΙΝΣΤΙΤΟΥΤΟΥ ΕΠΙΜΟΡΦΩΣΗΣ	16
3.1	ΓΕΝΙΚΑ.....	16
3.2	ΑΝΙΧΝΕΥΣΗ ΕΚΠΑΙΔΕΥΤΙΚΩΝ ΑΝΑΓΚΩΝ ΕΠΙΜΟΡΦΩΣΗΣ ΦΟΡΕΩΝ.....	16
3.3	ΟΡΓΑΝΩΣΗ ΤΩΝ ΣΕΜΙΝΑΡΙΩΝ	17
3.4	ΑΞΙΟΛΟΓΗΣΗ-ΑΠΟΤΙΜΗΣΗ ΤΩΝ ΣΕΜΙΝΑΡΙΩΝ.....	18
4	ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΣΕΜΙΝΑΡΙΑ ΚΑΙ ΑΝΑΛΥΣΗ ΜΕ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ 19	
4.1	ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΗ ΒΑΣΗ ΔΕΔΟΜΕΝΩΝ	19
4.2	DATA PREPARATION	21
4.3	ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ	22
4.3.1	<i>Περίπτωση Εφαρμογής 1: Πρόβλεψη υπέρβασης προϋπολογισμένου ποσού σεμιναρίου</i> <i>22</i>	
4.3.2	<i>Περίπτωση Εφαρμογής 2: Ανίχνευση αναγκών με βάση τα σχόλια των εκπαιδευόμενων</i> <i>40</i>	
4.3.3	<i>Περίπτωση Εφαρμογής 3: Δημιουργία Κανόνων Συσχέτισης με βάση το ιστορικό</i> <i>αιτήσεων εκπαιδευόμενου.....</i>	<i>45</i>
4.3.4	<i>Περίπτωση Εφαρμογής 4: Ανάλυση της Αξιολόγησης σεμιναρίων και εξαγωγή</i> <i>συμπερασμάτων.....</i>	<i>52</i>
4.3.5	<i>Περίπτωση Εφαρμογής 5: Μοντέλο πρόβλεψης Εκπαιδευόμενων που θα ολοκληρώσουν</i> <i>την επιμόρφωση.....</i>	<i>58</i>
5	ΣΥΜΠΕΡΑΣΜΑΤΑ	65
6	ΑΝΑΦΟΡΕΣ	66

Κατάλογος σχημάτων

ΕΙΚΟΝΑ 2.1 ΒΑΣΙΚΑ ΒΗΜΑΤΑ ΔΙΑΔΙΚΑΣΙΑΣ ΑΝΑΚΑΛΥΨΗΣ ΓΝΩΣΗΣ	12
ΕΙΚΟΝΑ 2.2 ΟΙ ΦΑΣΕΙΣ ΤΟΥ ΜΟΝΤΕΛΟΥ CRISP-DM (HTTPS://COMMONS.WIKIMEDIA.ORG/WIKI/FILE:CRISP-DM_PROCESS_DIAGRAM.PNG , KENNETH JENSEN, CC BY-SA 3.0 < HTTPS://CREATIVECOMMONS.ORG/LICENSES/BY-SA/3.0 >, VIA WIKIMEDIA COMMONS)	13
ΕΙΚΟΝΑ 3.1 ΧΑΡΤΗΣ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΙΝ.ΕΠ.	16
ΕΙΚΟΝΑ 3.2 ΘΕΜΑΤΙΚΟΙ ΤΙΤΛΟΙ ΚΑΙ ΚΥΚΛΟΙ ΣΕΜΙΝΑΡΙΩΝ ΙΝ.ΕΠ.....	18
ΕΙΚΟΝΑ 4.1 DBEAVER-ΣΥΝΔΕΣΗ ΜΕ ΤΟ ΟΠΣ ΤΟΥ ΦΟΡΕΑ	19
ΕΙΚΟΝΑ 4.2 ΠΙΝΑΚΕΣ ΤΟΥ ΟΠΣ.....	20
ΕΙΚΟΝΑ 4.3 ΕΞΑΓΩΓΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟ ΟΠΣ	20
ΕΙΚΟΝΑ 4.4 ΚΑΤΑΜΕΡΙΣΜΟΣ ΤΟΥ ΧΡΟΝΟΥ ΣΤΗΝ ΑΝΑΛΥΣΗ BIG DATA	22
ΕΙΚΟΝΑ 4.5 ΠΡΟΫΠΟΛΟΓΙΣΜΟΣ - ΑΠΟΛΟΓΙΣΜΟΣ ΣΕΜΙΝΑΡΙΟΥ	23
ΕΙΚΟΝΑ 4.6 ΠΙΝΑΚΑΣ ΜΕ ΣΤΟΙΧΕΙΑ ΣΕΜΙΝΑΡΙΩΝ	24
ΕΙΚΟΝΑ 4.7 ΠΙΝΑΚΑΣ ΘΕΜΑΤΙΚΩΝ ΚΥΚΛΩΝ (ΛΕΚΤΙΚΟ)	24
ΕΙΚΟΝΑ 4.8 ΠΙΝΑΚΑΣ ΤΙΤΛΩΝ ΣΕΜΙΝΑΡΙΩΝ	25
ΕΙΚΟΝΑ 4.9 ΣΥΝΕΝΩΣΗ ΠΙΝΑΚΩΝ ΤΟΥ ΟΠΣ	26
ΕΙΚΟΝΑ 4.10 ΔΙΕΡΓΑΣΙΑ ΠΡΟΕΤΟΙΜΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ.....	27
ΕΙΚΟΝΑ 4.11 ΦΙΛΤΡΑΡΙΣΜΑ ΔΕΔΟΜΕΝΩΝ	27
ΕΙΚΟΝΑ 4.12 ΜΕΤΟΝΟΜΑΣΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ.....	28
ΕΙΚΟΝΑ 4.13 ΔΗΜΙΟΥΡΓΙΑ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ	28
ΕΙΚΟΝΑ 4.14 ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΝΑΙΒΕ ΒAYES.....	29
ΕΙΚΟΝΑ 4.15 ΑΠΟΤΕΛΕΣΜΑΤΑ ΣΥΝΟΛΟΥ ΠΡΟΒΛΕΨΗΣ	30
ΕΙΚΟΝΑ 4.16 ΠΡΟΒΛΕΨΗ ΥΠΕΡΒΑΣΗΣ ΠΡΟΫΠΟΛΟΓΙΣΜΟΥ ΑΝΑ ΘΕΜΑΤΙΚΟ ΚΥΚΛΟ	31
ΕΙΚΟΝΑ 4.17 ΠΡΟΒΛΕΨΗ ΥΠΕΡΒΑΣΗΣ ΠΡΟΫΠΟΛΟΓΙΣΜΟΥ ΑΝΑ ΤΙΤΛΟ ΣΕΜΙΝΑΡΙΟΥ	32
ΕΙΚΟΝΑ 4.18 80% ΤΟΥ ΔΕΙΓΜΑΤΟΣ, ΑΚΡΑΙΕΣ ΤΙΜΕΣ	33
ΕΙΚΟΝΑ 4.19 100% ΤΟΥ ΔΕΙΓΜΑΤΟΣ, ΑΚΡΑΙΕΣ ΤΙΜΕΣ	33
ΕΙΚΟΝΑ 4.20 80% ΤΟΥ ΔΕΙΓΜΑΤΟΣ, ΧΩΡΙΣ ΑΚΡΑΙΕΣ ΤΙΜΕΣ	34
ΕΙΚΟΝΑ 4.21 100% ΤΟΥ ΔΕΙΓΜΑΤΟΣ, ΧΩΡΙΣ ΑΚΡΑΙΕΣ ΤΙΜΕΣ.....	34
ΕΙΚΟΝΑ 4.22 DISCRETIZE ATTRIBUTES (DIFFERENCE, HOURS, STUDENTS)	35
ΕΙΚΟΝΑ 4.23 DISCRETIZE BY FREQUENCY	36
ΕΙΚΟΝΑ 4.24 ΠΡΟΒΛΕΨΗ ΥΠΕΡΒΑΣΗΣ ΜΕ ΒΑΣΗ ΤΗ ΔΙΑΡΚΕΙΑ ΣΕΜΙΝΑΡΙΟΥ	37
ΕΙΚΟΝΑ 4.25 ΠΡΟΒΛΕΨΗ ΥΠΕΡΒΑΣΗΣ ΜΕ ΒΑΣΗ ΤΟ ΠΛΗΘΟΣ ΣΠΟΥΔΑΣΤΩΝ.....	38
ΕΙΚΟΝΑ 4.26 ΑΠΟΔΟΣΗ ΜΟΝΤΕΛΟΥ - DISCRETIZATION.....	39
ΕΙΚΟΝΑ 4.27 ΦΟΡΜΑ ΑΝΙΧΝΕΥΣΗΣ ΕΚΠΑΙΔΕΥΤΙΚΩΝ ΑΝΑΓΚΩΝ	40
ΕΙΚΟΝΑ 4.28 Ο ΠΙΝΑΚΑΣ ΜΕ ΤΟ ΚΕΙΜΕΝΟ ΠΡΟΣ ΑΝΑΛΥΣΗ	41
ΕΙΚΟΝΑ 4.29 ΕΙΣΑΓΩΓΗ ΤΟΥ ΠΙΝΑΚΑ ΣΤΟ RAPIDMINER ΩΣ DATASET	41
ΕΙΚΟΝΑ 4.30 ΜΟΝΤΕΛΟ TEXT MINING	42
ΕΙΚΟΝΑ 4.31 ΟΜΑΔΟΠΟΙΗΣΗ ΟΛΩΝ ΤΩΝ ΣΧΟΛΙΩΝ ΑΝΑ ΤΙΤΛΟ ΣΕΜΙΝΑΡΙΟΥ	42
ΕΙΚΟΝΑ 4.32 ΔΙΕΡΓΑΣΙΑ ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ DOCUMENTS	42
ΕΙΚΟΝΑ 4.33 ΔΙΑΝΥΣΜΑ ΛΕΞΕΩΝ (TOKENS).....	43
ΕΙΚΟΝΑ 4.34 Ο ΤΕΛΕΣΤΗΣ PROCESS DOCUMENTS FROM DATA.....	43
ΕΙΚΟΝΑ 4.35 Ο ΤΕΛΕΣΤΗΣ N-GRAMS	43
ΕΙΚΟΝΑ 4.36 ΣΥΧΝΑ ΕΜΦΑΝΙΖΟΜΕΝΑ TOKENS ΑΝΑ ΣΕΜΙΝΑΡΙΟ (Α).....	44
ΕΙΚΟΝΑ 4.37 ΣΥΧΝΑ ΕΜΦΑΝΙΖΟΜΕΝΑ TOKENS ΑΝΑ ΣΕΜΙΝΑΡΙΟ (Β).....	44
ΕΙΚΟΝΑ 4.38 Ο ΠΙΝΑΚΑΣ ΤΩΝ ΑΙΤΗΣΕΩΝ ΣΕ ΜΟΡΦΗ .CSV.....	46
ΕΙΚΟΝΑ 4.39 ΤΟ ΜΟΝΤΕΛΟ BASKET ANALYSIS	47
ΕΙΚΟΝΑ 4.40 ΠΑΡΑΜΕΤΡΟΙ ΤΕΛΕΣΤΩΝ FP-GROWTH ΚΑΙ CREATE ASSOCIATION RULES	47
ΕΙΚΟΝΑ 4.41 FP-GROWTH, ΤΑ EXAMPLE SETS ΠΟΥ ΠΡΟΚΥΠΤΟΥΝ ΑΝΑ ΕΚΠΑΙΔΕΥΟΜΕΝΟ (CD).....	48
ΕΙΚΟΝΑ 4.42 FP-GROWTH, FREQUENT ITEM SETS.....	48
ΕΙΚΟΝΑ 4.43 ASSOCIATION RULES (CONFIDENCE).....	49
ΕΙΚΟΝΑ 4.44 ASSOCIATION RULES (ΑΝΑ ΤΙΤΛΟ ΣΕΜΙΝΑΡΙΟΥ Ι).....	50

ΕΙΚΟΝΑ 4.45 ASSOCIATION RULES (ΑΝΑ ΤΙΤΛΟ ΣΕΜΙΝΑΡΙΟΥ II).....	50
ΕΙΚΟΝΑ 4.46 ASSOCIATION RULES (DESCRIPTION)	51
ΕΙΚΟΝΑ 4.47 ΑΞΙΟΛΟΓΗΣΗ ΣΕΜΙΝΑΡΙΟΥ ΑΠΟ ΕΚΠΑΙΔΕΥΟΜΕΝΟΥΣ.....	52
ΕΙΚΟΝΑ 4.48 ΣΥΝΟΛΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΣΕΜΙΝΑΡΙΟΥ	52
ΕΙΚΟΝΑ 4.49 ΠΙΝΑΚΑΣ ΑΞΙΟΛΟΓΗΣΕΩΝ ΣΕΜΙΝΑΡΙΩΝ	53
ΕΙΚΟΝΑ 4.50 RAPIDMINER - ΜΟΝΤΕΛΟ ΠΙΝΑΚΑ ΣΥΣΧΕΤΙΣΕΩΝ	54
ΕΙΚΟΝΑ 4.51 ΑΞΙΟΛΟΓΗΣΕΙΣ ΣΕΜΙΝΑΡΙΩΝ - ΠΙΝΑΚΑΣ ΣΥΣΧΕΤΙΣΕΩΝ.....	55
ΕΙΚΟΝΑ 4.52 ΑΞΙΟΛΟΓΗΣΕΙΣ ΣΕΜΙΝΑΡΙΩΝ - ΙΣΧΥΡΕΣ ΣΥΣΧΕΤΙΣΕΙΣ.....	56
ΕΙΚΟΝΑ 4.53 ΑΞΙΟΛΟΓΗΣΕΙΣ ΣΕΜΙΝΑΡΙΩΝ - MATRIX VISUALIZATION	57
ΕΙΚΟΝΑ 4.54 ΈΝΔΕΙΞΗ ΕΠΙΤΥΧΟΥΣ ΟΛΟΚΛΗΡΩΣΗΣ ΣΕΜΙΝΑΡΙΟΥ	59
ΕΙΚΟΝΑ 4.55 ΔΗΜΙΟΥΡΓΙΑ ΕΠΙΠΛΕΟΝ ΧΑΡΑΚΤΗΡΙΣΤΙΚΟΥ	59
ΕΙΚΟΝΑ 4.56 ΑΠΟΘΗΚΕΥΣΗ ΕΧΑΜPLESET	60
ΕΙΚΟΝΑ 4.57 ΔΗΜΙΟΥΡΓΙΑ ΕΧΑΜPLESET	60
ΕΙΚΟΝΑ 4.58 RAPIDMINER - ΜΟΝΤΕΛΟ DECISION TREE	61
ΕΙΚΟΝΑ 4.59 SET ROLE "LABEL"	62
ΕΙΚΟΝΑ 4.60 ΕΠΙΤΥΧΗΣ / ΑΝΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	62
ΕΙΚΟΝΑ 4.61 DECISION TREE, ΟΛΟΚΛΗΡΩΣΗ ΣΕΜΙΝΑΡΙΟΥ	63
ΕΙΚΟΝΑ 4.62 DECISION TREE, DESCRIPTIVE	63
ΕΙΚΟΝΑ 4.63 ΑΠΟΔΟΣΗ ΜΟΝΤΕΛΟΥ (ACCURACY).....	64

1 ΕΙΣΑΓΩΓΗ

1.1 Αντικείμενο και Στόχοι της Εργασίας

Ολοένα και περισσότερο τα τελευταία χρόνια, η Εξόρυξη Δεδομένων βρίσκει εφαρμογή σε πλήθος τομέων όπως η Οικονομία, οι Τηλεπικοινωνίες, η Ιατρική, η Ασφάλεια Δικτύων, η Ενέργεια και φυσικά η Εκπαίδευση/Επιμόρφωση.

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η εφαρμογή τεχνικών εξόρυξης δεδομένων σε δεδομένα που τηρούνται στη βάση δεδομένων ενός επιμορφωτικού φορέα του δημόσιου τομέα. Στη βάση δεδομένων του φορέα υπάρχουν αποθηκευμένα δεδομένα που αφορούν στην επιμόρφωση περισσότερων από 300.000 υπαλλήλων του δημόσιου και του ευρύτερου δημόσιου τομέα κατά τα 10 τελευταία χρόνια. Τα στοιχεία που τηρούνται σε αυτήν αφορούν μεταξύ άλλων:

- δημογραφικά στοιχεία (ηλικία, φύλο, τόπο κατοικίας κ.α.),
- στοιχεία σχετικά με το μορφωτικό επίπεδο,
- στοιχεία σχετικά με το επαγγελματικό προφίλ,
- στοιχεία σχετικά με το εργασιακό περιβάλλον και τις αρμοδιότητες του υπαλλήλου,
- στοιχεία σχετικά το ιστορικό επιμόρφωσης του υπαλλήλου στον φορέα (σεμινάρια που έχει παρακολουθήσει, βαθμολογίες, αξιολογήσεις, κ.α.)

Γίνεται αντιληπτό ότι τόσο ο όγκος, όσο και η ποικιλομορφία των δεδομένων, προσφέρονται για ανάλυση με τεχνικές εξόρυξης δεδομένων, η οποία (ανάλυση) θα μπορούσε να αποκαλύψει γνώση η οποία ως τώρα παραμένει κρυμμένη από τις τεχνικές που ακολουθούνται από τον φορέα για την επεξεργασία τους.

Αρχικά θα συλλεγούν και θα αναλυθούν στοιχεία σχετικά με τα σεμινάρια και πιο συγκεκριμένα με την ανίχνευση εκπαιδευτικών αναγκών επιμόρφωσης φορέων του δημοσίου, την οργάνωση των σεμιναρίων και την αξιολόγησή τους από τους εκπαιδευόμενους.

Στη συνέχεια θα αντληθούν δεδομένα από το Ολοκληρωμένο Πληροφοριακό Σύστημα του φορέα και με εφαρμογή τεχνικών εξόρυξης δεδομένων, θα γίνει προσπάθεια εξαγωγής γνώσης, η αξιοποίηση της οποίας θα μπορέσει να συνεισφέρει στη λήψη αποφάσεων και στη βελτίωση της εκπαιδευτικής και οργανωτικής διαδικασίας.

Στόχοι της εργασίας είναι:

- η βιβλιογραφική αναζήτηση προς εύρεση τεχνικών εξόρυξης δεδομένων που θα μπορούσαν να εφαρμοσθούν στην περίπτωση που εξετάζεται
- η σύντομη περιγραφή των τεχνικών αυτών
- η σύντομη περιγραφή του τρόπου διαχείρισης των επιμορφωτικών σεμιναρίων από τον φορέα
- η αναλυτική περιγραφή εφαρμογής επιλεγμένων τεχνικών εξόρυξης στα δεδομένα του φορέα με χρήση συγκεκριμένου εργαλείου λογισμικού (Rapidminer)
- Η εξαγωγή συμπερασμάτων

2 ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

2.1 Γενικά για την Εξόρυξη Δεδομένων

Με τον όρο Εξόρυξη Δεδομένων (Data Mining) εννοούμε γενικά την ανακάλυψη και εξαγωγή πληροφορίας και προτύπων (patterns) μέσα από μεγάλης κλίμακας σύνολα δεδομένων, με χρήση τεχνικών που συνδυάζουν χαρακτηριστικά μηχανικής μάθησης, στατιστικής και βάσεων δεδομένων (Soumen Chakrabarti, 2006). Υπό αυτήν την έννοια, ο όρος «Δεδομένα» μπορεί να ισχυριστεί κάποιος ότι δεν είναι απολύτως ακριβής, μιας και δεν είναι δεδομένα που τελικά εξορύσσονται, αλλά πρότυπα και, μέσω αυτών, γνώση (Han & Kamber, 2001).

Στην πραγματικότητα, το αντικείμενο της Εξόρυξης Δεδομένων είναι η ανάλυση μεγάλου όγκου δεδομένων με αυτόματο ή ημιαυτόματο τρόπο, με σκοπό την ανακάλυψη προτύπων που ενδεχομένως παρουσιάζουν ενδιαφέρον, όπως για παράδειγμα ομάδες εγγραφών με κοινά χαρακτηριστικά (συστάδες - clusters), εγγραφών με ασυνήθιστα ακραίες τιμές, αλλά και κανόνες συσχετίσεων μεταξύ δεδομένων ή μεταβλητών (association rule mining). Τα πρότυπα αυτά στη συνέχεια μπορεί να χρησιμοποιηθούν για περαιτέρω ανάλυση, για παράδειγμα μηχανική μάθηση (machine learning), ανάλυση προγνωστικών (predictive analytics) ή ως μέρος ενός συστήματος λήψης αποφάσεων (Wikipedia, 2021).

Ενδεικτικά παρατίθενται κάποιοι τομείς στους οποίους γίνεται πλέον ευρεία χρήση της Εξόρυξης Δεδομένων:

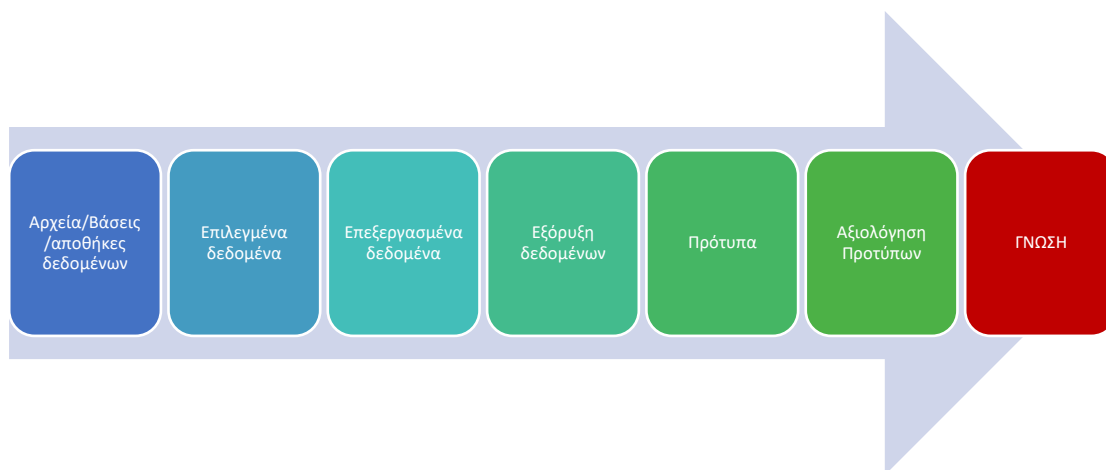
- **Υγεία:** Πρόβλεψη του αριθμού των ασθενών σε κάθε κατηγορία, διασφάλιση ότι ο ασθενής θα λάβει την κατάλληλη αγωγή στον κατάλληλο χρόνο, εντοπισμός απάτης και κατάχρησης σε ό,τι αφορά την ασφαλιστική κάλυψη
- **Εμπόριο:** Κατανόηση της καταναλωτικής συμπεριφοράς με βάση τη θεωρία ότι εάν ένας καταναλωτής αγοράσει ένα σύνολο προϊόντων, είναι πιθανότερο να αγοράσει κάποιο άλλο (συγκεκριμένο) σύνολο προϊόντων. Προσαρμογή των προϊόντων στις ανάγκες των πελατών. Διαχείριση σχέσεων εταιρείας-πελάτη (Customer Relationship Management): Κατανόηση της εμπορικής λειτουργίας μιας επιχείρησης, διατήρηση και διεύρυνση πελατειακής βάσης, εξατομικευμένη προώθηση προϊόντων, πολιτικές για την αύξηση της εμπιστοσύνης των πελατών προς την επιχείρηση
- **Εκπαίδευση:** Ανακάλυψη γνώσης μέσα από τη μαθησιακή συμπεριφορά των σπουδαστών. Πρόβλεψη των αποτελεσμάτων του σπουδαστή, αναπροσαρμογή της εκπαιδευτικής ύλης και των εκπαιδευτικών τεχνικών
- **Ασφάλεια συναλλαγών:** Εντοπισμός απάτης με εκπαίδευση μοντέλων επιβλεπόμενης μάθησης ώστε να αναγνωρίζουν συναλλαγές που αφορούν σε παράνομες δραστηριότητες
- **Ασφάλεια συστημάτων:** Ανίχνευση εισβολής σε συστήματα και πόρους με ανάλυση της δικτυακής κίνησης
- **Τραπεζική και Χρηματοοικονομική τεχνολογία:** Ανάλυση τραπεζικών συναλλαγών για ανακάλυψη προτύπων και συσχετίσεων που αφορούν σε χρηματιστηριακούς δείκτες, συμπεριφορά των αγορών κ.λπ

- **Εγκληματολογία:** Ανάλυση δεδομένων από καταχωρισμένες υποθέσεις εγκλημάτων και αξιοποίησή τους για επίλυση νέων υποθέσεων

2.2 Φάσεις και Μεθοδολογία ανάλυσης Δεδομένων Μεγάλης Κλίμακας

Η ανάλυση δεδομένων μεγάλης κλίμακας διαφέρει από την κλασσική ανάλυση δεδομένων κυρίως λόγω του όγκου και της ποικιλομορφίας των χαρακτηριστικών των υπό επεξεργασία δεδομένων. Η διαδικασία ανακάλυψης γνώσης, μέρος της οποίας αποτελεί και η εξόρυξη δεδομένων, περιλαμβάνει συνήθως τα παρακάτω βασικά βήματα (B. Μεγαλοοικονόμου - Πανεπιστήμιο Πατρών, Άνοικτά Ακαδημαϊκά Μαθήματα ΠΠ, 2015):

- Επιλογή δεδομένων από ενδεχομένως διαφορετικές πηγές
- Προεπεξεργασία: Στο βήμα αυτό γίνεται τακτοποίηση εσφαλμένων, προβληματικών ή ελλιπόντων δεδομένων. Στο βήμα αυτό συνήθως αφιερώνεται το 50-60% της συνολικής προσπάθειας
- Μετασχηματισμός: Ετερογενή δεδομένα μετατρέπονται σε κοινή τυποποίηση ώστε να είναι κατάλληλα για επεξεργασία
- Εξόρυξη δεδομένων: Εφαρμόζονται τεχνικές εξόρυξης για την παραγωγή αντίστοιχων μοντέλων
- Διερμηνεία/Αξιολόγηση: Τα παραγόμενα των μοντέλων παρουσιάζονται, συνήθως με χρήση μεθόδων οπτικοποίησης, στους χρήστες για αξιολόγηση



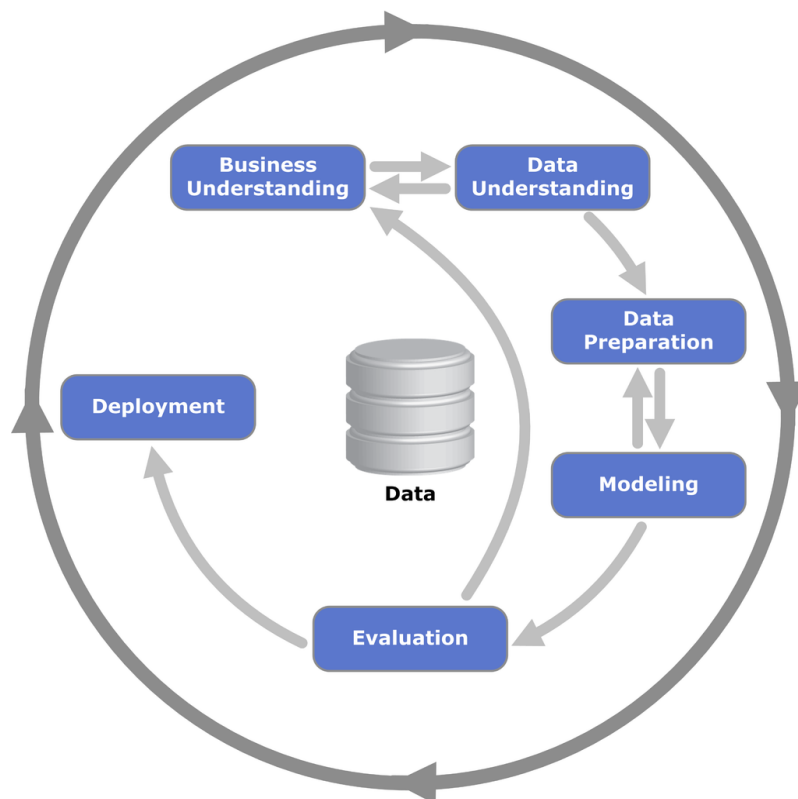
Εικόνα 2.1 Βασικά Βήματα Διαδικασίας Ανακάλυψης Γνώσης

Προκειμένου να ικανοποιηθούν οι ιδιαίτερες απαιτήσεις που προκύπτουν, απαιτείται μια βήμα προς βήμα μεθοδολογία, για την οργάνωση των εργασιών που σχετίζονται με την ανάκτηση, την επεξεργασία, την ανάλυση και την επαναχρησιμοποίηση των δεδομένων (Thomas Erl, 2016).

Η δημοφιλέστερη ίσως τέτοια μεθοδολογία, η οποία δημοσιεύθηκε το 1999 για την προτυποποίηση των διεργασιών εξόρυξης δεδομένων, είναι το Cross Industry Process for

Data Mining (CRISP-DM). Πρόκειται για ένα μοντέλο διεργασιών που περιλαμβάνει έξι φάσεις οι οποίες περιγράφουν τον κύκλο ζωής ενός έργου της επιστήμης της Πληροφορικής και είναι οι εξής (Data Science Process Alliance, n.d.), (North, 2012):

1. Business understanding: Η κατανόηση του επιχειρηματικού μοντέλου. Τι ανάγκες έχει ο οργανισμός/επιχείρηση; Θα μπορούσε η ανάλυση δεδομένων μεγάλης κλίμακας να βελτιώσει τις διαδικασίες που ακολουθούνται;
2. Data understanding: Η κατανόηση του είδους των δεδομένων που διαθέτουμε και που χρειαζόμαστε για την ανάλυση.
3. Data preparation: Η προετοιμασία και ο μετασχηματισμός των δεδομένων σε μορφή τέτοια, ώστε να επιδέχονται μοντελοποίηση.
4. Modeling: Η εφαρμογή αλγορίθμων και τεχνικών που αναζητούν, εντοπίζουν και εμφανίζουν πρότυπα ή μηνύματα μέσα στα δεδομένα.
5. Evaluation: Η διερευνητική ανάλυση και η αξιολόγηση των αποτελεσμάτων του μοντέλου ως προς την επίτευξη των στόχων του. Πιθανός επανασχεδιασμός ή/και διορθώσεις του μοντέλου.
6. Deployment: Η εφαρμογή του μοντέλου και η οπτικοποίηση των αποτελεσμάτων του. Η εξαγωγή συμπερασμάτων και λήψη αποφάσεων για ενέργειες με βάση τη γνώση που εξάγεται από το μοντέλο.



Εικόνα 2.2 Οι φάσεις του μοντέλου CRISP-DM (https://commons.wikimedia.org/wiki/File:CRISP-DM_Process_Diagram.png, Kenneth Jensen, CC BY-SA 3.0 <<https://creativecommons.org/licenses/by-sa/3.0/>>, via Wikimedia Commons)

2.3 Τεχνικές Εξόρυξης Δεδομένων

Κατά την τέταρτη φάση εφαρμογής του μοντέλου CRISP-DM, καλούμαστε να επιλέξουμε μεταξύ των διαθέσιμων τεχνικών εξόρυξης, εκείνες που μπορούν να επιφέρουν καλύτερα αποτελέσματα εάν εφαρμοστούν στην περίπτωση που εξετάζουμε. Ακολουθεί μια σύντομη περιγραφή των δημοφιλέστερων τεχνικών: (SOFTWARETESTINGHELP, 2021) (ANALYTICS INDIA MAGAZINE PVT LTD, 2020), (North, 2012), (Usama Fayyad, 1996)

2.3.1 Association Analysis / Frequent Pattern Mining

Η τεχνική αυτή εφαρμόζεται για την ανακάλυψη συσχετίσεων μεταξύ διαφορετικών χαρακτηριστικών ενός συνόλου δεδομένων, με σκοπό την ανάδειξη ενός προτύπου στη «συμπεριφορά» των χαρακτηριστικών αυτών. Κλασικό παράδειγμα της τεχνικής αυτής είναι η «Shopping Basket Analysis» ή αλλιώς, η ανακάλυψη των προϊόντων ενός καταστήματος που ένας πελάτης είναι πιθανότερο να αγοράσει μαζί.

2.3.2 Clustering Analysis

Πρόκειται για μια τεχνική διαμερισμού (partitioning) ενός συνόλου δεδομένων σε ομάδες (clusters). Με χρήση αλγορίθμων, τα δεδομένα χωρίζονται σε ομάδες/κλάσεις χωρίς να γνωρίζουμε από πριν σε ποια ομάδα ανήκει το κάθε αντικείμενο (unsupervised learning), ανάλογα με τις ομοιότητες που παρουσιάζουν στα χαρακτηριστικά τους. Η τεχνική αυτή είναι πολύ χρήσιμη για την καλύτερη κατανόηση των δεδομένων, αλλά και για την ανακάλυψη ακραίων τιμών (για παράδειγμα, σε δεδομένα που αφορούν ηλεκτρονικές αγορές με πιστωτική κάρτα). Κάποιοι από τους δημοφιλείς αλγόριθμους ομαδοποίησης είναι ο K-Means, ο fuzzy C-means, οι αλγόριθμοι Συσσωρευτικής ιεραρχικής ομαδοποίησης (Hierarchical Clustering Algorithms) κ.λπ.

2.3.3 Correlation Analysis

Πρόκειται για μία τεχνική που χρησιμοποιείται για την εξαγωγή συμπερασμάτων για το βαθμό αλλά και για το είδος (θετική/αρνητική) συσχέτισης μεταξύ των χαρακτηριστικών ενός συνόλου δεδομένων. Χρησιμοποιείται ευρέως για την εύρεση γραμμικών συσχετίσεων μεταξύ των χαρακτηριστικών.

2.3.4 Classification

Πρόκειται για μια επίσης πολύ διαδεδομένη τεχνική, η οποία πολλές φορές αναφέρεται και ως επιβλεπόμενη μάθηση (supervised learning). Χρησιμοποιείται για την πρόβλεψη της κατηγορίας στην οποία ανήκει κάθε ένα από τα υποσύνολα ενός συνόλου δεδομένων με χρήση δεδομένων εκπαίδευσης, δεδομένων δηλαδή τα οποία είναι εκ των προτέρων γνωστό ότι ανήκουν σε κάποια συγκεκριμένη κατηγορία.

2.3.5 Decision Tree Induction

Πρόκειται για αλγόριθμο επιβλεπόμενης μάθησης, που δημιουργεί μοντέλα δενδρικής δομής (αποτελούμενα δηλαδή από κόμβους, ακμές και φύλλα), τα οποία χρησιμοποιούνται είτε για την ταξινόμηση στοιχείων σε κλάσεις/κατηγορίες (prediction), είτε για την πρόβλεψη μιας τιμής (regression). Τα δένδρα απόφασης

μας δίνουν τη δυνατότητα οπτικοποίησης των κριτηρίων και των αποφάσεων (If/Then) με βάση τις οποίες διασπάται το δένδρο.

2.3.6 Regression Analysis

Οι αλγόριθμοι παλινδρόμησης ανήκουν στην οικογένεια των αλγορίθμων επιβλεπόμενης μάθησης και χρησιμοποιούνται κυρίως για την ανακάλυψη συσχετίσεων μεταξύ μεταβλητών ενός συνόλου δεδομένων, αλλά και για την πρόβλεψη τιμών σε μεταβλητές νέων δεδομένων με βάση τις συσχετίσεις μεταξύ των μεταβλητών σε ήδη γνωστά δεδομένα. Χρησιμοποιείται ευρέως σε πεδία όπως οικονομικές προβλέψεις, ανάλυση τάσεων, πρόβλεψη χρονοσειρών κ.α.

2.4 Δεδομένα Μεγάλης Κλίμακας στην Εκπαίδευση

Η Εκπαίδευση είναι ένας από τους τομείς όπου υπάρχει αποθηκευμένος τεράστιος όγκος δεδομένων αποθηκευμένων σε συστήματα διαχείρισης μάθησης, σε ολοκληρωμένα πληροφοριακά συστήματα ή σε αποθήκες δεδομένων. Λόγω του τεράστιου όγκου των δεδομένων, αλλά και της ανομοιογένειας που συνήθως τα διακρίνει, οι συνήθεις τεχνικές που εφαρμόζονται με χρήση των εργαλείων διαχείρισης βάσεων δεδομένων δεν επαρκούν. Η εξόρυξη και ανάλυση αυτών των δεδομένων με κατάλληλα εργαλεία και τεχνικές, μπορεί να αποφέρει μεγάλα οφέλη στα εκπαιδευτικά ιδρύματα. Οι αποφάσεις άλλωστε που βασίζονται στην ανάλυση δεδομένων, αποδεδειγμένα οδηγούν σε βελτίωση τόσο του επιχειρηματικού και οργανωτικού μοντέλου ενός φορέα, όσο και της παραγωγικότητάς του (H.J.Watson, 2011). Η γνώση που μπορεί να αποκτηθεί, μπορεί να βοηθήσει σημαντικά σε διαδικασίες που αφορούν στον μελλοντικό προγραμματισμό ή στη κατανομή πόρων. Ταυτόχρονα, μπορεί να συμβάλει στην ανάδειξη προβληματικών σημείων των διαδικασιών.

Ειδικότερα, η εξόρυξη δεδομένων ενός εκπαιδευτικού ιδρύματος μπορεί να βοηθήσει σε πεδία όπως (B.Tulasi, 2013):

- Λήψη αποφάσεων με χρήση σεναρίων «εάν/τότε»
- Αποτελεσματικότερη διαχείριση πόρων
- Βελτίωση των τελικών ποσοστών επιτυχίας των σπουδαστών με αναγνώριση αδυναμιών και κινδύνων που διατρέχουν σε αρχικό στάδιο της φοίτησής τους
- Ανάλυση της αποτίμησης εκ μέρους των συμμετεχόντων σε ένα εκπαιδευτικό πρόγραμμα (εκπαιδευτών και των εκπαιδευομένων), με σκοπό τη συνεχή βελτίωσή του από τους σχεδιαστές τους
- Καλύτερο οικονομικό σχεδιασμό με χρήση προβλεπτικών μοντέλων

3 ΤΑ ΕΠΙΜΟΡΦΩΤΙΚΑ ΣΕΜΙΝΑΡΙΑ ΤΟΥ ΙΝΣΤΙΤΟΥΤΟΥ ΕΠΙΜΟΡΦΩΣΗΣ

3.1 Γενικά

Το Ινστιτούτο Επιμόρφωσης (ΙΝ.ΕΠ.) του Εθνικού Κέντρου Δημόσιας Διοίκησης και Αυτοδιοίκησης (Ε.Κ.Δ.Δ.Α.), στο πλαίσιο της πολιτικής για την ανάπτυξη του ανθρώπινου δυναμικού υλοποιεί ένα ολοκληρωμένο πρόγραμμα εκπαίδευσης στελεχών του δημόσιου τομέα, με στόχο να συμβάλει στη βελτίωση της αποτελεσματικότητας και της ποιότητας στη Δημόσια Διοίκηση και την Τοπική Αυτοδιοίκηση.

Για το σκοπό αυτό έχει καθιερώσει: α) την Ανίχνευση Εκπαιδευτικών Αναγκών στη Δημόσια Διοίκηση, καθώς και β) στην Αποτίμηση της αποτελεσματικότητας της επιμόρφωσης στους φορείς του δημοσίου.

Σχηματικά, ο χάρτης επιχειρησιακών διαδικασιών που ακολουθείται από το ΙΝ.ΕΠ σε ετήσια βάση για την υλοποίηση των επιμορφωτικών του δράσεων απεικονίζεται ως εξής (Χάλαρης, 2019):



Εικόνα 3.1 Χάρτης Επιχειρησιακών διαδικασιών ΙΝ.ΕΠ.

Στην παρούσα εργασία, θα επιχειρηθεί η ανάλυση δεδομένων και η εφαρμογή τεχνικών εξόρυξης για την εξαγωγή συμπερασμάτων κυρίως στις φάσεις της Ανίχνευσης εκπαιδευτικών αναγκών, του Προγραμματισμού των σεμιναρίων, και της Αξιολόγησης των σεμιναρίων από την πλευρά των εκπαιδευομένων και των εκπαιδευτών.

3.2 Ανίχνευση εκπαιδευτικών αναγκών επιμόρφωσης Φορέων

Η Ανίχνευση εκπαιδευτικών αναγκών των φορέων του δημόσιου και του ευρύτερου δημόσιου τομέα, ώστε ανάλογα να προκύψει ο σχεδιασμός των προγραμμάτων του ΙΝ.ΕΠ για το επόμενο εξάμηνο, γίνεται με συγκεκριμένη μεθοδολογία και είναι προσανατολισμένη στις Δημόσιες Πολιτικές.

Χρησιμοποιούνται εργαλεία όπως οι συνεντεύξεις, οι ομάδες εργασίας και τα ατομικά ερωτηματολόγια ώστε να καλύπτεται η ανίχνευση των εκπαιδευτικών αναγκών από πάνω προς τα κάτω και από κάτω προς τα πάνω. Το σύστημα ανίχνευσης στηρίζεται σε μέθοδο, η οποία συνδυάζει αφενός, τις ανάγκες της διοίκησης, όπως αυτές προκύπτουν τόσο από τις ασκούμενες κάθετες (τομεακές) και οριζόντιες πολιτικές, όσο και από τις επιμέρους πολιτικές για την ανάπτυξη του ανθρώπινου δυναμικού (πρόσληψη, εξέλιξη, μεταβολές, εκπαίδευση) και αφετέρου τις εξατομικευμένες ανάγκες των υπαλλήλων, όπως αυτές

προκύπτουν μέσα από το ατομικό τους υπηρεσιακό και προσωπικό προφίλ και από τις προσωπικές τους εκτιμήσεις για τη βελτίωση της απόδοσής τους στην εργασία τους.

Με βάση τα πορίσματα της Ανίχνευσης Εκπαιδευτικών Αναγκών καταρτίζεται το επιχειρησιακό σχέδιο εκπαίδευσης του κάθε φορέα της Δημόσιας Διοίκησης και προγραμματίζεται η εκπαίδευση σε ετήσια βάση για το σύνολο του Δημόσιου Τομέα και τους Οργανισμούς Τοπικής Αυτοδιοίκησης. Είναι επομένως προφανής ο σημαντικότερος ρόλος της φάσης αυτής στο γενικότερη διαδικασία διοργάνωσης και υλοποίησης της επιμορφωτικής διαδικασίας.

3.3 Οργάνωση των σεμιναρίων

Το ΙΝ.ΕΠ, κατά την τελευταία δεκαετία, έχει επιμορφώσει περί τις 300.000 υπαλλήλους του δημόσιου και του ευρύτερου δημόσιου τομέα. Ενδεικτικά, για το Α' εξάμηνο του 2021 υλοποιήθηκαν 157 διαφορετικοί Τίτλοι επιμορφωτικών δράσεων, στους εξής 7 διαφορετικούς Θεματικούς Τομείς (ΕΚΔΔΑ, 2021):

- ΑΝΘΡΩΠΙΝΑ ΔΙΚΑΙΩΜΑΤΑ & ΚΟΙΝΩΝΙΚΗ ΠΟΛΙΤΙΚΗ (27 Τίτλοι Προγραμμάτων),
- ΒΙΩΣΙΜΗ ΑΝΑΠΤΥΞΗ (25 Τίτλοι Προγραμμάτων),
- ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ & ΔΙΑΚΥΒΕΡΝΗΣΗ (41 Τίτλοι Προγραμμάτων),
- ΕΙΣΑΓΩΓΙΚΗ ΕΚΠΑΙΔΕΥΣΗ (δεν περιλαμβάνει θεματικούς τομείς, πρόκειται για την υποχρεωτική εκπαίδευση των νεοδιορισμένων δημοσίων υπαλλήλων)
- ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟΜΙΚΗ ΠΟΛΙΤΙΚΗ (25 Τίτλοι Προγραμμάτων),
- ΠΛΗΡΟΦΟΡΙΚΗ & ΨΗΦΙΑΚΕΣ ΥΠΗΡΕΣΙΕΣ (34 Τίτλοι Προγραμμάτων),
- ΠΟΛΙΤΙΣΤΙΚΗ & ΤΟΥΡΙΣΤΙΚΗ ΑΝΑΠΤΥΞΗ (5 Τίτλοι Προγραμμάτων)

Σε αυτά έλαβαν μέρος 11.136 εκπαιδευόμενοι (381.603 ανθρωπόωρες) οι οποίοι υπέβαλαν συνολικά 8.335 αξιολογήσεις.

Κάθε σεμινάριο κωδικοποιείται στη βάση δεδομένων του φορέα με έναν μοναδικό αναγνωριστικό κωδικό της μορφής 8χγγγζηη (π.χ. 80135Δ20) όπου:

- χ: [0-9], ανάλογα με το αν το σεμινάριο υλοποιείται από το ΙΝ.ΕΠ (Αθήνα) ή από κάποιο από τα Περιφερειακά ΙΝ.ΕΠ.
- γγγ: [001-999]: αύξων αριθμός σεμιναρίου
- ζ: [Κ|Ρ|Δ|Β|Ο|Τ|Λ] ανάλογα με τον Θεματικό Κύκλο στον οποίο ανήκει το σεμινάριο
- ηη: [09-21]: ανάλογα με το έτος υλοποίησης του σεμιναρίου

Σεμινάρια με ίδιο Θεματικό Τίτλο π.χ. «ΤΟΠΙΚΗ ΠΟΛΙΤΙΣΤΙΚΗ ΑΝΑΠΤΥΞΗ ΚΑΙ ΤΠΕ» μπορούν να εμφανίζονται φυσικά πολλές φορές μέσα στο ΟΠΣ, ο κωδικός σεμιναρίου όμως είναι διαφορετικός και μοναδικός για το καθένα, αφού ακόμα και όλα τα υπόλοιπα στοιχεία του κωδικού τους να είναι ίδια, σίγουρα θα διαφέρει ο αύξων αριθμός τους (γγγ). Από την άλλη πλευρά, ένα κοινό που σίγουρα θα έχουν δύο σεμινάρια ίδιου Θεματικού Κύκλου στον κωδικό τους, είναι ο χαρακτήρας που δηλώνει τον Θεματικό Κύκλο στον οποίο ανήκουν.

Είδος Θεματικού Κύκλου	Τίτλος	Εναρξη	Λήξη	Διάρκεια (Ωρες)	Κωδικός
• ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΔΙΑΚΥΒΕΡΝΗΣΗ	ΕΚΠΑΙΔΕΥΣΗ ΕΚΠΑΙΔΕΥΤΩΝ/ΤΡΙΩΝ	02/09/2021	24/09/2021	(60)	80167Δ21
• ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΔΙΑΚΥΒΕΡΝΗΣΗ	ΕΞ ΑΠΟΣΤΑΣΕΩΣ ΕΚΠΑΙΔΕΥΣΗ ΣΤΗΝ ΗΓΕΣΙΑ ΚΑΙ ΔΙΟΙΚΗΣΗ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ	03/09/2021	17/12/2021	(118)	80168Δ21
• ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΔΙΑΚΥΒΕΡΝΗΣΗ	ΕΞ ΑΠΟΣΤΑΣΕΩΣ ΕΚΠΑΙΔΕΥΣΗ ΣΤΗΝ ΗΓΕΣΙΑ ΚΑΙ ΔΙΟΙΚΗΣΗ ΣΤΗΝ ΕΚΠΑΙΔΕΥΣΗ	03/09/2021	21/12/2021	(118)	81032Δ21
• ΕΙΣΑΓΩΓΙΚΗ ΕΚΠΑΙΔΕΥΣΗ	ΕΙΣΑΓΩΓΙΚΗ ΕΚΠΑΙΔΕΥΣΗ	06/09/2021	17/09/2021	(70)	80030B21
• ΑΝΘΡΩΠ. ΔΙΚΑΙΩΜ. ΚΑΙ ΚΟΙΝΩΝΙΚΗΣ ΠΟΛΙΤΙΚΗΣ	ΤΟ ΑΝΗΛΙΚΟ ΘΥΜΑ ΣΤΗΝ ΠΟΙΝΙΚΗ ΔΙΚΑΙΟΣΥΝΗ	06/09/2021	10/09/2021	(35)	80092Κ21
• ΟΙΚΟΝΟΜΙΑ ΚΑΙ ΔΗΜΟΣΙΟΝΟΜΙΚΗ ΠΟΛΙΤΙΚΗ	ΕΞΕΙΔΙΚΕΥΣΗ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΣΤΗ ΜΕΘΟΔΟΛΟΓΙΑ ΔΙΑΧΕΙΡΙΣΤΙΚΩΝ ΕΛΕΓΧΩΝ (ΠΡΟΧΩΡΗΜΕΝΟ ΕΠΙΠΕΔΟ)	06/09/2021	10/09/2021	(35)	80110Q21

Εικόνα 3.2 Θεματικοί Τίτλοι και Κύκλοι σεμιναρίων ΙΝ.ΕΠ

3.4 Αξιολόγηση-Αποτίμηση των σεμιναρίων

Η αξιολόγηση ενός εκπαιδευτικού προγράμματος υποβάλλεται ηλεκτρονικά αμέσως μετά την ολοκλήρωσή του, μέσω αντίστοιχων ερωτηματολογίων που συμπληρώνουν τόσο οι εκπαιδευόμενοι, όσο και οι εκπαιδευτές.

Η αποτίμηση των αποτελεσμάτων της επιμόρφωσης, που αφορά στην εφαρμογή και αξιοποίηση των γνώσεων και δεξιοτήτων που απέκτησαν οι υπάλληλοι των φορέων της δημόσιας διοίκησης από τη συμμετοχή τους στα επιμορφωτικά προγράμματα στην καθημερινή εργασία τους, αλλά και τις αλλαγές που προκλήθηκαν στη λειτουργία της υπηρεσίας που εργάζονται, υποβάλλεται από τους εκπαιδευόμενους επίσης ηλεκτρονικά, έως και 2 μήνες μετά τη λήξη του προγράμματος.

Ειδικότερα, υποβάλλονται ηλεκτρονικά:

- Αξιολόγηση του επιμορφωτικού προγράμματος από τους συμμετέχοντες/χουσες ως προς τη δομή του προγράμματος, το εκπαιδευτικό υλικό, την υλικοτεχνική υποδομή, τους εκπαιδευτές
- Αξιολόγηση των εκπαιδευτών για το επιμορφωτικό πρόγραμμα στο οποίο συμμετείχαν
- Αποτίμηση από τους εκπαιδευόμενους των αποτελεσμάτων της επιμόρφωσης, όσον αφορά στην εφαρμογή όσων διδάχθηκαν, στην καθημερινή τους εργασία.

Οι ηλεκτρονικές εφαρμογές μέσω των οποίων υλοποιούνται τα παραπάνω, ενσωματώνονται στο Ολοκληρωμένο Πληροφοριακό Σύστημα του ΕΚΔΔΑ.

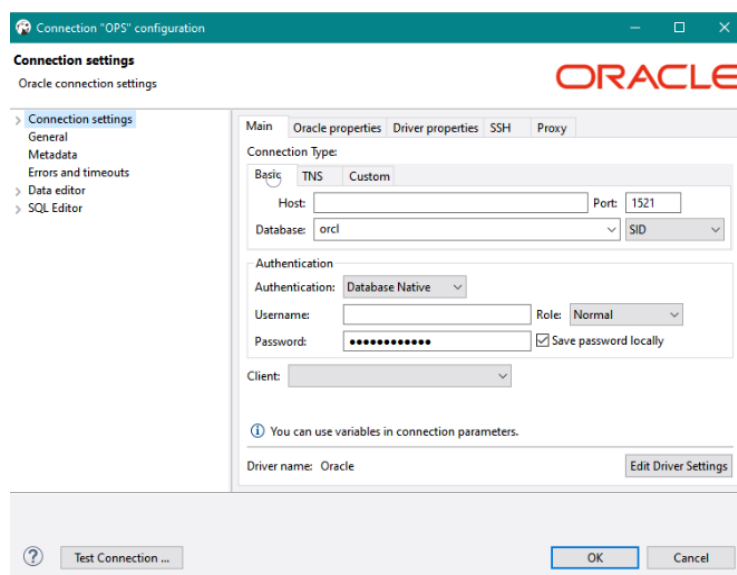
4 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΣΕΜΙΝΑΡΙΑ ΚΑΙ ΑΝΑΛΥΣΗ ΜΕ ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ

Για την παρούσα εργασία, ακολουθήθηκε το μοντέλο CRISP-DM, οι φάσεις του οποίου αναφέρθηκαν στην ενότητα 2.2 *Φάσεις και Μεθοδολογία ανάλυσης Δεδομένων Μεγάλης Κλίμακας*. Όπως περιγράφεται στην πρώτη φάση του μοντέλου, επιχειρήθηκε αρχικά η προσέγγιση του επιχειρηματικού μοντέλου του φορέα. Η κατανόηση δηλαδή του τρόπου με τον οποίο ο φορέας υλοποιεί τα σεμινάρια, από τη φάση της ανίχνευσης των εκπαιδευτικών αναγκών, ως τη φάση της υλοποίησης και της αξιολόγησής του από εκπαιδευτές και εκπαιδευόμενους.

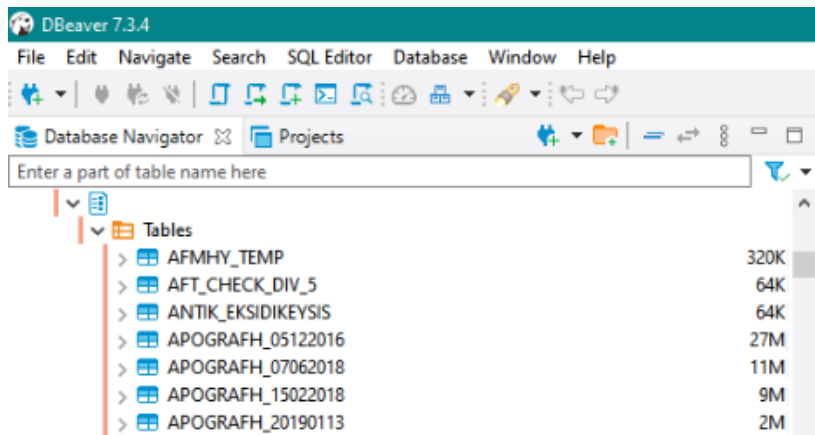
Το πρώτο και βασικό ερώτημα που θα έπρεπε να απαντηθεί, ήταν εάν και σε ποιον βαθμό, θα μπορούσε η χρήση τεχνικών εξόρυξης δεδομένων να βοηθήσει στην κατεύθυνση της βελτίωσης των διαδικασιών του φορέα. Από την εξέταση των στοιχείων, οικονομικών και άλλων, προέκυψε το συμπέρασμα ότι η ανάλυση των δεδομένων των σεμιναρίων θα μπορούσε να οδηγήσει σε χρήσιμα συμπεράσματα, τόσο σε θέματα οικονομικής διαχείρισης, όσο και σε θέματα ανίχνευσης αναγκών των εκπαιδευόμενων. Επίσης, με βάση την ανάλυση των αξιολογήσεων των εκπαιδευόμενων, θα μπορούσαν να εξαχθούν χρήσιμα συμπεράσματα για τη βελτίωση του σχεδιασμού της επόμενης εκπαιδευτικής περιόδου.

4.1 Συλλογή Δεδομένων από τη Βάση Δεδομένων

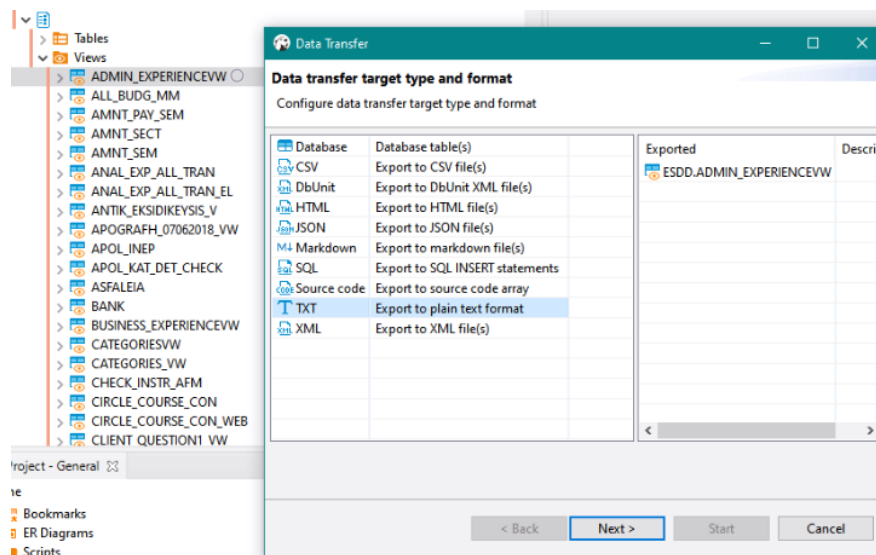
Για τη συλλογή των δεδομένων από τη βάση δεδομένων του φορέα, χρησιμοποιήθηκε το δωρεάν, ανοικτού κώδικα λογισμικό DBEAVER Community edition (<https://dbeaver.io/>). Δημιουργήθηκε διασύνδεση με το Ολοκληρωμένο Πληροφοριακό Σύστημα, και τα περιεχόμενα των επιλεγμένων πινάκων ή όψεων (views) της Βάσης εξαγόταν ως αρχεία κειμένου, είτε σε μορφή .txt, είτε σε μορφή comma-separated values (.csv).



Εικόνα 4.1 DBEAVER-Σύνδεση με το ΟΠΣ του φορέα



Εικόνα 4.2 Πίνακες του ΟΠΣ



Εικόνα 4.3 Εξαγωγή δεδομένων από το ΟΠΣ

Κατά την περίοδο συλλογής των δεδομένων το σχήμα της βάσης περιείχε 2033 πίνακες και 596 views. Όπως είναι προφανές, η επιλογή των πινάκων από τους οποίους θα έπρεπε να αντληθούν δεδομένα ήταν μια αρκετά επίπονη και χρονοβόρα διαδικασία η οποία, εκτός των άλλων, απαιτούσε και συνεργασία με τους διαχειριστές της βάσης, οι οποίοι γνώριζαν σε βάθος τόσο το επιχειρηματικό μοντέλο του φορέα, όσο και τη δομή της βάσης. Βασικό κριτήριο για το ποιοί πίνακες τελικά επιλέχθηκαν, αποτέλεσαν όπως είναι φυσικό οι στόχοι που τέθηκαν κατά το πρώτο στάδιο της ανάλυσης.

Πιο συγκεκριμένα, κρίθηκε ότι παρουσίαζε ενδιαφέρον να διερευνηθεί η σκοπιμότητα εφαρμογής τεχνικών εξόρυξης για τις παρακάτω πέντε περιπτώσεις:

- Πρόβλεψη υπέρβασης προϋπολογισμένου ποσού σεμιναρίου
- Ανίχνευση αναγκών επιμόρφωσης με βάση τα σχόλια των εκπαιδευόμενων
- Δημιουργία Κανόνων Συσχέτισης Θεματικών Τίτλων (σεμιναρίων) με βάση το ιστορικό αιτήσεων εκπαιδευόμενου

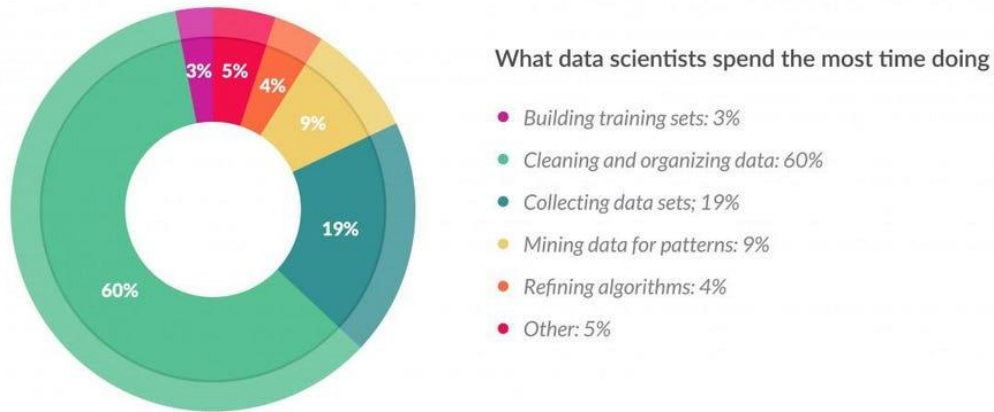
- Ανάλυση της Αξιολόγησης σεμιναρίων από τους εκπαιδευόμενους και εξαγωγή συμπερασμάτων
- Πρόβλεψη ποσοστών επιτυχούς ολοκλήρωσης της επιμόρφωσης ανά ηλικιακή ομάδα, εκπαιδευτική κατηγορία, θέση ευθύνης.

Στην παράγραφο 4.3 θα περιγραφούν αναλυτικότερα οι στόχοι που τέθηκαν, οι τεχνικές που κρίθηκε σκόπιμο να εφαρμοστούν, καθώς και ο τρόπος υλοποίησής τους με χρήση λογισμικού εξόρυξης δεδομένων.

4.2 Data Preparation

Στο 3^ο βήμα του μοντέλου CRISP-DM λαμβάνει χώρα η προετοιμασία των δεδομένων. Τα δεδομένα που έχουν συλλεχθεί στο προηγούμενο βήμα θα πρέπει να μορφοποιηθούν με τέτοιο τρόπο, ώστε να μπορούν στη συνέχεια να χρησιμοποιηθούν από το λογισμικό εξόρυξης για παραγωγή μοντέλων. Στο στάδιο αυτό συνήθως απαιτείται επίσης καθαρισμός ή/και διόρθωση των δεδομένων ώστε να απομακρυνθούν για παράδειγμα λανθασμένες ή ακραίες τιμές (outliers), ή να συμπληρωθούν, με χρήση συγκεκριμένων τεχνικών, τιμές που απουσιάζουν από το δείγμα (missing values). Είναι εξαιρετικά σημαντικό, τα δεδομένα που τελικά θα χρησιμοποιηθούν, να είναι υψηλής ποιότητας, διότι μόνο τότε τα συμπεράσματα που θα προκύψουν θα έχουν πραγματικά αξία.

Όπως αναφέρεται και στη σχετική βιβλιογραφία, διαπιστώθηκε ότι πρόκειται πράγματι για το πιο χρονοβόρο, αλλά και πιο επίπονο μέρος της διαδικασίας. Σύμφωνα με έρευνα (Press, 2016), το 60% του χρόνου ενός επιστήμονα της Πληροφορικής που ασχολείται με δεδομένα μεγάλης κλίμακας καταναλώνεται στο στάδιο της προετοιμασίας των δεδομένων. Μαζί με ένα 19% του χρόνου το οποίο σύμφωνα με την ίδια έρευνα αφιερώνεται στη συλλογή των δεδομένων, σημαίνει ότι ένα 80% περίπου του συνολικού χρόνου σε έναν κύκλο συλλογής-εξόρυξης-μοντελοποίησης δεδομένων μεγάλης κλίμακας, έχει να κάνει με τη συλλογή και προετοιμασία των δεδομένων. Μάλιστα, το 76% των συμμετεχόντων στην έρευνα, δήλωσαν ότι το στάδιο της προετοιμασίας, αποτελεί το λιγότερο ευχάριστο μέρος της διαδικασίας. Εντύπωση προκαλεί επίσης, ότι μόλις το 9% του συνολικού χρόνου αφιερώνεται στην εξόρυξη δεδομένων για εύρεση προτύπων (Εικόνα 4.4).



Εικόνα 4.4 Καταμερισμός του χρόνου στην ανάλυση Big Data

Τα βήματα της προετοιμασίας των δεδομένων είναι συνοπτικά τα παρακάτω (Talend, n.d.):

- Ανακάλυψη / αναγνώριση των δεδομένων. Η κατανόηση δηλαδή των δεδομένων ενός συνόλου και του τι ακριβώς αντιπροσωπεύουν, έτσι ώστε να μπορέσουμε να αποφασίσουμε πού και με ποιον τρόπο θα μπορέσουμε να τα αξιοποιήσουμε.
- Καθαρισμός και επικύρωση. Ιδιαίτερα χρονοβόρο στάδιο, κατά το οποίο:
 - αφαιρούνται προβληματικές τιμές,
 - συμπληρώνονται ελλιπείς τιμές,
 - αφαιρούνται ή αποκρύπτονται τιμές που σχετίζονται με ευαίσθητα δεδομένα
- Μετασχηματισμός των δεδομένων ώστε να είναι καλύτερα κατανοητά για κάποιον τρίτο, αλλά και εμπλουτισμός τους συνδέοντάς τα με άλλες σχετικές πληροφορίες, ώστε τελικά να παρέχουν βαθύτερη γνώση.
- Αποθήκευση των δεδομένων σε κάποια εφαρμογή (π.χ. Rapidminer repository)

Τα βήματα αυτά ακολουθήθηκαν όπως θα δούμε στις επόμενες παραγράφους, σε μεγαλύτερο ή μικρότερο βαθμό και για την προετοιμασία των δεδομένων που χρησιμοποιήθηκαν στις περιπτώσεις που εξετάσαμε στο πλαίσιο της παρούσας εργασίας.

4.3 Ανάλυση δεδομένων με τεχνικές εξόρυξης

4.3.1 Περίπτωση Εφαρμογής 1: Πρόβλεψη υπέρβασης προϋπολογισμένου ποσού σεμιναρίου

Business Understanding

Κατά το στάδιο της κατανόησης του επιχειρηματικού μοντέλου του φορέα, και πιο συγκεκριμένα κατά την εξέταση των οικονομικών στοιχείων ενός σεμιναρίου, έγιναν μεταξύ

άλλων κάποιες παρατηρήσεις, οι οποίες θεωρήθηκε ότι παρουσίαζαν ενδιαφέρον στο πλαίσιο της συγκεκριμένης εργασίας. Συγκεκριμένα, παρατηρήθηκε πως για κάθε προγραμματισμένο σεμινάριο, υπήρχαν τα παρακάτω τρία βασικά οικονομικά μεγέθη:

- το προϋπολογισμένο ποσό που έχει εγκριθεί από το Διοικητικό Συμβούλιο του φορέα για κάθε συγκεκριμένο σεμινάριο,
- ο αναλυτικός προϋπολογισμός για το ίδιο σεμινάριο, όπως έχει καταστρωθεί από το ΙΝ.ΕΠ,
- το τελικό, πραγματικό ποσό που δαπανήθηκε για το συγκεκριμένο σεμινάριο με βάση τα απολογιστικά στοιχεία.

Στην πλειονότητα των περιπτώσεων, το τελικό ποσό που δαπανήθηκε για ένα σεμινάριο παρουσίαζε μικρότερη ή μεγαλύτερη απόκλιση σε σχέση με το ποσό το οποίο είχε αρχικά εγκριθεί από το Δ.Σ., αλλά και από εκείνο που είχε προϋπολογισθεί από το ΙΝ.ΕΠ. (Εικόνα 4.5).

The screenshot shows a web-based form titled 'Πληροφοριακό Έντυπο Σεμιναρίου'. It contains various fields for seminar details. The 'Οικονομικά' (Financial) tab is selected. The following table summarizes the financial data shown in the form:

Field	Value
Προϋπ. Δ.Σ.	2,200.00
Αναλυτικός Προϋπ.	1,416.80
Απολογισμός	1,052.48

Εικόνα 4.5 Προϋπολογισμός - Απολογισμός σεμιναρίου

Κάτι τέτοιο είναι απολύτως λογικό, καθώς το τελικό ποσό που δαπανάται για ένα σεμινάριο μπορεί να εξαρτηθεί από παράγοντες οι οποίοι δεν είναι δυνατόν να προβλεφθούν με ακρίβεια εκ των προτέρων, όπως η συμμετοχή (πλήθος) των εκπαιδευομένων, αλλά και άλλες παράμετροι των οποίων η επίδραση δεν είναι δυνατόν να είναι γνωστή κατά τη φάση του σχεδιασμού ενός σεμιναρίου.

Εντούτοις, κρίθηκε σκόπιμο να διερευνηθεί εάν, και σε ποιον βαθμό, υπήρχε κάποιο πρότυπο με βάση το οποίο το τελικό ποσό του απολογισμού ενός σεμιναρίου διαμορφώνεται άνω του αρχικά προϋπολογισθέντος ποσού. Να δημιουργηθεί επομένως ένα μοντέλο το οποίο, χρησιμοποιώντας τα δεδομένα από ήδη πραγματοποιηθέντα σεμινάρια ως δεδομένα εκπαίδευσης, θα μπορεί να προβλέψει τα χαρακτηριστικά των

σεμιναρίων τα οποία τελικά θα υπερβούν τον προϋπολογισμό τους, εστιάζοντας ειδικότερα στον Θεματικό Κύκλο στον οποίο αυτά ανήκουν (π.χ. Πληροφορική, Δημόσια Διοίκηση κ.λπ), ή στον Τίτλο Σεμιναρίου (π.χ. «Βασικές Ψηφιακές Δεξιότητες»), στον αριθμό των συμμετεχόντων και στη διάρκεια του σεμιναρίου σε ώρες.

Κάτι τέτοιο θα ήταν χρήσιμο ως γνώση στη φάση του σχεδιασμού, καθώς αναμφίβολα θα βοηθούσε στην όσο το δυνατόν πιο ακριβή και με μικρότερες αποκλίσεις κατάστρωση του προϋπολογισμού του φορέα.

Data Understanding

Τα δεδομένα που χρειαζόμασταν για την δημιουργία του μοντέλου, βρίσκονταν σε 3 πίνακες του ΟΠΣ του φορέα. Ο βασικός πίνακας (I_SEMINARS) περιείχε τα οικονομικά στοιχεία ενδιαφέροντος, στα πεδία BUDGET_PR (Προϋπολογισμός ΔΣ), BUDGET (Προϋπολογισμός ΙΝΕΠ) και REPORT (Απολογισμός), το πλήθος εκπαιδευόμενων κάθε σεμιναρίου και το πλήθος των ωρών του σεμιναρίου. Επιπλέον, σε κωδικοποιημένη μορφή, τον Θεματικό Κύκλο στον οποίο ανήκε κάθε σεμινάριο, καθώς και τον Τίτλο του σεμιναρίου. Για τον λόγο αυτό, και για να είναι το αποτέλεσμα πιο κατανοητό, χρειάστηκε να γίνει εισαγωγή δεδομένων και από 2 άλλους πίνακες (I_SECTIONS, I_SEM_THEMES), οι οποίοι περιείχαν και το αντίστοιχο λεκτικό.

Η συλλογή των δεδομένων έγινε με το εργαλείο DBeaver όπως αναφέρθηκε στην παρ. 4.1. Τα δεδομένα αποθηκεύτηκαν σε 3 διαφορετικά .csv αρχεία (Εικόνα 4.6, Εικόνα 4.7, Εικόνα 4.8), τα οποία στη συνέχεια εισήχθησαν στο εργαλείο Rapidminer και αποθηκεύτηκαν στο αποθετήριο ως datasets.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	CD,"TITLE",	"CD_SEM_TYPE",	"DT_START",	"DT_END",	"HOURS",	"STUDENTS",	"NUM_APPROVE",	"FLG",	"STATUS",	"CD_ORG_UNIT",	"CD_SECTION",			
19885	80003Δ15,,	"1",	2015-01-29 00:00:00,	2015-02-13 00:00:00,	28,21,,	"1",	"1",	"30",	"9",	"003547",	2674.51,,	2387.18,,	"6483",	"2",
19886	80004Δ15,,	"1",	2015-01-29 00:00:00,	2015-02-13 00:00:00,	28,24,,	"1",	"1",	"30",	"9",	"003547",	1330.88,,	1274.89,,	"2067",	"2",
19887	80006Δ15,,	"1",	2015-06-05 00:00:00,	2015-07-11 00:00:00,	96,25,,	"2",	"1",	"30",	"9",	"013467",	"7833",	"2",	"013467",	"1",
19888	80007Δ15,,	"1",	2015-05-15 00:00:00,	2015-06-20 00:00:00,	96,25,,	"2",	"1",	"30",	"9",	"003547",	"7833",	"2",	"010218",	"1",
19889	80008Δ15,,	"1",	2015-06-22 00:00:00,	2015-06-23 00:00:00,	16,25,,	"2",	"1",	"30",	"9",	"003705",	"1140",	"2",	"004008",	"2",

Εικόνα 4.6 Πίνακας με στοιχεία σεμιναρίων

	A	B	C	D	E	F	G
1	CD,"DSCR",	"FGL",	"LET",	"CD_LMS",	"FLG_WEB",	"DT_END"	
2	32,ΠΟΛΙΤΙΣΤΙΚΗ & ΤΟΥΡΙΣΤΙΚΗ ΑΝΑΠΤΥΞΗ,"0",	Λ,32,"1",					
3	33,ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ & ΔΙΑΚΥΒΕΡΝΗΣΗ,"0",	Δ,33,"1",					
4	26,ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦΑΡΜ. ΤΕΧΝ.,	"0",	T,,	"0",	2017-12-31		
5	27,ΒΙΩΣΙΜΗ ΑΝΑΠΤΥΞΗ,"0",	P,,	"1",				
6	28,ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟΜΙΚΗ ΠΟΛΙΤΙΚΗ,"0",	O,,	"1",				
7	29,ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝΕΙΑ ΣΤΗ Δ.Δ.,	"0",	M,,	"0",	2017-12-31		
8	30,ΔΙΟΙΚΗΤΙΚΗ ΜΕΤΑΡΡΥΘΜΙΣΗ ΚΑΙ ΑΠΟΚΕΝΤΡΩΣΗ,"0",	Δ,,	"0",	2017-12-31			
9	31,ΚΟΙΝΩΝΙΚΗ ΠΟΛΙΤΙΚΗ,"0",	K,,	"0",	2017-12-31 00:00:00			
10	2,ΔΗΜΟΣΙΟΥ ΜΑΝΑΤΖΜΕΝΤ,"0",	M,1,"0",	2010-07-31 00:00:00				

Εικόνα 4.7 Πίνακας Θεματικών Κύκλων (λεκτικό)

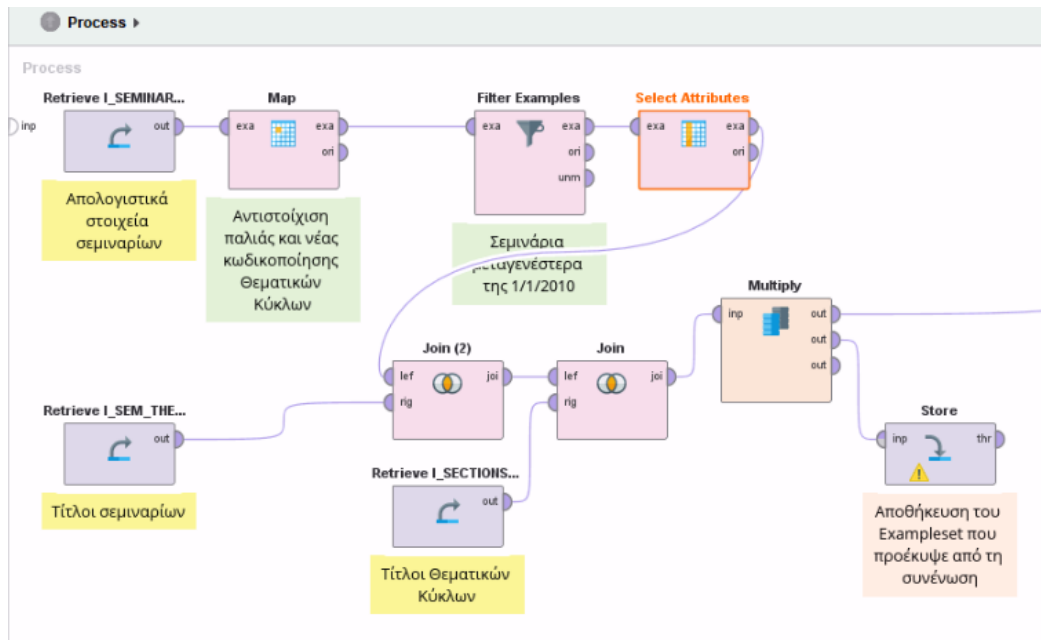
	A	B	C	D	E	F	G	H	I
1	CD,"DSCR","CD_ORG_UNIT","CD_SECTION","CD_UNIT_CAT","FLG_VALID","CD_CERTIF_THEME",								
1813	4570,"ΕΦΑΡΜΟΓΕΣ ΑΥΤΟCAD ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ","1","21","","1","","",""								
1814	4571,"WORD-EXCEL","1","21","","1","","",""								
1815	4572,"ΑΥΤΟCAD MAP","1","21","","1","","",""								
1816	4573,"ΕΦΑΡΜΟΓΕΣ EXCEL ΣΤΟ ΠΕΡΙΒΑΛΛΟΝ","1","21","","1","","",""								
1817	4574,"ΔΙΑΛΕΙΤΟΥΡΓΙΚΟΤΗΤΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ","1","21","","1","",""								
1818	4522,"ΑΝΑΠΤΥΞΗ ΕΦΑΡΜΟΓΩΝ WEB ΠΟΛΙΤΙΣΤΙΚΟΥ ΚΑΙ ΕΚΠΑΙΔΕΥΤΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ","1","21","","1","",""								
1819	4523,"AutoCAD ΕΠΙΠΕΔΟ ΙΙ","1","21","","1","","",""								
1820	4524,"ΓΛΩΣΣΑ ΠΡΟΓΡΑΜΜΑΤΙΣΜΟΥ XML","1","21","","1","","",""								

Εικόνα 4.8 Πίνακας Τίτλων σεμιναρίων

Data preparation

Όπως ήταν αναμενόμενο, κατά την εισαγωγή των δεδομένων στο Rapidminer, χρειάστηκαν αρκετές παρεμβάσεις, καθώς παρουσιάστηκαν διάφορα προβλήματα σε σχέση με τον τύπο ή την εγκυρότητα των δεδομένων, τα οποία οφείλονταν είτε σε σφάλματα κατά την καταχώριση, ή κατά την μετάπτωση δεδομένων από παλαιότερα σχήματα βάσεων στο τρέχον. Κάποιες στήλες που ούτως ή άλλως δεν παρουσίαζαν ενδιαφέρον, παραλείφθηκαν εντελώς.

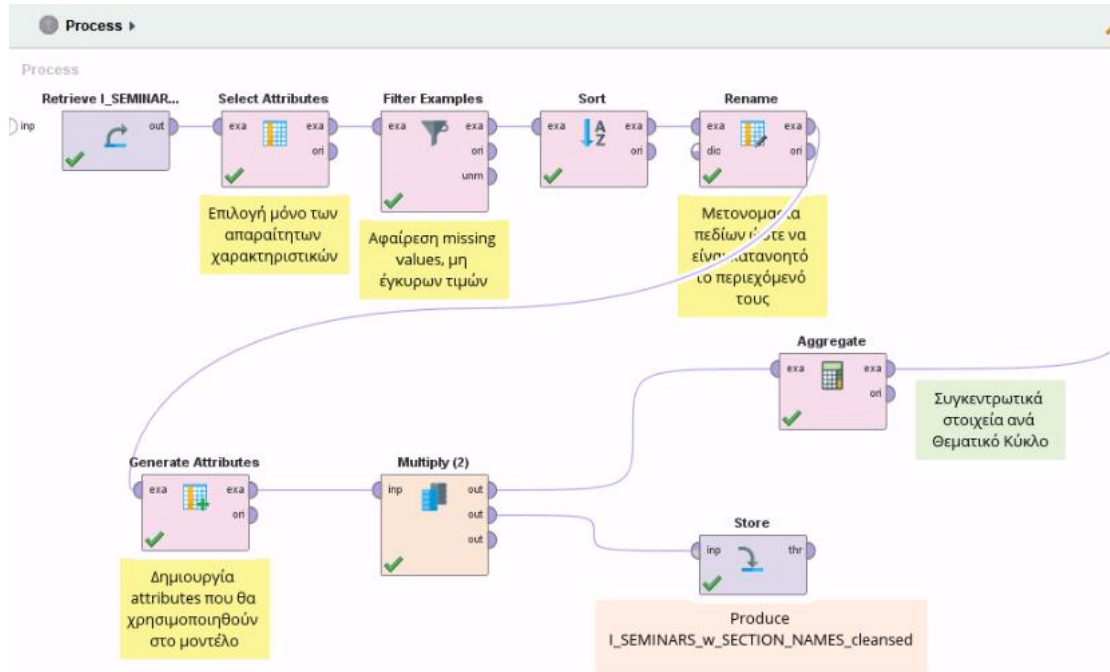
Μετά την εισαγωγή των αρχείων και τη δημιουργία των αντίστοιχων Examplesets στο Repository, γίνεται η συνένωση (Join) των τριών ExampleSets σε ένα, το οποίο περιέχει τα δεδομένα των τριών πινάκων. Γίνεται επίσης μια αντιστοίχιση παλιάς και νέας κωδικοποίησης για κάποιους Θεματικούς Κύκλους σεμιναρίων (Map). Για παράδειγμα, ο κωδικός Θεματικού Κύκλου (CD_SECTION) για τα σεμινάρια της Πληροφορικής ήταν έως κάποιο έτος ο '26', αλλά αργότερα άλλαξε σε '35'. Στους πίνακες όμως υπάρχουν εγγραφές που, παρότι αναφέρονται σε σεμινάρια του ίδιου Κύκλου («Πληροφορική»), έχουν διαφορετική τιμή στο αντίστοιχο πεδίο. Επίσης, γίνεται ένα πρώτο φιλτράρισμα των δεδομένων και επιλέγονται σεμινάρια από 1/1/2010 και μετά, καθώς για τα προγενέστερα σεμινάρια, το ποσοστό των τιμών που λείπουν σε χαρακτηριστικά που μας ενδιαφέρουν είναι ιδιαίτερα υψηλό, πιθανώς λόγω μη καταχώρισης:



Εικόνα 4.9 Συνένωση Πινάκων του ΟΠΣ

Στη συνέχεια γίνεται επεξεργασία του ExampleSet που προκύπτει. Επιλέγονται μόνο τα attributes που θα χρησιμοποιηθούν στο μοντέλο, αφαιρούνται ελλιπείς και μη έγκυρες τιμές, μετονομάζονται πεδία ώστε το όνομά τους να αποδίδει σαφώς το περιεχόμενό τους και τέλος, παράγονται νέα attributes (διαφορά Αναλυτικού Προϋπολογισμού – Απολογισμού και ένα Boolean χαρακτηριστικό «overbudget» το οποίο παίρνει την τιμή True ή False ανάλογα με το αν ο Απολογισμός υπερέβη τον Προϋπολογισμό ή όχι. Το χαρακτηριστικό αυτό είναι που θα χρησιμοποιηθεί μετέπειτα στο μοντέλο ως ετικέτα (Label) για την ταξινόμηση των δεδομένων δοκιμής σε μία από τις δύο πιθανές κλάσεις. Επίσης παράγονται συγκεντρωτικά στοιχεία (aggregate) σε ό,τι αφορά τα σύνολα των ποσών ανά Θεματικό Κύκλο.

Τέλος, το ExampleSet που προκύπτει (I_SEMINARS_w_SECTION_NAMES_cleansed), αποθηκεύεται στο αποθετήριο του RapidMiner ώστε να χρησιμοποιηθεί για τη δημιουργία του μοντέλου:



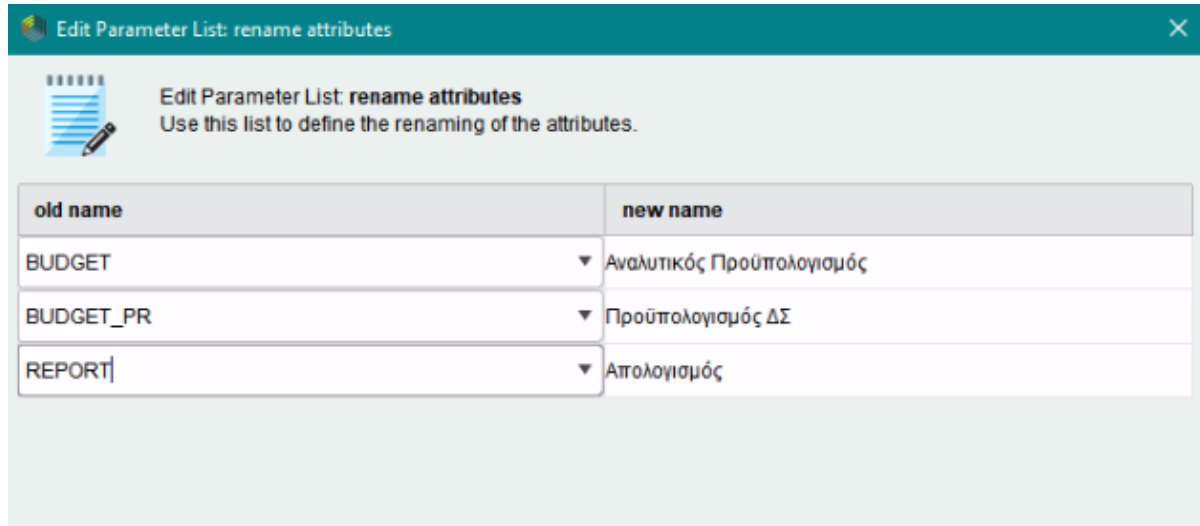
Εικόνα 4.10 Διεργασία προετοιμασίας δεδομένων

The 'Create Filters: filters' dialog box contains the following filter rules:

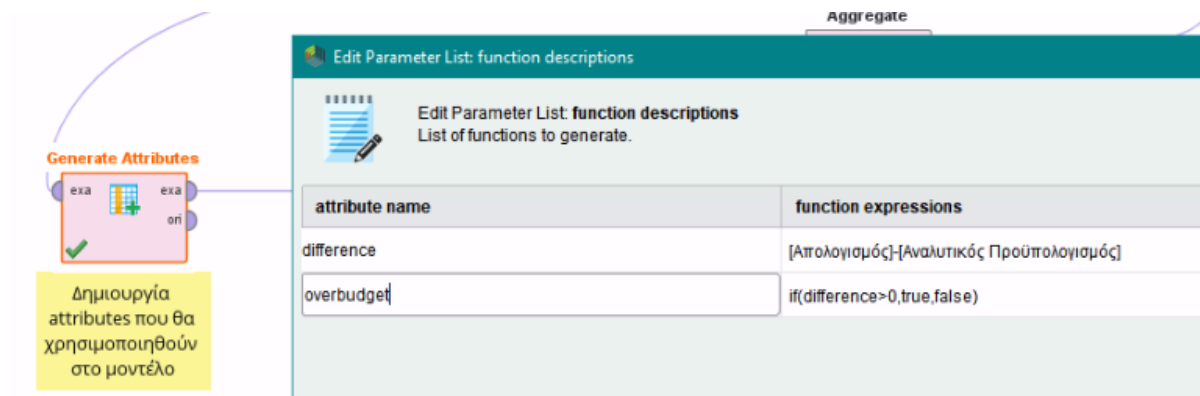
Attribute	Comparison	Value
BUDGET	is not missing	
REPORT	is not missing	
BUDGET_PR	is not missing	
REPORT	≠	0.0
DT_START	≥	01/01/2011 10:45:05 PM
REPORT	≠	0.0
BUDGET	≠	0.0
BUDGET_PR	≠	0.0
DSCR_from_ES2	is not missing	
DSCR	is not missing	

Options at the bottom: Match all, Match any, Preselect comparators. Buttons: Add Entry, OK, Cancel.

Εικόνα 4.11 Φιλτράρισμα δεδομένων



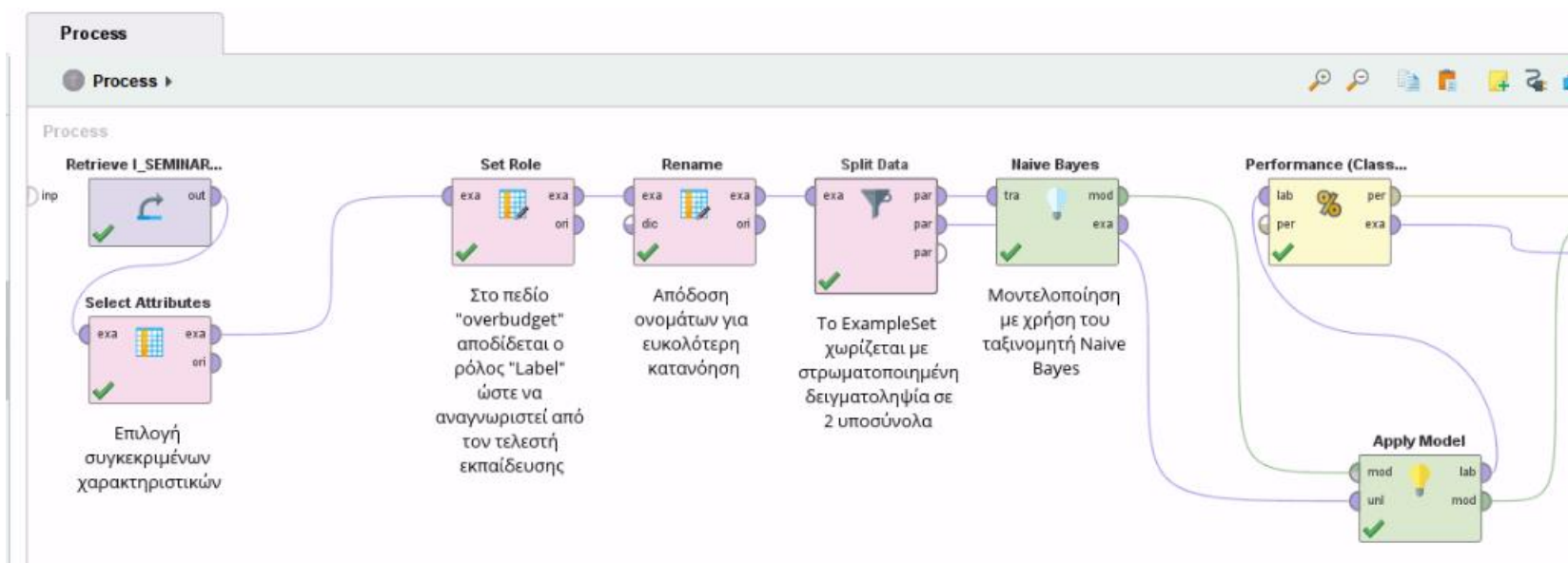
Εικόνα 4.12 Μετονομασία χαρακτηριστικών



Εικόνα 4.13 Δημιουργία χαρακτηριστικών

Modeling

Το μοντέλο επιχειρεί εφαρμόζοντας τον ταξινομητή Naïve Bayes, να προβλέψει την κατηγορία-κλάση κάθε ενός από τα στιγμιότυπα των δεδομένων δοκιμής, χρησιμοποιώντας δεδομένα εκπαίδευσης τα οποία περιλαμβάνουν ήδη την πληροφορία της κλάσης στην οποία ανήκουν. Πρόκειται δηλαδή για μια περίπτωση μάθησης με επίβλεψη, όπου τα δεδομένα δοκιμής ταξινομούνται σε 2 κατηγορίες: σε εκείνα των οποίων ο Απολογισμός προβλέπεται να ξεπεράσει τον Αναλυτικό Προϋπολογισμό (Κατηγορία `overbudget=true`) και σε εκείνα για τα οποία ο Απολογισμός είναι τελικά μικρότερος του Αναλυτικού Προϋπολογισμού (Κατηγορία `overbudget=false`). Οι τελεστές που χρησιμοποιήθηκαν περιγράφονται στην Εικόνα 4.14:



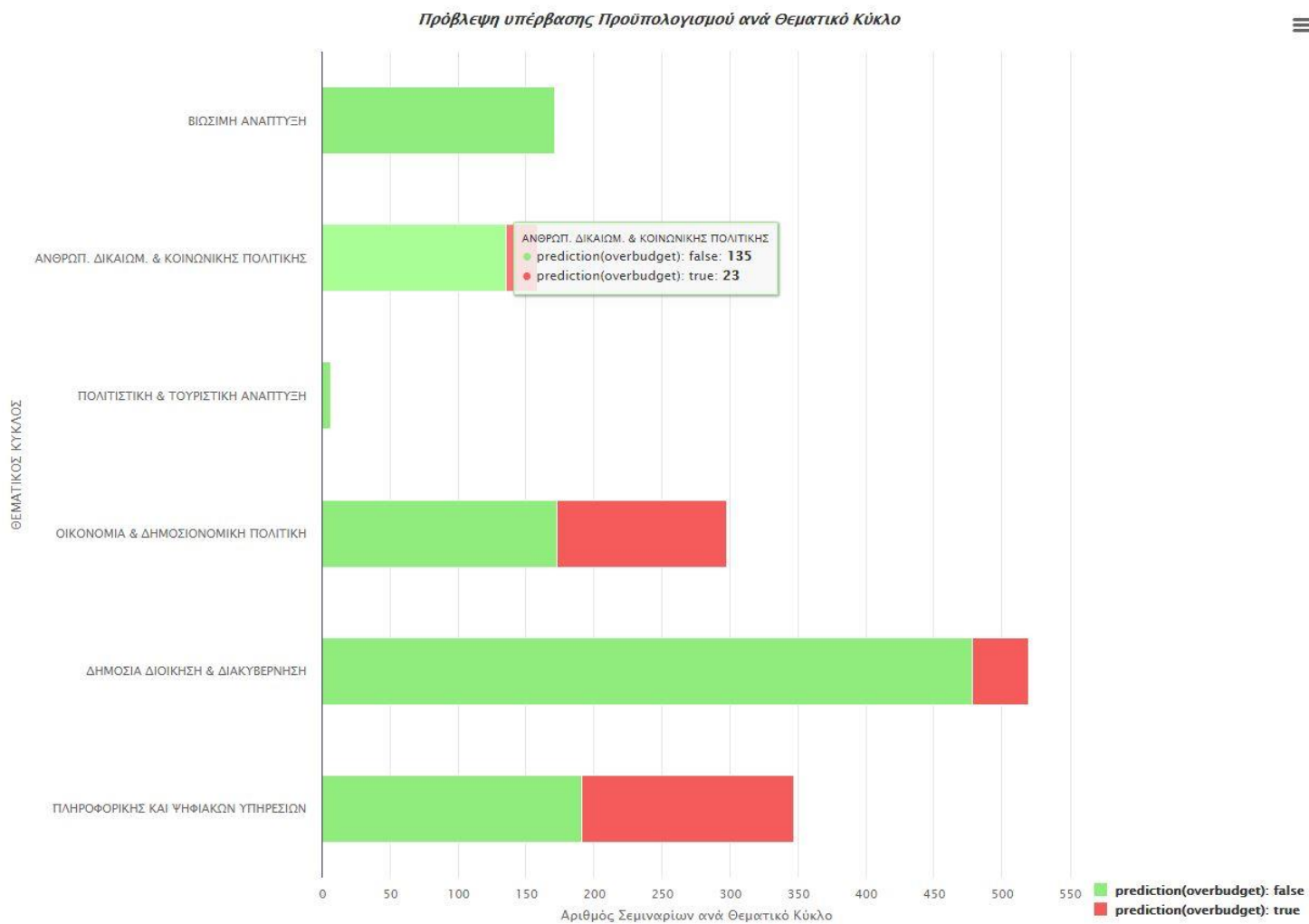
Εικόνα 4.14 Μοντέλο πρόβλεψης Naive Bayes

Evaluation

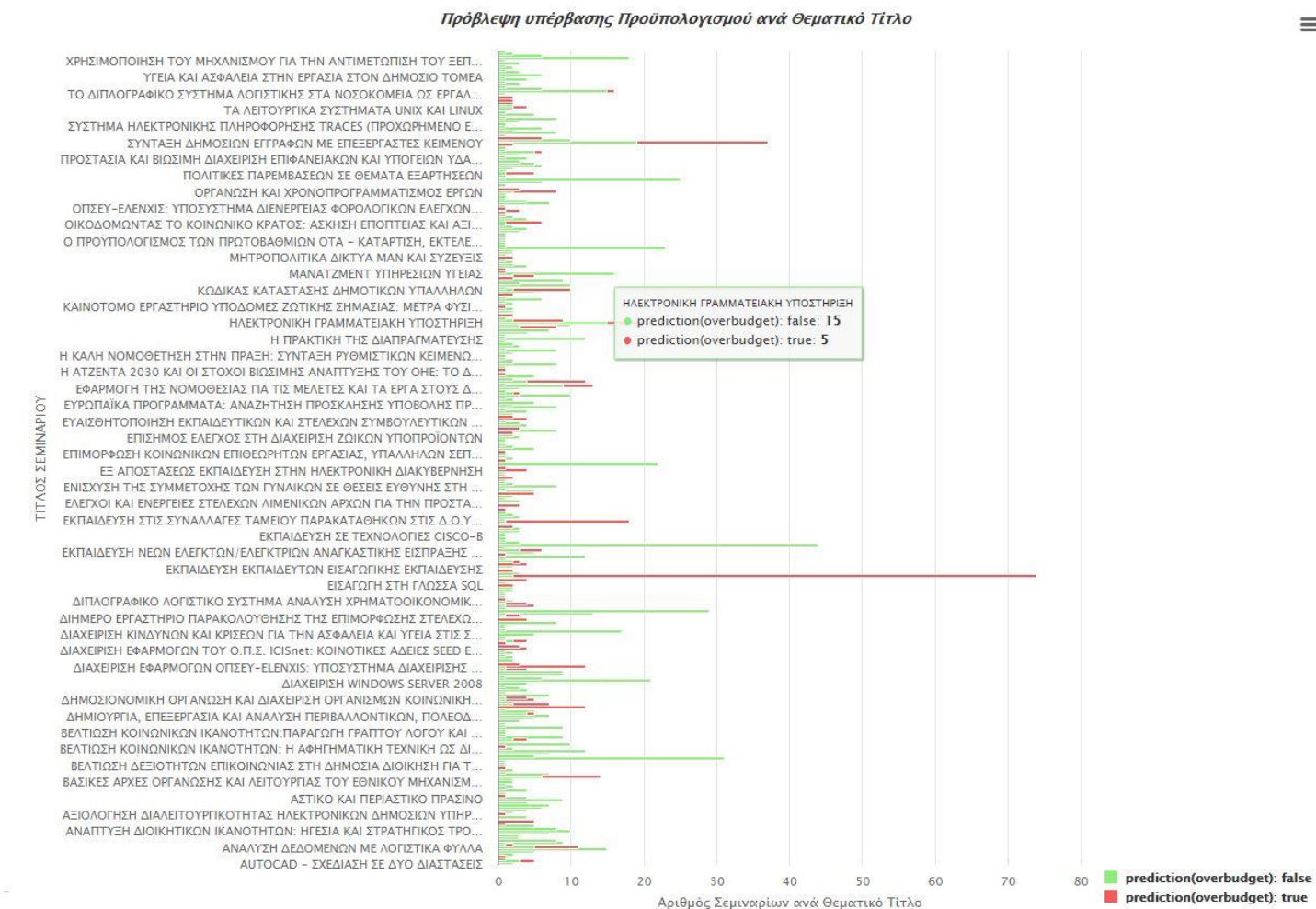
Μετά την εφαρμογή του μοντέλου,

Row No.	overbudget	prediction(o...	confidence(f...	confidence(t...	CD_SEM_TY...	ΤΙΤΛΟΣ ΣΕΜΙΝΑΡΙΟΥ	CD_POL_SE...	ΘΕΜΑΤΙΚΟΣ ΚΥΚΛΟΣ	difference
1	true	true	0.445	0.555	1	ΑΥΤΟCAD-ΣΧΕΔΙΑΣΗ ΣΕ ΔΥΟ ΔΙ...	05	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΨΗΦΙΑ...	60
2	false	false	0.603	0.397	1	ΑΥΤΟCAD-ΣΧΕΔΙΑΣΗ ΣΕ ΔΥΟ ΔΙ...	05	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΨΗΦΙΑ...	-244
3	false	false	0.529	0.471	1	ΑΥΤΟCAD-ΣΧΕΔΙΑΣΗ ΣΕ ΔΥΟ ΔΙ...	05	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΨΗΦΙΑ...	-125
4	false	false	0.501	0.499	1	ΑΥΤΟCAD-ΣΧΕΔΙΑΣΗ ΣΕ ΔΥΟ ΔΙ...	05	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΨΗΦΙΑ...	-73
5	false	false	0.952	0.048	1	ΑΥΤΟCAD-ΣΧΕΔΙΑΣΗ ΣΕ ΔΥΟ ΔΙ...	05	ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΨΗΦΙΑ...	-844
6	false	true	0.233	0.767	1	COACHING ΣΥΝΕΡΓΑΣΙΑΣ ΚΑΙ Α...	01	ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ & ΔΙΑΚΥ...	-193
7	false	false	1.000	0.000	0	ΑΓΟΡΑ	08	ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟ...	-267
8	true	true	0.077	0.923	1	ΑΔΕΙΟΔΟΤΗΣΕΙΣ ΠΡΩΤΟΒΑΘΜΙ...	08	ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟ...	20
9	false	false	1.000	0.000	1	ΑΝΑΖΗΤΗΣΗ ΠΟΡΩΝ ΣΤΟ ΤΟΜ...	11	ΠΟΛΙΤΙΣΤΙΚΗ & ΤΟΥΡΙΣΤΙΚΗ ...	-255
10	false	false	0.773	0.227	1	ΑΝΑΘΕΣΗ ΚΑΙ ΕΚΤΕΛΕΣΗ ΔΗΜ...	08	ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟ...	-5

Εικόνα 4.15 Αποτελέσματα συνόλου πρόβλεψης



Εικόνα 4.16 Πρόβλεψη υπέρβασης Προϋπολογισμού ανά Θεματικό Κύκλο



Εικόνα 4.17 Πρόβλεψη υπέρβασης Προϋπολογισμού ανά Τίτλο Σεμιναρίου

Αξιολογώντας την απόδοση του μοντέλου ως προς την ακρίβεια πρόβλεψης δειγμάτων που θα υπερβούν τον προϋπολογισμό (Εικόνα 4.18, Εικόνα 4.19), φαίνεται ότι το πλήθος δειγμάτων που εισάγεται στο μοντέλο (80% έναντι 100%) δε βοηθάει ιδιαίτερα στην ακριβέστερη πρόβλεψη:

Table View Plot View

accuracy: 85.55%

	true false	true true	class precision
pred. false	2563	236	91.57%
pred. true	226	173	43.36%
class recall	91.90%	42.30%	

Εικόνα 4.18 80% του δείγματος, ακραίες τιμές

Table View Plot View

accuracy: 85.89%

	true false	true true	class precision
pred. false	3213	302	91.41%
pred. true	262	221	45.76%
class recall	92.46%	42.26%	

Εικόνα 4.19 100% του δείγματος, ακραίες τιμές

Αντίθετα, η παράλειψη με χρήση φίλτρου, παραδειγμάτων του συνόλου δεδομένων των οποίων οι τιμές του πεδίου difference (διαφορά μεταξύ Προϋπολογισμού-Απολογισμού) θεωρήθηκαν ακραίες, φαίνεται να βελτιώνει την ακρίβεια της κλάσης true κατά έως και 8 ποσοστιαίες μονάδες:

Table View Plot View

accuracy: 87.51%

	true false	true true	class precision
pred. false	2547	183	93.30%
pred. true	209	200	48.90%
class recall	92.42%	52.22%	

Εικόνα 4.20 80% του δείγματος, χωρίς ακραίες τιμές

Table View Plot View

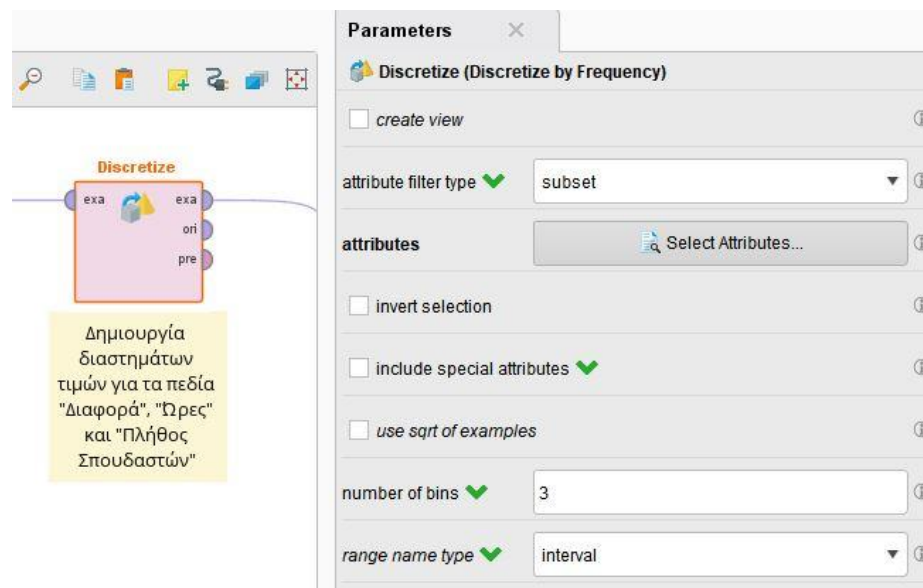
accuracy: 87.80%

	true false	true true	class precision
pred. false	3185	234	93.16%
pred. true	245	261	51.58%
class recall	92.86%	52.73%	

Εικόνα 4.21 100% του δείγματος, χωρίς ακραίες τιμές

Εφόσον το μοντέλο φαίνεται να λειτουργεί και να οδηγεί σε κάποια συμπεράσματα, θεωρείται σκόπιμο να εμπλουτιστεί περαιτέρω, και να εισαχθούν σε αυτό τα χαρακτηριστικά της διαφοράς προϋπολογισμένου και ποσού απολογισμού (difference), του πλήθους των σπουδαστών (STUDENTS) και της διάρκειας του σεμιναρίου σε ώρες (HOURS).

Η επιλογή των χαρακτηριστικών γίνεται μέσω του τελεστή “Select Attributes”. Επειδή τόσο το ύψος της διαφοράς, όσο και το πλήθος των ωρών και το πλήθος των σπουδαστών κάθε σεμιναρίου δεν είναι πάντα σταθερό, στο μοντέλο προστίθεται επίσης ο τελεστής “Discretize (by Frequency)” (Εικόνα 4.22) ώστε τα αριθμητικά αυτά χαρακτηριστικά να μετατραπούν σε 3 nominal διαστήματα τιμών (bins). Ο τελεστής ανιχνεύει τις συχνότητες των τιμών όπως αυτές εμφανίζονται για κάθε χαρακτηριστικό στο σύνολο των δειγμάτων και δημιουργεί ένα εύρος τιμών που χωρίζεται σε τόσα διαστήματα, όσα εμείς του ορίσουμε.



Εικόνα 4.22 Discretize attributes (difference, HOURS, STUDENTS)

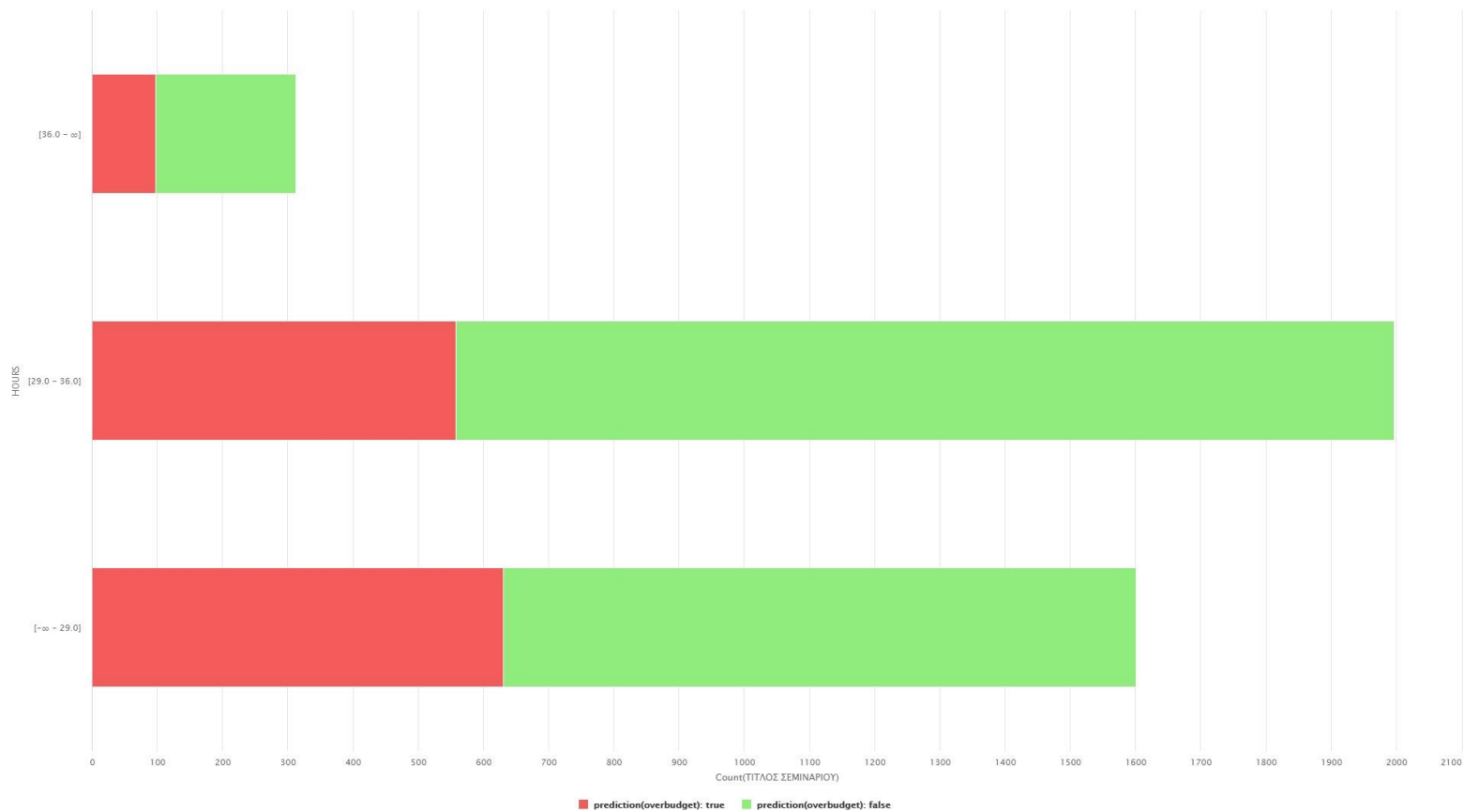
Στη συνέχεια, η πραγματική τιμή του κάθε χαρακτηριστικού αντικαθίσταται από το εύρος στο οποίο ανήκει η τιμή αυτή (Εικόνα 4.23):

Row No.	overbudget	prediction(o...	confidence(fal...	confidence(true)	ΤΙΤΛΟΣ ΣΕΜΙΝΑΡΙΟΥ	HOURS	STUDENTS	difference	CD_SEM_
1	false	true	0.264	0.736	ΜΕΡΙΚΟΙ ΕΠΙΤΟΠΙΟΙ ΕΛΕ...	$[-\infty - 29.0]$	$[24.5 - \infty]$	$[-7.5 - \infty]$	0
2	false	false	1.000	0.000	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟ...	$[-\infty - 29.0]$	$[18.5 - 24.5]$	$[-174.5 - -7.5]$	1
3	false	false	1.000	0.000	ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ Τ...	$[-\infty - 29.0]$	$[24.5 - \infty]$	$[-174.5 - -7.5]$	0
4	true	true	0.089	0.911	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟ...	$[-\infty - 29.0]$	$[-\infty - 18.5]$	$[-7.5 - \infty]$	1
5	false	false	1.000	0.000	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟ...	$[-\infty - 29.0]$	$[-\infty - 18.5]$	$[-174.5 - -7.5]$	1
6	false	false	1.000	0.000	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟ...	$[-\infty - 29.0]$	$[-\infty - 18.5]$	$[-174.5 - -7.5]$	1
7	false	false	1.000	0.000	ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟ...	$[-\infty - 29.0]$	$[18.5 - 24.5]$	$[-174.5 - -7.5]$	1
8	false	true	0.306	0.694	ΚΑΝΟΝΙΣΜΟΣ ΒΟΥΛΗΣ Κ...	$[-\infty - 29.0]$	$[18.5 - 24.5]$	$[-7.5 - \infty]$	0
9	false	true	0.426	0.574	ΠΙΛΟΤΙΚΗ ΕΠΙΜΟΡΦΩΤΙ...	$[-\infty - 29.0]$	$[24.5 - \infty]$	$[-7.5 - \infty]$	2

Εικόνα 4.23 Discretize by Frequency

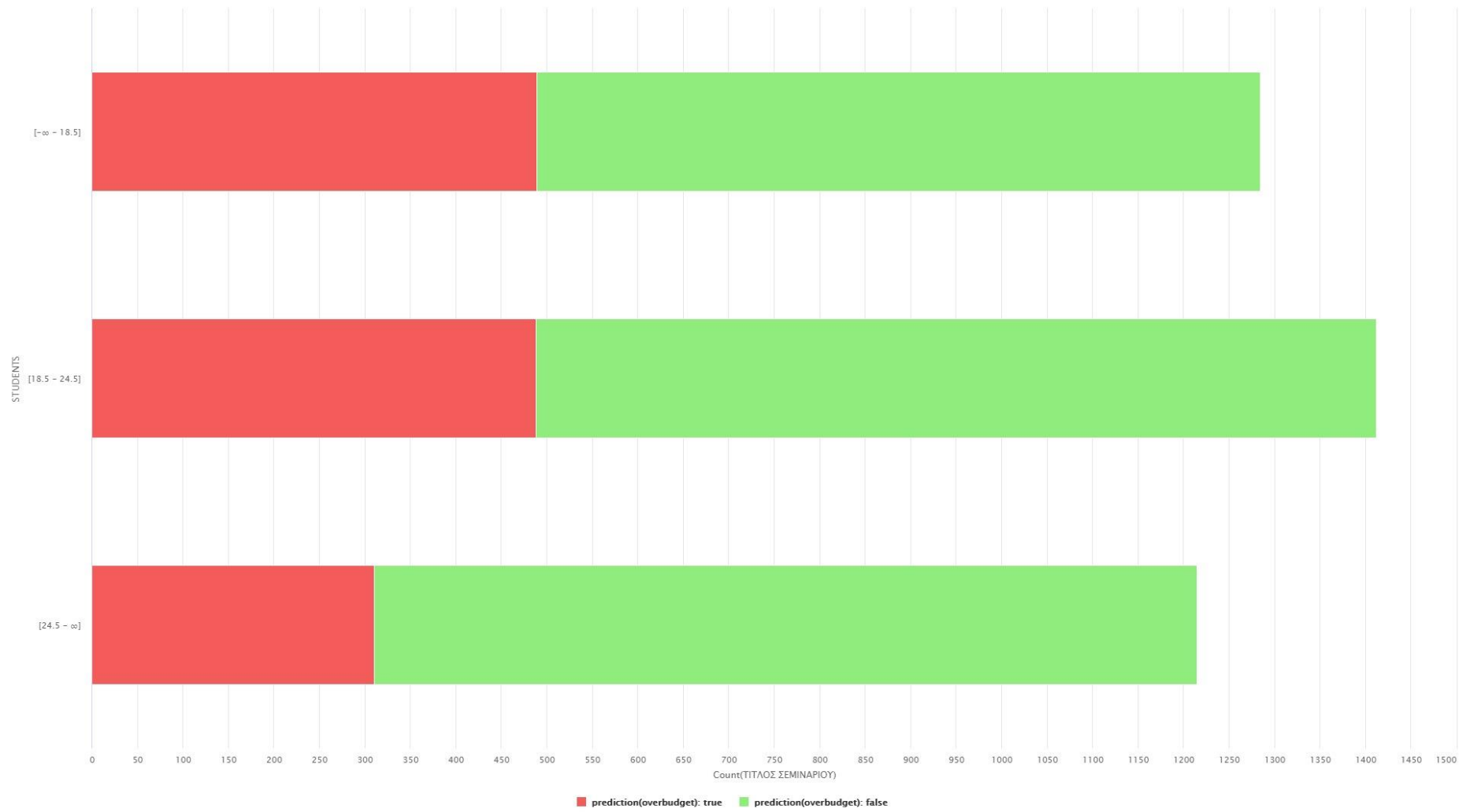
Deployment

Τρέχοντας το μοντέλο εκ νέου, μπορούμε πλέον να έχουμε μια εικόνα για το πόσα σεμινάρια προβλέπεται να υπερβούν τον προϋπολογισμό, ανάλογα με τη διάρκεια ή το πλήθος των σπουδαστών (Εικόνα 4.24, Εικόνα 4.25):



Εικόνα 4.24 Πρόβλεψη υπέρβασης με βάση τη διάρκεια σεμιναρίου

Εξόρυξη Δεδομένων και Σεμινάρια Επιμόρφωσης Φορέων του Δημόσιου Τομέα



Εικόνα 4.25 Πρόβλεψη υπέρβασης με βάση το πλήθος σπουδαστών

Η απόδοση του μοντέλου φαίνεται μάλιστα να έχει βελτιωθεί (100% του δείγματος, χωρίς ακραίες τιμές), ειδικά μάλιστα σε ό,τι αφορά στην πρόβλεψη της κλάσης «True» (Εικόνα 4.26):

Table View Plot View

accuracy: 88.21%

	true false	true true	class precision
pred. false	2624	0	100.00%
pred. true	461	826	64.18%
class recall	85.06%	100.00%	

Εικόνα 4.26 Απόδοση μοντέλου - Discretization

4.3.2 Περίπτωση Εφαρμογής 2: Ανίχνευση αναγκών με βάση τα σχόλια των εκπαιδευόμενων

Business Understanding

Όπως αναφέρθηκε και νωρίτερα, οι ενδιαφερόμενοι φορείς μπορούν μέσα από σχετικές φόρμες και διαδικασίες, να υποβάλλουν τα αιτήματά τους για επιμόρφωση επάνω σε πεδία ενδιαφέροντός τους. Ανάμεσα στα πεδία που καλούνται να συμπληρώσουν, υπάρχει και ένα πεδίο στο οποίο καλούνται να περιγράψουν, με ελεύθερο κείμενο, την σκοπιμότητα του σεμιναρίου. Να παραθέσουν δηλαδή τους λόγους για τους οποίους θεωρούν ότι οι υπάλληλοι του φορέα τους θα έπρεπε να επιμορφωθούν σε κάποιον συγκεκριμένο τίτλο σεμιναρίου του ΙΝ.ΕΠ.

Αιτήματα για Υλοποίηση Επιμ. Προγραμμάτων									
Φορέας	A/A	Κωδικός Αίτησης	Ημ/νία	Επιχ. Στόχοι	Οριστικό ποσοστό	Ημ/νία Οριστ.	Τίτλος		
β53	ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ	321	381	22/08/2015	ΑΝΑΣΧΕΔΙΑΣΜΟΣ ΔΙΟΙΚΗΤΙΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΤΟΥ ΔΡ	ΝΑΙ	22/08/2015	ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ Β ΕΞΑΜΗΝΟ 2015 ΠΡΟΣ	
β500	ΔΗΜΟΣ ΧΑΛΚΙΔΕΩΝ	342	441	07/08/2015	ΣΤΟΧΟΣ ΕΙΝΑΙ Η ΠΡΟΣΦΟΡΑ ΟΛΩΝ ΤΩΝ ΑΠΑΡΑΙΤΗΤ	ΝΑΙ	07/08/2015	ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ	
β500	ΔΗΜΟΣ ΧΑΛΚΙΔΕΩΝ	343			ΕΝΗΜΕΡΩΣΗ ΚΑΙ ΚΑΤΑΡΤΙΣΗ ΣΕ ΘΕΩΡΗΤΙΚΟ ΚΑΙ ΠΡ	ΟΧΙ		ΟΙΚΟΝΟΜΙΑ ΚΑΙ ΔΗΜΟΣΙΟΝΟΜΙΚΗ ΠΟΛΙΤΙΚΗ	
β7	ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ	381	401	02/07/2015	ΒΕΛΤΙΩΣΗ ΤΩΝ ΓΝΩΣΕΩΝ ΚΑΙ ΤΩΝ ΔΕΞΙΟΤΗΤΩΝ ΤΩ	ΝΑΙ	02/07/2015	ΑΙΤΗΜΑΤΑ ΕΠΙΜΟΡΦΩΣΗΣ ΤΩΝ ΣΤΕΛΕΧΩΝ ΤΟΥ Π	
2067	ΥΠΟΥΡΓΕΙΟ ΟΙΚΟΝΟΜΙΚΩΝ	422	461	26/08/2015	ΣΔΟΕ ΑΝΑΒΑΘΜΙΣΗ ΕΛΕΓΚΤΙΚΟΥ ΕΡΓΟΥ ΣΔΟΕ, ΣΤΗ	ΝΑΙ	26/08/2015	ΠΡΟΓΡ/ΣΜΟΣ ΕΠΙΜΟΡΦΩΣΗΣ ΥΠΑΛΛΗΛΩΝ ΤΗΣ Ε	

Εκπαιδευτικοί Τίτλοι									
Κύκλος	Υφιστ. Τίτλος	Άλλος Τίτλος	Θεσμικές Αλλαγές	Δομικές ή Οργαν. Αλλαγές	Άλλη περίπτωση	Εφαρ. ή Πε			
30	ΔΙΟΙΚΗΤΙΚΗ ΜΕΤΑΦΡΩΜΙΣΗ Κ	10429	ΑΝΑΣΧΕΔΙΑΣΜΟΣ ΔΙΟΙΚΗΤΙΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΣΤΟ ΔΗΜΟΣΙΟ ΤΟΜΕΑ				ΟΧΙ	ΟΧΙ	
	Σκοπιμότητα	Τίτλος Ομάδας-Στόχου	Περιγραφή Ομάδας-Στόχου	ΠΕ	ΤΕ	ΔΕ	Αριθμός Προγράμ		
	ΕΠΙΜΟΡΦΩΣΗ ΣΧΕΤΙΚΑ ΜΕ ΤΙΣ ΝΕΕΣ ΔΙΟΙΚΗΤΙΚΕΣ ΔΙΑΔΙΚΑΣΙΕΣ ΣΤΟ ΔΗΜΟΣΙΟ	ΥΠΟΥΡΓΕΙΟ ΟΙΚΟΝΟΜΙΚΩΝ-ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ & ΔΙΟΙΚΗΤΙΚΗΣ	ΤΜΗΜΑ Ε (ΟΡΓΑΝΩΣΗΣ & ΠΟΙΟΤΗΤΑΣ) ΤΗΣ ΔΙΝΣΗΣ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ &	3	0	0			

Εικόνα 4.27 Φόρμα Ανίχνευσης Εκπαιδευτικών Αναγκών

Θεωρήθηκε ότι παρουσίαζε ενδιαφέρον η ανάλυση, με χρήση τεχνικών εξόρυξης δεδομένων, των απαντήσεων αυτών και της πιθανότητας μέσα από αυτήν να ανακαλυφθούν κάποιες συχνά εμφανιζόμενες λέξεις ή φράσεις οι οποίες θα μπορούσαν να οδηγήσουν σε εξαγωγή συμπερασμάτων για την καλύτερη οργάνωση των σεμιναρίων, με προσαρμογή του εκπαιδευτικής ύλης κάθε προγράμματος σύμφωνα με τις συγκεκριμένες ανάγκες επιμόρφωσης των φορέων.

Data Understanding

Με τη βοήθεια των διαχειριστών της βάσης δεδομένων, εντοπίστηκε ο πίνακας με τα δεδομένα που μας ενδιέφεραν. Τα σχόλια των αιτούντων ως προς την σκοπιμότητα του σεμιναρίου που επιθυμούν να παρακολουθήσουν φυλάσσεται ως κείμενο στο πεδίο «Purpose». Ο πίνακας εξήχθη με χρήση του εργαλείου DBeaver σε μορφή .xlsx (Εικόνα 4.28):

KIN_APP_CD	CD_SECTION	SECTION_DSCR	CD_SEM_1SEM_TITLE_DSCROTHERITIT	PURPOSE
168	175			ΑΠΟΤΕΛΕΙ 1) ΚΑΛΥΤΕΡΗ ΕΦΑΡΜΟΓΗ ΤΟΥ ΕΘΝΙΚΟΥ ΚΑΙ ΚΟΙΝΟΤΙΚΟΥ ΔΙΚΑΙΟΥ, 2) ΑΠΟΔΟΤΙΚΟΤΕΡΗ ΧΡΗΣΗ ΤΩΝ ΣΧΕΤΙΚΩΝ ΜΗΧΑ/
169	176			ΑΠΟΤΕΛΕΙ ΑΠΟΚΤΗΣΗ ΙΚΑΝΟΤΗΤΩΝ ΗΓΗΣΙΑΣ, ΣΥΜΜΕΤΟΧΙΚΟΥ ΜΑΝΑΤΖΜΕΝΤ, ΔΙΑΧΕΙΡΙΣΗΣ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΚΑΙ ΥΠΟΚ/
166	173	29 ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝ	10489 ΑΝΑΠΤΥΞΗ Δ/ΝΣ	ΕΝΔΥΝΑΜΩΣΗ ΚΑΙ ΠΡΟΣΑΡΜΟΓΗ ΤΩΝ ΔΕΣΙΟΤΗΤΩΝ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΣΤΙΣ ΔΟΜΙΚΕΣ Η ΟΡΓΑΝΩΣΙΑΚΕΣ /
167	174	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10314 ΑΝΑΠΤΥΞΗ ΕΦΔ	ΚΑΛΥΤΕΡΗ ΟΡΓΑΝΩΣΗ ΤΗΣ ΥΠΗΡΕΣΙΑΣ
170	177	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10388 ΑΣΦΑΛΕΙΑ ΥΠΟ	ΕΝΔΥΝΑΜΩΣΗ ΚΑΙ ΠΡΟΣΑΡΜΟΓΗ ΤΩΝ ΔΕΣΙΟΤΗΤΩΝ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΣΤΙΣ ΑΝΑΓΚΕΣ ΤΟΥ Ε.Π. ΨΗΦΙΑΚΗ
171	178	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10317 ΒΑΣΙΚΕΣ ΓΝΩΣΕΙ	ΑΠΟΚΤΗΣΗ ΒΑΣΙΚΩΝ ΓΝΩΣΕΩΝ ΚΑΙ ΔΕΣΙΟΤΗΤΩΝ Η/Υ ΓΙΑ ΤΗΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΚΑΙ ΤΑΧΥΤΕΡΗ ΔΙΑΧΕΙΡΑΙΟΣΗ ΕΙ
172	179	29 ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝ	10676 ΒΕΛΤΙΩΣΗ ΓΛΩΣΣ	1.ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΑΝΑΠΤΥΞΗ ΑΚΤΟΠΛΟΙΚΩΝ ΣΥΓΚΟΙΝΩΝΙΩΝ ΚΑΙ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ 2.ΕΥΡΥΘΜΗ, ΣΥΝΤΟ
214	234	30 ΔΙΟΙΚΗΤΙΚΗ ΜΕΤΑΡΡΥΘΜΙΣΗ ΚΑΙ ΔΓ	10632 ΕΞ ΑΠΟΣΤΑΣΕΩΣ	ΑΥΞΗΣΗ ΠΑΡΑΓΩΓΙΚΟΤΗΤΑΣ - ΕΠΙΤΕΥΞΗ ΚΑΛΥΤΕΡΩΝ ΑΠΟΤΕΛΕΣΜΑΤΩΝ ΚΑΙ ΒΕΛΤΙΩΣΗ ΟΡΓΑΝΩΣΗΣ 4
218	238	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10378 ΕΚΠΑΙΔΕΥΣΗ ΣΕ	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΚΑΤΑΝΟΗΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΛΕΙΤΟΥΡΓΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΔΙΚΤΥΩΝ ΔΕΔΟΜΕΝΩΝ Η/Υ - Β
221	242	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10473 ΕΚΠΑΙΔΕΥΣΗ ΣΕ	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΚΑΤΑΝΟΗΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΛΕΙΤΟΥΡΓΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ ΚΑΙ ΔΙΚΤΥΩΝ ΔΕΔΟΜΕΝΩΝ Η/Υ - Β
223	246	28 ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟΜΙΚΗ Γ	10732 ΕΚΠΑΙΔΕΥΣΗ ΣΤΙ	ΚΑΘΥΛΗΝ ΑΡΜΟΔΙΟΙ ΓΙΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΕΣΗΔΗΣ (Ν.4281/20111 ΚΑΙ ΠΔ 103/2014) 4
226	248	29 ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝ	10331 Η ΠΡΑΚΤΙΚΗ ΤΗΣ	Η ΒΕΛΤΙΩΣΗ ΤΩΝ ΓΝΩΣΕΩΝ ΚΑΙ ΤΩΝ ΔΕΣΙΟΤΗΤΩΝ ΤΩΝ ΣΥΜΜΕΤΕΧΟΝΤΩΝ ΣΕ ΘΕΜΑΤΑ ΔΙΑΠΡΑΓΜΑΤΕΥΣΗΣ, ΩΣΤΕ ΝΑ Ε
227	249	31 ΚΟΙΝΩΝΙΚΗ ΠΟΛΙΤΙΚΗ	10467 ΔΗΜΟΣΙΑ ΔΙΟΗ	ΝΟΣΗΛΕΙΑ, ΗΔΙΚΑ, ΚΑΤΑΣΚΗΝΩΣΕΙΣ, ΠΑΡΟΧΟΙ ΥΓΕΙΑΣ, ΕΞΥΠΗΡΕΤΗΣΗ ΠΟΛΙΤΩΝ, ΑΣΦΑΛΙΣΗ ΤΟΥ ΚΛΑΔΟΥ ΥΓΕΙΑΣ, ΣΥΝΤ
228	250	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10682 ΕΞ ΑΠΟΣΤΑΣΕΩΣ	ΔΗΜΙΟΥΡΓΙΑ ΚΑΙ ΧΡΗΣΗ ΠΛΑΤΦΟΡΜΩΝ ΕΞ ΑΠΟΣΤΑΣΕΩΣ ΕΚΠΑΙΔΕΥΣΗΣ, ΕΞΟΙΚΟΝΟΜΗΣΗ ΠΟΡΩΝ - ΕΜΠΛΟΥΤΙΣΜΟΣ 1
230	251	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ	10681 ΕΞ ΑΠΟΣΤΑΣΕΩΣ	ΔΗΜΙΟΥΡΓΙΑ ΚΑΙ ΧΡΗΣΗ ΠΛΑΤΦΟΡΜΩΝ ΕΞ ΑΠΟΣΤΑΣΕΩΣ ΕΚΠΑΙΔΕΥΣΗΣ, ΕΞΟΙΚΟΝΟΜΗΣΗ ΠΟΡΩΝ - ΕΜΠΛΟΥΤΙΣΜΟΣ 1
229	252	29 ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝ	10489 ΑΝΑΠΤΥΞΗ Δ/ΝΣ	ΚΑΤΑΛΛΗΛΟΤΕΡΗ ΔΙΑΧΕΙΡΙΣΗ ΤΟΥ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΜΕ ΑΠΟΤΕΛΕΣΜΑ ΤΗΝ ΑΥΞΗΣΗ ΤΟΥ ΠΑΡΑΓΩΓΜΕΝΟΥ ΕΡΓ
231	253	28 ΟΙΚΟΝΟΜΙΑ & ΔΗΜΟΣΙΟΝΟΜΙΚΗ Γ	10428 ΟΙΚΟΝΟΜΙΚΟΣ	4 ΒΕΛΤΙΩΣΗ ΕΠΟΠΤΙΚΩΝ & ΕΛΕΓΚΤΙΚΩΝ ΔΙΑΔΙΚΑΣΙΩΝ 4
232	254	29 ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝ		ΚΑΤΑΡΤΙΣΗ ΒΕΛΤΙΩΣΗ ΜΗΧΑΝΙΣΜΩΝ ΚΑΤΑΠΟΛΕΜΗΣΗΣ ΦΑΙΝΟΜΕΝΩΝ ΔΙΑΦΘΟΡΑΣ
233	255	29 ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΗΤΑ & ΔΙΑΦΑΝ	10363 ΕΦΑΡΜΟΓΗ ΤΗΣ	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΕΚΔΟΣΗ ΠΕΡΙΦΕΡΕΙΑΚΩΝ ΑΠΟΦΑΣΕΩΝ
234	256	27 ΒΙΩΣΙΜΗ ΑΝΑΠΤΥΞΗ	10354 ΘΕΣΜΙΚΟ ΠΛΑΙ	ΒΕΛΤΙΩΣΗ ΣΤΗΝ ΑΝΤΙΜΕΤΩΠΙΣΗ ΘΕΜΑΤΩΝ ΕΝΕΡΓΕΙΑΚΗΣ ΑΠΟΔΟΤΙΚΟΤΗΤΑΣ ΤΩΝ ΚΤΙΡΙΑΚΩΝ ΕΓΚΑΤΑΣΤΑΣΕΩΝ ΤΟΥ ΤΑ
235	257	27 ΒΙΩΣΙΜΗ ΑΝΑΠΤΥΞΗ		ΟΡΓΑΝΩΣΗ ΒΕΛΤΙΩΣΗ ΣΤΗΝ ΕΦΑΡΜΟΓΗ ΔΙΑΔΙΚΑΣΙΩΝ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΣΥΜΒΑΣΕΩΝ ΤΕΧ. ΕΡΓΩΝ
236	258	27 ΒΙΩΣΙΜΗ ΑΝΑΠΤΥΞΗ	10398 ΕΦΑΡΜΟΓΗ ΤΟΥ	ΒΕΛΤΙΩΣΗ ΤΗΣ ΕΝΕΡΓΕΙΑΚΗΣ ΑΠΟΔΟΣΗΣ ΤΩΝ ΑΚΙΝΗΤΩΝ ΤΟΥ ΤΑΜΕΙΟΥ
237	259	27 ΒΙΩΣΙΜΗ ΑΝΑΠΤΥΞΗ	10355 ΚΩΔΙΚΟΠΟΙΗΣΗ	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΕΦΑΡΜΟΓΗ ΤΩΝ ΔΙΑΔΙΚΑΣΙΩΝ ΣΥΝΤΑΞΗΣ & ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΣΥΜΒΑΣΕΩΝ
142	122	26 ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ & ΕΦ		ΑΝΟΙΚΤΕΣ ΑΞΙΟΠΟΙΗΣΗ ΑΝΟΙΚΤΩΝ ΒΑΣΕΩΝ ΔΕΔΟΜΕΝΩΝ ΣΤΟ ΣΧΕΔΙΑΣΜΟ ΟΡΙΖΟΝΤΙΩΝ ΕΛΕΓΚΤΙΚΩΝ ΔΡΑΣΕΩΝ

Εικόνα 4.28 Ο πίνακας με το κείμενο προς ανάλυση

Data preparation

Ο Πίνακας περιέχει 18 χαρακτηριστικά, εκ των οποίων ενδιαφέρον για την περίπτωση που εξετάζουμε παρουσιάζουν ουσιαστικά 2 από αυτά: ο Τίτλος του σεμιναρίου (SEM_TITLE_DSCR) και το σχόλιο που έχει υποβληθεί από τον αιτούντα για το συγκεκριμένο σεμινάριο (PURPOSE) σχετικά με τη σκοπιμότητα παρακολούθησής του (Εικόνα 4.29):

SECTION_D...	CD_SEM_TL...	SEM_TITLE_DSCR ↑	PURPOSE
ΔΗΜΟΣΙΑ Δ...	11142	ΔΗΜΟΣΙΟΓΡΑΦΙΚΟΣ ΚΩΔΙΚΑΣ	ΕΠΙΚΑΙΡΟΠΟΙΗΣΗ ΓΝΩΣΕΩΝ
ΔΗΜΟΣΙΑ Δ...	11142	ΔΗΜΟΣΙΟΓΡΑΦΙΚΟΣ ΚΩΔΙΚΑΣ	ΕΝΗΜΕΡΩΣΗ ΓΙΑ ΤΙΣ ΝΕΟΤΕΡΕΣ ΑΛΛΑΓΕΣ
ΟΙΚΟΝΟΜΙΑ ...	10921	ΔΙΑΔΙΚΑΣΙΕΣ ΕΙΣΠΡΑΞΗΣ ΛΗΣΙΠΡΟΘΕΣΜΩΝ ΟΦΕΙΛΩΝ ΚΑΙ ΣΤ...	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΚΑΙ ΑΞΙΟΠΙΣΤΗ ΦΟΡΟΛΟΓΙΚΗ ΚΑΙ ΤΕΛΩΝΕΙΑΚΗ ΔΙΟΙΚΗΣΗ 4ΒΕΛΤΙΩΣΗ ΤΩΝ ΠΑΡΕΧΟΜ...
ΟΙΚΟΝΟΜΙΑ ...	10921	ΔΙΑΔΙΚΑΣΙΕΣ ΕΙΣΠΡΑΞΗΣ ΛΗΣΙΠΡΟΘΕΣΜΩΝ ΟΦΕΙΛΩΝ ΚΑΙ ΣΤ...	ΣΚΟΠΟΣ ΤΟΥ ΕΠΙΜΟΡΦΩΤΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ ΕΙΝΑΙ ΟΙ ΕΚΠΑΙΔΕΥΟΜΕΝΟΙ/ΝΕΣ ΝΑ ΑΠΟΚΤΗΣΟΥΝ ΤΙΣ ΑΠΑΡ...
ΟΙΚΟΝΟΜΙΑ ...	10921	ΔΙΑΔΙΚΑΣΙΕΣ ΕΙΣΠΡΑΞΗΣ ΛΗΣΙΠΡΟΘΕΣΜΩΝ ΟΦΕΙΛΩΝ ΚΑΙ ΣΤ...	ΣΚΟΠΟΣ ΤΟΥ ΕΠΙΜΟΡΦΩΤΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ ΕΙΝΑΙ ΟΙ ΕΚΠΑΙΔΕΥΟΜΕΝΟΙ/ΝΕΣ ΝΑ ΑΠΟΚΤΗΣΟΥΝ ΤΙΣ ΑΠΑΡ...
ΑΠΟΤΕΛΕΣΜ...	10465	ΔΙΑΜΟΡΦΩΣΗ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΣΧΕΔΙΩΝ	ΑΝΤΑΠΟΚΡΙΣΗ ΣΕ ΘΕΣΜΙΚΕΣ ΑΛΛΑΓΕΣ 4
ΑΠΟΤΕΛΕΣΜ...	10465	ΔΙΑΜΟΡΦΩΣΗ ΕΠΙΧΕΙΡΗΣΙΑΚΩΝ ΣΧΕΔΙΩΝ	ΑΠΟΚΤΗΣΗ ΓΝΩΣΕΩΝ ΚΑΙ ΔΕΣΙΟΤΗΤΩΝ ΣΤΙΣ ΒΑΣΙΚΕΣ ΕΝΝΟΙΕΣ ΚΑΙ ΔΙΑΔΙΚΑΣΙΕΣ ΣΥΝΘΕΣΗΣ ΕΝΟΣ ΕΠΙΧΕΙΡΗΣΙ...

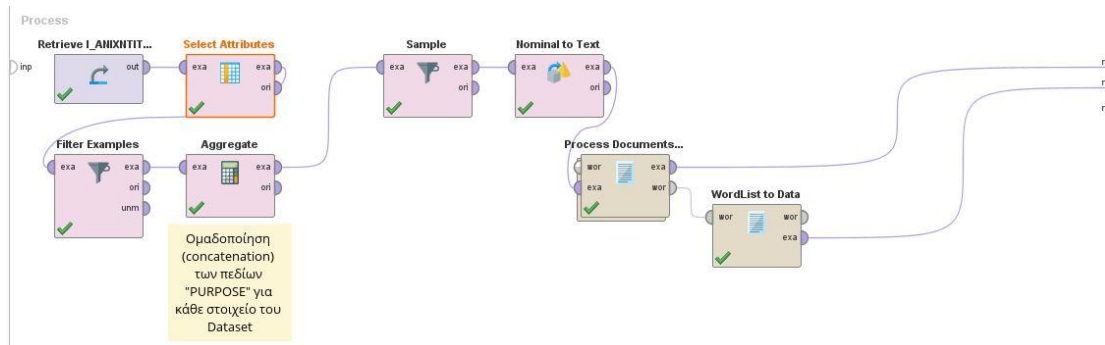
Εικόνα 4.29 Εισαγωγή του πίνακα στο Rapidminer ως dataset

Κάποια ακόμα χαρακτηριστικά συμπεριλήφθηκαν στο dataset για πιθανή αξιοποίησή τους μελλοντικά, ενώ όσα εμφανώς δεν παρουσίαζαν κανένα ενδιαφέρον, παραλείφθηκαν από την αρχή για μείωση του υπολογιστικού κόστους.

Modeling

Το μοντέλο (Εικόνα 4.30) δημιουργήθηκε με τον εξής σκοπό: τα σχόλια των αιτούντων (στήλη PURPOSE) να ομαδοποιηθούν ως ένα ενιαίο string, κατά Τίτλο σεμιναρίου (SEM_TITLE_DESCR). Στη συνέχεια, κάθε ένα από αυτά τα strings να αναλυθεί με χρήση των σχετικών τελεστών για ανάλυση κειμένου που παρέχει το εργαλείο, ώστε να ανιχνευθούν μέσα σε αυτά συχνά εμφανιζόμενες λέξεις ή φράσεις οι οποίες, συσχετιζόμενες με τον τίτλο του Σεμιναρίου που αφορούν, θα μπορούσαν να οδηγήσουν σε εξαγωγή συμπερασμάτων.

Το dataset που έχει προηγουμένως δημιουργηθεί εισάγεται στο μοντέλο και επιλέγονται τα χαρακτηριστικά που μας ενδιαφέρουν αρχικά (Τίτλος σεμιναρίου και σχόλιο). Αποκλείονται τα missing values (Filter Examples) και στη συνέχεια ο τελεστής Aggregate συνενώνει τις τιμές του πεδίου "PURPOSE" για όλα τα σεμινάρια με τον ίδιο τίτλο (Εικόνα 4.31).

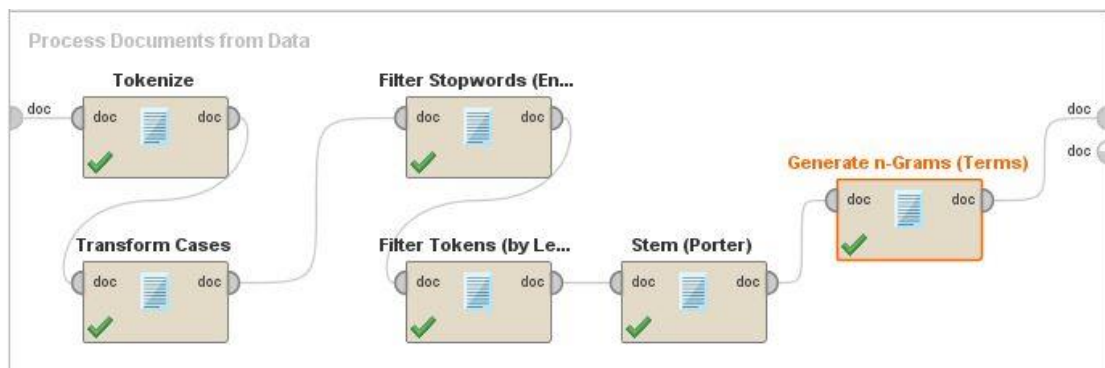


Εικόνα 4.30 Μοντέλο text mining

SEM_TITLE_DSCR	concat(PURPOSE)
ΒΑΣΙΚΕΣ ΓΝΩΣΕΙΣ ΣΤΟΥΣ ΗΛΕΚΤΡΟΝΙΚΟΥΣ ΥΠΟΛ...	ΑΝΑΠΤΥΞΗ ΓΝΩΣΕΩΝ, ΚΑΛΩΠΕΡΓΕΙΑ ΔΕΞΙΟΤΗΤΩΝ ΣΕ ΒΑΣΙΚΕΣ ΓΝΩΣΕΙΣ ΗΥ, ΕΞ ΟΙΚΕΙΩΣΗ ΤΩΝ ΥΠΑΛΛΗΛΩΝ ΜΕ ΤΙΣ ΤΠΕ/ΒΕΛΤΙΩΣΗ ΤΩΝ ΠΑΡΕΧΟΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ ΠΡΟΣ ΤΟΥΣ ΠΟΛΙΤΕΣ ΤΗΣ
ΒΑΣΙΚΕΣ ΨΗΦΙΑΚΕΣ ΓΝΩΣΕΙΣ ΚΑΙ ΔΕΞΙΟΤΗΤΕΣ	ΕΚΠΑΙΔΕΥΣΗ ΣΕ ΝΕΕΣ ΤΕΧΝΟΛΟΓΙΕΣ/ΒΕΛΤΙΩΣΗ ΤΩΝ ΒΑΣΙΚΩΝ ΓΝΩΣΕΩΝ ΚΑΙ ΔΕΞΙΟΤΗΤΩΝ ΠΟΥ ΕΙΝΑΙ ΑΠΑΡΑΙΤΗΤΕΣ ΓΙΑ ΟΛΕΣ ΤΙΣ ΔΡΑΣΤΗΡΙΟΤΗΤΕΣ ΤΟΥ ΥΠΟΥΡΓΕΙΟΥ ΜΑΣ /ΑΠΟΚΤΗΣΗ ΓΝΩΣΙ
ΒΑΣΙΚΗ ΚΑΤΑΡΤΙΣΗ ΣΤΑ ΟΙΚΟΝΟΜΙΚΑ ΤΕΛΩΝΕΙΑΚ...	ΤΑ ΟΙΚΟΝΟΜΙΚΑ ΤΕΛΩΝΕΙΑΚΑ ΟΙΚΟΝΟΜΙΚΑ ΚΑΘΕΣΤΩΤΑ ΚΑΙ ΑΤΕΛΕΙΕΣ ΑΦΟΡΟΥΝ ΤΗΝ ΔΙΑΚΙΝΗΣΗ ΚΑΙ ΠΑΡΑΔΟΣΗ ΕΜΠΟΡΕΥΜΑΤΩΝ ΜΕ ΑΠΛΩΑΓΗ ΑΠΟ ΔΑΣΜΟΥΣ Φ.Π.Α., ΤΕΛΗ Η ΑΛΛΕΣ ΕΠΙΒΑ
ΒΑΣΙΚΗ ΚΑΤΑΡΤΙΣΗ ΣΤΟ ΤΕΛΩΝΕΙΑΚΟ ΚΑΘΕΣΤΩΣ ...	ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΚΑΙ ΑΞΙΟΠΙΣΤΗ ΦΟΡΟΛΟΓΙΚΗ ΚΑΙ ΤΕΛΩΝΕΙΑΚΗ ΔΙΟΙΚΗΣΗ
ΒΑΣΙΚΗ ΚΑΤΑΡΤΙΣΗ ΣΤΟ ΦΟΡΟ ΠΡΟΣΤΙΘΕΜΕΝΗΣ ...	ΦΟΡΟΛΟΓΙΚΟΣ ΕΛΕΓΧΟΣ /ΜΕΓΙΣΤΟΠΟΙΗΣΗ ΔΗΜΟΣΙΩΝ ΕΣΟΔΩΝ/ΑΠΟΤΕΛΕΣΜΑΤΙΚΟΤΕΡΗ ΑΝΤΑΠΟΚΡΙΣΗ ΣΕ ΑΝΑΓΚΕΣ ΤΜΗΜΑΤΟΣ/ΒΕΛΤΙΩΣΗ ΤΩΝ ΠΑΡΕΧΟΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ /ΑΠΟΤΕΛΕΣ
ΒΕΛΤΙΩΣΗ ΓΛΩΣΣΙΚΩΝ ΔΕΞΙΟΤΗΤΩΝ ΚΑΙ ΕΝΙΣΧΥΣΧ...	ΔΙΕΘΝΟΠΟΙΗΣΗ/ΜΡΥΜΑΤΟΣ/ΚΑΥΤΕΡΗ ΕΠΙΚΟΙΝΩΝΙΑ/ΒΕΛΤΙΩΣΗ ΕΠΙΚΟΙΝΩΝΙΑΣ/ΒΕΛΤΙΩΣΗ ΓΛΩΣΣΙΚΩΝ ΔΕΞΙΟΤΗΤΩΝ ΣΤΗΝ ΑΓΓΛΙΚΗ ΓΛΩΣΣΑ/ΒΕΛΤΙΩΣΗ ΔΕΞΙΟΤΗΤΩΝ ΣΤΕΛΕΧΩΝ/ΑΠΟΤΕΛΕΣΜ

Εικόνα 4.31 Ομαδοποίηση όλων των σχολίων ανά τίτλο Σεμιναρίου

Η διεργασία Process Documents from Data (Εικόνα 4.32) δημιουργεί διανύσματα λέξεων, με τη συχνότητα που αυτά εμφανίζονται σε κάθε δείγμα του dataset (σε κάθε σεμινάριο) (Εικόνα 4.33):



Εικόνα 4.32 Διεργασία επεξεργασία των documents

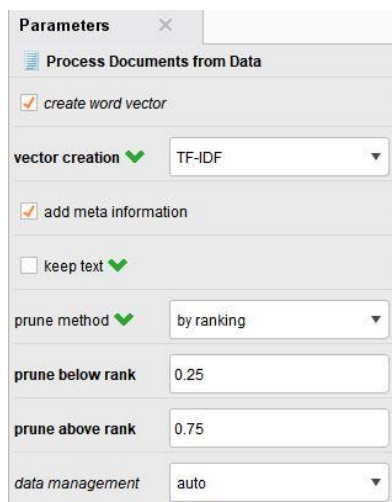
SEM_TITLE_DSCR ↑	ACCESS_ΠΛ...	ACCESS	ACCESS_ΠΛΑΙΣΙΟ...	ACCESS_ΠΛΑΙΣΙΟ_ΑΝΑΔΙΟΡΓΑΝΩΣΗΣ_ΥΠΗΡΕΣΙΩΝ
ΑΝΑΠΤΥΞΗ ΔΙΟΙΚΗΤΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΕΠΙΛΥΣΗ ΠΡΟΒΛΗΜΑΤΟΣ (PROBLEM SOLVING) ΚΑΙ...	0	0	0	0
ΑΝΑΠΤΥΞΗ ΔΙΟΙΚΗΤΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΗΓΕΣΙΑ ΚΑΙ ΣΤΡΑΤΗΓΙΚΟΣ ΤΡΟΠΟΣ ΣΚΕΨΗΣ- ΟΡΓΑ...	0	0	0	0
ΑΝΑΠΤΥΞΗ ΕΦΑΡΜΟΓΩΝ ΜΕ ΣΧΕΣΙΑΚΕΣ ΒΑΣΕΙΣ ΔΕΔΟΜΕΝΩΝ	0.056	0.056	0.056	0.056

Εικόνα 4.33 Διάνυσμα λέξεων (tokens)

Evaluation

Κατά τις διαδοχικές εκτελέσεις του μοντέλου έγιναν διάφορες δοκιμές ως προς τις παραμέτρους κυρίως 2 τελεστών:

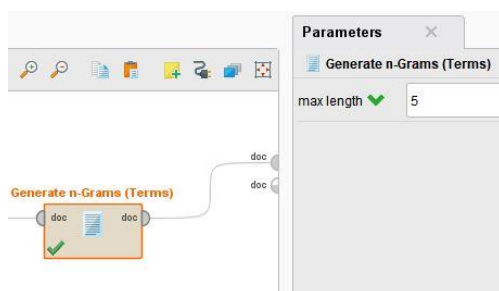
- του τελεστή Process Documents from Data:



Εικόνα 4.34 Ο τελεστής Process Documents from Data

Επιλέχθηκε να αποκλειστούν με τη μέθοδο του «κλαδέματος» (pruning) tokens (λέξεις) που εμφανίζονταν με συχνότητα μεγαλύτερη του 75% (άρα εξαιρετικά κοινές για όλα τα documents) και μικρότερη του 25% (άρα αρκετά σπάνιες)

- του τελεστή Generate n-Grams:



Εικόνα 4.35 Ο τελεστής n-Grams

Ο τελεστής παράγει ακολουθίες συνεχόμενων tokens που υπάρχουν σε ένα Document μέγιστου μήκους n . Για παράδειγμα, η ακολουθία «έκδοση_ηλεκτρονικού_εγγράφου» είναι ένα n-Gram μήκους 3. Επιλέχθηκε η τιμή 5 για τον συγκεκριμένο τελεστή, ως μία «χρυσή τομή» μεταξύ χρησίων αποτελεσμάτων και χαμηλού υπολογιστικού κόστους.

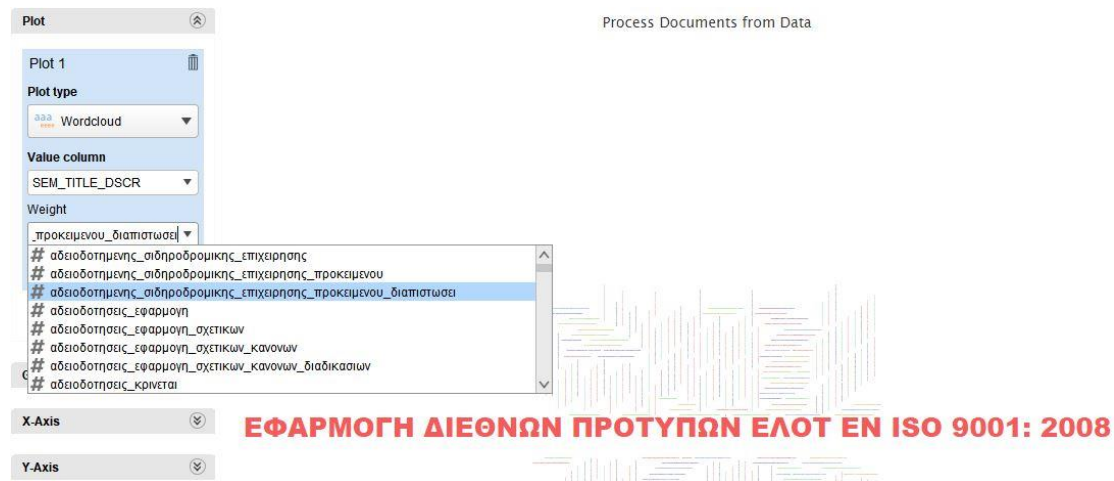
Deployment

Μετά την εκτέλεση του μοντέλου στην τελική του μορφή, μπορούμε αναλύοντας το διάγραμμα που προκύπτει, να έχουμε μια εικόνα σχετικά με τα πιο συχνά εμφανιζόμενα σχόλια και το σεμινάριο στο οποίο αντιστοιχούν. Για παράδειγμα, (Εικόνα 4.36) Εικόνα 4.36, για το Σεμινάριο με τίτλο «Διαχείριση Εκτάκτων Αναγκών», οι λέξεις που συναντώνται πιο συχνά στα σχόλια των αιτούντων είναι «έκτακτης_ανάγκης_πυρκαγιά_σεισμού_εκκένωσης»:



Εικόνα 4.36 Συχνά εμφανιζόμενα tokens ανά Σεμινάριο (α)

Αντίστοιχα, στο Σεμινάριο με τίτλο «Εφαρμογή Διεθνών Προτύπων ΕΛΟΤ EN ISO 9001:2008», συναντάται πολύ συχνά η ακολουθία λέξεων «αδειοδοτημένης_σιδηροδρομικής_επιχείρησης_προκειμένου_διαπιστώσει» (Εικόνα 4.37).



Εικόνα 4.37 Συχνά εμφανιζόμενα tokens ανά Σεμινάριο (β)

Η πληροφορία αυτή θα μπορούσε να χρησιμοποιηθεί από τους υπεύθυνους σχεδιασμού των σεμιναρίων, ώστε η ύλη του εκάστοτε σεμιναρίου να προσαρμόζεται ανάλογα με γνώμονα τις προσδοκίες των εκπαιδευόμενων, όπως αυτές αναδεικνύονται μέσα από τα σχόλιά τους.

4.3.3 Περίπτωση Εφαρμογής 3: Δημιουργία Κανόνων Συσχέτισης με βάση το ιστορικό αιτήσεων εκπαιδευόμενου

Business Understanding

Μια άλλη περίπτωση που θεωρήθηκε σκόπιμο να εξεταστεί, ήταν η πιθανότητα εξαγωγής συμπερασμάτων μέσα από την ανάλυση των δεδομένων των αιτήσεων για παρακολούθηση σεμιναρίων. Η δημιουργία δηλαδή ενός μοντέλου το οποίο θα δείχνει τη συσχέτιση όσον αφορά τους θεματικούς τίτλους σεμιναρίων, με βάση το ιστορικό των αιτήσεων που έχει κάνει κάθε ενδιαφερόμενος. Ή αλλιώς, ποια άλλα (κοινά) σεμινάρια έχουν παρακολουθήσει εκπαιδευόμενοι που παρακολούθησαν ένα συγκεκριμένο σεμινάριο. Για παράδειγμα, θα μπορούσε να αναδειχθεί, με σχετική βεβαιότητα, ότι εκπαιδευόμενοι που έκαναν αίτηση να παρακολουθήσουν το σεμινάριο «Α», έκαναν επίσης αίτηση να παρακολουθήσουν το σεμινάριο «Β». Με βάση την ανάλυση και ερμηνεία αυτών των, πιθανών, συσχετίσεων, θα μπορούσε να γίνει ακριβέστερη επιλογή της ομάδας-στόχου, για το επόμενο προγραμματισμένο σεμινάριο «Α».

Data Understanding

Καταρχάς εντοπίστηκε ο πίνακας της βάσης δεδομένων ο οποίος περιείχε καταχωρισμένες τις αιτήσεις των εκπαιδευόμενων. Στον πίνακα αυτό, υπήρχε καταγεγραμμένο το πλήρες ιστορικό των αιτήσεων που έχει κάνει ο κάθε εκπαιδευόμενος, για οποιοδήποτε Τίτλο Σεμιναρίου, οποιουδήποτε Θεματικού Κύκλου, από το 2010 και μετά. Τα κύρια πεδία ενδιαφέροντος ήταν ο μοναδικός κωδικός εκπαιδευόμενου (CD), ο μοναδικός κωδικός σεμιναρίου (CD_SEM) για το οποίο έχει κάνει αίτηση, καθώς και ο μοναδικός κωδικός του Τίτλου του Σεμιναρίου (CD_SEM_THEME).

Data preparation

Τα περιεχόμενα του πίνακα εξήχθησαν αρχικά από τη βάση δεδομένων με χρήση του εργαλείου DBeaver και αποθηκεύτηκαν σε μορφή .csv. Στη συνέχεια έγινε εισαγωγή τους στο αποθετήριο του λογισμικού Rapidminer, και δημιουργία αντίστοιχα ενός ExampleSet. Κατά την εισαγωγή έγιναν, όπου κρίθηκε αναγκαίο, μετατροπές στον τύπο κάποιων χαρακτηριστικών (π.χ. από numerical σε nominal). Επίσης, κάποια χαρακτηριστικά που εμφανώς δεν παρουσίαζαν ενδιαφέρον παραλείφθηκαν εξ αρχής, προς μείωση του υπολογιστικού κόστους.

	A	B	C	G	J	K	L	M	AK	AL		
1	CD	ENAMI	FNAME	BIRTH_YR	SPOUDES	EIDIKOTIT.	CD_SERV	SERV_DSC	CD_SEM	CD_SEM	THEME	SEM_THEME_DSCR
2	3121,0		ΠΑΝΤΕΛ	1964,0	ΠΤΥΧΙΟ ΑΕ	ΔΑΣΚΑΛΟ	7889,0	ΕΘΝΙΚΟ Κ	80028Δ20		11342	ΒΕΛΤΙΩΣΗ ΔΕΞΙΟΤΗΤΩΝ ΕΠΙΚΟΙΝΩΝΙΑΣ ΣΤΗΝ ΑΠ
3	263231,0		ΙΩΑΝΝΑ	1972,0	ΔΕΥΤΕΡΟΒ	ΜΗΧΑΝΟΙ	31,0	ΠΕΡΙΦΕΡΕ	88001Δ17		10979	ΑΝΑΠΤΥΞΗ ΔΙΟΙΚΗΤΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΔΙΑΧΕΙΡ
4	2148,0		ΚΩΝΣΤΑ	1947,0	ΠΤΥΧΙΟ Α.	ΠΟΛΙΤΙΚΩ	7886,0	ΓΕΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑ			3880	ΑΡΧΕΣ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΑΞΙΟΛΟΓΗΣΗΣ ΕΡΓΟΥ. ΕΦΑ
5	2149,0		ΓΕΩΡΓΙΑ	1960,0	ΔΕΥΤΕΡΟΒΑΘΜΙΑ	ΕΚ	7824,0	ΓΕΝΙΚΟ ΕΠΙΤΕΛΕΙΟ ΑΕ			3637	ΙΤΑΛΙΚΑ ΕΞΕΙΔΙΚΕΥΜΕΝΑ: ΠΡΟΕΤΟΙΜΑΣΙΑ ΓΙΑ ΤΟ
6	2412,0		ΦΕΡΕΝΙΚ	1959,0	ΠΤΥΧΙΟ Τ.Ι	ΚΟΙΝΩΝΙΚ	9,0	ΥΠΟΥΡΓΕΙΟ ΥΓΕΙΑΣ			2632	Γ.Κ.Π.Σ. (2000-2006) - ΕΦΑΡΜΟΓΗ
7	2423,0		ΕΛΕΝΗ	1968,0	ΔΕΥΤΕΡΟΒ	ΠΑΙΔΟΚΟΙ	6848,0	ΔΗΜΟΤΙΚΙ	80070Κ16		10840	ΗΜΕΡΙΔΑ ΕΥΑΙΣΘΗΤΟΠΟΙΗΣΗΣ ΣΤΕΛΕΧΩΝ ΚΑΙ ΕΚ
8	2431,0		ΜΑΡΙΑ	1966,0	ΔΕΥΤΕΡΟΒ	ΠΑΙΔΟΚΟΙ	229,0	ΔΗΜΟΣ ΑΘΗΝΑΙΩΝ			6584	ΒΕΛΤΙΩΣΗ ΚΟΙΝΩΝΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΕΠΙΚΟΙΝ
9	263115,0		ΜΗΝΑΣ	1953,0	ΠΤΥΧΙΟ ΑΕ	ΠΤΥΧΙΟ Μ	2631,0	ΔΙΕΥΘΥΝΣ	80056Κ11		10468	Ο ΡΟΛΟΣ ΤΗΣ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ ΣΤΗΝ ΠΡΟΣ
10	263331,0		ΒΑΡΒΑΡ	1974,0	ΔΕΥΤΕΡΟΒ	ΝΟΣΗΛΕΥ	477,0	ΘΕΡΑΠΕΥΤ	81043Β11		2591	ΕΙΣΑΓΩΓΙΚΗ ΕΚΠΑΙΔΕΥΣΗ
11	263338,0		ΜΑΡΙΑ	1977,0	ΔΕΥΤΕΡΟΒΑΘΜΙΑ	ΕΚ	1699,0	ΔΗΜΟΣ Π	81015Μ15		10334	ΤΥΠΟΠΟΙΗΣΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ-ΕΚΔΟΣΗ ΚΑ
12	2433,0		ΑΛΕΞΑΝ	1960,0	ΠΤΥΧΙΟ Α.Ε.Ι		7892,0	ΓΕΝΙΚΗ ΓΡ	30038Π01		687	INTERNET-E-MAIL
13	2434,0		ΓΕΩΡΓΙΑ	1963,0	ΔΕΥΤΕΡΟΒΑΘΜΙΑ	ΕΚ	229,0	ΔΗΜΟΣ Α	81010Μ09		10050	ΝΟΜΙΜΟΤΗΤΑ ΔΙΟΙΚΗΤΙΚΗΣ ΔΡΑΣΗΣ ΣΤΗΝ ΤΟΠΗ
14	263340,0		ΓΕΩΡΓΙΟ	1964,0	ΔΙΔΑΚΤΟΡ	ΜΗΧΑΝΟ	7,0	ΥΠΟΥΡΓΕΙ	80024Δ19		11130	ΑΝΑΠΤΥΞΗ ΔΙΟΙΚΗΤΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΔΙΟΙΚΗ
15	2457,0		ΕΛΕΥΘΕΡ	1961,0	ΔΕΥΤΕΡΟΒΑΘΜΙΑ	ΕΚ	7854,0	ΟΡΓΑΝΙΣΜ	87002Μ10		6579	Ο ΔΗΜΟΣΙΟΓΥΠΑΛΛΗΛΙΚΟΣ ΚΩΔΙΚΑΣ
16	1420,0		ΜΑΡΙΑ	1952,0	ΔΕΥΤΕΡΟΒΑΘΜΙΑ	ΕΚ	229,0	ΔΗΜΟΣ Α	ΑΘΗΝΑΙΩΝ		3598	ΔΕΞΙΟΤΗΤΕΣ ΣΤΗΝ ΑΓΓΛΙΚΗ ΓΛΩΣΣΑ ΚΑΙ ΣΤΟΝ ΑΓ
17	2170,0		ΝΙΚΟΛΑ	1955,0	ΠΤΥΧΙΟ Τ.Ε.Ι.		30,0	ΠΕΡΙΦΕΡΕΙΑ	ΑΤΤΙΚΗΣ		191	WINDOWS-EXCEL

Εικόνα 4.38 Ο πίνακας των αιτήσεων σε μορφή .csv

Modeling

Η περίπτωση που εξετάζουμε, μπορεί να αναλυθεί με εφαρμογή της τεχνικής Market Basket Analysis, αφού αφορά στην ανάλυση υποσυνόλων αντικειμένων (στην περίπτωσή μας Τίτλων Σεμιναρίων) τα οποία επιλέχθηκαν από τον ίδιο «πελάτη» (εκπαιδευόμενο) μεταξύ ενός μεγαλύτερου πλήθους Σεμιναρίων. Στοχεύει δηλαδή στην ανακάλυψη συσχετίσεων με βάση τις επιλογές των εκπαιδευόμενων και τη δημιουργία κανόνων συσχέτισης (Association Rules) του τύπου:

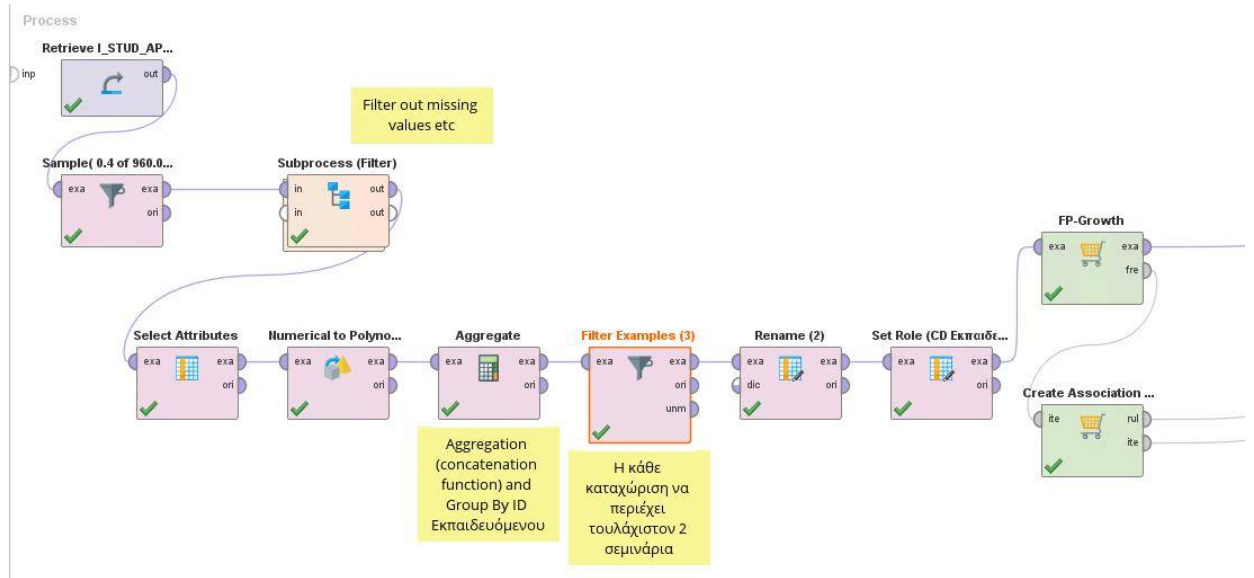
- **Εάν {έχει επιλεγεί το Σεμινάριο Α} Τότε {θα επιλεγεί το Σεμινάριο Β, Γ,...,η}**

οι οποίοι θα χρησιμοποιηθούν για εξαγωγή συμπερασμάτων.

Για τη δημιουργία του μοντέλου ακολουθήθηκαν τα παρακάτω βήματα:

- Ανάκτηση από το αποθετήριο του ExampleSet που περιείχε τα δεδομένα
- Απομόνωση δείγματος των δεδομένων (25% από 960.000 εγγραφές) για γρηγορότερη επεξεργασία
- Επιλογή μόνο εκείνων των χαρακτηριστικών που απαιτούνται για τη δημιουργία του μοντέλου (3 από 46)
- Καθάρισμα (Filter) των δεδομένων από missing values, ή από τιμές που θα μπορούσαν να οδηγήσουν σε παραποίηση του μοντέλου. Συγκεκριμένο παράδειγμα αποτελεί ο Τίτλος Σεμιναρίου «Εισαγωγική Εκπαίδευση». Το συγκεκριμένο σεμινάριο αφορά στην επιμόρφωση που λαμβάνουν όλοι οι δημόσιοι υπάλληλοι υποχρεωτικά μετά τον διορισμό τους. Παρόλα αυτά, συμπεριλαμβάνεται στον ίδιο πίνακα αιτήσεων με τα υπόλοιπα σεμινάρια, και μάλιστα το πλήθος των εγγραφών που αφορά είναι εξαιρετικά υψηλό, κάτι το οποίο θα οδηγούσε σε αλλοίωση των κανόνων συσχέτισης.
- Μετασχηματισμός των δεδομένων, με χρήση του τελεστή Aggregate, έτσι ώστε τα δεδομένα που αφορούν τον ίδιο εκπαιδευόμενο να καταχωρηθούν σε ένα item set, διαχωρισμένα μεταξύ τους από το pipe symbol "|".

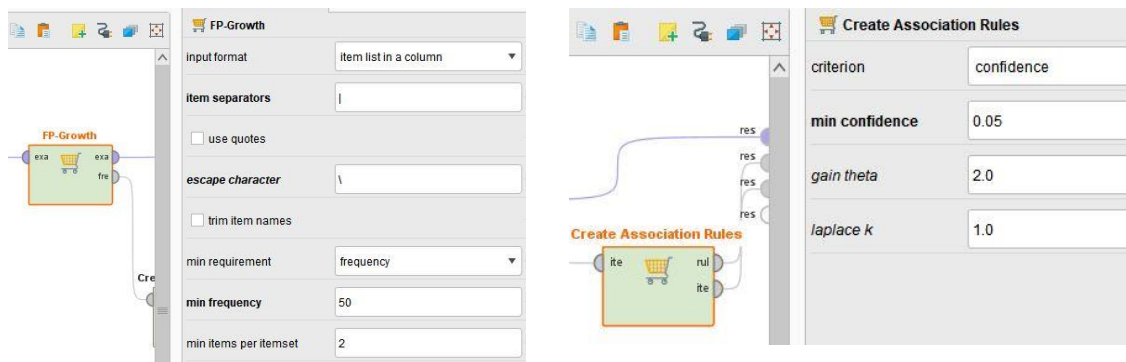
- Χρήση του τελεστή FP-Growth για την ανακάλυψη συχνά εμφανιζόμενων item sets.
- Χρήση του τελεστή Association Rules, για δημιουργία συσχετίσεων μεταξύ των σεμιναρίων που συχνά επιλέγουν μαζί οι εκπαιδευόμενοι.



Εικόνα 4.39 Το μοντέλο Basket Analysis

Evaluation

Κατά τις διαδοχικές εκτελέσεις του μοντέλου, δοκιμάζονται διάφοροι συνδυασμοί για τις παραμέτρους των τελεστών FP-Growth και Create Association Rules, ώστε να μπορούν να εξαχθούν χρήσιμα αποτελέσματα, ανάλογα με το dataset:



Εικόνα 4.40 Παράμετροι τελεστών FP-Growth και Create Association Rules

Deployment

Τα σεμινάρια που έχει παρακολουθήσει ο κάθε εκπαιδευόμενος ομαδοποιούνται:

ID_Εκπαιδευόμενου	Selected_Seminars_per_Student
102045	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ/ΕΚΠΑΙΔΕΥΣΗ ΣΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΔΗΜΟΣΙΕΣ ΣΥΜΒΑΣΕΙΣ/ΕΚΠΑΙΔΕΥΣΗ ΠΡΟΪΣΤΑΜΕΝΩΝ ΤΜΗΜΑΤΩΝ
102058	ΑΡΧΕΣ ΚΑΙ ΚΑΛΕΣ ΠΡΑΚΤΙΚΕΣ ΓΙΑ ΤΗΝ ΠΟΙΟΤΙΚΗ ΕΞΥΠΗΡΕΤΗΣΗ ΤΟΥ ΠΟΛΙΤΗ/Η ΠΡΑΚΤΙΚΗ ΤΗΣ ΔΙΑΠΡΑΓΜΑΤΕΥΣΗΣ
102059	ΔΙΟΙΚΗΣΗ ΜΕΣΩ ΣΤΟΧΩΝ/ΗΜΕΡΙΔΑ ΕΥΑΙΣΘΗΤΟΠΟΙΗΣΗΣ ΓΙΑ ΤΟ ΝΕΟ ΝΟΜΙΚΟ ΠΛΑΙΣΙΟ ΠΟΥ ΑΦΟΡΑ ΤΙΣ ΔΗΜΟΣΙΕΣ ΣΥΜΒΑΣΕΙΣ/ΕΚΠΑΙΔΕΥΣΗ ΣΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΔΗΜΟΣΙΕΣ ΣΥΜΒΑΣΕΙΣ
102068	ΛΟΓΙΣΤΙΚΟ ΔΗΜΟΣΙΟΥ- ΠΡΟΫΠΟΛΟΓΙΣΜΟΣ, ΕΛΕΓΧΟΣ, ΛΟΓΙΣΤΙΚΗ ΝΟΜΙΚΩΝ ΠΡΟΣΩΠΩΝ ΔΗΜΟΣΙΟΥ ΔΙΚΑΙΟΥ/ΕΚΠΑΙΔΕΥΣΗ ΠΡΟΪΣΤΑΜΕΝΩΝ ΤΜΗΜΑΤΩΝ
102069	ΔΙΟΙΚΗΣΗ ΜΕΣΩ ΣΤΟΧΩΝ/Η ΠΡΑΚΤΙΚΗ ΤΗΣ ΔΙΑΠΡΑΓΜΑΤΕΥΣΗΣ/Η ΠΡΑΚΤΙΚΗ ΤΗΣ ΔΙΑΠΡΑΓΜΑΤΕΥΣΗΣ
102086	ΑΞΙΟΠΟΙΗΣΗ ΤΩΝ ΜΕΣΩΝ ΚΟΙΝΩΝΙΚΗΣ ΔΙΚΤΥΩΣΗΣ (SOCIAL MEDIA) ΩΣ ΕΡΓΑΛΕΙΟΥ ΕΠΙΚΟΙΝΩΝΙΑΣ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ/ΒΕΛΤΙΩΣΗ ΚΟΙΝΩΝΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΕΠΙΚΟΙΝΩΝΙΑΚΕΣ ΔΥΣΛΕΙΤ...
102094	ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ/ΧΤΙΖΟΝΤΑΣ ΤΗΝ ΑΠΟΤΕΛΕΣΜΑΤΙΚΗ ΟΜΑΔΑ/ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ...

Εικόνα 4.41 FP-Growth, Τα Example Sets που προκύπτουν ανά Εκπαιδευόμενο (CD)

Ο τελεστής FP-Growth ανιχνεύει τα συχνά εμφανιζόμενα Itemsets (Τίτλους Σεμιναρίων):

Size	Support ↓	Item 1	Item 2	Item 3	Item 4
3	0.002	ΑΝΑΘΕΣΗ ΚΑΙ ΕΚΤΕΛΕΣΗ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩΝ ΠΡΟΜΗΘΕΙΩΝ ΚΑΙ ΓΕΝΙΚΩΝ Υ...	ΕΦΑΡΜΟΓΗ ΤΟΥ ΚΕΝΤΡΙΚΟΥ ΗΛΕΚΤΡΟΝΙΚΟΥ ΜΗΤΡΩΟΥ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩ...	ΗΛΕΚΤΡΟΝΙΚΕΣ ΔΗΜΟΣΙΕΣ ΣΥΜΒΑΣΕΙΣ ΠΡΟΜΗΘΕΙΩΝ ΚΑΙ ΥΠΗΡΕΣΙΩΝ ΜΕΣΩ ΕΣΗΔΗΣ	
3	0.002	ΔΙΑΧΕΙΡΙΣΗ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΣΤΙΣ ΥΠΗΡΕΣΙΕΣ ΥΓΕΙΑΣ	ΜΑΝΑΤΖΜΕΝΤ ΥΠΗΡΕΣΙΩΝ ΥΓΕΙΑΣ	ΟΡΓΑΝΩΣΗ ΚΑΙ ΔΙΟΙΚΗΣΗ ΝΟΣΗΛΕΥΤΙΚΩΝ ΥΠΗΡΕΣΙΩΝ	
3	0.002	ΔΙΑΧΕΙΡΙΣΗ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΣΤΙΣ ΥΠΗΡΕΣΙΕΣ ΥΓΕΙΑΣ	ΜΑΝΑΤΖΜΕΝΤ ΥΠΗΡΕΣΙΩΝ ΥΓΕΙΑΣ	ΔΙΟΙΚΗΣΗ ΟΛΙΚΗΣ ΠΟΙΟΤΗΤΑΣ ΣΤΙΣ ΥΠΗΡΕΣΙΕΣ ΥΓΕΙΑΣ	
4	0.002	ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚ...	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ	ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ΠΟΛΙΤΗ	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ

Εικόνα 4.42 FP-Growth, Frequent Item sets

Ο τελεστής Create Association Rules δημιουργεί ένα σύνολο από κανόνες συσχέτισης, οι οποίοι αναδεικνύουν σχέσεις του τύπου «με προϋπόθεση ότι κάποιος εκπαιδευόμενος έχει αιτηθεί να παρακολουθήσει τα σεμινάρια A και B και...,ν (Premises), τότε θα αιτηθεί να παρακολουθήσει το σεμινάριο X (Conclusion) με πιθανότητα (confidence) p , όπου $0 \leq p \leq 1$:

Premises	Conclusion	Confiden... ↓
ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΜΕ ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ, ΤΥΠΟΠΟΙΗΣΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ-ΕΚΔΟΣΗ ΚΑΙ ΔΗΜΟΣΙΕΥΣΗ ΔΙΟΙ...	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ	0.547
ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ, ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ EXCEL-ΒΑΣΙΚΕΣ ΔΕΞΙΟΤΗΤΕΣ	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ	0.523
ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ, ΤΥΠΟΠΟΙΗΣΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ-ΕΚΔΟΣΗ ΚΑΙ ΔΗΜΟΣΙΕΥΣΗ ΔΙΟΙΚΗΤΙ...	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ	0.507
MANATZMENT ΥΠΗΡΕΣΙΩΝ ΥΓΕΙΑΣ, ΔΙΟΙΚΗΣΗ ΟΛΙΚΗΣ ΠΟΙΟΤΗΤΑΣ ΣΤΙΣ ΥΠΗΡΕΣΙΕΣ ΥΓΕΙΑΣ	ΔΙΑΧΕΙΡΙΣΗ ΑΝΘΡΩΠΙΝΟΥ ΔΥΝΑΜΙΚΟΥ ΣΤΙΣ ΥΠΗΡΕΣΙΕΣ ΥΓΕΙΑΣ	0.504
ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ, ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΜΕ ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ	0.488
ΒΕΛΤΙΩΣΗ ΚΟΙΝΩΝΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΣΥΓΧΡΟΝΑ ΠΕΡΙΒΑΛΛΟΝΤΑ ΚΑΙ ΑΝΤΙΜΕΤΩΠΙΣΗ ΕΡΓΑΣΙΑΚΟΥ ΣΤΡΕΣ, ΕΠΕΞΕΡΓΑ...	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ	0.486

Εικόνα 4.43 Association Rules (confidence)

Ο υπεύθυνος κάποιου συγκεκριμένου προγράμματος (π.χ. Ηλεκτρονική Γραμματειακή Υποστήριξη) μπορεί να έχει εικόνα για το ιστορικό αιτήσεων των εκπαιδευόμενων που έχουν αιτηθεί να παρακολουθήσουν το συγκεκριμένο σεμινάριο, καθώς και με ποια πιθανότητα, εκπαιδευόμενοι που έχουν κάνει αίτηση για συγκεκριμένα σεμινάρια, να αιτηθούν να παρακολουθήσουν και αυτό (Εικόνα 4.44, Εικόνα 4.45):

No.	Premises	Conclusion	Confiden... ↓
351	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΤΥΠΟΠΟΙΗΣΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ-ΕΚΔΟΣΗ ΚΑΙ ΔΗΜΟΣ...	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.425
339	ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ, ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ...	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.392
332	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΣΥΝΤΑΞΗ, ΕΚΔΟΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.368
326	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ EXCEL-ΒΑΣΙΚΕΣ ΔΕΞΙΟΤΗΤΕΣ	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.361
321	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.355
316	ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΓΕΝΙΚΟΣ ΚΑΝΟΝΙΣΜΟΣ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ: ΟΙ ΥΠ...	ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.349

Εικόνα 4.44 Association Rules (ανά Τίτλο Σεμιναρίου I)

No.	Premises	Conclusion	Confiden... ↑
282	ΕΚΠΑΙΔΕΥΣΗ ΣΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΔΗΜΟΣΙΕΣ ΣΥΜΒΑΣΕΙΣ, ΕΦΑΡΜΟΓΗ ΤΟΥ ΚΕΝΤΡΙΚΟΥ ΗΛΕΚΤΡΟΝΙΚΟΥ ΜΗΤΡΩΟΥ Δ...	ΑΝΑΘΕΣΗ ΚΑΙ ΕΚΤΕΛΕΣΗ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩΝ ΠΡΟΜΗΘΕΙΩΝ ...	0.300
295	ΕΚΠΑΙΔΕΥΣΗ ΣΤΙΣ ΗΛΕΚΤΡΟΝΙΚΕΣ ΔΗΜΟΣΙΕΣ ΣΥΜΒΑΣΕΙΣ, ΔΙΑΧΕΙΡΙΣΗ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩΝ/ΕΦΑΡΜΟΓΗ ΝΟΜΙΚΟΥ ...	ΑΝΑΘΕΣΗ ΚΑΙ ΕΚΤΕΛΕΣΗ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩΝ ΠΡΟΜΗΘΕΙΩΝ ...	0.324
353	ΕΦΑΡΜΟΓΗ ΤΟΥ ΚΕΝΤΡΙΚΟΥ ΗΛΕΚΤΡΟΝΙΚΟΥ ΜΗΤΡΩΟΥ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩΝ (ΚΗΜΔΗΣ), ΗΛΕΚΤΡΟΝΙΚΕΣ ΔΗΜΟΣΙ...	ΑΝΑΘΕΣΗ ΚΑΙ ΕΚΤΕΛΕΣΗ ΔΗΜΟΣΙΩΝ ΣΥΜΒΑΣΕΩΝ ΠΡΟΜΗΘΕΙΩΝ ...	0.435

Εικόνα 4.45 Association Rules (ανά Τίτλο Σεμιναρίου II)

Τέλος, ο υπεύθυνος των προγραμμάτων, μπορεί να έχει περιγραφική απεικόνιση των κανόνων που παράγονται από τον τελεστή (Εικόνα 4.46):

AssociationRules

Association Rules

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΑΝΑΠΤΥΞΗ ΔΙΟΙΚΗΤΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΗΡΕΣΙΑ ΚΑΙ ΣΤΡΑΤΗΓΙΚΟΣ ΤΡΟΠΟΣ ΕΚΕΦΗΣ- ΟΡΓΑΝΩΣΙΑΚΕΣ ΑΔΑΓΓΕΣ] (confidence: 0.06)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΓΕΝΙΚΟΣ ΚΑΝΟΝΙΣΜΟΣ ΠΡΟΣΤΑΣΙΑΣ ΔΕΔΟΜΕΝΩΝ: ΟΙ ΥΠΟΧΡΕΩΣΕΙΣ ΤΗΣ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ] (confidence: 0.062)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ EXCEL:ΒΑΣΙΚΕΣ ΔΕΞΙΟΤΗΤΕΣ] (confidence: 0.065)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ, ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ΠΟΛΙΤΗ] (confidence: 0.069)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΕΚΠΑΙΔΕΥΣΗ ΠΡΟΪΣΤΑΜΕΝΩΝ ΤΜΗΜΑΤΩΝ] (confidence: 0.069)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [Η ΚΑΤΑΠΟΛΕΜΗΣΗ ΤΩΝ ΦΑΙΝΟΜΕΝΩΝ ΔΙΑΦΘΟΡΑΣ - ΕΛΕΓΚΤΙΚΟΙ ΜΗΧΑΝΙΣΜΟΙ] (confidence: 0.070)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΜΕ ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ] (confidence: 0.070)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΕΚΔΟΣΗ ΗΛΕΚΤΡΟΝΙΚΟΥ ΕΓΓΡΑΦΟΥ ΜΕ ΤΗ ΧΡΗΣΗ ΨΗΦΙΑΚΩΝ ΥΠΟΓΡΑΦΩΝ] (confidence: 0.072)

[ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ, ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ΠΟΛΙΤΗ] --> [ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΕΥΠΛΩΞΙΣ ΚΟΙΝΩΝΙΚΕΣ ΟΜΑΔΕΣ: ΠΟΙΟΤΗΤΑ ΠΑΡΕΧΟΜΕΝΩΝ ΥΠΗΡΕΣΙΩΝ] (confidence: 0.083)

[ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ, ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ΠΟΛΙΤΗ] --> [ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] (confidence: 0.083)

[ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ, ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ΠΟΛΙΤΗ] --> [ΑΝΑΠΤΥΞΗ ΔΙΟΙΚΗΤΙΚΩΝ ΙΚΑΝΟΤΗΤΩΝ: ΕΠΙΛΥΣΗ ΠΡΟΒΛΗΜΑΤΟΣ (PROBLEM SOLVING) ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΜΕ ΛΟΓΙΣΤΙΚΑ ΦΥΛΛΑ] (confidence: 0.083)

[ΝΟΜΙΜΟΤΗΤΑ ΚΑΙ ΔΙΑΦΑΝΕΙΑ ΣΤΗ ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ ΚΑΙ ΤΗΝ ΤΟΠΙΚΗ ΑΥΤΟΔΙΟΙΚΗΣΗ, ΠΟΙΟΤΙΚΕΣ ΥΠΗΡΕΣΙΕΣ ΤΟΥ ΔΗΜΟΣΙΟΥ ΤΟΜΕΑ ΠΡΟΣ ΤΟΝ ΠΟΛΙΤΗ] --> [ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ] (confidence: 0.083)

[ΣΥΝΤΑΞΗ ΔΗΜΟΣΙΩΝ ΕΓΓΡΑΦΩΝ ΜΕ ΕΠΕΞΕΡΓΑΣΤΕΣ ΚΕΙΜΕΝΟΥ, ΗΛΕΚΤΡΟΝΙΚΗ ΓΡΑΜΜΑΤΕΙΑΚΗ ΥΠΟΣΤΗΡΙΞΗ] --> [ΗΛΕΚΤΡΟΝΙΚΗ ΔΙΑΚΥΒΕΡΝΗΣΗ ΚΑΙ ΠΟΛΙΤΕΣ] (confidence: 0.08)

Εικόνα 4.46 Association Rules (description)

4.3.4 Περίπτωση Εφαρμογής 4: Ανάλυση της Αξιολόγησης σεμιναρίων και εξαγωγή συμπερασμάτων

Business Understanding

Όπως αναφέρθηκε και νωρίτερα, μετά την ολοκλήρωση της επιμόρφωσης, οι εκπαιδευόμενοι καλούνται να αξιολογήσουν και να βαθμολογήσουν σε κλίμακα από 1 έως 10 το σεμινάριο που παρακολούθησαν μέσα από μια σειρά ερωτήσεων, χωρισμένες σε τέσσερις κατηγορίες ενδιαφέροντος: α) τις εκπαιδευτικές συνιστώσες του σεμιναρίου, β) την οργάνωσή του, γ) τις υποδομές και την αίθουσα διεξαγωγής του δ) (εφόσον αφορά σε σεμινάριο που προβλέπει μετακίνηση εκπαιδευόμενων) το ξενοδοχείο διαμονής (εικ. 4.28):

ΕΚΠΑΙΔΕΥΤΙΚΕΣ ΣΥΝΙΣΤΩΣΕΣ ΤΟΥ ΕΠΙΜΟΡΦ. ΠΡΟΓΡΑΜΜΑΤΟΣ				ΟΡΓΑΝΩΣΗ ΤΟΥ ΕΠΙΜΟΡΦΩΤΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ			
ΕΚΠΑΙΔΕΥΤΙΚΟ ΥΛΙΚΟ ΠΟΥ ΣΑΣ ΔΟΘΗΚΕ			10	ΕΓΚΑΙΡΗ ΕΝΗΜΕΡΩΣΗ ΓΙΑ ΤΗ ΣΥΜΜΕΤΟΧΗ			10
ΕΝΗΜΕΡΩΣΗ ΓΙΑ ΤΟΥΣ ΣΤΟΧΟΥΣ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ			9	ΔΙΟΙΚ. ΥΠΟΣΤΗΡΙΞΗ ΑΠΟ ΤΗ ΓΡΑΜΜΑΤΕΙΑ ΤΟΥ ΙΝΕΠ			10
ΧΡΗΣΗ ΕΚΠΑΙΔΕΥΤΙΚΩΝ ΤΕΧΝΙΚΩΝ			10	ΟΡΓΑΝ. ΥΠΟΣΤΗΡΙΞΗ ΚΑΤΑ ΤΗ ΔΙΑΡΚΕΙΑ ΤΟΥ ΠΡΟΓΡ.			10
ΠΡΟΣΘΕΤΕΣ ΓΝΩΣΕΙΣ/ΔΕΞΙΟΤΗΤΕΣ ΠΟΥ ΑΠΟΚΤΗΣΑΤΕ			9	ΕΝΗΜΕΡΩΣΗ ΣΧΕΤΙΚΑ ΜΕ ΤΙΣ ΑΠΑΙΤΗΣΕΙΣ			10
ΣΑΦΗΝΕΙΑ ΣΤΟΧΩΝ			9	ΤΗΡΗΣΗ ΩΡΟΛΟΓΙΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ			10
ΣΥΝΑΦΕΙΑ ΣΤΟΧΩΝ ΚΑΙ ΠΕΡΙΕΧΟΜΕΝΟΥ			9				
ΣΥΝΟΛΙΚΗ ΔΙΑΡΚΕΙΑ ΕΠΙΜΟΡΦΩΤΙΚΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ			10				
ΣΥΝΟΛΙΚΗ ΔΟΜΗ			10				
ΥΠΟΔΟΜΕΣ-ΑΙΘΟΥΣΑ ΕΠΙΜΟΡΦΩΣΗΣ				ΞΕΝΟΔΟΧΕΙΟ ΔΙΑΜΟΝΗΣ			
ΧΩΡΗΤΙΚΟΤΗΤΑ	10	ΤΕΧΝΟΛΟΓΙΚΟΣ ΕΞΟΠΛΙΣΜΟΣ	9	ΕΞΥΠΗΡΕΤΗΣΗ ΠΡΟΣΩΠ.		ΔΙΑΤΡΟΦΗ	
ΦΩΤΕΙΝΟΤΗΤΑ ΑΙΘΟΥΣΑΣ	10	ΚΑΘΑΡΙΟΤΗΤΑ	10	ΔΙΑΜΟΝΗ		ΣΥΝΔΕΣΗ ΜΕ ΜΜΜ	
ΔΙΑΤΑΞΗ ΑΙΘΟΥΣΑΣ	8	ΚΛΙΜΑΤΙΣΜΟΣ/ΘΕΡΜΑΝΣΗ	10				
ΗΧΟΜΟΝΩΣΗ	9	ΠΡΟΣΒΑΣΙΜΟΤΗΤΑ ΣΕ ΑΜΕΑ	9				

Εικόνα 4.47 Αξιολόγηση σεμιναρίου από εκπαιδευόμενους

Στο τέλος της αξιολόγησης υπάρχει επίσης σχετικό πεδίο (εικ. 4.29), όπου οι εκπαιδευόμενοι αξιολογούν, επίσης σε κλίμακα 1 έως 10, το σεμινάριο στο σύνολό του:

ΣΧΟΛΙΑ - ΠΑΡΑΤΗΡΗΣΕΙΣ ΓΙΑ ΒΕΛΤΙΩΣΗ ΕΠΙΜΟΡΦ. ΠΡΟΓΡΑΜΜΑΤΟΣ	ΣΥΝΟΛΙΚΗ ΑΞΙΟΛΟΓΗΣΗ
<input type="text"/>	9

Εικόνα 4.48 Συνολική αξιολόγηση σεμιναρίου

Το κομμάτι της Αξιολόγησης αποτελεί για το ΙΝ.ΕΠ, όπως και για κάθε οργανισμό που παρέχει υπηρεσίες ή προϊόντα κάθε είδους, ένα εξαιρετικά σημαντικό εργαλείο, καθώς αποτυπώνει τη γνώμη των ενδιαφερόμενων ως προς το επίπεδο της ποιότητας της παρεχόμενης υπηρεσίας ή προϊόντος. Με κατάλληλη ανάλυση των δεδομένων της ανάδρασης αυτής, ο φορέας μπορεί να οδηγηθεί σε εξαγωγή συμπερασμάτων, τα οποία με τη σειρά τους μπορούν να οδηγήσουν σε ενέργειες για τη βελτίωση της ποιότητας της εκπαιδευτικής διαδικασίας.

Θεωρήθηκε λοιπόν ότι παρουσίαζε ιδιαίτερο ενδιαφέρον η ανάλυση των δεδομένων αυτών με χρήση τεχνικών εξόρυξης δεδομένων, ώστε να βρεθούν, πιθανώς, μέσα από αυτήν, συσχετίσεις και πρότυπα που επηρεάζουν τη συνολική αξιολόγηση ενός σεμιναρίου από έναν εκπαιδευόμενο.

Data Understanding

Τα δεδομένα των αξιολογήσεων διατηρούνται σε έναν πίνακα του ΟΠΣ, σε κάθε εγγραφή του οποίου υπάρχει καταχωρισμένη η αξιολόγηση κάθε εκπαιδευόμενου, για κάθε ένα σεμινάριο το οποίο παρακολούθησε (εικ. 4.30):

	A	B	C	D	M	N	O	P	Q	R	S	T	U	V
1	CD	CD_SEM	CD_STUDE	DURATION	INFORMT.	EDUCATIC	ADDITION	OBJECTIVI	CONTENTI	STRUCTUR	INFORMP,	ADMINSU	ORGANSU	INFORMRIT
2	186	80042Δ11	20894		9	9	5	6	9	9	10	9	10	10
3	187	81014Κ11	3860		6	9	8	8	8	8	9	10	10	10
4	202	80058Τ11	18365		10	10	10	10	10	10	9	10	10	10
5	203	80043Δ11	8166		10	10	9	9	9	9	9	9	9	9
6	206	80023Ρ11	2606		3	8	8	6	8	8	7	10	10	10
7	233	80023Ρ11	3666		6	7	7	6	6	6	7	6	6	6
8	235	80043Δ11	11122		9	9	10	9	10	8	9	10	10	10
9	237	80042Δ11	2943		10	10	10	10	10	10	10	10	10	10
10	241	80043Δ11	19436		7	7	6	7	7	6	6	8	8	8
11	245	80060Μ11	19182		9	9	10	9	10	10	10	10	10	10

Εικόνα 4.49 Πίνακας Αξιολογήσεων Σεμιναρίων

Data preparation

Ο συγκεκριμένος πίνακας περιέχει πάνω από 300.000 εγγραφές. Για λόγους ευελιξίας, κρίθηκε σκόπιμο να εισαχθούν στο εργαλείο μόνο οι εγγραφές που αφορούσαν σε αξιολόγηση σεμιναρίων έτους 2019 (~25.000 εγγραφές), με βάση τα εξής δεδομένα:

- Αποτελούν επαρκές δείγμα για εξαγωγή συμπερασμάτων
- Είναι, με εξαίρεση τα δεδομένα για το 2020, τα πιο πρόσφατα
- Τα δεδομένα του 2020 δεν προτιμήθηκαν για ανάλυση, αφού, λόγω της υγειονομικής κατάστασης στη χώρα, τα σεμινάρια υλοποιούνται από τον 3/2020 μέσω τηλεκπαίδευσης και επομένως δεν υπάρχουν δεδομένα για την μία από τις τέσσερις συνιστώσες της αξιολόγησης (υποδομές και αίθουσα διεξαγωγής).

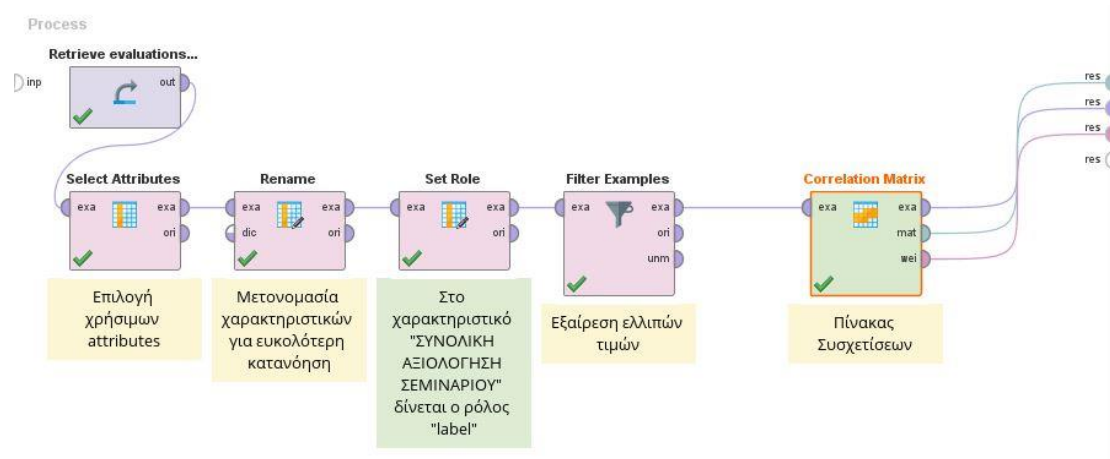
Μετά την αφαίρεση κάποιων χαρακτηριστικών που δεν παρουσίαζαν κανένα ενδιαφέρον (π.χ. ημερομηνία υποβολής της αξιολόγησης), δημιουργήθηκε ένα dataset με 24.700 εγγραφές και 21 χαρακτηριστικά (attributes).

Modeling

Για τη δημιουργία του μοντέλου χρησιμοποιήθηκε ο τελεστής Correlation Matrix του εργαλείου Rapidminer. Ο τελεστής υπολογίζει τον βαθμό συσχέτισης μεταξύ κάθε ενός από τα ζεύγη των ανεξάρτητων μεταβλητών του συνόλου δεδομένων. Στη συνέχεια κατασκευάζει τον σχετικό πίνακα συσχετίσεων.

Στόχος ήταν να αναδειχθούν πιθανές συσχετίσεις μεταξύ των συνιστωσών που οι εκπαιδευόμενοι καλούνται να αξιολογήσουν και κυρίως μεταξύ του συνολικού βαθμού αξιολόγησης του σεμιναρίου και όλων των υπόλοιπων χαρακτηριστικών. Δεδομένου ότι οι μεταβλητές είναι ανεξάρτητες μεταξύ τους, οι συσχετίσεις που θα προέκυπταν δεν αναμενόταν να αποδείξουν κάποια αιτιώδη σχέση μεταξύ δύο χαρακτηριστικών. Εάν δηλαδή για παράδειγμα η βαθμολογία του εκπαιδευτικού υλικού και του συνολικού βαθμού αξιολόγησης σεμιναρίου παρουσίαζαν υψηλό βαθμό συσχέτισης μεταξύ τους, αυτό δε θα σήμαινε ότι καλύτερο εκπαιδευτικό υλικό θα οδηγεί απαραίτητα σε υψηλότερη βαθμολόγηση του σεμιναρίου στο σύνολό του (Συνολική Αξιολόγηση). Παρόλα αυτά, θα

αποτελούσε μια πολύ καλή ένδειξη ότι αυτά τα δύο χαρακτηριστικά παρουσιάζουν υψηλό βαθμό αλληλεξάρτησης και είναι επομένως αρκετά πιθανό ότι, εάν η βαθμολόγηση του ενός βελτιωθεί, το ίδιο θα συμβεί και με τη βαθμολόγηση του άλλου.



Εικόνα 4.50 Rapidminer - Μοντέλο Πίνακα Συσχετίσεων

Evaluation

Το εργαλείο Rapidminer οπτικοποιεί τα αποτελέσματα με τρόπο τέτοιο, ώστε η εξαγωγή κάποιων πρώτων συμπερασμάτων να είναι σχετικά εύκολη. Για παράδειγμα, στην εικόνα Εικόνα 4.51 είναι φανερό με μια πρώτη ματιά, ότι 2 από τα χαρακτηριστικά («Εγκαιρη Ενημέρωση για τη συμμετοχή» και «Προσβασιμότητα ΑΜΕΑ») δε φαίνεται να συσχετίζονται ισχυρά (θετικά ή αρνητικά) με κανένα από τα υπόλοιπα χαρακτηριστικά και επομένως δεν παρουσιάζουν ιδιαίτερο ενδιαφέρον για περαιτέρω ανάλυση.

Attributes	ΣΥΝΟΛΙ...	ΕΚΠΑΙΔ...	ΠΡΟΣΘ...	ΣΑΦΗΝ...	ΣΥΝΑΦ...	ΕΓΚΑΙΡ...	ΔΙΟΙΚΗ...	ΟΡΓΑΝ...	ΤΗΡΗΣΗ...	ΤΕΧΝΟ...	ΠΡΟΣΒ...	ΕΚΠΑΙΔ...	ΣΥΝΟΛΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΣΕΜΙ...
ΣΥΝΟΛΙΚΗ ΔΙΑΡΚΕΙΑ ΠΡΟΓΡΑΜΜΑΤΟΣ	1	0.265	0.248	0.280	0.284	0.039	0.090	0.117	0.099	0.088	0.057	0.181	0.287
ΕΚΠΑΙΔΕΥΤΙΚΕΣ ΜΕΘΟΔΟΙ	0.265	1	0.474	0.509	0.510	0.063	0.143	0.196	0.171	0.147	0.086	0.400	0.450
ΠΡΟΣΘΕΤΕΣ ΔΕΞΙΟΤΗΤΕΣ ΠΟΥ ΑΠΟΚΤΗΘΗΚ...	0.248	0.474	1	0.554	0.551	0.053	0.125	0.162	0.132	0.102	0.077	0.299	0.519
ΣΑΦΗΝΕΙΑ ΣΤΟΧΩΝ	0.280	0.509	0.554	1	0.775	0.069	0.168	0.213	0.176	0.129	0.099	0.330	0.508
ΣΥΝΑΦΕΙΑ ΣΤΟΧΩΝ και ΠΕΡΙΕΧΟΜΕΝΟΥ	0.284	0.510	0.551	0.775	1	0.063	0.154	0.208	0.167	0.129	0.090	0.331	0.537
ΕΓΚΑΙΡΗ ΕΝΗΜΕΡΩΣΗ ΓΙΑ ΤΗ ΣΥΜΜΕΤΟΧΗ	0.039	0.063	0.053	0.069	0.063	1	0.351	0.199	0.123	0.042	0.034	0.060	0.062
ΔΙΟΙΚΗΤΙΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.090	0.143	0.125	0.168	0.154	0.351	1	0.532	0.216	0.100	0.092	0.125	0.142
ΟΡΓΑΝΩΤΙΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.117	0.196	0.162	0.213	0.208	0.199	0.532	1	0.260	0.180	0.134	0.179	0.194
ΤΗΡΗΣΗ ΩΡΟΛΟΓΙΟΥ ΠΡΟΓΡ.	0.099	0.171	0.132	0.176	0.167	0.123	0.216	0.260	1	0.076	0.062	0.142	0.146
ΤΕΧΝΟΛΟΓΙΚΟΣ ΕΞΟΠΛΙΣΜΟΣ	0.088	0.147	0.102	0.129	0.129	0.042	0.100	0.180	0.076	1	0.223	0.102	0.147
ΠΡΟΣΒΑΣΙΜΟΤΗΤΑ ΑΜΕΑ	0.057	0.086	0.077	0.099	0.090	0.034	0.092	0.134	0.062	0.223	1	0.060	0.104
ΕΚΠΑΙΔΕΥΤΙΚΟ ΥΛΙΚΟ	0.181	0.400	0.299	0.330	0.331	0.060	0.125	0.179	0.142	0.102	0.060	1	0.292
ΣΥΝΟΛΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΣΕΜΙΝΑΡΙΟΥ	0.287	0.450	0.519	0.508	0.537	0.062	0.142	0.194	0.146	0.147	0.104	0.292	1

Εικόνα 4.51 Αξιολογήσεις Σεμιναρίων - Πίνακας Συσχετίσεων

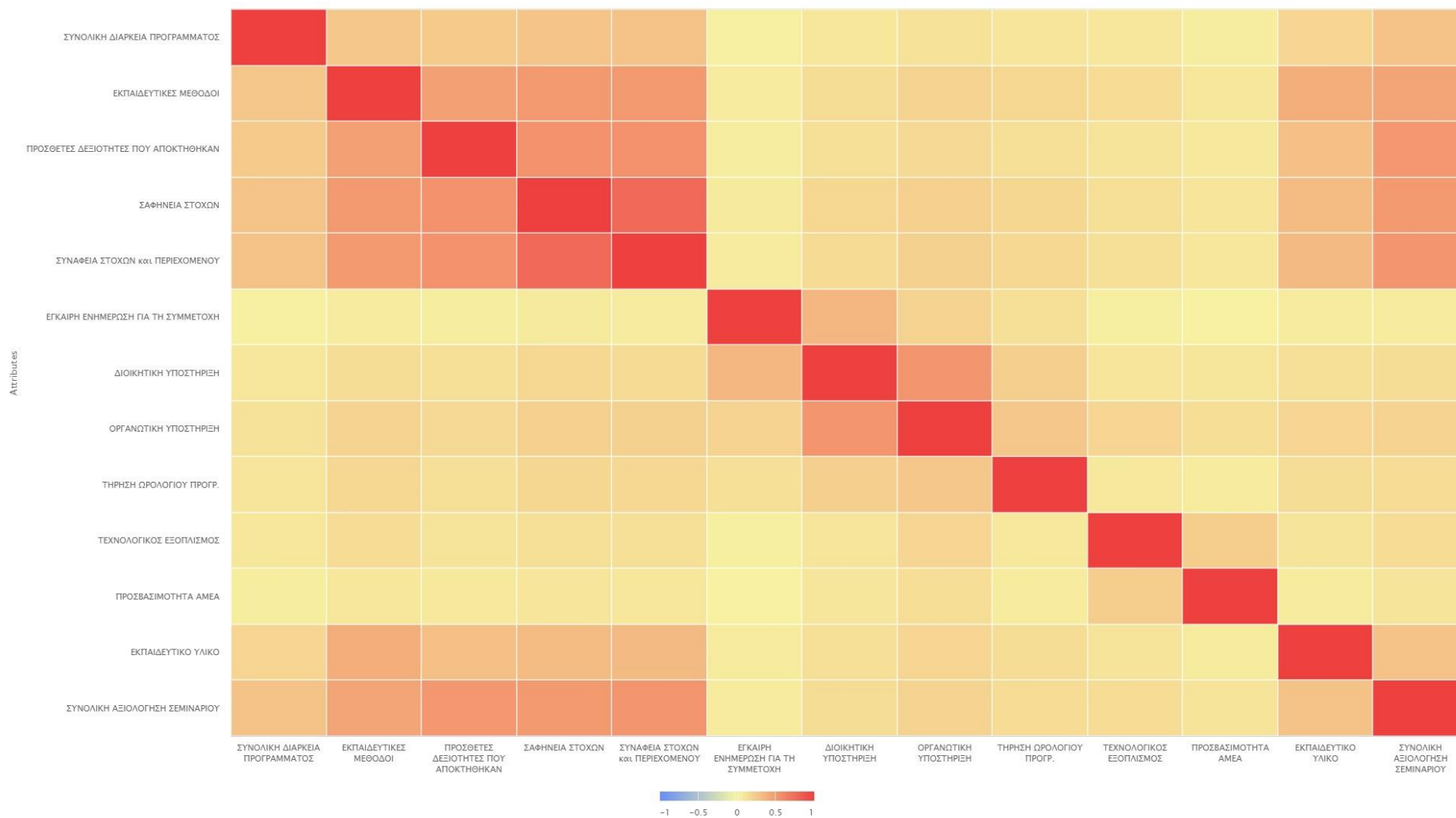
Αντίστοιχα, φαίνεται ότι τέσσερα συγκεκριμένα χαρακτηριστικά (Εκπαιδευτικές Μέθοδοι, Πρόσθετες Δεξιότητες που αποκτήθηκαν, Σαφήνεια Στόχων, Συνάφεια Στόχων και Περιεχομένου) παρουσιάζουν σχετικά ισχυρή (θετική) συσχέτιση με το χαρακτηριστικό της συνολικής αξιολόγησης του σεμιναρίου

που μας ενδιαφέρει εν προκειμένω. Το γεγονός αυτό υπονοεί ότι, καλύτερες βαθμολογίες σε ένα ή περισσότερα από αυτά τα χαρακτηριστικά, ενδέχεται να οδηγήσουν σε υψηλότερους βαθμούς στη Συνολική Αξιολόγηση:

Attributes	ΣΥΝΟΛΙΚΗ ΔΙΑΡΚΕΙΑ ...	ΕΚΠΑΙΔΕΥΤΙΚΕΣ ΜΕΘΟΔΟΙ	ΠΡΟΣΘΕΤΕΣ ΔΕΞΙΟΤΗΤΕΣ ΠΟΥ...	ΣΑΦΗΝΕΙΑ ΣΤΟΧΩΝ	ΣΥΝΑΦΕΙΑ ΣΤΟΧΩΝ και ΠΕΡΙΕΧΟΜΕΝΟΥ
ΣΥΝΟΛΙΚΗ ΔΙΑΡΚΕΙΑ ΠΡΟΓΡΑΜΜΑΤΟΣ	1	0.265	0.248	0.280	0.284
ΕΚΠΑΙΔΕΥΤΙΚΕΣ ΜΕΘΟΔΟΙ	0.265	1	0.474	0.509	0.510
ΠΡΟΣΘΕΤΕΣ ΔΕΞΙΟΤΗΤΕΣ ΠΟΥ ΑΠΟΚΤΗΘΗΚΑΝ	0.248	0.474	1	0.554	0.551
ΣΑΦΗΝΕΙΑ ΣΤΟΧΩΝ	0.280	0.509	0.554	1	0.775
ΣΥΝΑΦΕΙΑ ΣΤΟΧΩΝ και ΠΕΡΙΕΧΟΜΕΝΟΥ	0.284	0.510	0.551	0.775	1
ΕΓΚΑΙΡΗ ΕΝΗΜΕΡΩΣΗ ΓΙΑ ΤΗ ΣΥΜΜΕΤΟΧΗ	0.039	0.063	0.053	0.069	0.063
ΔΙΟΙΚΗΤΙΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.090	0.143	0.125	0.168	0.154
ΟΡΓΑΝΩΤΙΚΗ ΥΠΟΣΤΗΡΙΞΗ	0.117	0.196	0.162	0.213	0.208
ΤΗΡΗΣΗ ΩΡΟΛΟΓΙΟΥ ΠΡΟΓΡ.	0.099	0.171	0.132	0.176	0.167
ΤΕΧΝΟΛΟΓΙΚΟΣ ΕΞΟΠΛΙΣΜΟΣ	0.088	0.147	0.102	0.129	0.129
ΠΡΟΣΒΑΣΙΜΟΤΗΤΑ ΑΜΕΑ	0.057	0.086	0.077	0.099	0.090
ΕΚΠΑΙΔΕΥΤΙΚΟ ΥΛΙΚΟ	0.181	0.400	0.299	0.330	0.331
ΣΥΝΟΛΙΚΗ ΑΞΙΟΛΟΓΗΣΗ ΣΕΜΙΝΑΡΙΟΥ	0.287	0.450	0.519	0.508	0.537

Εικόνα 4.52 Αξιολογήσεις Σεμιναρίων - Ισχυρές συσχετίσεις

Εξόρυξη Δεδομένων και Σεμινάρια Επιμόρφωσης Φορέων του Δημόσιου Τομέα



Εικόνα 4.53 Αξιολογήσεις Σεμιναρίων - Matrix Visualization

Deployment

Με βάση τα αποτελέσματα που προκύπτουν, οι υπεύθυνοι των σεμιναρίων έχουν μια αρκετά ξεκάθαρη εικόνα ως προς τα συγκεκριμένα χαρακτηριστικά που συσχετίζονται περισσότερο με τον βαθμό συνολικής αξιολόγησης ενός σεμιναρίου. Με αυτό το δεδομένο, μπορούν για παράδειγμα να προσαρμόζουν το υλικό που διανέμεται στους εκπαιδευόμενους έτσι ώστε οι στόχοι κάθε σεμιναρίου να περιγράφονται σαφώς πριν την έναρξή του και να είναι απόλυτα συναφείς με το περιεχόμενό του. Επίσης, να εφιστούν την προσοχή στους εκπαιδευτές ώστε να ακολουθούν τις ενδεδειγμένες εκπαιδευτικές μεθόδους και να δίνουν περισσότερο βάρος στην απόκτηση πρόσθετων δεξιοτήτων από μέρους των εκπαιδευομένων.

4.3.5 Περίπτωση Εφαρμογής 5: Μοντέλο πρόβλεψης Εκπαιδευόμενων που θα ολοκληρώσουν την επιμόρφωση

Business Understanding

Μία άλλη περίπτωση που παρουσιάζει ενδιαφέρον στο πλαίσιο της εκπαιδευτικής διαδικασίας, είναι η δυνατότητα πρόβλεψης του ποσοστού των εκπαιδευομένων που, ενώ έχουν αρχικά επιλεγεί να παρακολουθήσουν ένα σεμινάριο, τελικά για διάφορους λόγους δεν το ολοκληρώνουν επιτυχώς. Μια τέτοια γνώση μπορεί να αποδειχθεί αξιοποιήσιμη κατά το σχεδιασμό μιας επιμορφωτικής δράσης, πολύ περισσότερο δε, εάν συνοδεύεται από ποιοτικά χαρακτηριστικά του προφίλ των εκπαιδευομένων, όπως ηλικία, εκπαιδευτική βαθμίδα, θέση ευθύνης ή ακόμα και η γεωγραφική κατανομή ανά περιφέρεια.

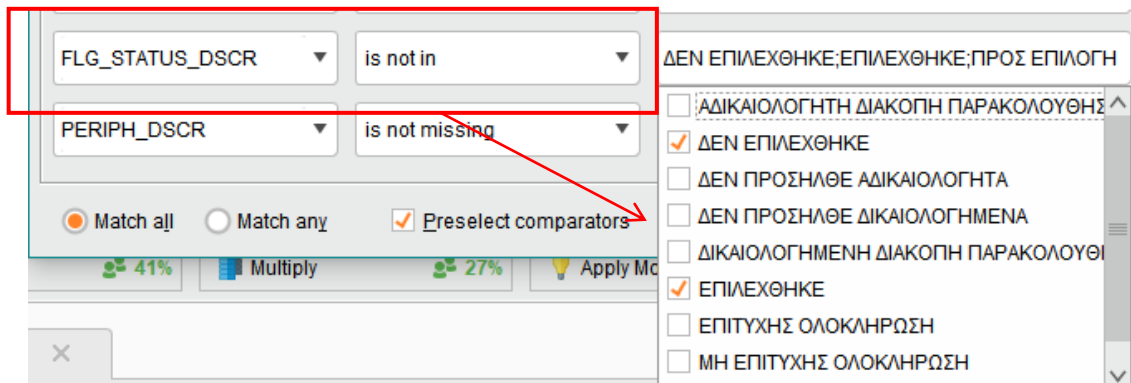
Για την ανάλυση της συγκεκριμένης περίπτωσης κρίθηκε ως βέλτιστη τεχνική, η τεχνική των δένδρων απόφασης, με στόχο την κατάταξη των εκπαιδευομένων με συγκεκριμένα χαρακτηριστικά σε 2 κατηγορίες, σε αυτούς που προβλέπεται ότι θα ολοκληρώσουν επιτυχώς ένα σεμινάριο και σε εκείνους που δε θα το ολοκληρώσουν.

Data Understanding

Όλα τα χαρακτηριστικά που μας ενδιέφεραν στη συγκεκριμένη περίπτωση, δηλαδή Ηλικία, Εκπαιδευτική βαθμίδα, Θέση ευθύνης και Ένδειξη επιτυχούς ολοκλήρωσης, υπήρχαν στον πίνακα αιτήσεων των εκπαιδευομένων ο οποίος είχε ήδη εξεταστεί στην προηγούμενη περίπτωση. Επομένως, ήταν εύκολος ο εντοπισμός των μεταβλητών που μας ενδιέφεραν μέσα στο dataset το οποίο μάλιστα δε χρειάστηκε να δημιουργηθεί εκ νέου στο αποθετήριο του Rapidminer, αφού ήδη υπήρχε.

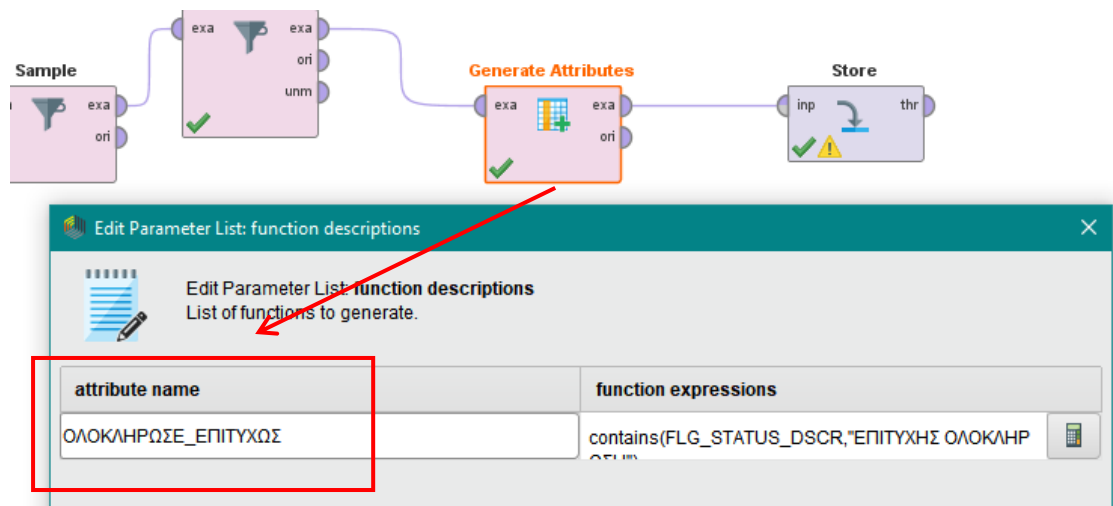
Data preparation

Παρόλα αυτά, το χαρακτηριστικό του πίνακα που υποδήλωνε την κατάσταση ολοκλήρωσης ενός σεμιναρίου (FLG_STATUS_DSCR, δεν περιείχε μόνο 2 τιμές (ναι/όχι), αλλά περισσότερες:

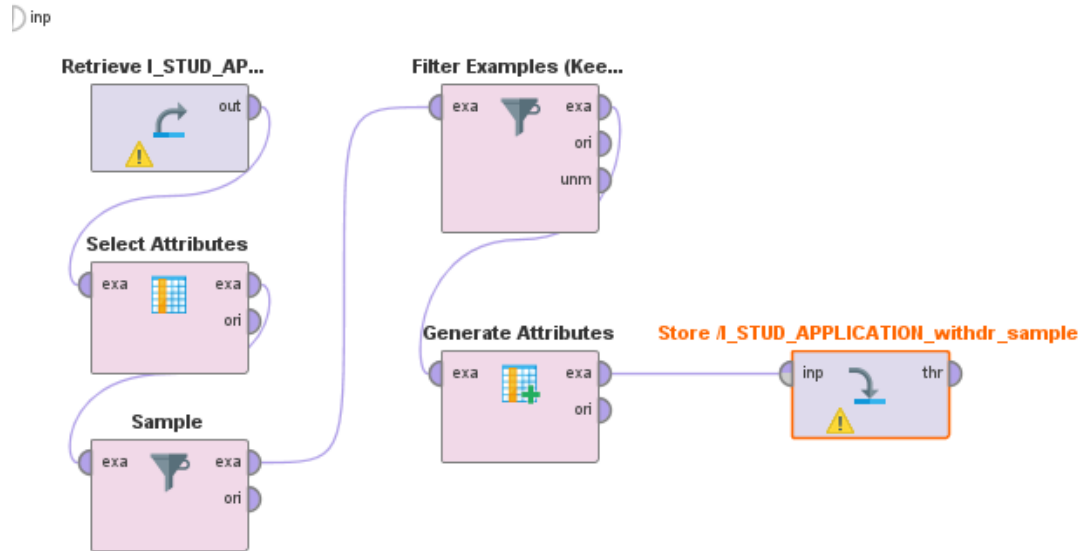


Εικόνα 4.54 Ένδειξη επιτυχούς ολοκλήρωσης σεμιναρίου

Για τον λόγο αυτό, έπρεπε αφενός να επιλεγούν μόνο οι τιμές που σήμαιναν επιτυχή ή όχι ολοκλήρωση, άρα να μη συμπεριληφθούν στο δείγμα όσες εγγραφές είχαν $FLG_STATUS_DSCR = \text{“ΕΠΙΛΕΧΘΗΚΕ/ΔΕΝ ΕΠΙΛΕΧΘΗΚΕ/ΠΡΟΣ ΕΠΙΛΟΓΗ”}$, και αφετέρου να ομαδοποιηθούν τα πεδία των οποίων η τιμή υποδείκνυε την για διάφορους λόγους μη επιτυχή ολοκλήρωση, ώστε στο τέλος να υπάρχουν 2 πιθανές τιμές, άρα 2 διαφορετικές κλάσεις/κατηγορίες ώστε να είναι δυνατή η δημιουργία του Decision Tree. Για τον μετασχηματισμό των δεδομένων, χρησιμοποιήθηκε ο τελεστής *Generate Attributes*, ώστε να δημιουργηθεί ένα επιπλέον χαρακτηριστικό τύπου Boolean («ΟΛΟΚΛΗΡΩΣΕ_ΕΠΙΤΥΧΩΣ») το οποίο λαμβάνει τιμή «True» για όσα Item sets έχουν ένδειξη «ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ» και τιμή «False» σε όλες τις άλλες περιπτώσεις:



Εικόνα 4.55 Δημιουργία επιπλέον χαρακτηριστικού



Εικόνα 4.56 Αποθήκευση Exampleset

Result History ExampleSet (//Local Repository/4_3_3_Ass [...] cations/I_STUD_APPLIC

Open in Turbo Prep Auto Model

Row No.	BIRTH_YR	STUD_CAT	STUD_POS	FLG_STATUS_DSCR	ΟΛΟΚΛΗΡΩΣΕ_ΕΠΙΤΥΧΩΣ
28	1955	ΠΕ	Προϊστάμενος...	ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	true
29	1967	ΠΕ	Υπάλληλος	ΔΕΝ ΠΡΟΣΗΛΘΕ ΑΔΙΚΑΙΟΛΟΓΗΤΑ	false
30	1968	ΠΕ	Προϊστάμενος...	ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	true
31	1956	ΠΕ	Προϊστάμενος...	ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	true
32	1967	ΔΕ	Προϊστάμενος...	ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	true
33	1960	ΤΕ	Προϊστάμενος...	ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	true
34	1960	ΤΕ	Προϊστάμενος...	ΔΕΝ ΠΡΟΣΗΛΘΕ ΔΙΚΑΙΟΛΟΓΗΜΕΝΑ	false
35	1957	ΔΕ	Προϊστάμενος...	ΕΠΙΤΥΧΗΣ ΟΛΟΚΛΗΡΩΣΗ	true

Εικόνα 4.57 Δημιουργία Exampleset

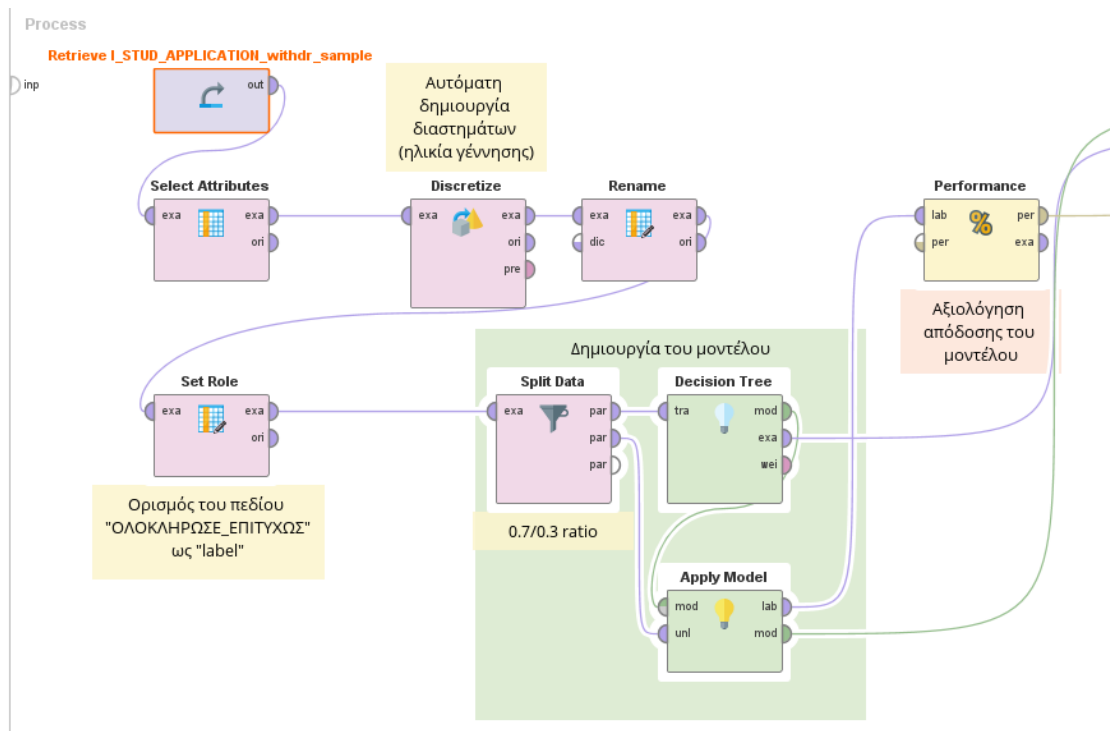
Στο πεδίο αυτό αποδόθηκε σε επόμενο στάδιο ο ρόλος «Label» ώστε να αποτελέσει την κατηγορία ταξινόμησης του δένδρου απόφασης.

Modeling

Για τη δημιουργία του μοντέλου ακολουθήθηκαν τα παρακάτω βήματα:

- Το dataset εισήχθη στο μοντέλο
- Επιλέχθηκαν τα χαρακτηριστικά ενδιαφέροντος (Ημ/νια γέννησης, Εκπαιδευτική βαθμίδα, Θέση ευθύνης, Ένδειξη επιτυχούς ολοκλήρωσης)
- Με χρήση του τελεστή Discretize γίνεται διαχωρισμός, ανάλογα με την ημ/νια γέννησης, σε 2 σύνολα

- Τα χαρακτηριστικά μετονομάζονται ώστε να είναι περιγραφικά του περιεχομένου τους
- Στο πεδίο ένδειξης ολοκλήρωσης αποδίδεται ο ρόλος "Label" αφού είναι η κλάση που μας ενδιαφέρει
- Το σύνολο δεδομένων διαχωρίζεται σε 2 υποσύνολα δεδομένων εκπαίδευσης και δεδομένων επαλήθευσης (σε αναλογία 0.7/0.3) ώστε να εισαχθούν στον τελεστή Decision Tree ο οποίος θα δημιουργήσει το μοντέλο

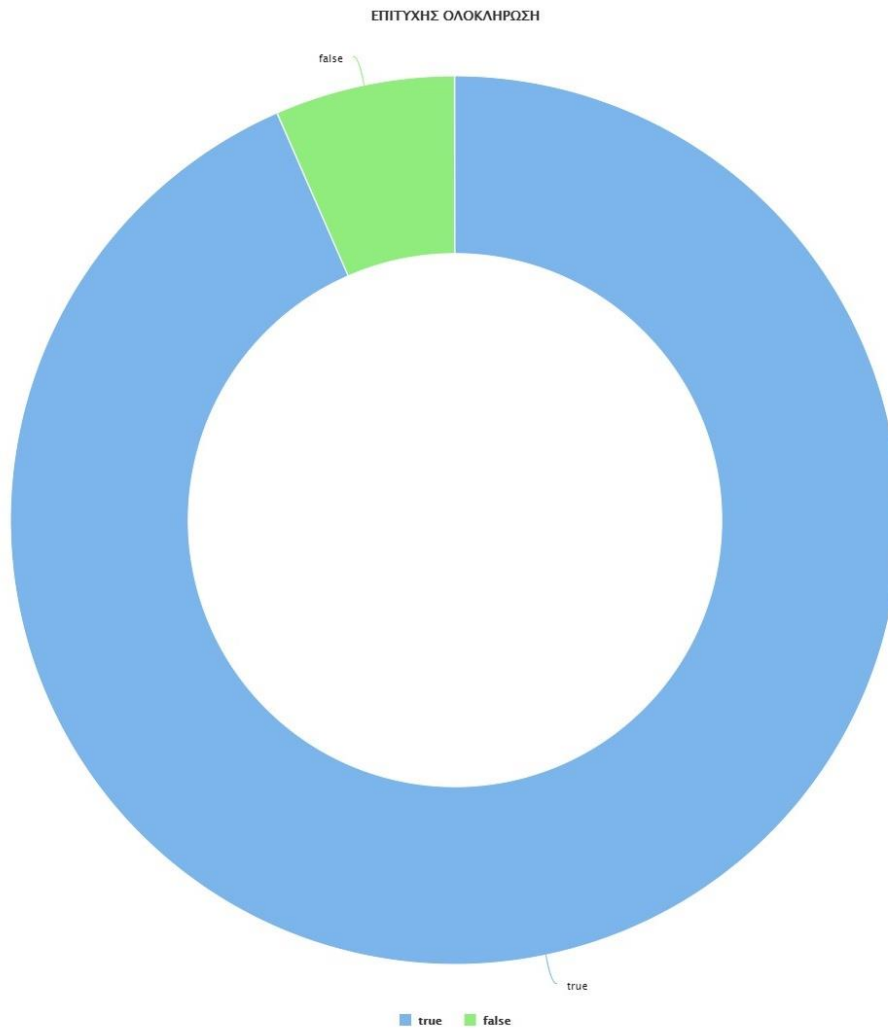


Εικόνα 4.58 Rapidminer - Μοντέλο Decision Tree

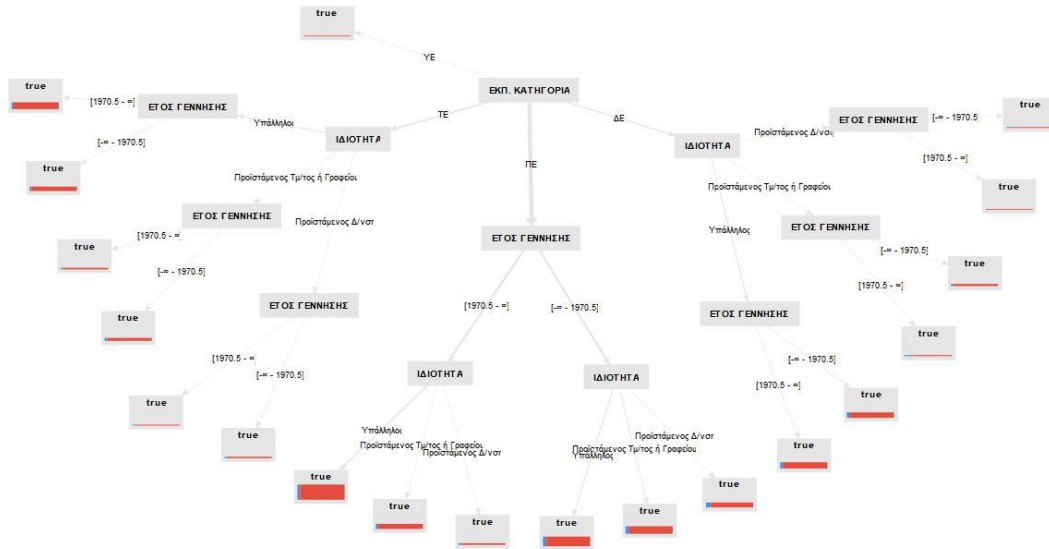
Evaluation

Row No.	ΟΛΟΚΛΗΡΩΣΕ_ΕΠΙΤΥΧΩΣ	ΕΤΟΣ ΓΕΝΝΗ...	ΕΚΠ. ΚΑΤΗΓ...	ΙΔΙΟΤΗΤΑ
28	true	[-∞ - 1970.5]	ΠΕ	Προϊστάμενος...
29	true	[-∞ - 1970.5]	ΠΕ	Υπάλληλος
30	false	[-∞ - 1970.5]	ΠΕ	Προϊστάμενος...
31	true	[-∞ - 1970.5]	ΔΕ	Προϊστάμενος...
32	false	[-∞ - 1970.5]	ΠΕ	Προϊστάμενος...
33	true	[-∞ - 1970.5]	ΠΕ	Προϊστάμενος...

Εικόνα 4.59 Set Role "Label"



Εικόνα 4.60 Επιτυχής / Ανεπιτυχής Ολοκλήρωση



Εικόνα 4.61 Decision Tree, ολοκλήρωση σεμιναρίου

Tree

```

ΕΚΠ. ΚΑΤΗΓΟΡΙΑ = ΔΕ
|  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Δ/νσης
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]: true {false=1, true=13}
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]: true {false=0, true=7}
|  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Τμ/τος ή Γραφείου
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]: true {false=6, true=110}
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]: true {false=6, true=26}
|  ΙΔΙΟΤΗΤΑ = Υπάλληλος
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]: true {false=29, true=382}
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]: true {false=32, true=447}
ΕΚΠ. ΚΑΤΗΓΟΡΙΑ = ΠΕ
|  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]
|  |  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Δ/νσης: true {false=37, true=320}
|  |  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Τμ/τος ή Γραφείου: true {false=46, true=531}
|  |  ΙΔΙΟΤΗΤΑ = Υπάλληλος: true {false=58, true=708}
|  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]
|  |  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Δ/νσης: true {false=6, true=72}
|  |  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Τμ/τος ή Γραφείου: true {false=20, true=352}
|  |  ΙΔΙΟΤΗΤΑ = Υπάλληλος: true {false=72, true=1171}
ΕΚΠ. ΚΑΤΗΓΟΡΙΑ = ΤΕ
|  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Δ/νσης
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]: true {false=4, true=56}
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]: true {false=1, true=14}
|  ΙΔΙΟΤΗΤΑ = Προϊστάμενος Τμ/τος ή Γραφείου
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]: true {false=12, true=185}
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]: true {false=5, true=73}
|  ΙΔΙΟΤΗΤΑ = Υπάλληλος
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [-∞ - 1970.5]: true {false=12, true=306}
|  |  ΕΤΟΣ ΓΕΝΝΗΣΗΣ = [1970.5 - ∞]: true {false=23, true=502}
ΕΚΠ. ΚΑΤΗΓΟΡΙΑ = ΥΕ: true {false=0, true=2}
    
```

Εικόνα 4.62 Decision Tree, descriptive

The screenshot shows a software interface with a 'Performance' tab selected. On the left, there are icons for 'Performance' (a percentage sign) and 'Description' (a document icon). The main area displays 'accuracy: 93.47%' and a table with the following data:

	true false	true true	class precision
pred. false	0	0	0.00%
pred. true	158	2262	93.47%
class recall	0.00%	100.00%	

Εικόνα 4.63 Απόδοση μοντέλου (accuracy)

Στο μοντέλο εισάγονται δεδομένα από όλους τους διαθέσιμους τίτλους σεμιναρίων, θα μπορούσε όμως να γίνει επιλεκτικά και εισαγωγή δεδομένων μόνο από συγκεκριμένους κύκλους ή/και Τίτλους σεμιναρίων, για ακόμα πιο εξειδικευμένες προβλέψεις.

Deployment

Με βάση τη γνώση που προκύπτει από την παρατήρηση των αποτελεσμάτων του μοντέλου, οι υπεύθυνοι σχεδιασμού των προγραμμάτων μπορούν να έχουν μία πολύ ακριβέστερη πρόβλεψη σε ό,τι αφορά τα ποσοστά ολοκλήρωσης στα σεμινάρια ενδιαφέροντός τους και να προσαρμόσουν τις ομάδες-στόχο ενός σεμιναρίου ή ενός Θεματικού Κύκλου σεμιναρίων, ανάλογα με τα ποσοστά επιτυχούς ολοκλήρωσης ανά ηλικιακή ομάδα, ανά εκπαιδευτική βαθμίδα ή ανά Ιδιότητα.

Αντίστοιχα, θα μπορούσαν να εξαγάγουν συμπεράσματα για την καταλληλότητα ή τον βαθμό δυσκολίας που παρουσιάζει ένα πρόγραμμα σε συγκεκριμένες ομάδες υπαλλήλων (π.χ. υψηλά ποσοστά ανεπιτυχούς ολοκλήρωσης συγκεκριμένων σεμιναρίων από υπαλλήλους Δευτεροβάθμιας εκπαίδευσης, θα μπορούσε να σημαίνει ότι η ομάδα στόχος θα πρέπει να περιοριστεί σε υπαλλήλους ΤΕ ή ΠΕ, ή ότι η ύλη του προγράμματος θα πρέπει να προσαρμοστεί ώστε να είναι ευκολότερα κατανοητή από υπαλλήλους ΔΕ).

5 ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην εποχή όπου η εξόρυξη δεδομένων αποτελεί πλέον έναν εξαιρετικά διαδεδομένο τρόπο για την ανάλυση, κατανόηση και αξιοποίηση δεδομένων μεγάλης κλίμακας που αλλιώς ίσως παρέμεναν αναξιοποίητα, στην παρούσα εργασία επιχειρήθηκε να αναδειχθούν τρόποι με τους οποίους θα μπορούσαν να αξιοποιηθούν δεδομένα που τηρεί κάποιος εκπαιδευτικός φορέας.

Εξετάστηκαν πέντε διαφορετικές περιπτώσεις πρακτικής εφαρμογής με συλλογή, προετοιμασία και ανάλυση δεδομένων, με βάση τα βήματα του μοντέλου CRISP-DM, από τη φάση της κατανόησης του επιχειρηματικού μοντέλου, ως τη φάση της οπτικοποίησης και της αξιολόγησης των αποτελεσμάτων της μοντελοποίησης προς εξαγωγή συμπερασμάτων και λήψη αποφάσεων.

Εφαρμόστηκαν κάποιες από τις πιο συχνά χρησιμοποιούμενες τεχνικές εξόρυξης και τα συμπεράσματα ήταν ιδιαίτερα ενθαρρυντικά ως προς τη σκοπιμότητα χρήσης παρόμοιων τεχνικών από έναν εκπαιδευτικό φορέα, αφού φάνηκε πως, ακόμα και με χρήση δωρεάν λογισμικού, και χωρίς ιδιαίτερα υψηλές απαιτήσεις σε υπολογιστική ισχύ, οι τεχνικές αυτές μπορούν να βοηθήσουν έναν φορέα να προσεγγίσει τα δεδομένα που ούτως ή άλλως τηρεί στις βάσεις δεδομένων του με διαφορετικό τρόπο και να ανακαλύψει μέσα σε αυτά γνώση η οποία παραμένει κρυμμένη, αλλά που μπορεί να βοηθήσει σημαντικά στην καλύτερη οργάνωση της εκπαιδευτικής διαδικασίας σε όλες τις φάσεις της.

6 Αναφορές

- ANALYTICS INDIA MAGAZINE PVT LTD, 2020. *Top 8 Data Mining Techniques In Machine Learning*, s.l.: s.n.
- B.Tulasi, 2013. Significance of Big Data and Analytics in Higher Education. *International Journal of Computer Applications*, 04.
- Data Science Process Alliance, n.d. *What is CRISP DM?*. [Ηλεκτρονικό]
Available at: <https://www.datascience-pm.com/crisp-dm-2>
- H.J.Watson, 2011. Business Analytics Insight : Hype or Here to stay?. *Buisness Intelligence Journal*.
- Han, J. & Kamber, M., 2001. *Data mining: concepts and techniques*. s.l.:Morgan Kaufmann.
- North, D. M. A., 2012. *Data Mining for the Masses*. s.l.:Global Text Project.
- Press, G., 2016. Cleaning Big Data: Most Time-Consuming, Least Enjoyable Data Science Task, Survey Says. *Forbes*, 23 03.
- SOFTWARETESTINGHELP, 2021. *SOFTWARE TESTING HELP*. [Ηλεκτρονικό]
Available at: https://www.softwaretestinghelp.com/data-mining-techniques/#List_Of_Data_Extraction_Techniques
[Πρόσβαση 21 08 2021].
- Soumen Chakrabarti, M. E. U. F. J. G. J. H. S. M. G. P.-S. W. W., 2006. *Data Mining Curriculum: A Proposal (Version 1.0)*. [Ηλεκτρονικό]
Available at: https://kdd.org/exploration_files/CURMay06.pdf
- Talend, n.d. *What is Data Preparation?*, s.l.: s.n.
- Thomas Erl, W. K. P. B., 2016. Big Data Fundamentals: Concepts, Drivers & Techniques. Στο: *Big Data Fundamentals: Concepts, Drivers & Techniques*. s.l.:Pearson.
- Usama Fayyad, G. P.-S. P. S., 1996. From data mining to knowledge discovery in databases. *AI magazine*, 15 03.
- Wikipedia, 2021. *Data mining*. [Ηλεκτρονικό]
Available at: https://en.wikipedia.org/wiki/Data_mining
- Β. Μεγαλοοικονόμου - Πανεπιστήμιο Πατρών, Ανοικτά Ακαδημαϊκά Μαθήματα ΠΠ, 2015. *Μελέτη Περιπτώσεων στη Λήψη Αποφάσεων: Εξόρυξη Δεδομένων και Ανακάλυψη Γνώσης*, Πάτρα: s.n.
- ΕΚΔΔΑ, 2021. *Εξαμηνιαία έκθεση αξιολόγησης εκπαιδευτικού έργου ΙΝ.ΕΠ - α εξάμηνο 2021*. [Ηλεκτρονικό]
Available at: <https://www.ekdd.gr/wp-content/uploads/2021/11/ΕΚΘΕΣΗ-ΑΞΙΟΛΟΓΗΣΗΣ-Α-ΕΞΑΜ-2021.pdf>
- Χάλαρης, Ι., 2019. *ΜΕΘΟΔΟΛΟΓΙΕΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΕΣ ΒΕΛΤΙΣΤΟΠΟΙΗΣΗΣ ΔΙΑΔΙΚΑΣΙΩΝ ΔΗΜΟΣΙΩΝ ΥΠΗΡΕΣΙΩΝ*, s.l.: s.n.