ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΑΤΡΩΝ
UNIVERSITY OF PATRAS

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΠΠΣ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ ΜΕΣΟΛΟΓΓΙ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΛΟΓΙΣΜΙΚΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΚΑΙ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ - ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ KNIME

Ιωάννου Αριστείδης ΑΜ: 15102

Φαζάκης Νικόλαος ΑΜ: 15327

Επιβλέπων καθηγητής

Αριστογγιάνης Γαρμπής

Μεσολόγγι 2022

UNIVERSITY OF PATRAS

SCHOOL OF ECONOMICS & BUSINESS

DEPARTMENT OF MANAGEMENT SCIENCE AND TECHNOLOGY

FORMER DEPARTMENT OF BUSINESS ADMINISTRATION AT MESSOLONGHI

THESIS

MACHINE LEARNING SOFTWARE AND DATA MINING - CASE STUDY KNIME

Ioannou Aristeidis AM : 15102

Fazakis Nickolas AM: 15327

Messolonghi 2022

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας του Πανεπιστημίου Πατρών δεν υποδηλώνει απαραιτήτως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Τμήματος.

ΠΕΡΙΛΗΨΗ

Στην σημερινή εποχή που ζούμε η τεχνολογία έχει εισβάλει για τα καλά μέσα στην ζωή μας, και ψάχνουμε συνέχεια νέους τρόπους για να την διευκολύνουμε. Ένας τρόπος είναι η δημιουργία μίας διαδικτυακής βάσης που θα μπορεί να επεξεργάζεται τον όγκο δεδομένων που υπάρχουν στο διαδίκτυο , να τα αποθηκεύει και να μας δίνει στατιστικές για μία συγκεκριμένη περίοδο ή ακόμα, να αναλύει κάθε είδους πληροφορία και να τις διαχωρίζει, κρατώντας μόνο τις πιο σημαντικές. Σε αυτή την εργασία θα δημιουργήσουμε μια πλατφόρμα που θα μας βοηθήσει να κάνουμε ακριβός αυτό η οποία ονομάζεται “μηχανική μάθηση και επεξεργασία δεδομένων ”.

Στην προσπάθεια μας αυτή να αναλύσουμε και να εξετάσουμε τι είναι η εξόρυξη δεδομένων και η μηχανική μάθηση, ανατρέξαμε στο διαδίκτυο (Google, Youtube, κ.τ.λ) ανακαλύπτοντας τον τρόπο λειτουργία τους, να καταλάβουμε το λόγο της δημιουργίας τους και τα σημεία που ίσως υστερούν. Ειδικότερα, κατεβάσαμε το πρόγραμμα KNIME αναλύοντας της ποιο σημαντικές ροές δεδομένων και όλες τις μεταβλητές που μπορούν να χρησιμοποιήσουν οι χρήστες. Στην συνέχεια διερευνήσαμε 2 άλλα παρόμοια προγράμματα WEKA & RapidMiner όπου αναφέρουμε αναλυτικά τα χαρακτηριστικά τους.

ABSTRACT

In this day and age technology has invaded our lives for good, and we are constantly looking for new ways to make it easier. One way is to create an online database that can process the amount of data available on the internet, store it and give us statics from a specific period or even analyses any kind of information and separate them, keeping only the most important. In this work we will create a platform that help us do this accurately and its called “machine learning and data processing”.

In this effort to analyse and examine what is data mining and machine learning, we looked on the internet (Google, Youtube ,etc.) to discover how they work, to understand the reason for their creation and the points that they may be lagging behind. In particular, we downloaded the KNIME program analysing its most important data streams and all the variables that users can use. Then we investigated 2 other similar programs WEKA & RapidMiner where we report in detail their features.

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΚΑΤΑΛΟΓΟΣ ΠΙΝΑΚΩΝ

ΚΑΤΑΛΟΓΟΣ ΔΙΑΓΡΑΜΜΑΤΩΝ

ΚΑΤΑΛΟΓΟΣ ΕΙΚΟΝΩΝ

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

[1] https://www.certara.com/

[2] https://www.schrodinger.com/ProductDescription.php?mID=6&sID=33&cID=0

[3] https://www.infocom.co.jp/ja/index.html

[4] http://www.treweren.com/index.php?Item=Home&Width=1920&Height=969&Browser=Netscape

[5] http://www.enalosplus.novamechanics.com/

ΑΠΟΔΟΣΗ ΟΡΩΝ

KNIME : Konstanz Information Miner

ETL : Extraction, Transformation, Loading

JDBC : Java Database Connections

RPA : Robotic process Automation

ERP : Διαχείριση Επιχειρηματικών Πόρων

AI : Τεχνική Νοημοσύνη

OCR : Οπτική Αναγνώριση χαρακτήρων

LDA : Dimensionality Reduction

ΕΙΣΑΓΩΓΗ

Στην παρούσα εργασία διερευνούμε τρία παρόμοια λογισμικά μηχανικής μάθησης και εξόρυξης δεδομένων, όπως είναι το WEKA , KNIME και το RapidMiner. Στο πρώτο κεφάλαιο αναλύουμε την εξόρυξη δεδομένων, τι είναι, πως δημιουργήθηκε και για ποιο σκοπό, ούτως ώστε ο αναγνώστης να αρχίσει να καταλαβαίνει τον σκοπό αυτής της πτυχιακή. Μέσα από την εξόρυξη δεδομένων δημιουργείται η εξόρυξη κειμένου και η εξόρυξη γνώσης/συναισθήματος. Στο δεύτερο κεφάλαιο αναφέρουμε την μηχανική μάθηση, την ιστορία πίσω από τη μηχανική μάθηση , τον ορισμό και σε ποιες κατηγορίες χωρίζεται. Στο τρίτο κεφάλαιο αναλύουμε το πρόγραμμα WEKA, πότε δημιουργήθηκε και από ποιους, τον σκοπό της δημιουργίας του, σε ποιους χρήστες απευθύνετε και τα πλεονεκτήματα-μειονεκτήματα του. Στο Τέταρτο κεφάλαιο περιγράφουμε το πρόγραμμα KNIME, την ιστορία πίσω από αυτό, τον σκοπό που δημιουργήθηκε και τα θετικά- αρνητικά του. Στο 5ο κεφάλαιο αναλύουμε το RapidMiner & AYLIEN , την ιστορία που κρύβεται πίσω από αυτό, τον σκοπό δημιουργίας και συνεργασία τους και τα πλεονεκτήματα-μειονεκτήματα που έχουν. Στο 6ο κεφάλαιο δείχνουμε την πλήρη εγκατάσταση του KNIME βήμα προς βήμα και εξηγώντας αναλυτικά τι κάνουμε. Στο 7ο κεφάλαιο αναφέρουμε τους πιο σημαντικούς κόμβους και μεταβλητές που υπάρχουν μέσα στο πρόγραμμα KNIME, για την κατανόηση του κάθε παραδείγματος που θα δημιουργήσουμε παρακάτω. Στο 8ο κεφάλαιο δημιουργούμε παραδείγματα ροής δεδομένων μέσα στην πλατφόρμα του KNIME. Τέλος κάνουμε μία σύγκριση - αξιολόγηση των προγραμμάτων που αναφέραμε παραπάνω, δηλώνοντας την προσωπική μας άποψη .

ΚΕΦΑΛΑΙΟ 1
Εισαγωγή Στην Θεωρία της Εξόρυξης Δεδομένων

1.0 Εισαγωγή Στην Εξόρυξης Δεδομένων

ΕΙΚΟΝΑ 1. Αποθήκες & Εξόρυξη Δεδομένων

1.1 ΕΙΣΑΓΩΓΗ

Σε αυτό το κεφάλαιο θα μιλήσουμε για κάποια λογισμικά εξόρυξη δεδομένων. Η εξόρυξη δεδομένων είναι μία σύγχρονη τεχνική για να αναλύσουμε ένα πολύ μεγάλο όγκο δεδομένων και να εξάγουμε χρήσιμες πληροφορίες μέσα από αυτά. Αυτό επιτυγχάνετε από την δημιουργία μιας πλατφόρμας όπου αντλεί πληροφορίες που βρίσκονται στις βάσεις δεδομένων και των πληροφοριακών συστημάτων όπως τα ERP τις κάθε επιχείρησης. Η πλατφόρμα αυτή συνήθως χρησιμοποιείται από επιχειρήσεις που έχουν μεγάλο όγκο πληροφορίας και

συσσωρεύονται εκεί συνεχώς. Αυτό συμβαίνει επειδή ο χρήστης πολλές φορές δεν γνωρίζει την δομή και την σημασία των τιμών που βρίσκονται στις βάσεις δεδομένων , ώστε να κάνουν στοχευμένες ερωτήσεις όπως γίνετε στην στατιστική.

Η εξόρυξη δεδομένων ή όπως πολλοί το αποκαλούν εξόρυξη γνώσης , δημιουργήθηκε για των εντοπισμό άγνωστων μέχρι στιγμής προτύπων και τη διαλογή δεδομένων με την χρήση ενός αλγόριθμου, φτιάχνοντας μοντέλα προβλέψεων και συσχετίσεων αναλύοντας τους παράγοντες που παίζουν ρόλο για την επίτευξη τον στόχων τον επιχειρήσεων και όχι μόνο .[1]

1.2 Τι είναι η Εξόρυξη δεδομένων (Data Mining)

ΕΙΚΟΝΑ 2. Data Mining Λογότυπο

Η εξόρυξη δεδομένων ή αλλιώς Data Mining παραπέμπει συνήθως σε μία πλατφόρμα με μεγάλη ποσότητα δεδομένων ή επεξεργασίας δεδομένων (Συλλογή , εξαγωγή δεδομένων , warehouse (αποθήκη) , ανάλυση δεδομένων , στατιστικής) . Για να επιτευχθεί αυτό γίνετε χρήση αλγόριθμων ομαδοποίησης, της μηχανικής μάθησης και των συστημάτων βάσεων δεδομένων. Αν θέλουν να

[1] https://el.wikipedia.org/wiki/Εξόρυξη_δεδομένων

https://repository.kallipos.gr/handle/11419/1227

δώσουμε την κατάλληλη έννοια της λέξης είναι η ανακάλυψη (δηλαδή η ανίχνευση κάτι καινούργιου).

Στην διαχείριση επιχειρηματικών πόρων (ERP), η εξόρυξη δεδομένων θεωρείται ως η στατιστική και λογική ανάλυση εκτεταμένων συνόλων από δεδομένα συναλλαγών και εργασιών για τον εντοπισμό επαναλαμβανόμενων μοτίβων ή τάσεων προκειμένου να βοηθήσουν στην λήψη αποφάσεων (Ellen Monk, Bret Wagner, 2006).[2]

1.2.1 Δημιουργία του Data Mining

Η συνεχής εξέλιξη της τεχνολογίας στην πληροφορική , παρέχει την δυνατότητα αποθήκευσης μεγάλου όγκου δεδομένων, σε αρχεία , βάσεις δεδομένων, στο διαδίκτυο κτλ. Αυτό έδωσε την δυνατότητα στις επιχειρήσεις, υπηρεσίες κ.τ.λ να στραφούν προς τα εκεί. Η εξόρυξη δεδομένων ή όπως πολλοί το αποκαλούν εξόρυξη γνώσης , δημιουργήθηκε για των εντοπισμό άγνωστων μέχρι στιγμής προτύπων και τη διαλογή δεδομένων με την χρήση ενός αλγόριθμου, φτιάχνοντας μοντέλα προβλέψεων και συσχετίσεων αναλύοντας τους παράγοντες που παίζουν ρόλο για την επίτευξη τον στόχων τον επιχειρήσεων και όχι μόνο .[3]

1.2.2 Στόχος του Data Mining

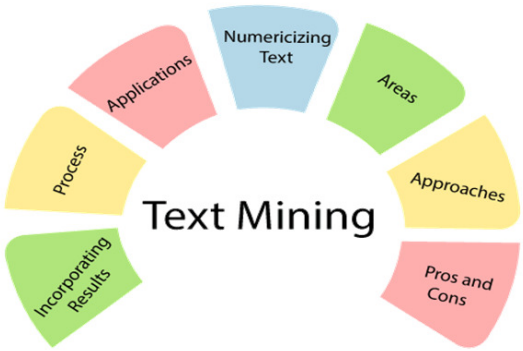
Στόχος της εξόρυξη δεδομένων είναι η ανάλυση μεγάλου όγκου δεδομένων επιλέγοντας τα ποιο σημαντικά στοιχεία τα οποία ήταν άγνωστα μέχρι στιγμής , όπως ομάδες από εγγραφές δεδομένων (συσταδοποίηση), ασυνήθιστες εγγραφές

[2] https://el.wikipedia.org/wiki/Συστήμα_ενδοεπιχειρησιακού_σχεδιασμού
https://el.wikipedia.org/wiki/Εξόρυξη_δεδομένων

[3] https://el.wikipedia.org/wiki/Εξόρυξη_δεδομένων

(anomaly detection) και εξαρτήσεις (κανόνες συσχετίσεων). Αυτές της πληροφορίες τις αντλούμε συνήθως από τις βάσεις δεδομένων (ERP) και μπορούν να θεωρηθούν ως περιγραφή των δεδομένων εισαγωγής, όπου μπορούν να χρησιμοποιηθούν για περαιτέρω ανάλυση. Με την εξόρυξη δεδομένων μπορούμε να προσδιορίσουμε και να εξασφαλίσουμε με μεγαλύτερη ακρίβεια αποτελέσματα από ένα σύστημα υποστήριξης αποφάσεων.[4]

1.3 Εργαλείο Text Mining(Εξόρυξη Κειμένου)



ΕΙΚΟΝΑ 3. Text Mining Λογότυπο

Η εξόρυξη κειμένου ή αλλιώς Text Mining βασίζεται στις παραπάνω τεχνικές και προσπαθεί να επιλύσει το πρόβλημα της υπερφόρτωσης πληροφορίων, βοηθώντας τους χρήστες να εξάγουν αυτόματα πληροφορία κυρίως μέσω της κατηγοριοποίησης και ομαδοποίησης εγγράφων.[5]

1.3.1 Στόχος της εξόρυξης κειμένου

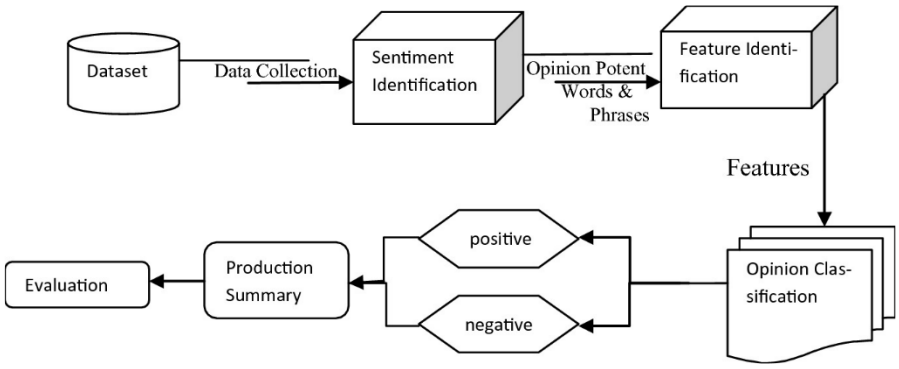
Στόχος της εξόρυξης κειμένου είναι η ανακάλυψη καινούριων πληροφοριών, άγνωστης μέχρι στιγμής και μη καταγεγραμμένης από κάποιο άλλο χρήστη στο

[4] https://el.wikipedia.org/wiki/Εξόρυξη_δεδομένων

[5] https://en.wikipedia.org/wiki/Text_mining

παρελθόν. Οι περισσότεροι χρήστες δεν μπορούν να ξεχωρίσουν πια πληροφορία είναι ήδη γνωστή και πια ανακαλύπτουν (Αυτό είναι και το μοναδικό πρόβλημα που υπάρχει για τον χρήστη). Η διαφορά στην εξόρυξη κειμένου είναι στο γεγονός ότι τα υποδείγματα γνώσης εξάγονται από κείμενο φυσικής γλώσσας και όχι από δομημένες βάσεις των γεγονότων.[6]

1.4 Opinion Mining / Sentiment Analysis(Εξόρυξη Γνώμης/ Συναισθήματος)



ΕΙΚΟΝΑ 4. Εξόρυξη Γνώσης/Συναισθήματος

Η εξόρυξη γνώσης/συναισθήματος είναι ένας πιο εξειδικευμένος κλάδος της επεξεργασίας φυσικής γλώσσας και της εξόρυξη κειμένου. Σημαντικό κομμάτι της συλλογής πληροφοριών που υπήρξε είναι η ανάγκη για γνώση σχετικά με το τι σκέφτονται οι άλλοι για ένα συγκεκριμένο θέμα. Μέσα από το διαδίκτυο Blogs-Forums και άλλα, εμφανίστηκαν νέες αντλήσεις γνώσης και ευκαιρίες προς την κατεύθυνση εξόρυξης συναισθήματος και κατανόησης απόψεων. Τα τελευταία χρόνια η αυτόματη εξόρυξη γνώμης κυρίως μέσω κειμένου προσελκύει περισσότερο την ακαδημαϊκή κοινότητα και τους εταιρικούς οργανισμούς.[4]

[6] https://en.wikipedia.org/wiki/Text_mining
https://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf

Ο όρος **Opinion Mining** εμφανίστηκε για πρώτη φορά το 2003 από μία δημοσίευση των Kushai Dave, Steve Lawrence, David M. Pennock. Σύμφωνα με τους παραπάνω επιστήμονες, η ιδανική χρήση της εξόρυξης συναισθήματος είναι να " μπορεί να επεξεργαστεί ένα σύνολο από δεδομένα αναζήτησης, δημιουργώντας μια λίστα των κύριων χαρακτηριστικών τους και συνοψίζοντας τις απόψεις που επικρατούν για κάθε ένα από αυτά τα χαρακτηριστικά σε θετικές, ουδέτερες και αρνητικές".[7]

Ο όρος **Sentiment Analysis** εμφανίστηκε για πρώτη φορά το 2001 από τους Sanjiv Das, Mike Chen και Richatd M. Tong αναφέροντας ότι συμπίπτει με την εξόρυξη γνώσης και χρησιμοποιείται αναφορικά με την αυτόματη ανάλυση κειμένου προς την πρόβλεψη και αξιολόγησης αυτών.[8]

1.4.1 Στόχος Εξόρυξης Γνώσης/ Συναισθήματος

Η Εξόρυξη γνώσης/ συναισθήματος στοχεύει στο να προσδιορίσει την υποκειμενική στάση του ομιλητή ή του γράφοντα σχετικά με ένα ζήτημα ή την συνολική άποψη που επικρατεί σε ένα έγγραφο. Το αποτέλεσμα κρίνεται από την κρίση ή την αξιολόγηση του ομιλούντα ή γράφοντα.[9]

[7] https://el.wikipedia.org/wiki/Εξόρυξη_γνώμης

[8] https://en.wikipedia.org/wiki/Sentiment_analysis

[9] https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/5306/kalyvae_egovernment.pdf?sequence=3

https://www.ebooks4greeks.gr/epixeirhmatikh-eyfyia-kai-eksoryksh-dedomenwn

ΚΕΦΑΛΑΙΟ 2
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

2.0 Μηχανική Μάθηση

Η μηχανική μάθηση είναι υποπέδιο της επιστήμης των υπολογιστών που αναπτύχθηκε από τη μελέτη της αναγνώρισης προτύπων και της υπολογιστικής θεωρίας μάθησης στην τεχνητή νοημοσύνη. Το 1959, ο **Άρθουρ Σάμουελ** ορίζει τη μηχανική μάθηση ως "Πεδίο μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν, χωρίς να έχουν ρητά προγραμματιστεί". Η μηχανική μάθηση διερευνά τη μελέτη και την κατασκευή αλγορίθμων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Τέτοιοι αλγόριθμοι λειτουργούν κατασκευάζοντας μοντέλα από πειραματικά δεδομένα, προκειμένου να κάνουν προβλέψεις βασιζόμενες στα δεδομένα ή να εξάγουν αποφάσεις που εκφράζονται ως το αποτέλεσμα.

Η Μηχανική μάθηση είναι στενά συνδεδεμένη και συχνά συγχέεται με υπολογιστική στατιστική, που επικεντρώνεται στην πρόβλεψη μέσω της χρήσης των υπολογιστών. Έχει ισχυρούς δεσμούς με την μαθηματική βελτιστοποίηση, η οποία παρέχει μεθόδους, τη θεωρία και τομείς εφαρμογής. Η μηχανική μάθηση εφαρμόζεται σε μία σειρά από υπολογιστικές εργασίες, όπου τόσο ο σχεδιασμός όσο και ο ρητός προγραμματισμός των αλγορίθμων είναι ανέφικτος. Κάποια παραδείγματα είναι τα φίλτρα Spam και η οπτική αναγνώριση χαρακτήρων(OCR), οι μηχανές αναζήτησης και η υπολογιστική όραση.

Στο πεδίο της ανάλυσης δεδομένων, η μηχανική μάθηση είναι μια μέθοδος που χρησιμοποιείται για την επινόηση πολύπλοκων μοντέλων και αλγορίθμων που οδηγούν στην πρόβλεψη. Τα αναλυτικά μοντέλα επιτρέπουν στους ερευνητές, τους επιστήμονες δεδομένων, τους μηχανικούς και τους αναλυτές να παράγουν

αξιόπιστες αποφάσεις και αποτελέσματα , ακόμα και να αναδείξουν την αλληλοσυσχέτιση μέσω της μάθησης από ιστορικές σχέσεις και τάσεις στα δεδομένα.[10]

2.1 Ορισμός

Ο Tom M.Mitchell πρότεινε έναν πιο επίσημο ορισμό που χρησιμοποιείται ευρέως. "Ένα πρόγραμμα υπολογιστή λέγεται ότι μαθαίνει από εμπειρία E ως προς μια κλάση εργασιών T και ένα μέτρο επίδοσης P, αν η επίδοση του σε εργασίες της κλάσης T, όπως αποτιμάται από το μέτρο P, βελτιώνεται με την εμπειρία E". Αυτός ο ορισμός είναι σημαντικός για τον καθορισμό της μηχανικής μάθησης σε βασικό λειτουργικό πλαίσιο παρά με γνωστικούς όρους. Ο Alan Turing έκανε μια πρόταση στην εργασία του " Υπολογιστικές Μηχανές και Νοημοσύνη" λέγοντας, αν οι μηχανές μπορούν να σκέφτονται!.[11]

2.2 Ιστορία

Η μηχανική μάθηση αναπτύχθηκε από την αναζήτηση για την τεχνητή νοημοσύνη , ως επιστημονικό εγχείρημα. Από την πρώιμη περίοδο της έρευνας στον τομέα της τεχνητής νοημοσύνης είχε τεθεί το ζήτημα κατασκευής μηχανών που θα μάθαιναν από δεδομένα . Προσπάθησαν να προσεγγίσουν το πρόβλημα με διάφορες μεθόδους όπως τα μοντέλα και Perceptron's, όπου διαπιστώθηκε αργότερα ότι ήταν επανεφευρέσεις των γενικευμένων γραμμικών μοντέλων της στατιστικής.

[10] https://el.wikipedia.org/wiki/Μηχανική_μάθηση

[11] https://el.wikipedia.org/wiki/Μηχανική_μάθηση

Η μηχανική μάθηση άρχισε να ακμάζει την δεκαετία του 1990. Πήρε αρκετές πληροφορίες από την τεχνητή νοημοσύνη, που στόχο είχαν την επίλυση προβλημάτων πρακτικής φύσης και δίνοντας έμφαση σε μοντέλα και μεθόδους της στατιστικής και της θεωρίας πιθανοτήτων.[12]

2.3 Τύποι Προβλημάτων και Εργασιών

Οι εργασίες μηχανικής μάθησης συνήθως ταξινομούνται σε 3 μεγάλες κατηγορίες (Επιτρεπόμενη Μάθηση- Μη Επιτρεπόμενη Μάθηση- Ενίσχυση Μάθησης), ανάλογα με την φύση του εκπαιδευτικού σήματος ή την ανατροφοδότηση που είναι διαθέσιμα σε ένα σύστημα εκμάθησης. Εμείς θα αναφέρουμε αναλυτικά όλες τις κατηγορίες-υποκατηγορίες που υπάρχουν:

- **Επιβλεπόμενη Μάθηση(Supervised Learning):** Το υπολογιστικό πρόγραμμα δέχεται τις παραδειγματικές εισόδους καθώς και τα επιθυμητά αποτελέσματα από ένα δάσκαλο, και ο στόχος είναι να μάθει έναν γενικό κανόνα προκειμένου να αντιστοιχίσει τις εισόδους με τα αποτελέσματα.
- **Μη Επιβλεπόμενη Μάθηση(Unsupervised Learning):** Χωρίς να παρέχει κάποια εμπειρία στον αλγόριθμο μάθησης, πρέπει να βρεις την δομή των δεδομένων εισόδου. Η μη επιτρεπόμενη μάθηση μπορεί να είναι αυτοσκοπός (ανακαλύπτοντας κρυμμένα μοτίβα σε δεδομένα) ή μέσο για ένα τέλος(χαρακτηριστικό της μάθησης).
- **Ενισχυτική μάθηση:** Ένα πρόγραμμα υπολογιστή αλληλοεπιδρά με ένα δυναμικό περιβάλλον στο οποίο πρέπει να επιτευχθεί ένας συγκεκριμένος στόχος, χωρίς κάποιο δάσκαλο να του λέει ρητά αν έχει φτάσει κοντά στον στόχο του.[13]

[12] https://el.wikipedia.org/wiki/Μηχανική_μάθηση

https://www.csc.com.gr/machine-learning-

[13] https://el.wikipedia.org/wiki/Μηχανική_μάθηση

- **Αναπτυξιακή Μάθηση(Developmental Robotics):** Δημιουργήθηκε για την εκμάθηση από ρομπότ, αναπτύσσει την δική της διαδικασία μαθησιακών καταστάσεων αντλώντας πληροφορίες από ανθρώπους εκπαιδευτές, χρησιμοποιώντας μηχανισμούς καθοδήγησης(Ενεργή μάθηση, Ωρίμανση, Μίμηση).
- **Διαδικασία Εκμάθησης(Meta Learning):** Αυτή η διαδικασία αναπτύσσει τις δικές της επαγωγικές μεθόδους και τις μεταφέρει στις μηχανές εκμάθησης με βάση την προηγούμενη εμπειρία.
- **Μεταγωγή(Transport):** Η μεταγωγή είναι μια παράμετρος τις **μη επιτρεπόμενης μάθησης** , καθώς κατά τον χρόνο εκμάθησης γνωρίζουμε το σύνολο των καταστάσεων του προβλήματος, όμως ένα κομμάτι των στόχων λείπουν.

Άλλο είδος κατηγοριοποίησης είναι:

- **Στην Ταξινόμηση :** Χωρίζει τα δεδομένα εισόδου σε δύο ή περισσότερες κλάσεις, όπου η μηχανή κατασκευάζει ένα μοντέλο που αντιστοιχεί τα δεδομένα σε δύο ή περισσότερες κλάσεις.
- **Στην Συσταδοποίηση :** Χωρίζει τα δεδομένα εισόδου σε ομάδες τις οποίες δεν γνωρίζουμε εξαρχής.
- **Στην Παλινδρόμηση :** τα αποτελέσματα είναι συνεχείς και όχι διακριτά.
- **Στην Εκτίμηση Πυκνότητας :** Μπορείς να βρίσκεις την κατανομή των δεδομένων εισόδου σε ένα συγκεκριμένο χώρο.

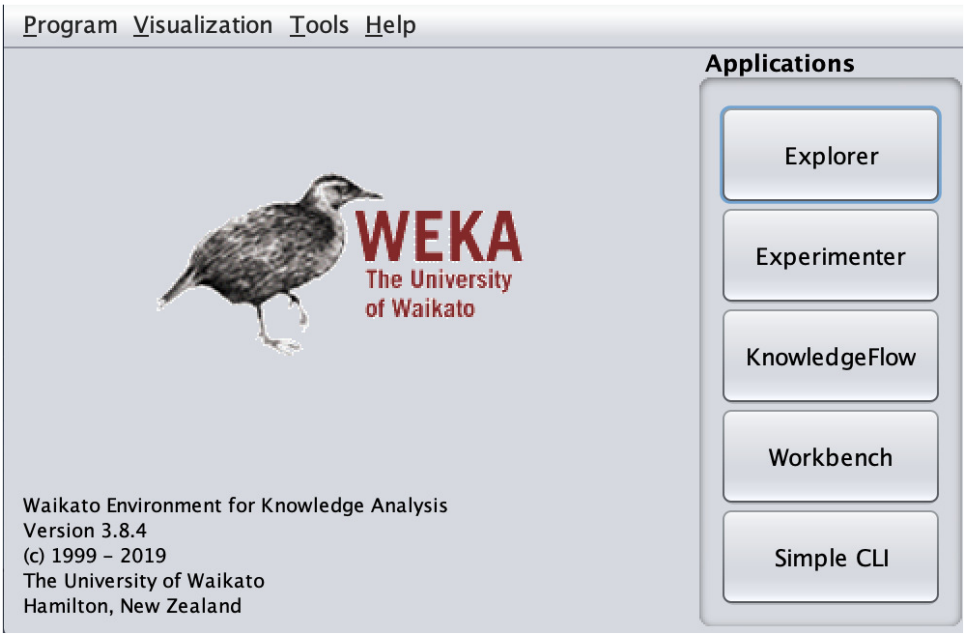
https://people.iee.ihu.gr/~kdiamant/MachineLearning/MachineLearningLesson01.pdf

https://repository.kallipos.gr/handle/11419/3382

- **Προβλήματα Μείωσης Διασπασιμότητας(Dimensionality Reduction) :** Μειώνει των χώρο των δεδομένων απλοποιώντας και αντιστοιχίζοντας τα δεδομένα εισόδου.
- **Στατιστικό Μοντέλο Θεμάτων(Statistic Topic Model):** Αυτό το μοντέλο έχει προσαρμοστεί στην εύρεση εγγράφων όπου καλύπτουν παρόμοια θέματα από ένα σύνολο εγγράφων , που είναι γραμμένα σε φυσική γλώσσα.

ΚΕΦΑΛΑΙΟ 3
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΠΕΡΙΓΡΑΦΗ ΤΟΥ ΠΡΟΓΡΑΜΜΑΤΟΣ WEKA

3. WEKA



ΕΙΚΟΝΑ 5. WEKA Κεντρικό Μενού

3.1 Τι είναι WEKA

Το WEKA είναι ένα λογισμικό για μηχανική μάθηση και εξόρυξη δεδομένων. Δημιουργήθηκε στο πανεπιστήμιο του Waikato της Ν. Ζηλανδίας και διατίθεται ως ελεύθερο λογισμικό(Freeware). Πήρε το όνομα του από το Weka , ένα μικρό και υπό εξαφάνιση πουλί της Ν. Ζηλανδίας. Η πληθώρα μεθόδων εξόρυξης

δεδομένων και η συνεχείς υποστήριξη και εξέλιξη του το έκανε πολύ δημοφιλές.[14]

3.2 Χαρακτηριστικά και Δυνατότητες WEKA

3.2.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ

- Έχει χρησιμοποιηθεί από πολλούς επιστήμονες για την υλοποίηση των εργασιών τους.
- Περιέχει πολλές μεθόδους κατηγοριοποίησης , παλινδρόμησης , ανάλυση συστάδων και κανόνες συσχέτισης , όπου μπορείς να τα επεξεργαστείς όπως εσύ θες.
- Μπορείς επιπλέων να τροποποιήσεις τον αλγόριθμο του αφού είναι λογισμικό ανοικτού κώδικα , ωστόσο αν ξέρεις πως να προγραμματίζεις και θέλεις να τον βελτιώσεις , τότε έχεις αυτή την επιλογή.
- Η γλώσσα που έχει χρησιμοποιηθεί για την κατασκευή του είναι η Java, κάτι που το κάνει να εγκαθίσταται εύκολα σε πλατφόρμες υλικού και λογισμικού.
- Υπάρχει μεγάλη ποικιλία βιβλιοθηκών για μηχανική μάθηση και εξόρυξης δεδομένων , απλά χρειάζεται να γραφτεί κώδικας. Διαθέτει όμως και το γραφικό του περιβάλλον στο οποίο δεν χρειάζονται γνώσεις προγραμματισμού.
- Διαθέτει λογισμικό ανοικτού κώδικα(Freeware), τον οποίο μπορείς να το επεξεργαστείς όπως θες εσύ.
- Το WEKA διαθέτη 2 εκδόσεις, την stable η οποία απευθύνεται σε απλούς χρήστες και την Development στην οποία προορίζεται για τους

[14] https://www.researchgate.net/figure/WEKA-GUI-Chooser-2-HISTORY-OF-WEKA-The-first-release-of-WEKA-was-brought-in-the-market_fig1_266593066

προγραμματιστές , οι οποίοι θέλουν να την διορθώσουν και να την εξελίξουν.

- Έχει διάφορες επιλογές λειτουργικών συστημάτων , Windows, Mac OS X και Linux.
- Επιπλέων σου δίνει την επιλογή να καταλάβεις καλύτερα τον κώδικα, εφόσον σου προσφέρει οδηγίες και απαντήσεις για τυχών προβλήματα που αντιμετωπίζεις.

3.2.2 ΜΕΙΟΝΕΚΤΗΜΑΤΑ

- Δεν επιτρέπει την ενσωμάτωση άλλων εργαλείων
- Δύσκολη επιλογή χαρακτηριστικών που θα λάβουν μέρος στην ανάλυση σε κάθε βήμα, καθώς γίνεται με βάση τον δείκτη όχι την ονομασία της στήλης (στον κόμβο Remove όταν γίνεται χειρωνακτικά)
- Εύχρηστο γραφιστικό περιβάλλον το οποίο χρήζει κάποιας βελτίωσης
- Δεν είναι προφανές ποιοι κόμβοι πρέπει να χρησιμοποιηθούν και σε ποιο σημείο και με ποιους κόμβους μπορούν να συνδεθούν.[15]

3.3 Περιβάλλον WEKA :

- WAIKATO Environment for Knowledge Analysis
- Εργαλείο ανοικτού κώδικα
- Διατίθεται δωρεάν
- Βασίζεται σε γλώσσα Java
- Υλοποιημένοι μέθοδοι για επεξεργασία δεδομένων (Data Pre-processing), ταξινόμηση (Classification)- Παλινδρόμηση (Regression), Συσταδοποίηση (Clustering), Εύρεση κανόνων συσχέτισης (Association

[15] https://www.cs.waikato.ac.nz/ml/weka/book.html

Rules), Επιλογή χαρακτηριστικών (Attribute Selection) και Οπτικοποίηση ( ).[16]

3.4 Πλατφόρμα WEKA

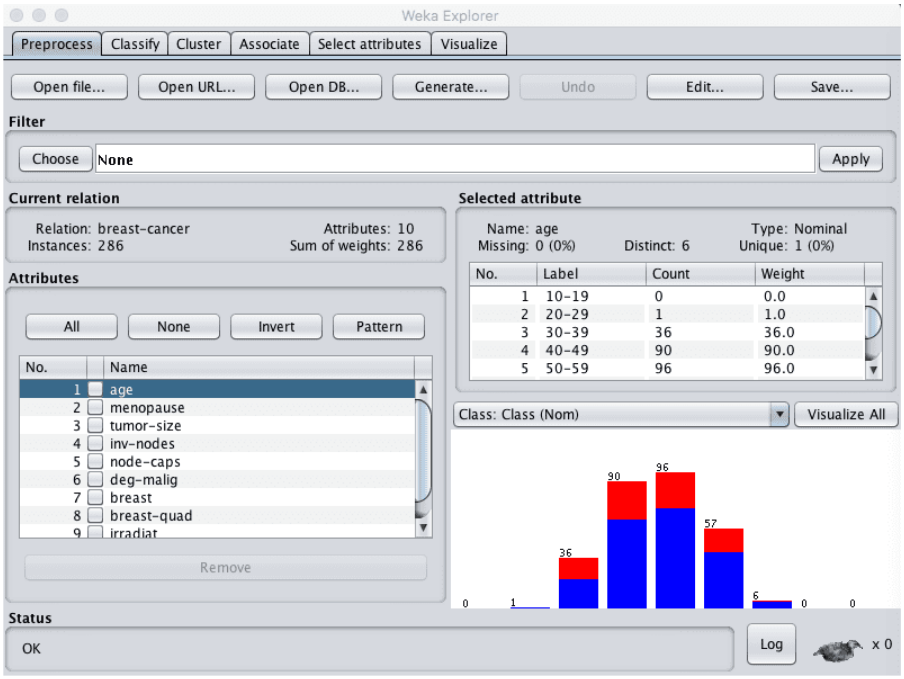
Το WEKA είναι διαθέσιμο από τη διεύθυνση http://old-www.cms.waikato.ac.nz/ml/weka/ . Ο διαχειριστής πακέτων παρουσιάζει μια λίστα πακέτων κοντά στην κορυφή του παραθύρου του και έναν πίνακα στο κάτω μέρος που εμφανίζει πληροφορίες για το τρέχον επιλεγμένο πακέτο στη λίστα. Ο χρήστης μπορεί να επιλέξει πακέτα που είναι διαθέσιμα αλλά δεν έχουν εγκατασταθεί ακόμα, μόνο πακέτα που είναι εγκαταστημένο ή όλα τα πακέτα. Η λίστα παρουσιάζει το όνομα, την κατηγορία, την έκδοση που είναι εγκατεστημένη και ένα πεδίο που υποδεικνύει εάν το πακέτο έχει φορτωθεί επιτυχώς από την WEKA ή όχι.[17]

[16] https://www.cs.waikato.ac.nz/ml/weka/book.html

[17] https://waikato.github.io/weka-wiki/downloading_weka/

https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/weka-gui-learn-machine-learning/

https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf



EIKONA 6. WEKA Explorer

ΚΕΦΑΛΑΙΟ 4
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΘΕΩΡΙΑ ΠΕΡΙΓΡΑΦΗΣ ΤΟΥ KNIME

4. KNIME

ΕΙΚΟΝΑ 7. Λογότυπο KNIME

4.1 Τι είναι το **Knime**

Το KNIME ή αλλιώς το Konstanz Information Miner, είναι μια δωρεάν και ανοιχτού κώδικα πλατφόρμα ανάλυσης δεδομένων, αναφοράς και ενοποίησης. Το KNIME ενσωματώνει διάφορα στοιχεία για τη μηχανική μάθηση και την εξόρυξη δεδομένων μέσω της έννοιας "Building Blocks of Analytics" της αρθρωτής διοχέτευσης δεδομένων. Η γραφική δ επαφή χρήστη και η χρήση του JDBC επιτρέπει τη συναρμολόγηση κόμβων που συνδυάζουν διαφορετικές πηγές δεδομένων, συμπεριλαμβανομένης της προ-επεξεργασίας (ETL: Extraction, Transformation, Loading), για μοντελοποίηση, ανάλυση δεδομένων και οπτικοποίηση χωρίς ή με ελάχιστο προγραμματισμό. Πρόσφατα έγιναν προσπάθειες να χρησιμοποιηθεί το KNIME ως εργαλείο ρομποτικής αυτοματοποίησης διεργασιών (RPA).[18]

[18] https://en.wikipedia.org/wiki/KNIME

4.2 Ιστορία

Η ανάπτυξη του KNIME ξεκίνησε τον Ιανουάριο του 2004 από μια ομάδα μηχανικών λογισμικού στο Πανεπιστήμιο της Konstanz ως ιδιόκτητο προϊόν. Η αρχική ομάδα προγραμματιστών με επικεφαλής τον Michael Berthold προερχόταν από μια εταιρεία στη Silicon Valley που παρείχε λογισμικό για τη φαρμακευτική βιομηχανία. Ο αρχικός στόχος ήταν να δημιουργηθεί μια αρθρωτή, εξαιρετικά επεκτάσιμη και ανοιχτή πλατφόρμα επεξεργασίας δεδομένων που επέτρεπε την εύκολη ενσωμάτωση διαφορετικών μονάδων φόρτωσης, επεξεργασίας, μετασχηματισμού, ανάλυσης και οπτικής εξερεύνησης δεδομένων χωρίς εστίαση σε κάποια συγκεκριμένη περιοχή εφαρμογής. Η πλατφόρμα προοριζόταν να είναι μια πλατφόρμα συνεργασίας και έρευνας και θα έπρεπε επίσης να χρησιμεύσει ως πλατφόρμα ολοκλήρωσης για διάφορα άλλα έργα ανάλυσης δεδομένων. Το 2006 κυκλοφόρησε η πρώτη έκδοση του KNIME και αρκετές φαρμακευτικές εταιρείες άρχισαν να χρησιμοποιούν το KNIME και αρκετοί προμηθευτές λογισμικού επιστήμης της ζωής άρχισαν να ενσωματώνουν τα εργαλεία τους στο KNIME. Αργότερα το ίδιο έτος, μετά από ένα άρθρο στο γερμανικό περιοδικό c't, χρήστες από διάφορες άλλες περιοχές εντάχθηκαν στο πλοίο.

Από το 2006, το KNIME χρησιμοποιείται στη φαρμακευτική έρευνα, χρησιμοποιείται επίσης σε άλλους τομείς όπως η ανάλυση δεδομένων πελατών CRM, η επιχειρηματική ευφυΐα, η εξόρυξη κειμένου και η ανάλυση οικονομικών δεδομένων. Πρόσφατα έγιναν προσπάθειες να χρησιμοποιηθεί το KNIME ως εργαλείο ρομποτικής αυτοματοποίησης διεργασιών (RPA).

Από το 2012, το KNIME χρησιμοποιείται από περισσότερους από 15.000 πραγματικούς χρήστες (δηλαδή χωρίς να υπολογίζονται οι λήψεις αλλά οι χρήστες ανακτούν τακτικά ενημερώσεις όταν γίνονται διαθέσιμες) όχι μόνο στις βιοεπιστήμες αλλά και σε τράπεζες, εκδότες, κατασκευαστές αυτοκινήτων,

τηλεπικοινωνίες, εταιρείες συμβούλων και διάφορες άλλες βιομηχανίες καθώς και σε μεγάλο αριθμό ερευνητικών ομάδων παγκοσμίως. Οι πιο πρόσφατες ενημερώσεις στον KNIME Server και στις επεκτάσεις KNIME Big Data, παρέχουν υποστήριξη για αποθήκευση Apache Spark 2.3, Parquet και HDFS.Για έκτη συνεχή χρονιά, το KNIME τοποθετείται ως ηγέτης για τις Πλατφόρμες Επιστήμης Δεδομένων και Μηχανικής Μάθησης στο Magic Quadrant της Gartner.[19]

4.2 Χαρακτηριστικά και Δυνατότητες **KNIME**

Στην KNIME κατασκευάζουμε λογισμικά για τη δημιουργία και την παραγωγή της επιστήμης δεδομένων, χρησιμοποιώντας ένα εύκολο και διαισθητικό περιβάλλον, επιτρέποντας σε κάθε χρήστη στη διαδικασία της επιστήμης δεδομένων να επικεντρωθεί σε αυτό που κάνει καλύτερα. [20]

Γιατί οι ομάδες χρησιμοποιούν το KNIME:

- Τμήματα FP&A και Ελέγχου: Αυτοματοποιήστε τις οικονομικές αναλύσεις για τη μη αυτόματη συγκέντρωση δεδομένων και το ανθρώπινο λάθος.
- Βιομηχανία Επιστημών Ζωής(Life Sciences Industry): Πρόσβαση, μεταμόρφωση και αλληλεπίδραση με μεγάλες ποσότητες δεδομένων επιστήμης της ζωής με ειδικά σχεδιασμένα εργαλεία.

[19] https://www.knime.com/knime-open-source-story

[20] https://www.knime.com/software-overview

- FSI και τραπεζικός κλάδος : Μεταμορφώστε τον τρόπο με τον οποίο εργάζονται οι ομάδες επιχειρήσεων και δεδομένων για να παρέχουν κορυφαίες υπηρεσίες στον χρηματοοικονομικό κλάδο.[21]

4.2.1 Πλεονεκτήματα

- Πιο εύχρηστο γραφιστικό περιβάλλον για τον μέσο χρήστη.
- Εύκολη σύνδεση κόμβων
- Παρέχει περισσότερη πληροφορία μέσω έτοιμων παραδειγμάτων
- Η τεκμηρίωση του κάθε κόμβου που εισάγεται στο Workflow, μας βοηθά να κατανοήσουμε ευκολότερα την εργασία που εκτελεί κάθε κόμβος.
- Προειδοποιήσεις, μηνύματα και χρώματα που εμφανίζονται απευθείας επάνω στον κόμβο που παρουσιάζει το πρόβλημα βοηθά στην κατανόηση και άμεση διόρθωση του προβλήματος.
- Η επιλογή των χαρακτηριστικών που θα λάβουν μέρος στην ανάλυση σε κάθε βήμα είναι πιο εύκολη καθώς γίνεται με βάση την ονομασία της στήλης.
- Διαισθητικό περιβάλλον που έχει ως αποτέλεσμα καλύτερη εμπειρία χρήσης.
- Υποστηρίζει την ενσωμάτωση εκατοντάδων εργαλείων (Επεκτάσεις) extensions: Plugins, Modules.
- Δίνει τη δυνατότητα στον χρήστη να περιορίσει τις εγγραφές που θα χρησιμοποιηθούν για την εκπαίδευση του αλγορίθμου, ώστε η διαδικασία να εκτελεστεί πιο γρήγορα.
- Έχει αρκετά μεγάλη χωρητικότητα σε σχέση με άλλα παρόμοια προγράμματα.

[21] https://www.knime.com/software-overview

- Έχει διάφορες επιλογές λειτουργικών συστημάτων , Windows, Mac OS X και Linux.
- Διαθέτει λογισμικό ανοικτού κώδικα(Freeware), τον οποίο μπορείς να το επεξεργαστείς όπως θες εσύ.
- Βασίζεται στην γλώσσα προγραμματισμού Java.[22]

4.2.2 Μειονεκτήματα

- Λιγότεροι αλγόριθμοι παλινδρόμησης
- Περιορίζονται σε εργασίες ταξινόμησης, αφού επιτρέπουν μόνο ονομαστικές μεταβλητές ως μεταβλητές στόχους
- Η γραμμική παλινδρόμηση, το Random Forest και το δέντρο αποφάσεων προσφέρουν λιγότερες δυνατότητες παραμετροποίησης.
- Υστερεί στις μεθόδους επιλογής χαρακτηριστικών
- Σημαντικός περιορισμός η μετατροπή της μεταβλητής Creator, από ονομαστική σε αριθμητική με τιμές 0,1.[23]

4.3 Περιβάλλον KNIME :

- Konstanz Information Miner
- Πλατφόρμα ανάλυσης δεδομένων ανοιχτού κώδικα
- Ενσωμάτωση νέων αλγόριθμων και μεθόδων επεξεργασίας δεδομένων
- Διαδραστική εκτέλεση μίας ροής δεδομένων και οπτική αναπαράσταση
- Πολλά διαφορετικά λειτουργικά συστήματα (32-64 bit) Windows, Linux & MAC

[22] https://www.knime.com/software-overview

[23] https://www.knime.com/software-overview

- Βασίζεται σε πλατφόρμα ανοικτού κώδικα Eclipse και την Java.[24]

4.4 KNIME Analytics Platform

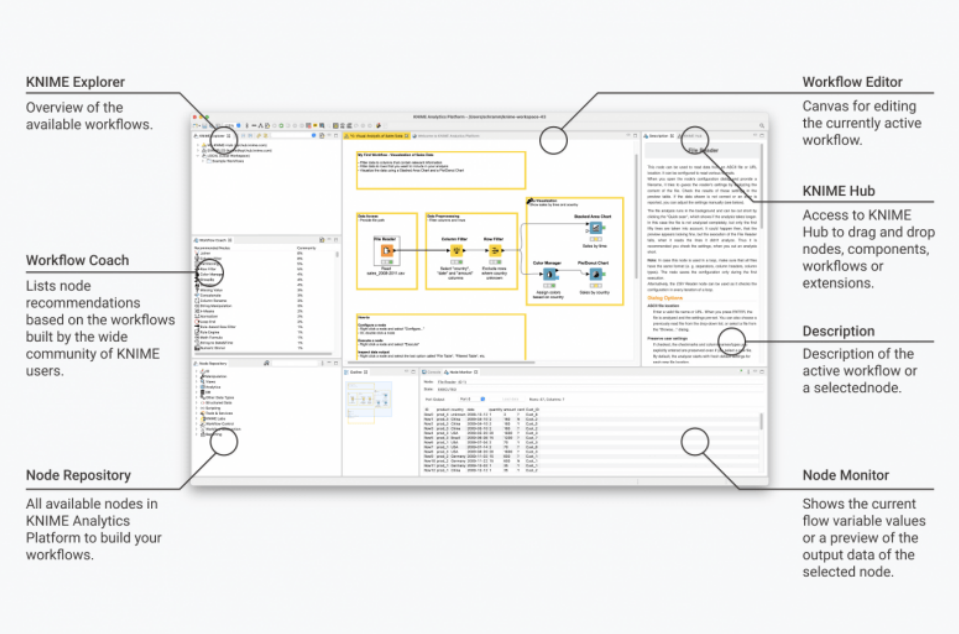
Η πλατφόρμα KNIME δημιουργεί τα επιστημονικά δεδομένα που σας βοηθά να ανακαλύψετε τις δυνατότητες που κρύβονται στα δεδομένα σας, τις δικές μου νέες πληροφορίες ή να προβλέψετε νέες δυνατότητες. Είναι γρήγορη στην ανάπτυξη και εύκολη στην κλίμακα και διαισθητική στην εκμάθηση. Είναι διαθέσιμη στην σελίδα https://www.knime.com/getting-started-guide και σας δίνει αρκετές λεπτομέρειες πώς να ξεκινήσετε.[8][25]

[24] https://polynoe.lib.uniwa.gr/xmlui/bitstream/handle/11400/156/Kotsaki_186682006_presentation.pdf?sequence=2&isAllowed=y

[25] https://www.knime.com/getting-started-guide

https://www.gartner.com/reviews/market/data-science-machine-learning-platforms/vendor/knime/product/knime-analytics-platform/likes-dislikes

https://www.knime.com/sites/default/files/2021-03/KNIME%20Beginner%27s%20Luck%204.3_20210219_sample.pdf



EIKONA 8. KNIME Analytics Platform

ΚΕΦΑΛΑΙΟ 5
ΕΙΣΑΓΩΓΗ ΣΤΗΝ ΘΕΩΡΙΑ ΤΟΥ RAPIDMINER & AYLIEN

5. RapidMiner + AYLIEN

ΕΙΚΟΝΑ 9. RapidMiner + Aylien Λογότυπο

Σε αυτό το κεφάλαιο θα αναφέρουμε τι είναι το RapidMiner , τι μπορεί να μας προσφέρει χαρακτηριστικά - πλεονεκτήματα- Μειονεκτήματα και μια γενική ιδέα για τις λειτουργίες του.[26]

27

5.1 Τι είναι RapidMiner + AYLIEN

[26] https://www.google.gr/search?q=rapidminer+release+date&sxsrf=APq-WBvOyk5fW1fESD8I7r5m92oV2h0q0w%3A1648750598114&ei=BvBFYufFBoWK9u8PrZCG2Ag&ved=0ahUKEwjn4OTN-vD2AhUFhf0HHS2IAYsQ4dUDCA4&uact=5&oq=rapidminer+release+date&gs_lcp=Cgdnd3Mtd2l6EAMyBwgjELADECcyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsANKBAhBGABKBAhGGABQAFgAYMUZaAFwAXgAgAEAiAEAkgEAmAEAyAEJwAEB&sclient=gws-wiz

Το RapidMiner είναι ένα δωρεάν λογισμικό ανοιχτού κώδικα για Data Science, Data Mining, Text Mining, Predictive analytics και πολλά άλλα. Έχει πολλές δυνατότητες και μπορείς να βρεις ακόμα περισσότερες επεκτάσεις μετά την εγκατάσταση του, πολλές από αυτές είναι δωρεάν. Το RapidMiner έχει την επέκταση από την εταιρεία AYLIEN που ειδικεύεται στη Sentiment analytics αλλά και άλλες λειτουργίες επεξεργασίας κειμένου π.χ Language detection, topic detection κ.τ.λ.[28]

5.2 Χαρακτηριστικά και Δυνατότητες RapidMiner

Η πλατφόρμα επιστήμης δεδομένων της RapidMiner προσφέρει μετασχηματιστικό επιχειρηματικό αντίκτυπο για περισσότερους από 40.000+ οργανισμούς σε κάθε κλάδο, με στόχο την μείωση του κόστους και την αποφυγή κινδύνων.

Η RapidMiner είναι μια λύση επιστήμης δεδομένων βασισμένη σε σύννεφο και εσωτερικής εγκατάστασης, η οποία βοηθά τους μικρούς έως και μεγάλους οργανισμούς να έχουν πρόσβαση, να φορτώνουν και να αναλύουν δομημένα και μη δομημένα δεδομένα. Τα βασικά χαρακτηριστικά περιλαμβάνουν αυτοματοποίηση διεργασιών, επικύρωση μοντέλου, σύνδεση δεδομένων και ανάμειξη.

Η εφαρμογή συνοδεύεται από ένα οπτικό εργαλείο μεταφοράς, απόθεσης και βιβλιοθήκη μηχανικής εκμάθησης, η οποία επιτρέπει στους προγραμματιστές να δημιουργούν και να αναπτύσσουν μοντέλα πρόβλεψης. Οι προγραμματιστές μπορούν να οπτικοποιήσουν την απόδοση/σταθερότητα των συνόλων δεδομένων

[28] https://www.softwareadvice.com/

και την αφαίρεση δεδομένων χαμηλής ποιότητας. Επιπλέων βοηθά τους χρήστες να εντοπίζουν προβλήματα δεδομένων, όπως συσχετίσεις, τιμές που λείπουν κτλ.

Άλλο ένα θετικό είναι ότι περιλαμβάνει την δυνατότητα βαθμολόγησης που επιτρέπει στους αναλυτές να εντοπίζουν κατασκευαστικά ελαττώματα στα μοντέλα και τους οικονομικούς κινδύνους. Βέβαια για να το αποκτήσεις , υπάρχει μια ετήσια συνδρομή και η υποστήριξη παρέχεται μέσω E-mail, τηλεφώνου, συνομιλίας και τεκμηρίωσης.

Σε σύγκριση με το Knime, το RapidMiner είναι πιο εύκολο στην χρήση , παρά τις ομοιότητες. Επίσης το RapidMiner δίνει αυτόματες λύσης και συμβουλές όταν κάτι δεν πάει καλά με την διαδικασία που κάνεις.[29]

Περιβάλλον RapidMiner :

- Μετασχηματιστικός επιχειρηματικός αντίκτυπος
- Αναβάθμιση του οργανισμού σας
- Βάθος για τους επιστήμονες, απλοποιημένο για όλους τους άλλους.
- Εύκολη εμπιστοσύνη, συντονισμός και εξήγηση
- Μελλοντική καινοτομία, φορητότητα και επεκτασιμότητα.

5.2.1 Πλεονεκτήματα

- Το RapidMiner είναι ένα Open-source software(Λογισμικό ανοικτού κώδικα) που χρησιμοποιείται για την εξόρυξη δεδομένων και κειμένου για την επιστημονική και εμπορική χρήση.

[29] https://rapidminer.com/why-rapidminer/

- Είναι ένα λογισμικό αναγνωρισμένο σε παγκόσμια κλίμακα. Μεταξύ των χρηστών είναι οι εταιρίες Ford, Honda, Nokia και πολλές άλλες μεγάλες και μεσαίες.
- Το RapidMiner χρησιμοποιείται και για έρευνα και διεργασίες σε πραγματικής φύσεως δεδομένα παγκόσμια.
- Είναι αρκετά εύκολο στην χρήση του, σε σύγκριση με άλλα παρόμοια προγράμματα.
- Επίσης έχει την δυνατότητα βαθμολόγησης του προγράμματος για να γνωρίζουν οι ερευνητές που υστερεί, τυχών κατασκευαστικά λάθη και ελαττώματα.
- Βασίζεται στην γλώσσα προγραμματισμού Java
- Περιλαμβάνεται μια εσωτερική xml αναπαράσταση ώστε να εξασφαλίζεται η τυποποιημένη μορφή ανταλλαγής εξόρυξης δεδομένων σε διάφορα πειράματα.
- Εξασφαλίζεται η αποτελεσματική διαχείριση των δεδομένων αφού υπάρχει δυνατότητα προβολής αυτών σε πολλά επίπεδα.
- GUI, γραμμή εντολών Mode (λειτουργία batch) και Java API για την χρήση του από άλλα προγράμματα.
- Περιλαμβάνει μεγάλη ποικιλία από Plugins
- Μια μεγάλη σειρά αναπαράστασης των δεδομένων με λεπτομερή διάσταση.[30]

5.2.2 Μειονεκτήματα

[30] https://en.wikipedia.org/wiki/RapidMiner

- Το RapidMiner δίνει αυτόματες λύσεις και συμβουλές στον χρήστη όταν η διαδικασία που κάνει είναι λανθασμένη, αλλά για να το αποκτήσεις υπάρχει μια ετήσια συνδρομή.
- Μειωμένη ικανότητα καταμερισμού
- Προαπαιτούμε γνώση βάσεων δεδομένων
- Δεν διαθέτει μεγάλο αποθηκευτικό χώρο, σε σύγκριση με άλλα παρόμοια προγράμματα.[31]

5.3 5 Artificial Intelligence (AI) Types,Defined

Η τεχνητή νοημοσύνη (AI) επαναπροσδιορίζει τις ιδέες της επιχείρησης σχετικά με την εξαγωγή πληροφοριών από δεδομένα. Η συντριπτική πλειοψηφία των στελεχών τεχνολογίας 91% και 84% του ευρύτερου κοινού πιστεύουν ότι η τεχνική νοημοσύνη είναι η επόμενη τεχνολογική επανάσταση σύμφωνα με την Έρευνα Τεχνητής Νοημοσύνης (AI) του Edelman το 2019.[32]

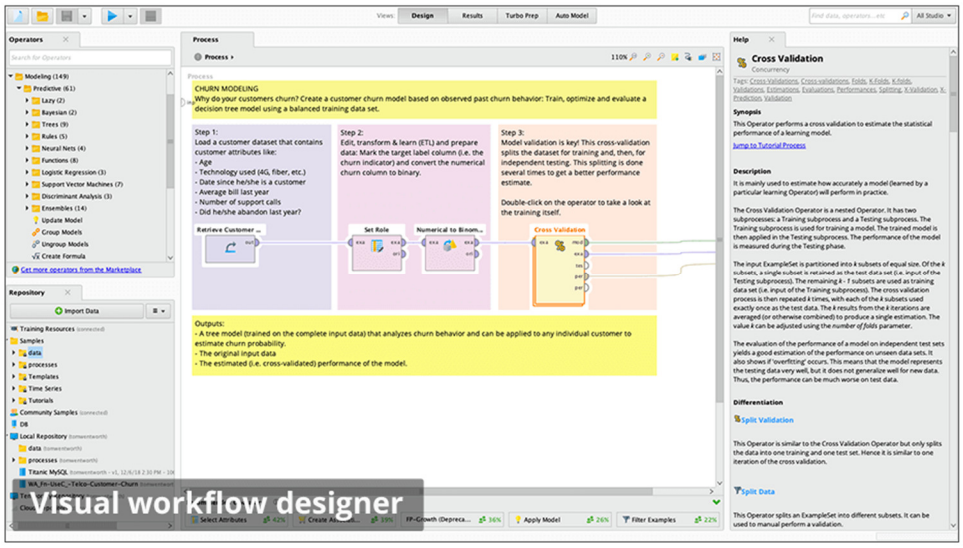
- Μηχανική Εκμάθηση (ML)
- Βαθιά Μάθηση (DP)
- Επεξεργασία Φυσικής Γλώσσας (NLP)
- Computer Vision (Όραση Υπολογιστή) (CV)
- Επεξηγήσημη Τεχνική Νοημοσύνη (XAI)

5.4 Πλατφόρμα RapidMiner

[31] https://en.wikipedia.org/wiki/RapidMiner

[32] https://en.wikipedia.org/wiki/RapidMiner

Η πλατφόρμα RapidMiner έχει μία πλούσια βιβλιοθήκη με 1500+ αλγόριθμους και λειτουργίες όπου σας εξασφαλίζει ένα από τα καλύτερα μοντέλα για κάθε περίπτωση χρήσης προκατασκευασμένων προτύπων για περιπτώσεις κοινής χρήσης και πολλών άλλων. Επιπλέων εμπεριέχει το Wisdom of Crowds το οποίο παρέχει προληπτικές συστάσεις σε κάθε βήμα για την διευκόλυνση των αρχάριων.[10][33]



ΕΙΚΟΝΑ 10. RapidMiner Πλατφόρμα

[33] https://rapidminer.com/products/

https://www.google.gr/search?q=rapidminer+release+date&sxsrf=APq-WBvOyk5fW1fESD8I7r5m92oV2h0q0w%3A1648750598114&ei=BvBFYufFBoWK9u8PrZCG2Ag&ved=0ahUKEwjn4OTN-vD2AhUFhf0HHS2IAYsQ4dUDCA4&uact=5&oq=rapidminer+release+date&gs_lcp=Cgdnd3Mtd2l6EAMyBwgjELADECcyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsAMyBwgAEEcQsANKBAhBGABKBAhGGABQAFgAYMUZaAFwAXgAgAEAiAEAkgEAmAEAyAEJwAEB&sclient=gws-wiz

https://docs.rapidminer.com/latest/studio/operators/rapidminer-studio-operator-reference.pdf

ΚΕΦΑΛΑΙΟ 6
ΕΓΚΑΤΑΣΤΑΣΗ KNIME

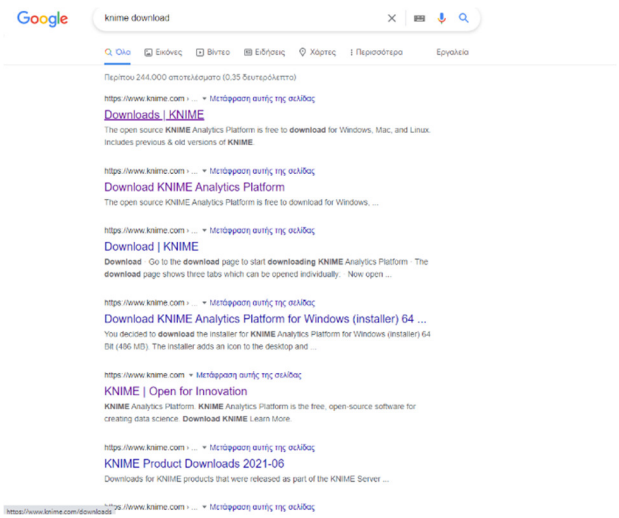
6.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα μιλήσουμε και θα δείξουμε πως γίνετε η εγκατάσταση του KNIME βήμα προς βήμα και αναλύοντας το καθένα από αυτά.

6.2 Εγκατάσταση Βήμα προς Βήμα του KNIME

ΒΗΜΑ 1ο. Google Search

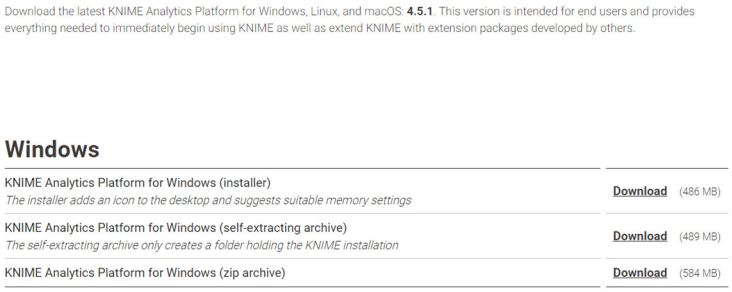
Πληκτρολογούμε σε ένα ιστό-τοπο πχ Google- Mozilla κτλ. Download KNIME, μας εμφανίζει τις σελίδες στις οποίες μπορούμε να μπούμε και να το κατεβάσουμε. Εμείς επιλέγουμε Download KNIME Analytics Platform.[11]



ΕΙΚΟΝΑ 11. Google Search KNIME Download

ΒΗΜΑ 2ο. KNIME Platform

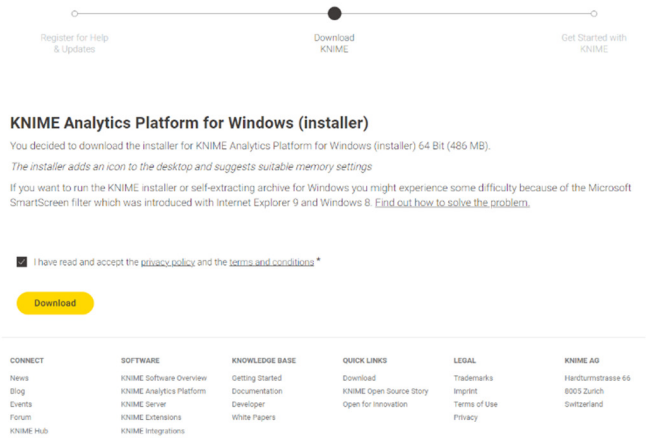
Σε αυτό το βήμα, εμφανίζονται 3 επιλογές για τον τρόπο που θέλουμε να το κατεβάσουμε. Επιπλέων αναφέρει την τελευταία έκδοση και σε ποια συστήματα λειτουργεί(Linux, Windows, macOS).[12]



EIKONA 12. KNIME Platform windows installer URL

ΒΗΜΑ 3ο. Installer

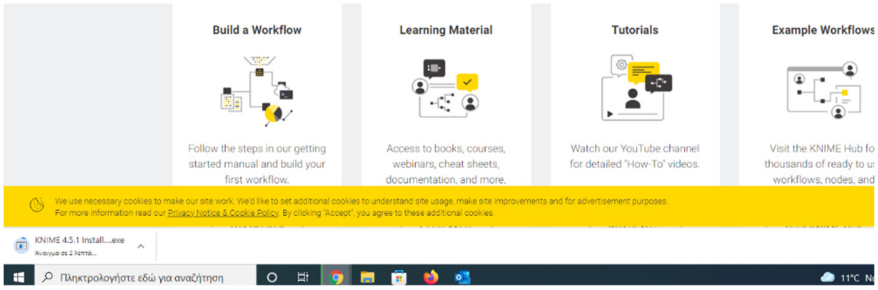
Στο τρίτο βήμα αφού έχουμε επιλέξει με ποιόν από τους 3 τρόπους θέλουμε, μας πετάει στην επόμενη σελίδα για να συμφωνήσουμε με τους όρους και πατάμε Download.[13]



EIKONA 13. KNIME Platform Installer

ΒΗΜΑ 4ο. KNIME.EXE Downloading

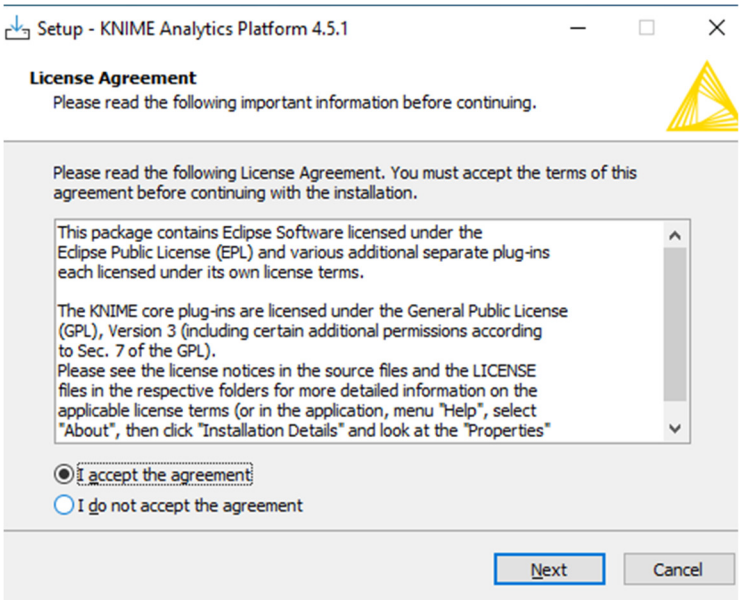
Στο τέταρτο βήμα αρχίζει να κατεβαίνει το αρχείο, και μόλις το κατεβάσει πατάμε πάνω για να το ανοίξουμε.[14]



ΕΙΚΟΝΑ 14. KNIME.exe

ΒΗΜΑ 5ο. Licence Agreement

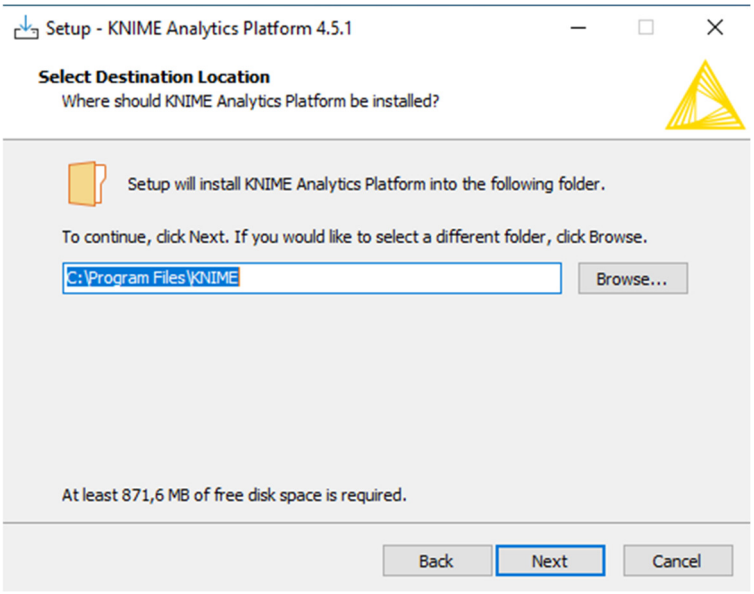
Σε αυτό το βήμα αποδεχόμαστε την συμφωνία και πατάμε Next.[15]



ΕΙΚΟΝΑ 15. Licence Agreement

ΒΗΜΑ 6ο. Setup

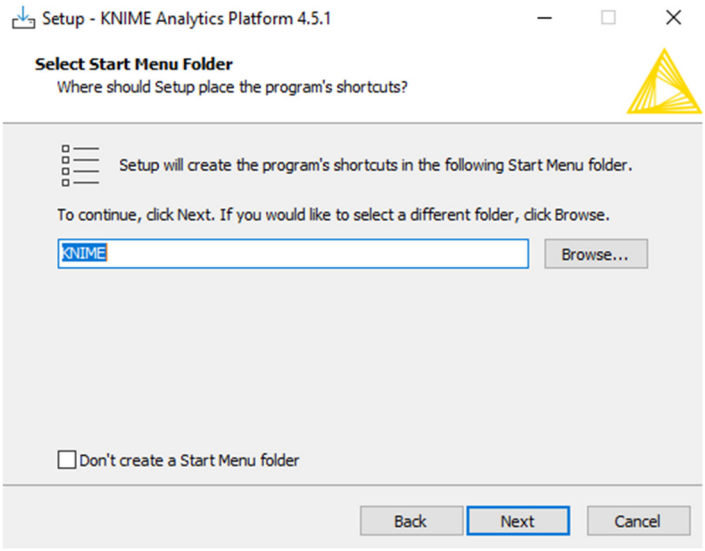
Στο έκτο βήμα επιλέγουμε την διαδρομή στην οποία θέλουμε να αποθηκευτή το αρχείο, για να μπορούμε να το βρίσκουμε εύκολα.[16]



ΕΙΚΟΝΑ 16. Setup KNIME Platform

ΒΗΜΑ 7ο . Επιλογή Ονόματος Αρχικού Μενού Φάκελου

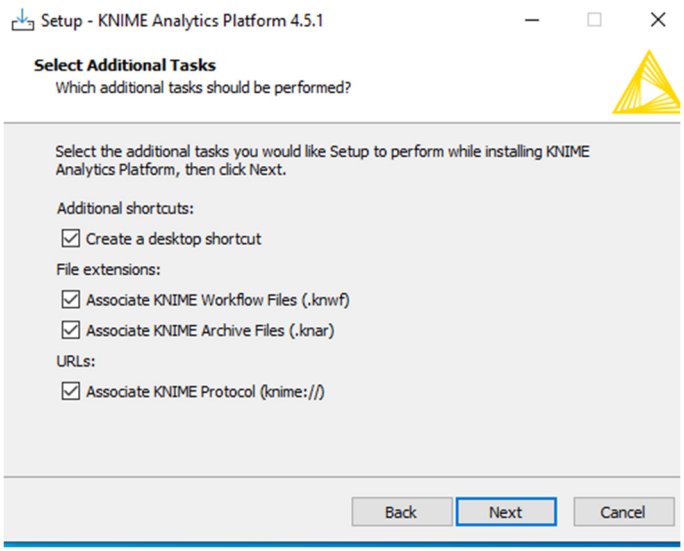
Σε αυτό το βήμα επιλέγουμε το όνομα που θέλουμε στον φάκελο της επιφάνειας εργασίας.[17]



ΕΙΚΟΝΑ 17. Επιλογή Ονόματος Φάκελου

ΒΗΜΑ 8ο. Επιλογή Συντομεύσεων

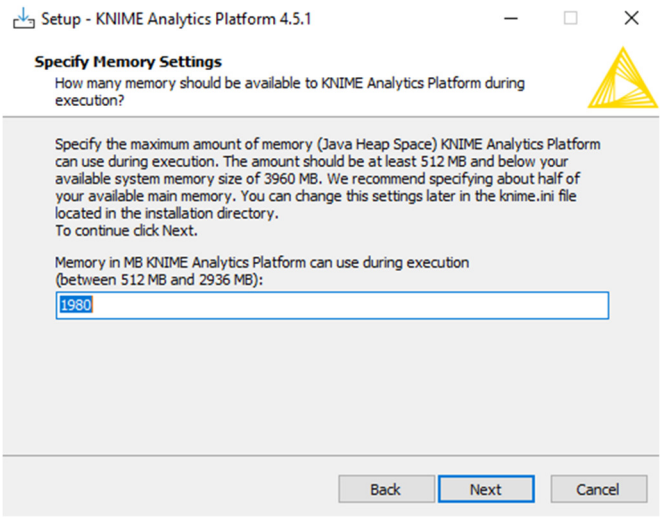
Στο 8ο βήμα επιλέγουμε αν θέλουμε συντομεύσεις και σε ποιους φακέλους.[18]



ΕΙΚΟΝΑ 18. Επιλογή Συντομεύσεων

ΒΗΜΑ 9ο. Επιλογή Χώρου Αποθήκευσης

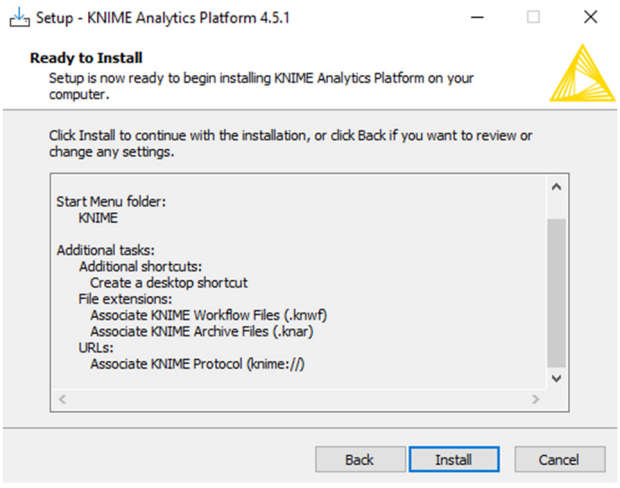
Σε αυτό το σημείο πόσο χώρο θέλουμε να διαθέσουμε σε κάθε εργασία που κάνουμε, αν και μπορούμε να το αλλάξουμε αφότου το έχουμε κάνει εγκατάσταση.[19]



ΕΙΚΟΝΑ 19. Επιλογή Χώρου αποθήκευσης

ΒΗΜΑ 10ο. Έναρξη Εγκατάσταση

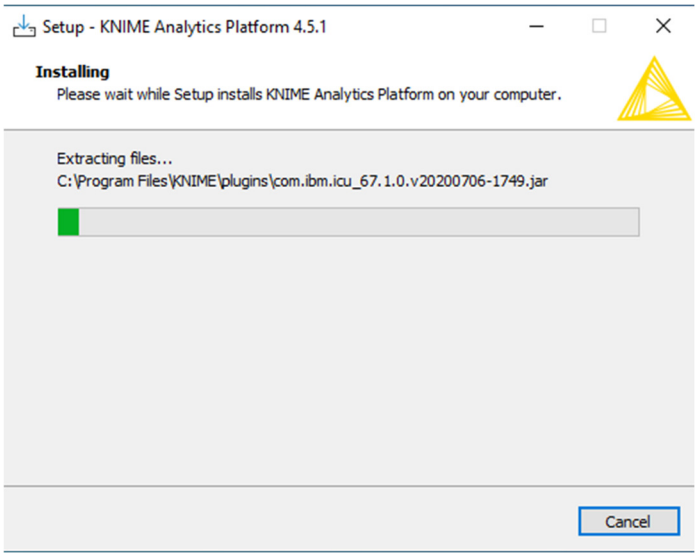
Στο βήμα αυτό απλά πατάμε Install KNIME.[20]



ΕΙΚΟΝΑ 20. Έναρξη Εγκατάστασης KNIME

ΒΗΜΑ 11ο. Extracting Files

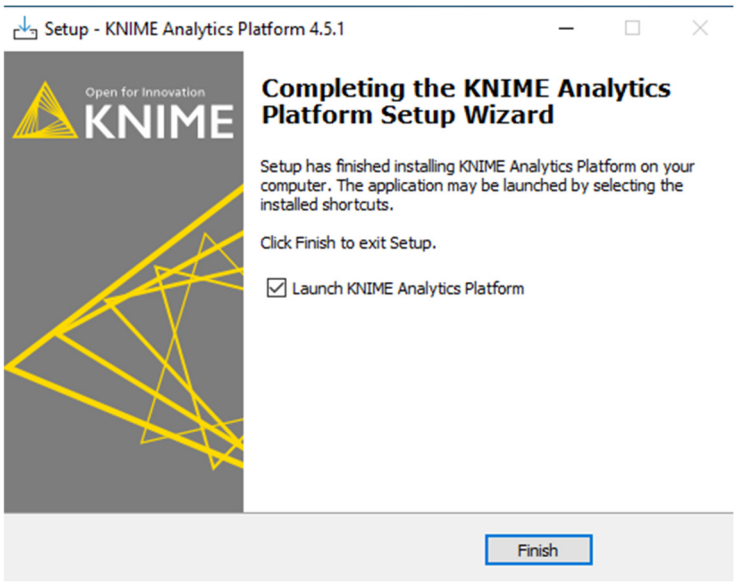
Σε αυτό το βήμα απλά περιμένουμε να φορτώσει η μπάρα.[21]



ΕΙΚΟΝΑ 21. Διαδικασία εγκατάστασης KNIME

ΒΗΜΑ 12ο. Τέλος Εγκατάστασης

Στο δωδέκατο βήμα απλά πατάμε Finish.[22]



ΕΙΚΟΝΑ 22. Τέλος Εγκατάστασης KNIME

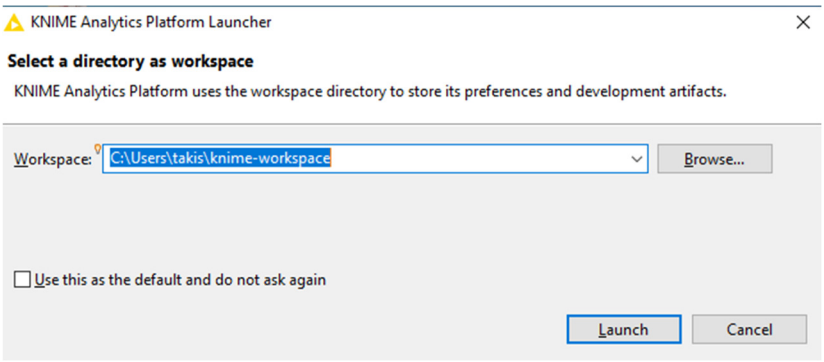
ΒΗΜΑ 13ο. KNIME Version + Original Image

Μόλις τελειώσει, και ξεκινάει να ανοίξει την πλατφόρμα, μας εμφανίζει την τελευταία έκδοση του και την ακριβή ημερομηνία.[23]

ΕΙΚΟΝΑ 23. KNIME Version

ΒΗΜΑ 14ο. Εκκίνηση

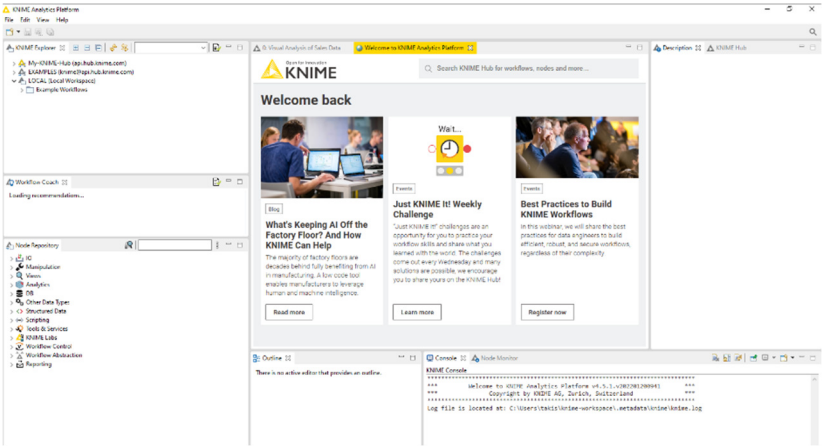
Σε αυτό το βήμα απλά πατάμε Launch και αν θέλουμε να μην ξαναρωτήσει, τσεκάρουμε το κουτάκι στην μέση αριστερά που λέει (Use this as the default and do not ask again).[24]



ΕΙΚΟΝΑ 24. Εκκίνηση της πλατφόρμας KNIME

ΒΗΜΑ 15ο. Platform KNIME

Στο τέλος απλά μας ανοίγει η πλατφόρμα KNIME.[25]



ΕΙΚΟΝΑ 25. Πλατφόρμα KNIME

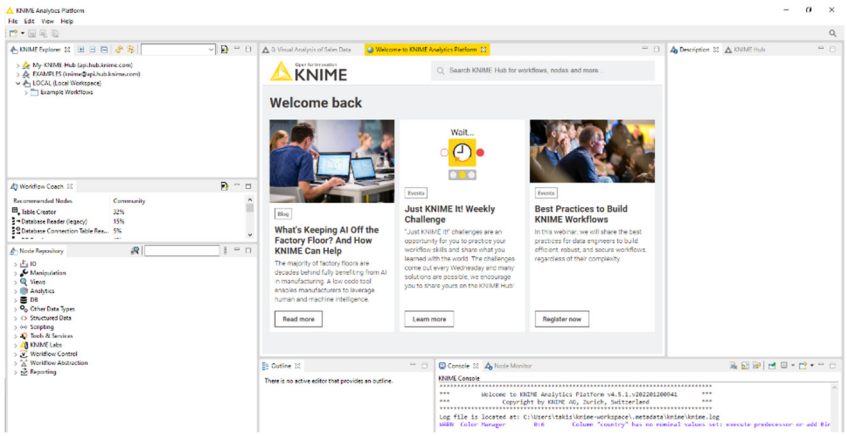
ΚΕΦΑΛΑΙΟ 7
ΟΔΗΓΙΕΣ ΧΡΗΣΗΣ ΤΗΣ ΠΛΑΤΦΟΡΜΑΣ KNIME

Εισαγωγή

Σε αυτό το κεφάλαιο θα αναλύσουμε το περιβάλλον και τις δυνατότητες του KNIME. Αναλυτικότερα θα αναφέρουμε τι μπορεί να μας παρέχει το μενού, και θα εξηγήσουμε τι κάνει η κάθε επιλογή από το μενού . Μπορούμε να το ονομάσουμε και εγχειρίδιο χρήσης.

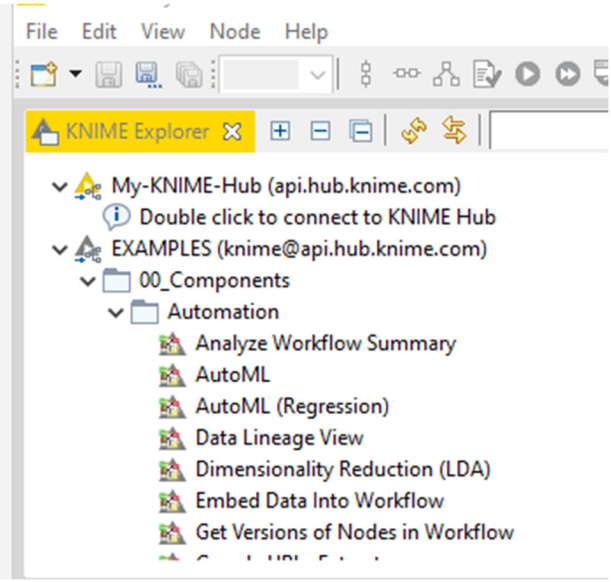
7.1 Αρχικό Μενού.

Στην εικόνα παρακάτω βλέπουμε το αρχικό μενού όταν το ανοίγουμε πρώτη φορά. Αρχικά μας δίνει τρείς επιλογές για να μας καλωσορίσει στην Πλατφόρμα. Το ένα είναι Blog και τα άλλα δύο είναι Events, τα οποία έχουν σκοπό να καταλάβουμε καλύτερα το πρόγραμμα.[26]



ΕΙΚΟΝΑ 26. KNIME Αρχικό Μενού

7.2 My-KNIME-Hub



ΕΙΚΟΝΑ 27. KNIME Παράδειγμα ροής δεδομένων

7.2.1 Παράδειγμα KNIME Μενού

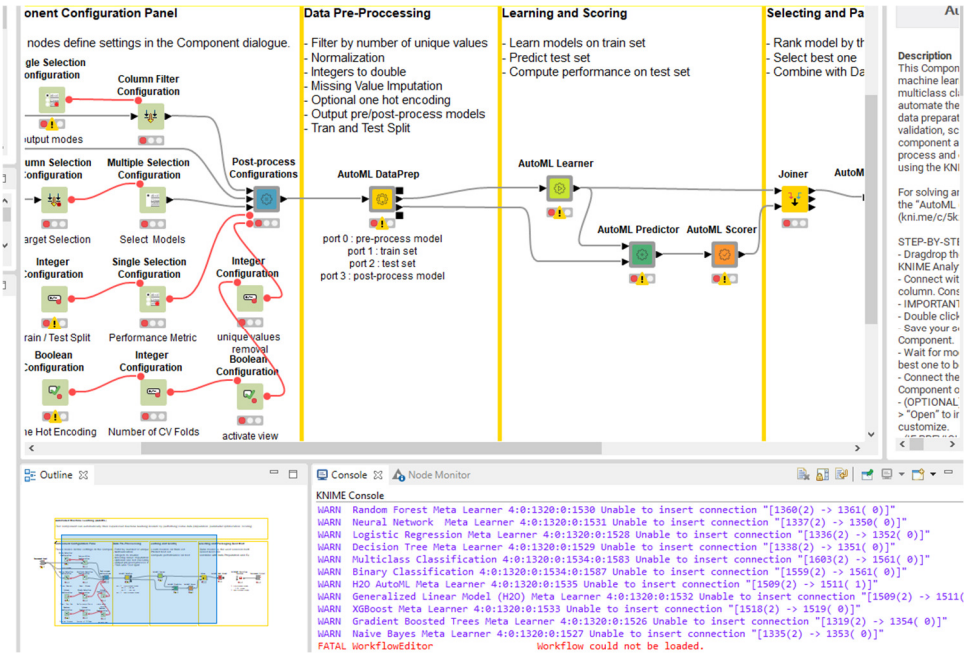
Στο μενού αριστερά υπάρχουν 3 κατηγορίες, η πρώτη κατηγορία είναι το My-KNIME-Hub όπως φαίνεται στην παρακάτω εικόνα. Σε αυτή την κατηγορία υπάρχουν έτοιμα παραδείγματα από την KNIME, για να μπορούμε να καταλάβουμε λίγο πως λειτουργεί.[27]

7.2.2 KNIME AutoML

Περιγραφή : Αυτό το στοιχείο εκπαιδεύει αυτόματα εποπτευόμενα μοντέλα μηχανικής εκμάθησης τόσο για δυαδική όσο και για πολύκλαδή ταξινόμηση. Το στοιχείο είναι σε θέση να αυτοματοποιήσει ολόκληρο τον κύκλο ML εκτελώντας κάποια προετοιμασία δεδομένων, βελτιστοποίηση παραμέτρων με διασταυρούμενη επικύρωση, βαθμολόγηση, αξιολόγηση και επιλογή. Το στοιχείο καταγράφει επίσης ολόκληρη τη διαδικασία από άκρο σε άκρο και

εξάγει τη ροή εργασιών ανάπτυξης χρησιμοποιώντας την ενσωματωμένη επέκταση ανάπτυξης KNIME. [28]

- **Component Configuration Panel :** Αυτοί οι κόμβοι ορίζουν τις ρυθμίσεις στο παράθυρο διαλόγου Component.
- **Data Pre-Processing :** Φιλτράρισμα κατά αριθμό μοναδικών τιμών, Κανονικοποίηση, Ακέραιοι για διπλασιασμό, Λείπει καταλογισμός τιμής, προαιρετική μια καυτή κωδικοποίηση, Έξοδος μοντέλων πριν/μετά τη διαδικασία, Διαχωρισμός και δοκιμή.
- **Learning and Scoring :** Μάθετε μοντέλα στο σετ τρένου, πρόβλεψη σετ δοκιμής και υπολογίστε την απόδοση στο δοκιμαστικό σύνολο.
- **Selecting and Packaging Best Model :** Κάνει κατάταξη μοντέλου από τη μέτρηση που έχει επιλέξει ο χρήστης, επιλέξτε το καλύτερο και συνδυάζει προετοιμασία και εξαγωγή δεδομένων.



ΕΙΚΟΝΑ 28. AutoML Workflow View

7.2.3 Visual Analysis of Sales Data

Η πρώτη ροή εργασίας οπτικοποίησης στις πωλήσεις δεδομένων.

- Φιλτράρετε δεδομένα σε στήλες που παρέχουν σχετικές πληροφορίες.
- Φιλτράρετε δεδομένα σε σειρές που θέλετε να συμπεριλάβετε στην ανάλυση σας.
- Οπτικοποιήστε τα δεδομένα χρησιμοποιώντας ένα γράφημα στοιβαγμένων περιοχών και ένα γράφημα πίτας/ντόνατ.

1) Data Access : εμπεριέχει το μονοπάτι του αρχείου.

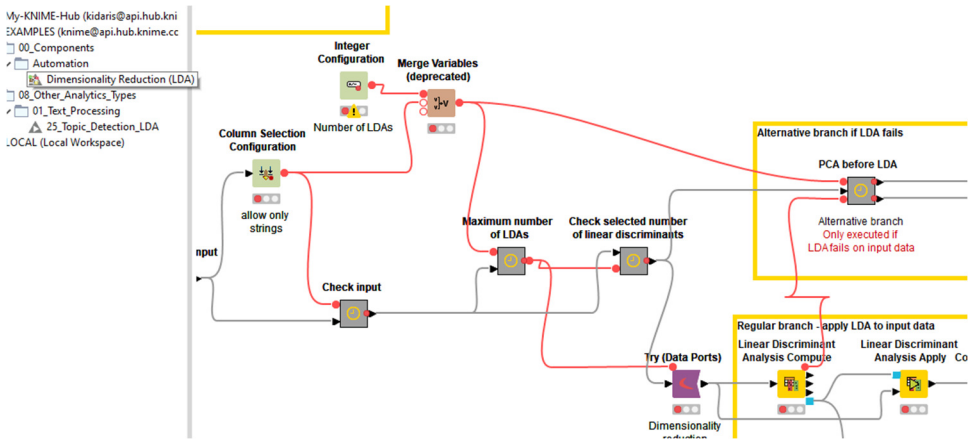
2) Data Pre-processing : Φιλτράρει τις στήλες και τις σειρές.

3) Data Visualization : Δείχνει τις πωλήσεις ανά ώρα και χώρα.

7.2.4 Μείωση Διαστάσεων (LDA)

Περιγραφή : Αυτό το στοιχείο μειώνει τον αριθμό των στηλών στα δεδομένα εισόδου με ανάλυση γραμμικής διάκρισης βασίζεται στον διαχωρισμό δύο ή περισσότερων κατηγοριών στα δεδομένα. Επομένως, μια στήλη συμβολοσειράς πρέπει να επιλέγει ως στήλη στόχος. Οι αριθμητικές στήλες προβάλλονται στους γραμμικούς συνδυασμούς τους, γραμμικά διαχωριστικά, που διαχωρίζουν καλύτερα τις διαφορετικές κατηγορίες-στόχους.

Χρήση : Το στοιχείο αυτό μπορεί να χρησιμοποιηθεί για μείωση διαστάσεων, για παράδειγμα, πριν από την εκπαίδευση ενός μοντέλου μηχανικής εκμάθησης. Η ανάλυση γραμμικής διάκρισης λειτουργεί επίσης ως ταξινομητής και χωρίζονται ως κατανεμημένα δεδομένα σε δύο ή περισσότερες κατηγορίες-στόχους.[29]

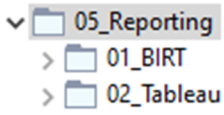


ΕΙΚΟΝΑ 29. LDA ROOT

7.2.5 Αναφορά (Reporting)

Περιγραφή BIRT : Η ροή εργασίας δείχνει τη χρήση κόμβων βάσεων δεδομένων Query Reader και DB Writer σε μία απλή ροή εργασίας KNIME. Χρησιμοποιεί τη βάση δεδομένων sqlite που βασίζεται σε αρχεία (όλη η βάση δεδομένων είναι γραμμένη σε ένα φάκελο στον σκληρό δίσκο).

Περιγραφή Tableau integration : Αυτή η ροή εργασίας χτίζει ένα απλό δέντρο αποφάσεων, και περνάει τα αποτελέσματα στο Tableau nodes για να γράφεις σε τοπικά αρχεία και να τα στέλνει στο Tableau server.[30]



ΕΙΚΟΝΑ 30. Reporting menu

7.2.6 Control Structures (Δομές Ελέγχου)

Streaming and Wrapped Nodes : Η εκτέλεση ροής (Streaming execution) είναι ένας άλλος τρόπος εκτέλεσης κόμβων και διαφέρει από την προεπιλεγμένη

εκτέλεση “κόμβος προς κόμβο”. Τα πλεονεκτήματα είναι λιγότερη I/O και ταχύτερος χρόνος εκτέλεσης σε βάρος της περιορισμένης εξερεύνησης και ιχνηλασιμότητας. Οι κόμβοι που έχουν δυνατότητα ροής εκτελούνται ταυτόχρονα.

Ο πρώτος τυλιγμένος κόμβος (που ενσωματώνει μια ροή εργασίας) διαβάζει ένα αρχείο CSV και κάνει βασική επεξεργασία σε αυτό. Ο δεύτερος τυλιγμένος κόμβος δημιουργεί πολλά δεδομένα και τα επεξεργάζεται. Εδώ η εκτέλεση ροής είναι πιο προφανής.

Components and Configuration Nodes : Αυτή η ροή εργασίας δείχνει πώς χρησιμοποιούνται οι κόμβοι διαμόρφωσης μαζί με τα στοιχεία.

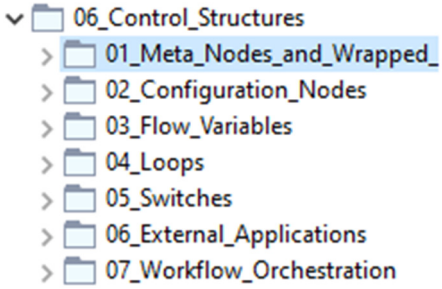
Flow Variables : Είναι μία ροή εργασίας που δείχνει πώς να χρησιμοποιείτε μεταβλητές ροής εργασίας σε έναν κόμβο Java Snippet για σκοπούς φιλτραρίσματος.

Αυτό το σενάριο αναλύει τόσο τις μεταβλητές που ορίζονται στη ροή εργασίας όσο και τα πεδία ημερομηνίας στον πίνακα εισαγωγής και τις μετατρέπει σε αντικείμενα Java Date. Τα αντικείμενα ημερομηνίας έχουν πρακτικές μεθόδους πριν και μετά για την αφαίρεση άσχετων εγγραφών.

Modified Variable Propagation :

- **Request Captured :** Αυτή η ροή εργασίας χρησιμοποιεί των βιβλίων Google API για να λάβει τον αριθμό των διαθέσιμων βιβλίων στο χώρο φύλαξης των βιβλίων Google για τους παρεχόμενους όρους αναζήτησης.
- **Request Loop :** Αυτή η ροή εργασίας χρησιμοποιεί το API των βιβλίων Google για να λάβει τους τίτλους και τους συγγραφείς για τα βιβλία που αντιστοιχούν στους παρεχόμενους όρους αναζήτησης.
- **Loops Parameter optimization :** Βελτιστοποιεί τις υπερ.-παραμέτρους ενός τυχαίου δάσους δέντρων απόφασης και το εκπαιδεύει με τις βελτιστοποιημένες υπερ.-παραμέτρους.

- **Cases Switch :** Αυτή η ροή εργασίας δείχνει τη χρήση του κόμβου “ Δεδομένα μεταγωγής περίπτωσης”. Οι ίδιες αρχές ισχύουν για κάθε “ Case Switch Model” & “ Case Switch Variable”.
- **External Application :** Επιδεικνύει τη χρήση του κόμβου “Εξωτερικό Εργαλείο”. Η ροή εργασίας ευρετηριάζει τον αρχικό κατάλογο του χρήστη (όλα τα αρχεία) και στη συνέχεια καλεί το a script ‘size.sh’ σε αυτή τη λίστα. Το σενάριο (Script) δημιουργεί ένα ενδιάμεσο αρχείο που περιέχει το μέγεθος σε KB των μεμονωμένων αρχείων. Αυτό το αρχείο διαβάζεται στο KNIME και οι πληροφορίες μεγέθους ενώνονται με τον πίνακα εισαγωγής.[31]



EIKONA 31. Control Structure Menu

7.2.7 Read

Excel Reader : Αυτός ο κόμβος διαβάζει αρχεία Excel(xlsx,xlsm,xlsb &xls). Μπορεί να διαβάσει ένα μεμονωμένο ή πολλά αρχεία ταυτόχρονα, διαβάζοντας ωστόσο μόνο ένα φύλλο ανά αρχείο. Οι υποστηριζόμενοι τύποι Excel που μπορούν να διαβαστούν είναι συμβολοσειρά, αριθμός, Boolean, ημερομηνία και ώρα, αλλά όχι εικόνες, διαγράμματα κ.λ.π.

File Reader : Διαβάζει τα πιο κοινά αρχεία κειμένου. Για να μαντέψετε αυτόματα τη δομή του αρχείου, κάντε κλικ στο κουμπί Μορφή Αυτόματης ανίχνευσης. Εάν αντιμετωπίζετε προβλήματα με λανθασμένους τύπους

δεδομένων, απενεργοποιήστε την επιλογή σάρωση σειρών δεδομένων περιορισμού στην καρτέλα ρυθμίσεις για προχωρημένους.

ARFF Reader : Αυτός ο κόμβος διαβάζει δεδομένα ARFF από μία διεύθυνση URL. Στο παράθυρο διαλόγου διαμόρφωσης, καθορίστε μια έγκυρη διεύθυνση URL και ορίστε ένα προαιρετικό πρόθεμα σειράς. Ένα αναγνωριστικό σειράς δημιουργείται από τον αναγνώστη με τη μορφή "Πρόθεμα + αριθμός σειράς". Τα αρχεία ARFF δεν υποστηρίζονται αυτήν τη στιγμή.

CSV Reader : Διαβάζει αρχεία CSV. Για να μαντέψετε αυτόματα τη δομή του αρχείου, κάντε κλικ στο κουμπί μορφή αυτόματης ανίχνευσης. Εάν αντιμετωπίζετε προβλήματα με λανθασμένους τύπους δεδομένων, απενεργοποιήστε την επιλογή σάρωση σειρών δεδομένων περιορισμού στην καρτέλα ρυθμίσεις για προχωρημένους. Εάν η δομή του αρχείου εισόδου αλλάζει μεταξύ διαφορετικών κλήσεων, ενεργοποιήστε την Υποστήριξη αλλαγής της επιλογής σχημάτων αρχείων στην καρτέλα προηγούμενες ρυθμίσεις.

Line Reader : Διαβάζει γραμμές από αρχείο ή URL. Κάθε γραμμή θα αντιπροσωπεύεται από ένα κελί δεδομένων συμβολοσειράς σε μια μόνο σειρά. Το πρόθεμα γραμμής και η κεφαλίδα στήλης μπορούν να καθοριστούν στο παράθυρο διαλόγου. Αυτός ο κόμβος μπορεί να έχει πρόσβαση σε μια ποικιλία διαφορετικών συστημάτων αρχείων.

Table Reader : Αυτός ο κόμβος διαβάζει αρχεία που έχουν γραφτεί χρησιμοποιώντας τον κόμβο Table Writer. Διατηρεί όλες τις μετα-πληροφορίες όπως τομέα, ιδιότητες, χρώματα, μέγεθος. Αυτός ο κόμβος μπορεί να έχει πρόσβαση σε μία ποικιλία διαφορετικών συστημάτων αρχείων.

PMML Reader : Αυτός ο κόμβος διαβάζει οποιοδήποτε μοντέλο PMML από ένα αρχείο XML συμβατό με PMML. Αυτός ο κόμβος μπορεί να έχει πρόσβαση σε μία ποικιλία διαφορετικών συστημάτων αρχείων.

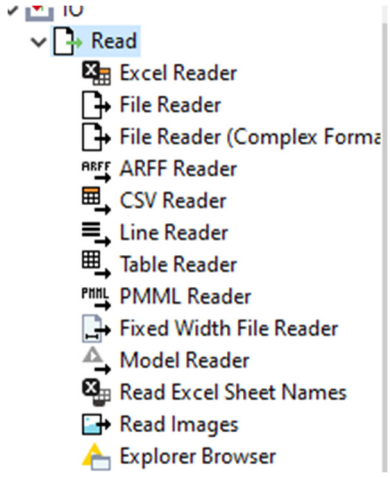
Fixed Width File Reader : Αυτός ο κόμβος μπορεί να χρησιμοποιηθεί για ανάγνωση δεδομένων από ένα αρχείο ASCII ή μια τοποθεσία URL. Διαβάζει αρχεία όπου μια στήλη ορίζεται από τον αριθμό των χαρακτήρων και όχι από έναν οριοθετεί στήλης. Όταν ανοίγετε το παράθυρο διαλόγου διαμόρφωσης του κόμβου και παρέχετε ένα όνομα αρχείου, ο αναγνώστης δημιουργεί μια στήλη. Αυτή η στήλη εμφανίζει μόνο τους υπόλοιπους χαρακτήρες στον πίνακα προεπισκόπησης και δεν θα προσαρτηθεί στον πίνακα εξόδου.

Model Reader : Αυτός ο κόμβος μπορεί να έχει πρόσβαση σε μία ποικιλία διαφορετικών συστημάτων αρχείων.

Read Excel Sheet Names : Η απόδοση του κόμβου ανάγνωσης είναι περιορισμένη. Η ανάγνωση μεγάλων αρχείων διαρκεί πολύ και χρησιμοποιεί πολλή μνήμη.

Read Images : Διαβάστε τις εικόνες από μία λίστα διευθύνσεων URL και προσαρτήστε τις ως νέα στήλη. Η λίστα URL είναι μια στήλη στον πίνακα εισαγωγής που περιέχει έγκυρες διευθύνσεις URL(π.χ. File:/tmp/image.png). Μπορείτε να χρησιμοποιήσετε τον κόμβο “List Files” για να σαρώσετε έναν κατάλογο που περιέχει αρχεία *.png ή *.svg.

Explorer Browser : Επιτρέπει την περιήγηση στις τοποθεσίες που έχουν προσαρτηθεί στον εξερευνητή KNIME και την έκθεση του αποτελέσματος ως διεύθυνση URL και απόλυτη διαδρομή αρχείου. Η διαδρομή του αρχείου που προκύπτει και η διεύθυνση URL εκτίθενται ως μεταβλητές ροής και μπορούν να χρησιμοποιηθούν σε κόμβους εγγραφής για εγγραφή.[32]



ΕΙΚΟΝΑ 32. Όλοι οι κόμβοι Read

7.2.8 Write

CSV Writer : Αυτός ο κόμβος καταγράφει τον πίνακα δεδομένων εισόδου σε ένα αρχείο ή σε απομακρυσμένη τοποθεσία που υποδηλώνεται με μια διεύθυνση URL. Ο πίνακας δεδομένων εισόδου πρέπει να περιέχει μόνο συμβολοσειρά ή αριθμητικές στήλες. Άλλοι τύποι στηλών δεν υποστηρίζονται. Επιπλέον, μπορεί να έχει πρόσβαση σε μια ποικιλία διαφορετικών συστημάτων αρχείων.

ARFF Writer : Αυτός ο κόμβος αποθηκεύει δεδομένα σε ένα αρχείο ή σε μια απομακρυσμένη τοποθεσία που υποδηλώνεται με μια διεύθυνση URL σε μορφή ARFF. Στο διάλογο Configuration, καθορίστε μια έγκυρη τοποθεσία προορισμού. Όταν εκτελείται, ο κόμβος εγγράφει τα δεδομένα, που προέρχονται από τη θύρα εισόδου του, στην καθορισμένη θέση. Σε αυτό το χρονικό σημείο, γράφει μόνο μη αραιά αρχεία ARFF.

Table Writer : Αυτός ο κόμβος γράφει έναν πίνακα δεδομένων στην εσωτερική μορφή του KNIME σε ένα αρχείο, το οποίο μπορεί να διαβαστεί χρησιμοποιώντας τον κόμβο ανάγνωσης πίνακα.

PMML Writer : Αυτός ο κόμβος εγγράφει ένα μοντέλο PMML από μια θύρα μοντέλου σε ένα αρχείο συμβατό με PMML v4.2 ή σε απομακρυσμένη τοποθεσία

που υποδηλώνεται με μία διεύθυνση URL. Εάν ένα αρχείο PMML από άλλη έκδοση διαβαστεί από το PMML Reader και γράφεται απευθείας από αυτόν τον κόμβο, μετατρέπεται σε PMML v4.2. Εάν το μοντέλο δεν είναι έγκυρο, δημιουργείται εξαίρεση κατά την εκτέλεση.

Excel Cell Update : Ο κόμβος ενημερώνει τα κελιά σε ένα υπάρχον υπολογιστικό φύλλο του excel. Οι διευθύνσεις κελιών και το νέο τους περιεχόμενο παρέχονται από έναν πίνακα δεδομένων εισόδου.

Excel Writer : Αυτός ο κόμβος γράφει τον πίνακα δεδομένων εισόδου σε ένα υπολογιστικό φύλλο ενός αρχείου excel, το οποίο στη συνέχεια μπορεί να διαβαστεί με άλλες εφαρμογές όπως το Microsoft Excel. Ο κόμβος μπορεί να δημιουργήσει εντελώς νέα αρχεία ή να προσθέσει δεδομένα σε ένα υπάρχον αρχείο excel. Κατά την προσάρτηση, τα δεδομένα εισόδου μπορούν να προσαρτηθούν ως νέο υπολογιστικό φύλλο ή μετά την τελευταία σειρά ενός υπάρχοντος υπολογιστικού φύλλου. Με την προσθήκη πολλαπλών θυρών εισαγωγής πίνακα δεδομένων, τα δεδομένα μπορούν να εγγραφούν/προσαρτηθούν σε πολλά υπολογιστικά φύλλα στο ίδιο αρχείο.

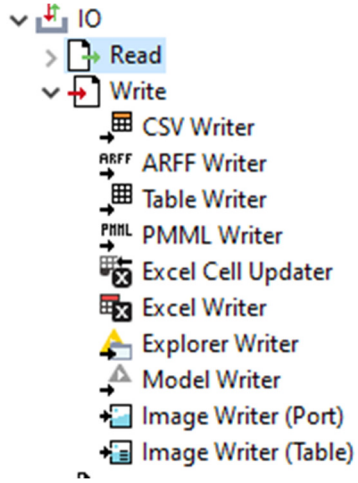
Explorer Writer : Γράφει/Αντιγράφει ένα αρχείο του οποίου η διαδρομή παρέχεται ως μεταβλητή ροής σε μια θέση προσαρτημένη στον KNIME Explorer. Το αρχείο που πρόκειται να αντιγραφεί πρέπει να υπάρχει και να είναι προ βάσιμο. Αυτός ο κόμβος χρησιμοποιείται ως επί το πλείστων σε συνδυασμό με κόμβους εγγραφής αρχείων, οι οποίοι δεν είναι σε θέση να γράψουν απευθείας στον εξερευνητή.

Model Writer : Αυτός ο κόμβος γράφει ένα μοντέλο KNIME σε ένα αρχείο που μπορεί να διαβαστεί με τον κόμβο Model Reader.

Image Writer (port) : Γράφει ένα αντικείμενο θύρας εικόνας σε ένα αρχείο ή μια απομακρυσμένη τοποθεσία που υποδηλώνεται με μια διεύθυνση URL. Το

αντικείμενο εισαγωγής εικόνας πρέπει να περιέχει μια έγκυρη εικόνα, διαφορετικά ο κόμβος θα αποτύχει κατά την εκτέλεση.

Image Writer (table) : Αυτός ο κόμβος παίρνει όλες τις εικόνες σε μία συγκεκριμένη στήλη του πίνακα εισόδου και τις γράφει, καθεμία ως ξεχωριστό αρχείο, σε έναν κατάλογο. Θα προσθέσει τις διαδρομές τον εγγεγραμμένων αρχείων στον πίνακα εισόδου καθώς και την αντίστοιχη κατάσταση εγγραφής(Δημιουργήθηκε-τροποποιήθηκε-αντικαταστάθηκε).[33]



ΕΙΚΟΝΑ 33. Όλοι οι κόμβοι Write

7.2.9 Connectors

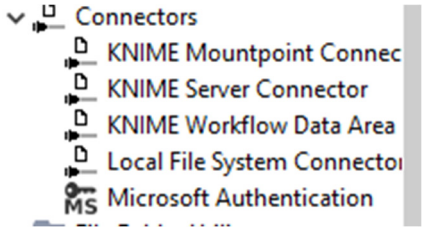
KNIME Mountpoint Connector : Παρέχει μια σύνδεση συστήματος αρχείων με πρόσβαση σε ένα σημείο προσάρτησης KNIME, για παράδειγμα “LOCAL” ή “My-KNIME-Hub". Μπορεί επίσης να παρέχει μια σύνδεση συστήματος αρχείων με πρόσβαση στο σημείο προσάρτησης που περιέχει την τρέχουσα ροή εργασίας. Η προκύπτουσα θύρα εξόδου επιτρέπει στους κατάντη κόμβους να έχουν πρόσβαση σε αρχεία, π.χ για ανάγνωση ή εγγραφή ή εκτέλεση άλλων λειτουργιών συστήματος αρχείων στο επιλεγμένο σημείο προσάρτησης.

KNIME Server Connector : Αυτός ο κόμβος συνδέεται με έναν διακομιστή KNIME μέσω REST. Η προκύπτουσα θύρα εξόδου επιτρέπει στους κατάντη κόμβους να έχουν πρόσβαση στο χώρο αποθήκευσης ροής εργασίας διακομιστή ως σύστημα αρχείων, π.χ για ανάγνωση ή εγγραφή αρχείων και φακέλων ή για εκτέλεση άλλων λειτουργιών του συστήματος αρχείων.

KNIME Workflow Data Area Connector : Παρέχει μια σύνδεση συστήματος αρχείων με πρόσβαση στην περιοχή δεδομένων της τρέχουσας ροής εργασίας KNIME. Η θύρα εξόδου που προκύπτει στους κατάντη κόμβους να έχουν πρόσβαση σε αρχεία, π.χ για ανάγνωση ή εγγραφή ή εκτέλεση άλλων λειτουργιών του συστήματος αρχείων(περιήγηση/λίστα αρχείων, αντιγραφή, μετακίνηση).

Local File System Connector : Αυτός ο κόμβος παρέχει πρόσβαση στο σύστημα αρχείων του τοπικού μηχανήματος. Η προκύπτουσα θύρα εξόδου επιτρέπει στους κατάντη κόμβους να έχουν πρόσβαση σε αρχεία, π.χ για ανάγνωση ή εγγραφή ή για εκτέλεση άλλων λειτουργιών του συστήματος αρχείων.

Microsoft Authentication : Αυτός ο κόμβος περιέχει έλεγχο ταυτότητας για πρόσβαση στις υπηρεσίες cloud του Microsoft Azure και του Office 365.[34]



ΕΙΚΟΝΑ 34. Όλοι οι κόμβοι Connectors

7.2.10 Column (Binning)

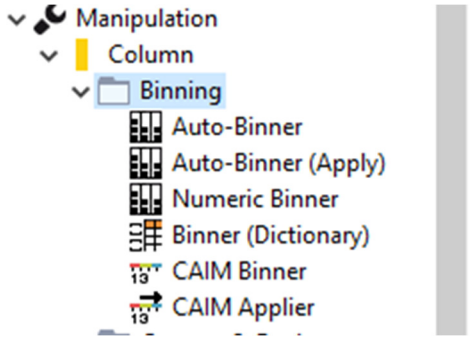
Auto Binner : Αυτός ο κόμβος επιτρέπει την ομαδοποίηση αριθμητικών δεδομένων σε διαστήματα - call bins. Υπάρχουν δύο επιλογές ονόματος για τα

bins και δύο μέθοδοι που καθορίζουν τον αριθμό και το εύρος των τιμών που εμπίπτουν σε ένα bin. Χρησιμοποιήστε τον κόμβο Numeric Binner, εάν θέλετε να ορίστε προσαρμοσμένους κάδους.

Numeric Binner : Για κάθε στήλη μπορεί να οριστεί ένας αριθμός διαστημάτων γνωστά ως bins. Διασφαλίζουν αυτόματα ότι τα εύρη ορίζονται με φθίνουσα σειρά και ότι τα διαστήματα είναι συνεπή. Επιπλέον, κάθε στήλη είτε αντικαθίσταται με τη στήλη δεσμευμένης, τύπου συμβολοσειράς, είτε προστίθεται μια νέα δεσμευμένη στήλη τύπου συμβολοσειράς.

CAIM Binner : Αυτός ο κόμβος υλοποιεί τον αλγόριθμο δέσμευσης CAIM (διακριτοποίησης) σύμφωνα με τους Kurgan & Cios (2004). Η δέσμευση CAIM εκτελείται σε σχέση με μια επιλεγμένη στήλη κλάσης, δημιουργεί όλα τα πιθανά όρια binning και επιλέγει αυτά που ελαχιστοποιούν το μέτρο αλληλεξάρτησης κλάσης. Μειώνοντας το χρόνο εκτέλεσης, αυτή η υλοποίηση δημιουργεί μόνο εκείνα τα όρια όπου αλλάζει η τιμή και η κλάση.

CAIM Appllier : Αυτός ο κόμβος λαμβάνει ένα μοντέλο binning και έναν πίνακα δεδομένων ως είσοδο και δεσμεύει τις στήλες των δεδομένων εισόδου σύμφωνα με το μοντέλο. Όλες οι στήλες που μπορούν να βρεθούν στο μοντέλο συνδυάζονται ανάλογα.[35]



ΕΙΚΟΝΑ 35. Όλοι οι κόμβοι Column

7.2.11 Column (Convert & Replace)

Category To Number : Αυτός ο κόμβος παίρνει στήλες με ονομαστικά δεδομένα και αντιστοιχίζει κάθε κατηγορία σε ακέραιο αριθμό. Για την διευκόλυνσή σας, μπορείτε να επεξεργαστείτε πολλές στήλες με αυτόν τον κόμβο. Ωστόσο, αυτές οι στήλες επεξεργάζονται ξεχωριστά σαν να χρησιμοποιούσατε έναν μόνο κόμβο κατηγορίας σε αριθμό για κάθε στήλη.

Cell Replacer : Αντικαθιστά τα κελιά σε μία στήλη σύμφωνα με τον πίνακα λεξικών. Ο κόμβος έχει δύο εισόδους, η πρώτη είσοδος περιέχει τη στήλη στόχου της οποίας οι τιμές πρέπει να αντικατασταθούν χρησιμοποιώντας τον πίνακα λεξικού. Η δεύτερη είσοδος είναι από τον πίνακα του λεξικού, επιλέξτε μία στήλη που χρησιμοποιείται ως κριτήριο αναζήτησης. Οποιαδήποτε εμφάνιση στη στήλη προορισμού (1η είσοδος) που αντιστοιχεί στην τιμή αναζήτησης αντικαθίσταται από την αντίστοιχη τιμή της στήλης εξόδου, η οποία είναι μια άλλη στήλη στον πίνακα του λεξικού.

Οι τιμές που λείπουν αντιμετωπίζονται ως συνήθεις τιμές, δηλαδή είναι έγκυρες ως τιμή αναζήτησης και αντικατάστασης. Εάν υπάρχουν διπλότυπα στη στήλη αναζήτησης στον πίνακα του λεξικού, η τελευταία εμφάνιση ορίζει το ζεύγος αντικατάστασης. Εάν η στήλη εισόδου/ αναζήτησης είναι τύπος συλλογής, καθεμία από τις τιμές που περιέχονται στη συλλογή είναι στοιχείο αναζήτησης.

Column Ayto Type Cast : Αυτός ο κόμβος καθορίζει τον πιο συγκεκριμένο τύπο στις διαμορφωμένες στήλες συμβολοσειράς και αλλάζει τους τύπους στηλών ανάλογα. Η σειρά τύπου είναι πρώτα να ελέγξετε αν οι τιμές είναι ημερομηνίες, μετά ακέραιος, μακρύς, διπλός και τέλος συμβολοσειρά. Για ημερομηνία μπορεί να καθοριστεί μια προσαρμοσμένη μορφή.

Column Rename : Μετονομάστε τα ονόματα στηλών ή αλλάξτε τους τύπους τους. Το παράθυρο διαλόγου σας επιτρέπει να αλλάξετε το όνομα μεμονωμένων στηλών επεξεργάζονται το πεδίο κειμένου ή να αλλάξετε τον τύπο στήλης

επιλέγοντας έναν από τους πιθανούς τύπους στο σύνθετο πλαίσιο. Συμβατοί τύποι είναι αυτοί στους οποίους τα κελιά σε μια στήλη μπορούν είτε να μετασχηματιστούν. Μια διαμόρφωση με κόκκινο περίγραμμα υποδεικνύει ότι η διαμορφωμένη στήλη δεν υπάρχει πλέον.

Math Formula : Αυτός ι κόμβος αξιολογεί μια μαθηματική έκφραση με βάση τις τιμές σε μια σειρά. Τα υπολογισμένα αποτελέσματα μπορούν είτε να προσαρτηθούν ως νέα στήλη είτε να χρησιμοποιηθούν για να αντικαταστήσουν μια στήλη εισόδου. Οι διαθέσιμες μεταβλητές είναι οι τιμές στην αντίστοιχη σειρά του πίνακα. Οι συναρτήσεις που χρησιμοποιούνται συνήθως εμφανίζονται στη λίστα "Μαθηματικές συναρτήσεις".

Number To String : Μετατρέπει του αριθμούς σε μια στήλη σε συμβολοσειρές. Σημειώστε ότι για μια προηγμένη διαμόρφωση, όπως στρογγυλοποίηση ή αναπαράσταση σε επιστημονική ειδοποίηση, μπορείτε επίσης να χρησιμοποιήσετε τον κόμβο Round Double.

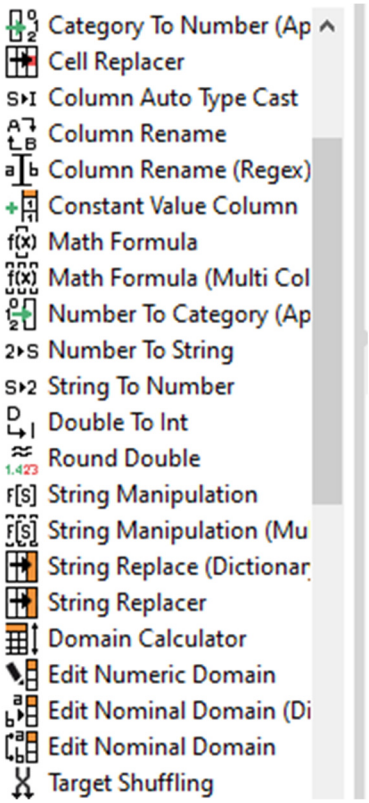
Round Double : Στρογγυλοποιεί διπλές τιμές στο καθορισμένο δεκαδικό ψηφίο ή σημαντικούς αριθμούς, εφαρμόζοντας την καθορισμένη μέθοδο στρογγυλοποίησης. Μπορούν να καθορίσουν οι στήλες που περιέχουν τις διπλές τιμές. Οι στρογγυλεμένες τιμές μπορούν να προστεθούν ως πρόσθετες στήλες ή οι παλιές τιμές αντικαθίστανται από τις στρογγυλεμένες τιμές. Εάν προσαρτηθούν στρογγυλεμένες τιμές και επιπλέον στήλες, πρέπει να καθοριστεί ένα επίθημα στήλης για της στήλες που θα προσαρτηθούν. Για να στρογγυλοποιήσετε τις τιμές είναι διαθέσιμες επτά διαφορετικές λειτουργίες : UP, DOWN, CEILING, FLOOR, HALF_UP, HALF_DOWN, HALF_EVEN.

String Manipulation : Χειρίζεται συμβολοσειρές όπως αναζήτηση και αντικατάσταση, κεφαλαία ή κατάργηση κενών διαστημάτων που προηγούνται και τελούν.

Domain Calculator : Σαρώνει τα δεδομένα και ενημερώνει τη λίστα πιθανών τιμών ή και τις ελάχιστες και τις μέγιστες τιμές των επιλεγμένων στηλών. Αυτός ο κόμβος είναι χρήσιμος όταν οι πληροφορίες τομέα των δεδομένων έχουν αλλάξει και πρέπει να ενημερωθούν στην προδιαγραφή του πίνακα, για παράδειγμα, οι πληροφορίες τομέα που περιέχονται σε μια προδιαγραφή πίνακα μπορεί να είναι άκυρες όταν πραγματοποιείται φιλτράρισμα σειρών.

Edit Numeric Domain : Ορίζει τα άνω/κάτω όρια των επιλεγμένων αριθμητικών στηλών όπως καθορίζονται από τον χρήστη. Κατά τη διάρκεια της εκτέλεσης τα δεδομένα ελέγχονται έναντι του πρόσφατα καθορισμένου τομέα και εκτελείται μια επιλέξιμη ενέργεια εάν τα δεδομένα δεν χωρούν στα δεδομένα όρια.

Target Shuffing : Αυτός ο κόμβος εκτελεί ανακάτεμα στόχου(Target Shuffing) μεταθέτοντας τυχαία τις τιμές σε μία στήλη του πίνακα εισόδου. Αυτό θα διακόψει οποιαδήποτε σύνδεση μεταξύ των μεταβλητών εισόδου και της μεταβλητής απόκρισης διατηρώντας παράλληλα τη συνολική κατανομή τις μεταβλητής στόχου. Target shuffing χρησιμοποιείται για την εκτίμηση της βασικής απόδοσης ενός προγνωστικού μοντέλου. Αναμένεται ότι η ποιότητα ενός μοντέλου θα μειωθεί δραστικά εάν οι τιμές στόχου ανακατεύονταν καθώς αφαιρέθηκε οποιαδήποτε σχέση μεταξύ εισόδου και στόχου. Συνιστάται να επαναλάβετε αυτή τη διαδικασία (ανακάτεμα στόχου + κατασκευή μοντέλων + αξιολόγηση μοντέλου) πολλές φορές και να καταγράψετε το ψεύτικο αποτέλεσμα, προκειμένου να λάβετε καλές εκτιμήσεις για το πόσο καλά αποδίδει το πραγματικό μοντέλο σε σύγκριση με τυχαιοποιημένα δεδομένα.[36]



ΕΙΚΟΝΑ 36. Όλοι οι κόμβοι Convert & Replace

7.2.12 Column (Filter)

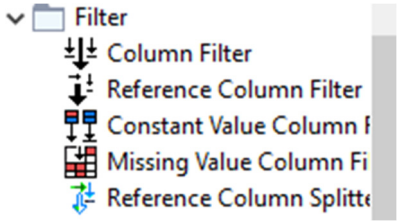
Column Filter : Αυτός ο κόμβος επιτρέπει το φιλτράρισμα στηλών από τον πίνακα εισόδου, ενώ μόνο οι υπόλοιπες στήλες μεταβιβάζονται στον πίνακα εξόδου. Μέσα στο παράθυρο διαλόγου, οι στήλες μπορούν να μετακινηθούν μεταξύ της λίστας συμπερίληψης και εξαίρεσης.

Reference Column Filter : Αυτός ο κόμβος επιτρέπει το φιλτράρισμα στηλών από τον πρώτο πίνακα χρησιμοποιώντας τον δεύτερο πίνακα ως πίνακα αναφοράς. Ανάλογα με τη ρύθμιση διαλόγου, είτε οι στήλες από τον πίνακα αναφοράς περιλαμβάνονται είτε εξαιρούνται στον πίνακα εξόδου.

Constant Value Filter : Αυτός ο κόμβος φιλτράρει στήλες που περιέχουν αποκλειστικά διπλότυπα της ίδιας τιμής από τον πίνακα δεδομένων εισόδου. Τα παραδείγματα περιλαμβάνουν μια στήλη που περιέχει μόνο μηδενικά, μια στήλη που περιέχει μόνο πανομοιότυπες συμβολοσειρές ή μια στήλη που περιέχει μόνο

κελιά που λείπουν. Σε ένα παράθυρο διαλόγου, μπορείτε να επιλέξετε σε ποιες στήλες θα εφαρμοστεί το φίλτρο. Από αυτές τις επιλεγμένες στήλες, μπορείτε να επιλέξετε να αφαιρέσετε είτε όλες τις στήλες σταθερών τιμών, είτε τις στήλες που περιέχουν μόνο συγκεκριμένες σταθερές αριθμητικές τιμές, συμβολοσειρά ή τιμές που λείπουν. Τέλος, μπορείτε επίσης να καθορίσετε τον ελάχιστο αριθμό γραμμών που πρέπει να ληφθούν υπόψη για φιλτράρισμα ενός πίνακα.

Reference Column Splitter : Αυτός ο κόμβος επιτρέπει τον διαχωρισμό στηλών από τον πρώτο πίνακα χρησιμοποιώντας τον δεύτερο πίνακα ως πίνακα αναφοράς.[37]



ΕΙΚΟΝΑ 37. Όλοι οι κόμβοι Filter

7.2.13 Column (Splite & Combine)

Cell Splitter : Αυτός ο κόμβος χρησιμοποιεί έναν χαρακτήρα οριοθέτη που καθορίζεται από το χρήστη για να χωρίσει το περιεχόμενο μιας επιλεγμένης στήλης σε μέρη. Προσθέτει είτε καθορισμένο αριθμό στηλών στον πίνακα εισόδου, καθεμία από τις οποίες φέρει ένα μέρος της αρχικής στήλης, είτε μια μεμονωμένη στήλη που περιέχει μια συλλογή κελιών με την διαίρεση της εξόδου. Μπορεί να καθοριστεί εάν η έξοδος αποτελείται από μία ή περισσότερες στήλες, μόνο μια στήλη που περιέχει κελιά λίστας ή μόνο μία στήλη που περιέχει κελιά συνόλου στα οποία έχουν αφαιρεθεί διπλότυπα. Εάν η στήλη περιέχει περισσότερους οριοθέτες από τους απαιτούμενους , οι πρόσθετοι οριοθέτες αγνοούνται .

Column Aggregator :

- Ομαδοποιεί τις επιλεγμένες στήλες ανά γραμμή και συγκεντρώνει τα κελιά τους χρησιμοποιώντας την επιλεγμένη μέθοδο συγκέντρωσης.
- Για να αλλάξετε το όνομα της νέας στήλης συγκέντρωσης που δημιουργήθηκε, κάντε διπλό κλικ στην στήλη ονόματος.
- Μπορείτε να βρείτε μια λεπτομερή περιγραφή των διαθέσιμων μεθόδων συγκέντρωσης στην καρτέλα "Περιγραφή" στο παράθυρο διαλόγου κόμβου.

Column Combiner : Συνδυάζει το περιεχόμενο ενός συνόλου στηλών και προσαρτά την ξεχωριστή στήλη των συνδυασμένων συμβολοσειρών στον πίνακα εισαγωγής. Ο χρήστης πρέπει να καθορίσει στο διάλογο τις στήλες που μας ενδιαφέρουν και κάποιες άλλες ιδιότητες, όπως οριοθετεί για να διαχωρίσει τα διαφορετικά περιεχόμενα κελιών και τις επιλογές εισαγωγικών.

Column Splitter : Αυτός ο κόμβος χωρίζει τις στήλες του πίνακα εξόδου. Καθορίζει στο διάλογο ποιες στήλες θα εμφανίζονται στον επάνω πίνακα (αριστερή λίστα) και στον κάτω πίνακα (Δεξιά λίστα). Χρησιμοποιήστε τα κουμπιά για να μετακινήσετε στήλες από μια λίστα στην άλλη.

Column Appender : Το Column Appender παίρνει δύο ή περισσότερους πίνακες και τους συνδυάζει γρήγορα προσαρτώντας τις στήλες τους σύμφωνα με τη σειρά των πινάκων στις θύρες εισόδου. Απλώς προσαρτά στήλες από τον δεύτερο πίνακα εισόδου στον πίνακα πρώτης εισαγωγής και κάνει το ίδιο για οποιαδήποτε επόμενο πίνακα εάν προστέθηκαν πρόσθετες θύρες. Ο κόμβος εκτελεί μια απλή λειτουργία σύνδεσης, αλλά μπορεί να είναι πιο γρήγορος εάν πληρούνται ορισμένες προϋποθέσεις.

Column to Grid : Διαχωρίζει μια επιλεγμένη στήλη σε νέες στήλες, έτσι ώστε να ευθυγραμμίζονται σε ένα πλέγμα. Αυτό είναι χρήσιμο για την εμφάνιση, για παράδειγμα, μια στήλη που περιέχει εικόνα σε ένα πλέγμα που μπορεί στην

συνέχεια να εμφανιστεί σε έναν πίνακα αναφοράς. Ο αριθμός των στηλών πλέγματος πρέπει να οριστεί στο παράθυρο διαλόγου, ο αριθμός των σειρών καθορίζεται ανάλογα.

Create Bit Vector : Δημιουργεί για κάθε γραμμή ενός δεδομένου πίνακα εισόδου ένα διάνυσμα bit. Τα διανύσματα bit είτε δημιουργούνται από πολλαπλές στήλες αριθμητικών ή συμβολοσειρών, μια στήλη συμβολοσειράς που περιέχει τις θέσεις bit προς ρύθμιση, δεκαεξαδικές - δυαδικές συμβολοσειρές ή μια στήλη συλλογής. Για να προσαρμόσετε τις ρυθμίσεις κόμβου, επιλέξτε το πρώτο αντικείμενο της στήλης πηγής. Ανάλογα με την επιλεγμένη επιλογή, τα στοιχεία διαλόγου είναι ενεργοποιημένα.

Create Byte Vectornode : Το Create Byte Vectornode δημιουργεί πληροφορίες ονόματος ως μεταδομένα στήλης. Μπορείτε να χρησιμοποιήσετε αυτές τις πληροφορίες σε αυτόν τον κόμβο.

Joiner : Αυτός ο κόμβος συνδυάζει δύο πίνακες παρόμοιους με έναν σύνδεσμο σε μια βάση δεδομένων. Συνδυάζει κάθε γραμμή από την επάνω θύρα εισόδου που έχει τις ίδιες τιμές σε επιλεγμένες στήλες. Σειρές που παραμένουν αταίριαστες μπορούν επίσης να εξάγονται.

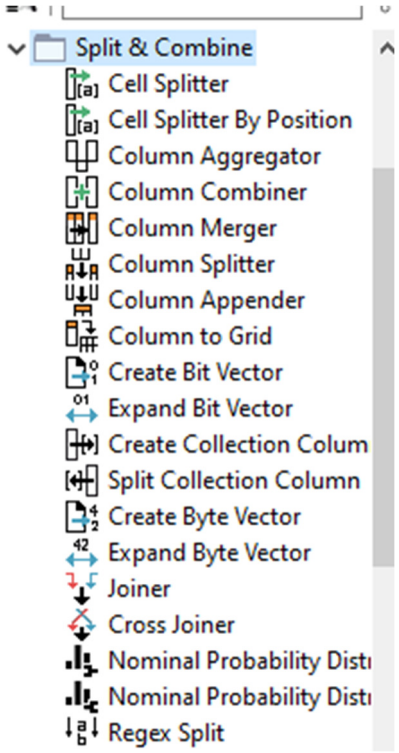
Cross Joiner : Πραγματοποιεί διασταυρούμενη ένωση δύο πινάκων. Κάθε σειρά του επάνω πίνακα ενώνεται με κάθε σειρά του κάτω πίνακα. Σημείωση, αυτή η λειτουργία είναι υπερβολικά ακριβή, καθώς ο αριθμός των σειρών στην έξοδο είναι το γινόμενο και των δύο μετρήσεων σειρών του πίνακα εισαγωγής, αυξάνοντας το μέγεθος του τεμαχίου θα επιταχυνθεί.

Nominal Probability Distribution Creator : Δημιουργεί μια στήλη που περιέχει μια κατανομή πιθανοτήτων είτε από αριθμητικές στήλες είτε από μια στήλη μονής συμβολοσειράς. Στην περίπτωση αριθμητικών στηλών μπορούν να επιλέγουν μία ή περισσότερες στήλες που περιέχουν τιμές πιθανότητας. Οι τιμές πιθανότητας πρέπει να είναι μη αρνητικές και να αθροίζονται στο 1. Σε

περίπτωση στήλης συμβολοσειράς, μπορεί να επιλεγεί μια μονή στήλη. Η κατανομή πιθανότητας της στήλης συμβολοσειράς παράγει μια μονο-κωδικοποίηση της στήλης συμβολοσειράς.

Nominal Probability Distribution Splitter : Διαχωρίζει μια στήλη που περιέχει μια κατανομή ονομαστικής πιθανότητας σε πολλές στήλες που περιέχουν τις μεμονωμένες πιθανότητες.

Regex Split : Αυτός ο κόμβος χωρίζει το περιεχόμενο συμβολοσειράς μια επιλεγμένης στήλης σε λογικές ομάδες χρησιμοποιώντας κανονικές εκφράσεις. Μια ομάδα προσδιορίζεται από ένα ζεύγος παρενθέσεων, όπου το μοτίβο σε τέτοιες παρενθέσεις είναι μια κανονική ομάδα επισυνάπτεται ως μεμονωμένη στήλη.[38]



ΕΙΚΟΝΑ 38. Split & Combine

7.2.14 Column (Transform)

f-F Case Converter : Αυτός ο κόμβος μετατρέπει αλφαριθμητικούς χαρακτήρες σε πεζούς ή κεφαλαίους.

Column Comparator : Συγκρίνει τις τιμές των κελιών δύο επιλεγμένων στηλών κατά σειρά. Είναι διαθέσιμος ένας αριθμός διαφορετικών μεθόδων σύγκρισης: ίσο, μη ίσο, μικρότερο, μεγαλύτερο, λιγότερο ίσο και μεγαλύτερο ίσο. Προστίθεται μια νέα στήλη κρατώντας είτε την τιμή της αριστερής είτε της δεξιάς στήλης, ένα κελί που λείπει ή μια ετικέτα που έχει καθοριστεί από το χρήστη. Ο τύπος της επισυναπτόμενης στήλης εξαρτάται από αυτό το επιλεγόμενο περιεχόμενο: Εάν επιλεγεί μια ετικέτα που ορίζεται από το χρήστη, η στήλη που προκύπτει είναι συμβολοσειράς τύπου, σε όλες τις άλλες περιπτώσεις ο τύπος καθορίζεται από τον τύπο των επιλεγμένων στηλών.

Column Resorter : Αυτός ο κόμβος αλλάζει τη σειρά των στηλών εισαγωγής, με βάση τις ρυθμίσεις που καθορίζονται από το χρήστη. Οι στήλες μπορούν να μετακινηθούν σε μεμονωμένα βήματα αριστερά ή δεξιά, ή εντελώς στην αρχή, ή στο τέλος του πίνακα εισόδου. Επιπλέον, οι στήλες μπορούν επίσης να ταξινομηθούν με βάση το όνομα τους. Παρέχεται ταξινομημένος πίνακας στο λιμάνι εξόδου. Μόλις διαμορφωθεί το παράθυρο διαλόγου του κόμβου, είναι δυνατό να συνδεθεί ένας νέος πίνακας εισόδου με διαφορετική δομή στο theode και να ρυθμίσετε ξανά τις παραμέτρους του διαλόγου.

Lag Column : Αντιγράφει τις τιμές στηλών από τις προηγούμενες σειρές στην τρέχουσα σειρά. Ο κόμβος μπορεί να χρησιμοποιηθεί σε :

- Δημιουργήστε ένα αντίγραφο της επιλεγμένης στήλης και μετακινήστε τα κελιά lsteps επάνω (l = διάστημα καθυστέρησης)
- Δημιουργήστε L αντίγραφα της επιλεγμένης στήλης και μετακινήστε τα κελιά κάθε αντίγραφου 1, 2, 3, …. L-1 βήματα προς τα επάνω (L= υστέρηση)

Η επιλογή καθυστέρησης Lin this mode είναι χρήσιμη για την πρόβλεψη χρόνο-σειρών. Εάν οι σειρές ταξινομούνται με αύξουσα χρονική σειρά, για να εφαρμόσετε μια καθυστέρηση L στην επιλεγμένη στήλη σημαίνει να τοποθετήσετε τις προηγούμενες τιμές L-1 της στήλης και την τρέχουσα τιμή της στήλης σε μία σειρά. Ο πίνακας δεδομένων μπορεί στη συνέχεια να χρησιμοποιηθεί για πρόβλεψη χρόνο-σειρών.

Reference Column Resorter : Αυτός ο κόμβος αλλάζει τη σειρά των στηλών εισόδου με βάση τη σειρά που παρέχεται στον δεύτερο πίνακα εισόδου. Ο τελευταίος πρέπει να περιέχει μια στήλη συμβολοσειράς με ονόματα στηλών όπως στον πρώτο πίνακα εισόδου. Στην συνέχεια, οι στήλες της πρώτης εισόδου ταξινομούνται σύμφωνα με τη σειρά αυτής της στήλης. Οι στήλες που δεν αποτελούν μέρος αυτής της στήλης ταξινόμησης ταξινομούνται στην αρχή ή στο τέλος ή απορρίπτονται εντελώς από την έξοδο. Η στήλη ταξινόμησης δεν πρέπει να περιέχει διπλότυπα. Τα άγνωστα αναγνωριστικά στηλών αγνοούνται.

Denormalizes : Αυτός ο κόμβος αποκανονικοποιεί τα δεδομένα εισόδου σύμφωνα με τις παραμέτρους κανονικοποίησης που δίνονται στην είσοδο του μοντέλου PMML. Επομένως, ο συγγενικός μετασχηματισμός αντιστρέφεται και οι αρχικές τιμές αναδημιουργούνται. Επιπλέον, χρησιμοποιείται συνήθως μετά την κανονικοποίηση των δεδομένων δοκιμής και την πιθανή μετατροπή άλλων μαθημάτων/προβλέψεων στο αρχικό εύρος.

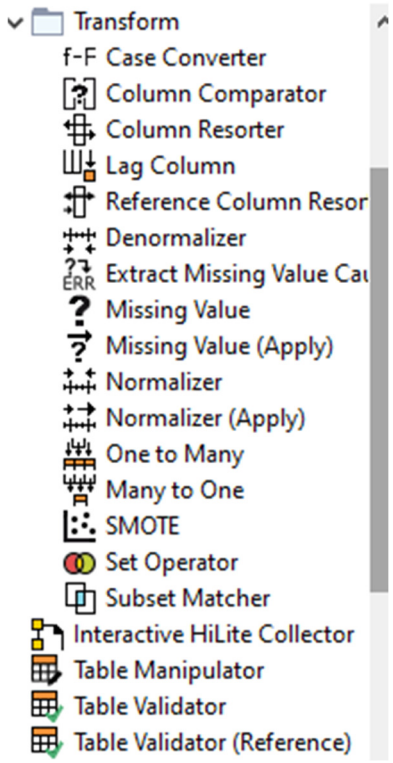
Extra Missing Value Causes : Αυτός ο κόμβος μπορεί να χρησιμοποιηθεί για την εξαγωγή των μηνυμάτων σφάλματος των missing value. Εάν δεν υπάρχει μήνυμα σφάλματος για μια συγκεκριμένη τιμή που λείπει, θα επιστραφεί μια κενή συμβολοσειρά.

Missing Value : Αυτός ο κόμβος βοηθά στο χειρισμό τιμών που λείπουν, που βρίσκονται στα κελιά του αρχείου εισόδου. Η πρώτη καρτέλα στο παράθυρο διαλόγου παρέχει προεπιλεγμένες επιλογές χειρισμού για όλες τις στήλες στον

πίνακα εισαγωγής που δεν αναφέρονται ρητά στη δεύτερη καρτέλα, με την ένδειξη “ Ατομικό”. Αυτή η δεύτερη καρτέλα επιτρέπει μεμονωμένες ρυθμίσεις για κάθε διαθέσιμη στήλη.

Normalizer : Αυτός ο κόμβος κανονικοποιεί τις τιμές όλων των στηλών. Στο παράθυρο διαλόγου, μπορείτε να επιλέξετε τις στήλες στις οποίες θέλετε να εργαστείτε. Οι ακόλουθες μέθοδοι κανονικοποίησης είναι διαθέσιμες στο παράθυρο στο παράθυρο διαλόγου.

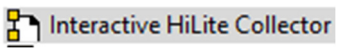
One to Many : μετατρέπει όλες τις πιθανές τιμές σε μια επιλεγμένη στήλη, η καθεμία σε μια νέα στήλη. Η τιμή ορίζεται ως το όνομα της νέας στήλης, οι τιμές των κελιών σε αυτήν τη στήλη είναι είτε 1, εάν αυτή η σειρά περιέχει αυτήν την πιθανή τιμή, είτε 0, εάν όχι. Ο κόμβος προσαρτά όσες στήλες είναι δυνατές που ορίζονται για τις επιλεγμένες στήλες. Εάν μια σειρά περιέχει μια τιμή που λείπει σε μια επιλεγμένη στήλη, όλες οι αντίστοιχες νέες στήλες περιέχουν την τιμή 0. Για την αποφυγή διπλών ονομάτων στηλών με πανομοιότυπες πιθανές τιμές σε διαφορετικές επιλεγμένες στήλες, το όνομα της στήλης που δημιουργείται περιλαμβάνει το αρχικό όνομα της στήλης σε αυτήν την περίπτωση. Το παράθυρο του κόμβου σας επιτρέπει μόνο να επιλέξετε στήλες με ονομαστικές τιμές.[39]



EIKONA 39. Transform

7.2.15 Interactive HiLite Collector

Το Node επιτρέπει την εφαρμογή σχολιασμών σειρών hilit εντός μιας δια δραστικής προβολής: Εντός της δια δραστικής προβολής, μπορεί να εισαχθεί μια συμβολοσειρά σχολιασμού η οποία στη συνέχεια προσαρτάται σε όλες τις σειρές hilit. Κατά την επανεκτέλεση αυτού του κόμβου, ο σχολιασμός προστίθεται ως νέα στήλη, στο τέλος του πίνακα εισαγωγής. Όταν γίνεται επαναφορά του κόμβου, όλοι οι σχολιασμοί απορρίπτονται.[40]



EIKONA 40. HiLite Collector

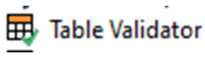
7.2.16 Table Manipulator

Επιτρέπει την εκτέλεση πολλών μετασχηματισμών στηλών σε οποιονδήποτε αριθμό πινάκων εισόδου, όπως μετονομασία, φιλτράρισμα, αναδιάταξη και αλλαγή τύπου των στηλών εισόδου. Εάν είναι διαθέσιμοι περισσότεροι από ένας πίνακες εισόδου, ο κόμβος συνενώνει όλες τις σειρές εισόδου σε έναν ενιαίο πίνακα αποτελεσμάτων. Εάν οι πίνακες εισόδου περιέχουν το ίδιο αναγνωριστικό γραμμής, ο κόμβος μπορεί είτε να δημιουργήσει ένα νέο αναγνωριστικό γραμμής είτε να προσαρτήσει το ευρετήριο του πίνακα εισόδου στο αρχικό αναγνωριστικό σειράς του αντίστοιχου πίνακα εισαγωγής.[41]

ΕΙΚΟΝΑ 41. Table Manipulator

7.2.17 Table Validator

Επιτρέπει την εκτέλεση πολλών μετασχηματισμών στηλών σε οποιανδήποτε αριθμό πινάκων εισόδου, όπως μετονομασία, φιλτράρισμα, αναδιάταξη και αλλαγή τύπου των στηλών εισόδου. Αυτός ο κόμβος εξασφαλίζει μια συγκεκριμένη δομή πίνακα και περιεχόμενο πίνακα χρησιμοποιώντας μια προδιαγραφή πίνακα δεδομένων αναφοράς που ορίζεται από τον χρήστη στο παράθυρο διαλόγου διαμόρφωσης. Η βάση για τη διαμόρφωση δίνεται από την προδιαγραφή πίνακα εισόδου κατά τη διαμόρφωση και παρέχει το βασικό πρότυπο για τον πίνακα εξόδου. Διασφαλίζεται ότι η δομή του πίνακα που προκύπτει είναι ως επί το πλείστων πανομοιότυπη με τις προδιαγραφές που ορίζει ο χρήστης. Αυτό γίνετε με την καταφυγή στηλών, την εισαγωγή στηλών που λείπουν και προαιρετική αφαίρεση πρόσθετων στηλών.[42]



ΕΙΚΟΝΑ 42. Table Validator

7.2.18 ROW Filter

Duplicate Row Filter : Αυτός ο κόμβος προσδιορίζει διπλές σειρές. Οι διπλότυπες σειρές έχουν πανομοιότυπες τιμές σε ορισμένες στήλες. Ο κόμβος επιλέγει μια μονή σειρά για κάθε σύνολο διπλότυπων. Μπορείτε είτε να αφαιρέσετε όλες τις διπλότυπες σειρές από τον πίνακα εισαγωγής και να διατηρήσετε μόνο τις μοναδικές και επιλεγμένες σειρές είτε να επισημάνετε τις σειρές με πρόσθετες πληροφορίες σχετικά με την κατάστασή τους.

Filter Apply : Αυτός ο κόμβος φιλτράρει την είσοδο σύμφωνα με τους ορισμούς του φίλτρου που είτε δίνονται στον ίδιο τον πίνακα εισόδου είτε προαιρετικά ως πρόσθετη είσοδος μοντέλου. Εάν ένα πρόσθετο μοντέλο δίνεται ως είσοδος, εφαρμόζονται μόνο οι ορισμοί του φίλτρου. Εάν η είσοδος περιέχει ένα φίλτρο που ορίζεται σε μία στήλη που δεν υπάρχει στον πίνακα εισαγωγής, ο κόμβος δεν θα αποτύχει, αλλά εμφανίζει ένα προειδοποιητικό μήνυμα.

Filter Apply Row Splitter : Αυτός ο κόμβος χωρίζει την είσοδο σύμφωνα με τους ορισμούς του φίλτρου που είτε δίνονται στον ίδιο τον πίνακα εισόδου είτε προαιρετικά ως πρόσθετη είσοδος μοντέλου. Εάν ένα πρόσθετο μοντέλο δίνεται ως είσοδος, εφαρμόζονται μόνο οι ορισμοί του φίλτρου.

Filter Definition Merger : Αυτός ο κόμβος συγχωνεύει έως και τρείς ορισμούς φίλτρων σε έναν ορισμό φίλτρου. Δεν τροποποιεί τους ορισμούς φίλτρων, αλλά μόνο τους συγχωνεύει. Οι περισσότερες είσοδοι έχουν προτεραιότητα και ορίζουν το φίλτρο μιας στήλης.

HiLite Row Splitter : Αυτός ο κόμβος διαχωρίζει τις σειρές που έχουν επιλεγεί από τις μη κατακερματισμένες σειρές στα δεδομένα εισόδου σύμφωνα με την κατάσταση τους κατά τη στιγμή της εκτέλεσης του κόμβου. Και οι δύο πίνακες μεταβιβάζονται στη συνέχεια στις θύρες εξόδου.

Nominal Value Row Filter : Φιλτράρει τις σειρές με βάση την επιλεγμένη τιμή ενός ονομαστικού χαρακτηριστικού. Μπορεί να επιλέγει μια ονομαστική στήλη και μια ή περισσότερες ονομαστικές τιμές αυτού του χαρακτηριστικού. Οι σειρές που έχουν αυτήν την ονομαστική τιμή στην επιλεγμένη στήλη περιλαμβάνονται στα δεδομένα εξόδου στη θύρα εξόδου 0.

Nominal Value Row Splitter : Διαχωρίζει τις σειρές με βάση την επιλεγμένη τιμή ενός ονομαστικού χαρακτηριστικού. Μπορεί να επιλέγει μια ονομαστική στήλη και μία ή περισσότερες ονομαστικές τιμές αυτού του χαρακτηριστικού. Οι σειρές που έχουν αυτήν την ονομαστική τιμή στην επιλεγμένη στήλη περιλαμβάνονται στα δεδομένα εξόδου στη θύρα εξόδου 0, οι υπόλοιπες στη θύρα εξόδου 1.

Numeric Row Splitter : Αυτός ο κόμβος χρησιμοποιεί ένα καλά καθορισμένο αριθμητικό εύρος για να χωρίσει τα δεδομένα εισόδου σε δύο μέρη. Ενώ η πρώτη θύρα εξόδου περιέχει τα δεδομένα που ταιριάζουν με τα κριτήρια , η δεύτερη περιέχει τα δεδομένα που δεν συμμορφώνονται με τις ρυθμίσεις. Μέσα στο παράθυρο διαλόγου ο χρήστης μπορεί να επιλέξει μια αριθμητική στήλη και προαιρετικά να καθορίσει ένα κάτω και ένα πάνω όριο σε αυτήν για να χωρίσει τα δεδομένα που ταιριάζουν/ δεν ταιριάζουν με τα κριτήρια.

Reference Row Filter : Αυτός ο κόμβος επιτρέπει το φιλτράρισμα σειρών από τον πρώτο πίνακα χρησιμοποιώντας τον δεύτερο πίνακα ως αναφορά. Ανάλογα με τη ρύθμιση διαλόγου, οι σειρές από τον πίνακα αναφοράς είτε περιλαμβάνονται είτε εξαιρούνται στον πίνακα εξόδου. Κατά τη διάρκεια της δοκιμής για in-/exclusion συγκρίνονται οι τιμές των επιλεγμένων στηλών και των δύο πινάκων.

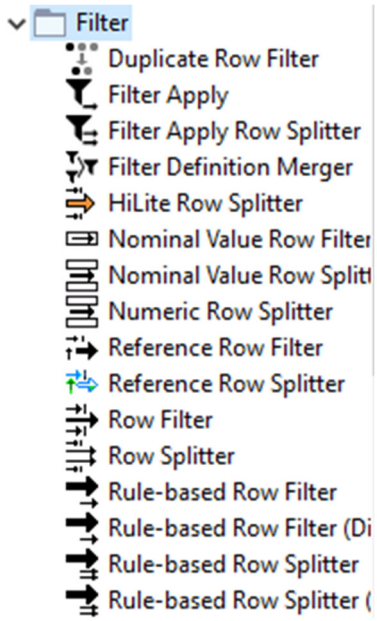
Row Filter : Ο κόμβος επιτρέπει το φιλτράρισμα σειρών σύμφωνα με ορισμένα κριτήρια. Μπορεί να περιλαμβάνει ή να εξαιρεί: ορισμένες περιοχές, σειρές με συγκεκριμένο αναγνωριστικό σειράς και σειρές με συγκεκριμένη τιμή σε μη

επιλέξιμη στήλη. Ακολουθούν τα βήματα για το πώς να ρυθμίσετε τις παραμέτρους του κόμβου στο πλαίσιο διαλόγου διαμόρφωσής του.

Row Splitter : Αυτός ο κόμβος έχει ακριβώς την ίδια λειτουργικότητα με τον κόμβο Row Filter, εκτός από το ότι έχει μια πρόσθετη έξοδο που παρέχει τις σειρές που φιλτράρονται.

Rule-based Row Filter : Αυτός ο κόμβος παίρνει μια λίστα κανόνων που καθορίζουν από το χρήστη και προσπαθεί να τους αντιστοιχίζει σε κάθε γραμμή στον πίνακα εισαγωγής. Εάν ο πρώτος κανόνας αντιστοίχισης έχει TRUE αποτέλεσμα, η σειρά θα επιλεγεί για συμπερίληψη. Διαφορετικά αν αποφέρει FALSE θα εξαιρεθεί.

Rule-based Row Splitter : Αυτός ο κόμβος παίρνει μια λίστα κανόνων που ορίζονται από το χρήστη και προσπαθεί να τους αντιστοιχίσει σε κάθε γραμμή στον πίνακα εισαγωγής με την καθορισμένη σειρά. Εάν ένας κανόνας ταιριάζει και το αποτέλεσμα είναι TRUE, η σειρά θα επιλεγεί για συμπερίληψη στον πρώτο πίνακα εξόδου. Εάν ο πρώτος κανόνας αντιστοίχισης αποφέρει FALSE, η σειρά θα συμπεριληφθεί στον δεύτερο πίνακα εξόδου. Εάν δεν ταιριάζει κανένας κανόνας, αυτή η σειρά θα τοποθετηθεί στον δεύτερο πίνακα εξόδου.[43]



ΕΙΚΟΝΑ 43. Row -Filter

7.2.19 Row Transform

Concatenate : Αυτός ο κόμβος συνενώνει δύο πίνακες. Ο πίνακας στην είσοδο 0 δίνεται ως πίνακας πρώτης εισαγωγής, ο πίνακας στην είσοδο 1 είναι ο δεύτερος πίνακας, αντιστοίχως. Οι στήλες με ίσα ονόματα συνδέονται . Εάν ένας πίνακας εισόδου περιέχει ονόματα στηλών που δεν περιέχει ο άλλος πίνακας, οι στήλες μπορούν είτε να συμπληρωθούν με τιμές που λείπουν είτε να φιλτραριστούν, δηλαδή δεν θα βρίσκονται στον πίνακα εξόδου.

GroupBy : Ομαδοποιεί τις σειρές ενός πίνακα με τις μοναδικές τιμές στις επιλεγμένες στήλες ομάδας. Δημιουργεί μια σειρά για κάθε μοναδικό σύνολο τιμών της επιλεγμένης στήλης συγκεντρώνονται με βάση τις καθορισμένες ρυθμίσεις συγκέντρωσης. Ο κάθε πίνακας εξόδου περιέχει μια γραμμή για κάθε συνδυασμό μοναδικών τιμών των στηλών της επιλεγμένης ομάδας.

Ungroup : Δημιουργεί για κάθε λίστα τιμών συλλογής μια λίστα σειρών με τις τιμές της συλλογής σε μια στήλη και όλες τις άλλες στήλες που δίνονται από την αρχική σειρά. Οι σειρές με κενή συλλογή παραλείπονται, καθώς και σειρές που περιέχουν μόνο τιμές που λείπουν στο κελί συλλογής με το Ενεργοποιήθηκε η επιλογή “Παράλειψη τιμών που λείπουν”.

Partitioning : Ο πίνακας εισόδου χωρίζεται σε δύο διαμερίσματα (δηλαδή κατά σειρά),π.χ δεδομένα τρένου και δοκιμών. Τα δύο διαμερίσματα είναι διαθέσιμα στις δύο θύρες εξόδου.

Pivoting : Εκτελεί μια περιστροφή στον δεδομένο πίνακα εισόδου χρησιμοποιώντας έναν επιλεγμένο αριθμό στηλών για ομαδοποίηση και περιστροφή. Οι στήλες της ομάδας θα οδηγήσουν σε μοναδικές σειρές, όπου οι συγκεντρωτικές τιμές μετατρέπονται σε στήλες για κάθε σύνολο συνδυασμών στηλών μαζί με κάθε μέθοδο συγκέντρωσης. Επιπλέον, ο κόμβος επιστρέφει τη

συνολική συνάθροιση με βάση(α) μόνο τις στήλες της ομάδας και (β) με βάση μόνο τις περιστρεφόμενες στήλες που καταλήγουν σε μία μόνο γραμμή. Προαιρετικά, με τη συνολική συγκέντρωση χωρίς περιστροφή.

Unpivoting : Αυτός ο κόμβος περιστρέφει τις επιλεγμένες στήλες από τον πίνακα εισόδου σε γραμμές και αντιγράφει ταυτόχρονα τις υπόλοιπες στήλες εισόδου προσαρτώντας τις σε κάθε αντίστοιχη γραμμή εξόδου.

Rank : Για κάθε ομάδα υπολογίζεται μια μεμονωμένη κατάταξη με βάση τα επιλεγμένα χαρακτηριστικά κατάταξης και τον τρόπο κατάταξης. Ο χρήστης πρέπει να περιέχει τουλάχιστον ένα χαρακτηριστικό βάσει του οποίου θα πρέπει να υπολογιστεί μια κατάταξη.

Row Sampling : Αυτός ο κόμβος εξάγει ένα δείγμα από τα δεδομένα εισόδου. Το παράθυρο διαλόγου σας δίνει τη δυνατότητα να καθορίσετε το μέγεθος του δείγματος. Οι ακόλουθες επιλογές είναι διαθέσιμες στο παράθυρο διαλόγου.

Bootstrap Sampling : δειγματοληψία των δεδομένων χρησιμοποιώντας το bootstrapping. Το bootstrapping είναι μια τεχνική δειγματοληψίας, η οποία αντλεί τυχαία σειρές από την είσοδο με αντικατάσταση. Επομένως, ο πίνακας εξόδου πιθανότατα θα περιέχει διπλότυπες σειρές ενώ άλλες σειρές δεν υπάρχουν καθόλου στην έξοδο.

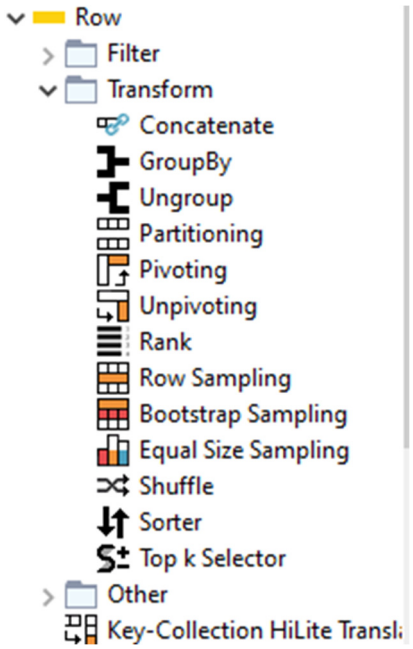
Equal Size Sampling : Καταργεί σειρές από το σύνολο δεδομένων εισόδου έτσι ώστε οι τιμές στην κατηγοριοποιημένη στήλη να κατανέμονται εξίσου. Αυτό μπορεί να είναι χρήσιμο, για παράδειγμα, εάν ένας αλγόριθμος εκμάθησης είναι επιρρεπής σε άνισες κατανομές κλάσεων και θέλετε να μειώσετε το μέγεθος του συνόλου δεδομένων έτσι ώστε τα χαρακτηριστικά κλάσης να εμφανίζονται εξίσου συχνά στο σύνολο δεδομένων. Ο κόμβος θα αφαιρέσει τυχαίες σειρές που ανήκουν στις κλάσεις πλειοψηφίας. Οι θέσεις που επιστρέφονται από αυτόν τον κόμβο θα περιέχουν όλες τις εγγραφές από τις κλάσεις μειοψηφίας και ένα τυχαίο

δείγμα από καθεμία από τις κλάσεις πλειοψηφίας, όπου κάθε δείγμα περιέχει τόσα αντικείμενα όσα περιέχει η κατηγορία μειοψηφίας.

Shuffle : Ανακατεύει τις σειρές των πινάκων εισαγωγής έτσι ώστε να είναι σε τυχαία σειρά.

Sorter : Αυτός ο κόμβος ταξινομεί τις σειρές σύμφωνα με κριτήρια που καθορίζονται από το χρήστη. Στο παράθυρο διαλόγου, επιλέξτε τις στήλες σύμφωνα με τις οποίες πρέπει να ταξινομηθούν τα δεδομένα. Επιπλέον μπορείτε να τα ταξινομήσετε σε αύξουσα ή φθίνουσα σειρά.

Top k Selector : Ο κόμβος συμπεριφέρεται όπως ένας συνδυασμός Sorter node ακολουθούμενο από ένα φίλτρο γραμμής που διατηρεί μόνο τις πρώτες K σειρές του πίνακα εκτός από τη σειρά των σειρών που εξαρτάται από τις ρυθμίσεις παραγγελιών εξόδου. Σημειώστε, ωστόσο, ότι η υλοποίηση αυτού του κόμβου είναι πιο αποτελεσματική στη συνέχεια, ο συνδυασμός κόμβων παραπάνω. Στο παράθυρο διαλόγου, επιλέξτε τις στήλες σύμφωνα με τις οποίες θα πρέπει να επιλεγούν τα δεδομένα. Για κάθε στήλη μπορείτε επίσης να καθορίσετε εάν μια μεγαλύτερη ή μικρότερη τιμή θεωρείται ανώτερη.[44]



ΕΙΚΟΝΑ 44. Row- Transform

7.2.20 Row Other

Add Empty Rows : Προσθέτει έναν ορισμένο αριθμό κενών σειρών με τιμές που λείπουν στον πίνακα εισαγωγής. Αυτό μπορεί να είναι χρήσιμο όταν χρησιμοποιείται σε έναν πίνακα αναφοράς για να διασφαλιστεί ότι ένας πίνακας έχει ελάχιστο αριθμό σειρών, οι οποίες στη συνέχεια εμφανίζονται ως κενές σειρές. Το περιεχόμενο των σειρών που επισυνάπτονται μπορεί να οριστεί στο παράθυρο διαλόγου, η προεπιλογή είναι να συμπληρώσετε τα αντίστοιχα κελιά που λείπουν αξίες. Σημειώστε ότι η μηχανή αναφοράς σας επιτρέπει να μορφοποιήσετε κελιά που περιέχουν τιμές που λείπουν χρησιμοποιώντας τη δυνατότητα χάρτης.

Extract Column Header : Δημιουργεί νέο πίνακα με μία μόνο γραμμή που περιέχει τα ονόματα των στηλών. Ο κόμβος έχει δύο εξόδους : Η πρώτη θύρα περιέχει τις κεφαλίδες στηλών και η δεύτερη θύρα περιέχει τα δεδομένα εισόδου, όπου τα ονόματα στηλών αλλάζουν σε προεπιλεγμένο μοτίβο.

Insert Column Header : Ενημερώνει τα ονόματα στηλών ενός πίνακα σύμφωνα με τον πίνακα αντιστοίχισης του δεύτερου λεξικού. Ο πίνακας λεξικού πρέπει να περιέχει δύο στήλες, η μία από τις οποίες περιέχει την αναζήτηση, η άλλη στήλη να περιέχει τα ονόματα των νέων στηλών. Η στήλη αναζήτησης μπορεί να είναι η στήλη RowID.

Εάν η εκχώρησα νέα τιμή στη στήλη τιμής λείπει, το αρχικό όνομα της στήλης θα διατηρηθεί. Εάν η στήλη αναζήτησης περιέχει διπλότυπα των αρχικών ονομάτων στηλών, ο κόμβος θα αποτύχει.

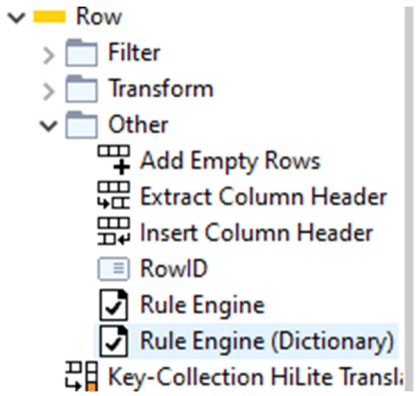
RowID : Αυτός ο κόμβος μπορεί να χρησιμοποιηθεί για να αντικαταστήσει το RowID των δεδομένων εισόδου με τις τιμές μιας άλλης στήλης ή να δημιουργήσει αναγνωριστικό σειράς της μορφής : row0, Row1, Row2, ….. ο

χρήστης έχει πρόσθετες επιλογές που διασφαλίζουν τη μοναδικότητα και τις αξίες που δεν επιτρέπουν τον χειρισμό. Μπορεί επίσης να χρησιμοποιηθεί για τη δημιουργία μιας νέας στήλης, η οποία περιέχει μια τιμή RowIDas.

Rule Engine : Αυτός ο κόμβος παίρνει μια λίστα κανόνων που καθορίζονται από το χρήστη και προσπαθεί να τους αντιστοιχίσει σε κάθε γραμμή στον πίνακα εισαγωγής. Εάν ένας κανόνας ταιριάζει, η τιμή του αποτελέσματος του προτίθεται σε μία νέα στήλη. Ο πρώτος κανόνας αντιστοίχισης κατά σειρά ορισμού καθορίζει το αποτέλεσμα.

Κάθε κανόνας αντιπροσωπεύεται από μια γραμμή. Για να προσθέσετε σχόλια, ξεκινήστε μια γραμμή με //. Οτιδήποτε μετά το // δεν θα ερμηνευτεί ως κανόνας. Οι κανόνες αποτελούνται από ένα μέρος συνθήκης , το οποίο πρέπει να αξιολογείται ως true ή false, και ένα αποτέλεσμα που τοποθετείται στη νέα στήλη εάν ο κανόνας ταιριάζει.

Το αποτέλεσμα ενός κανόνα μπορεί να είναι οποιοδήποτε από τα ακόλουθα : μια συμβολοσειρά, ένας αριθμός, μια σταθερά boole, μια αναφορά σε μια άλλη στήλη ή η τιμή μίας τιμής μεταβλητής ροής. Ο τύπος της στήλης αποτελέσματος είναι ο κοινός υπερ. τύπος όλων των πιθανών αποτελεσμάτων. Εάν κανένας κανόνας δεν ταιριάζει, το αποτέλεσμα λείπει.[45]



ΕΙΚΟΝΑ 45. Row -Other

7.2.21 Table

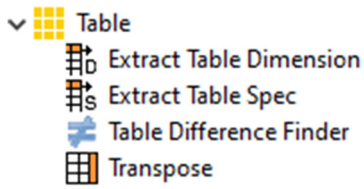
Extract Table Dimension : Ο αριθμός των γραμμών και των στηλών του πίνακα εισόδου εξάγεται και γράφεται στον πίνακα εξόδου. Ο πίνακας εξόδου αποτελείται από δύο σειρές και μία στήλη. Η πρώτη σειρά περιέχει τον αριθμό των σειρών και η δεύτερη σειρά περιέχει τον αριθμό των στηλών του πίνακα εισαγωγής. Επιπλέον, οι δύο αριθμοί ωθούνται προς τα έξω ως μεταβλητές ροής.

Extract Table Spec : Ο κόμβος εξάγει τις μετα-πληροφορίες από τον πίνακα εισαγωγής(ονόματα στηλών, τύποι κ.λ.π.). Οι μετα-πληροφορίες αναφέρονται ως προδιαγραφή δεδομένων πίνακα και περιέχουν πληροφορίες όπως ονόματα στηλών, τύπους, τιμές τομέα και κάτω και άνω όρια. Περιέχει επίσης τις πληροφορίες , ποιες από τις στήλες έχουν συσχετισμένο πρόγραμμα χειρισμού προβολής. Κάθε στήλη στον πίνακα εισόδου αναπαρίσταται ως μια γραμμή στην έξοδο. Αυτός ο κόμβος μπορεί να χρησιμοποιηθεί σε συνδυασμό με τον κόμβο εισαγωγής κεφαλίδας στήλης, όπου τα ονόματα στηλών εξάγονται από τον πίνακα εισόδου, στη συνέχεια τροποποιούνται στη ροή εργασίας και αργότερα συγχωνεύονται στον κύριο πίνακα.

Table Difference Finder : Προσφέρει τη δυνατότητα σύγκρισης δύο πινάκων με βάση τις τιμές και τις προδιαγραφές πίνακα. Πρώτον, οι τιμές στις επιλεγμένες στήλες συγκρίνονται και στους δύο πίνακες και εμφανίζονται οι πιθανές διαφορές για κάθε γραμμή και στήλη. Δεύτερον, οι θέσεις των επιλεγμένων στηλών συγκρίνονται και στους δύο πίνακες και τα αποτελέσματα εμφανίζονται για κάθε στήλη. Οι επιλεγμένες στήλες είναι είτε όλες οι στήλες και στους δύο πίνακες είτε ένα υποσύνολο στηλών στον πίνακα αναφοράς, δηλαδή η δεύτερη είσοδος.

Transpose : Μεταφέρει ολόκληρο τον πίνακα εισαγωγής εναλλάσσοντας σειρές και στήλες. Τα νέα ονόματα στηλών παρέχονται από τα προηγούμενα αναγνωριστικά σειρών και τα νέα αναγνωριστικά σειρών είναι τα προηγούμενα

ονόματα στηλών. Ο νέος τύπος στήλης είναι ο πιο συγκεκριμένος τύπος βάσης και ισχύει για όλα τα κελιά σε μία σειρά.[46]



ΕΙΚΟΝΑ 46. Menu Table

7.2.22 PMML

Column Filter(PMML) : Αυτός ο κόμβος επιτρέπει το φιλτράρισμα στηλών από τον πίνακα εισόδου, ενώ μόνο οι υπόλοιπες στήλες μεταβιβάζονται στον πίνακα εξόδου. Μέσα στο παράθυρο διαλόγου, οι στήλες μπορούν να μετακινηθούν μεταξύ της λίστας συμπερίληψης και εξαίρεσης.

Denormalizer (PMML) : Αυτός ο κόμβος αποκανονικοποιεί τα δεδομένα εισόδου σύμφωνα με τις παραμέτρους κανονικοποίησης που δίνονται στην είσοδο του μοντέλου PMML. Επομένως, ο συγγενικός μετασχηματισμός αντιστρέφεται και οι αρχικές τιμές αναδημιουργούνται.

Αυτός ο κόμβος χρησιμοποιείται συνήθως μετά την κανονικοποίηση των δεδομένων δοκιμής και την πιθανή μετατροπή άλλων μαθημάτων/προβλέψεων στο αρχικό εύρος.

Many to One (PMML) : Μετατρέπει τις τιμές πολλών στηλών σε μία στήλη ανάλογα με τη μέθοδο συμπερίληψης και τεκμηριώνει τον μετασχηματισμό στο PMML. Εάν οριστεί σε δυαδικό, η τιμή της συμπυκνωμένης στήλης ορίζεται στο όνομα της πρώτης στήλης με τιμή 1. Εάν η μέθοδος συμπερίληψης είναι μέγιστη ή ελάχιστη, η τιμή της συμπυκνωμένης στήλης ορίζεται στο όνομα της στήλης που, από όλες τις στήλες που περιλαμβάνονται, έχει τη μεγαλύτερη ή τη

μικρότερη τιμή. Εάν η μέθοδος συμπερίληψης είναι δυαδική και όλες οι στήλες είναι 0, η συμπυκνωμένη στήλη θα περιέχει μια τιμή που λείπει.

Normalizer (PMML) : Αυτός ο κόμβος κανονικοποιεί τις τιμές όλων των στηλών. Στο παράθυρο διαλόγου, μπορείτε να επιλέξετε τις στήλες στις οποίες θέλετε να εργαστείτε. Οι ακόλουθες μέθοδοι κανονικοποίησης είναι διαθέσιμες στο παράθυρο διαλόγου.

Number To String (PMML) : Μετατρέπει τους αριθμούς σε μια στήλη σε συμβολοσειρές. Σημειώστε ότι για μια προηγούμενη διαμόρφωση, όπως στρογγυλοποίηση ή αναπαράσταση σε επιστημονική ισοποίηση, μπορείτε επίσης να χρησιμοποιήσετε τον κόμβο “Round Double”. Εάν η προαιρετική είσοδος PMML είναι συνδεδεμένη και περιέχει λειτουργίες Προ επεξεργασίας στο Transformation Dictionary, οι πράξεις μετατροπής αυτού του κόμβου επισυνάπτονται.

Normalizer Apply (PMML) : Αυτός ο κόμβος κανονικοποιεί τα δεδομένα εισόδου σύμφωνα με τις παραμέτρους κανονικοποίησης που δίνονται στην είσοδο του μοντέλου PMML. Θα εφαρμόσει έναν συγγενικό μετασχηματισμό σε όλες τις στήλες στα δεδομένα εισόδου που περιέχονται στην είσοδο του μοντέλου.

Αυτός ο κόμβος χρησιμοποιείται συνήθως όταν τα δεδομένα δοκιμής πρέπει να κανονικοποιηθούν με τον ίδιο τρόπο που έχουν κανονικοποιηθεί τα δεδομένα εκπαίδευσης.

Numeric Binner (PMML) : Για κάθε στήλη μπορεί να οριστεί ένας αριθμός διαστημάτων - γνωστά ως bins. Διασφαλίζουν αυτόματα ότι τα εύρη ορίζονται με φθίνουσα σειρά και ότι τα διαστήματα είναι συνεπή. Επιπλέον, κάθε στήλη είτε αντικαθίσταται με τη στήλη δεσμευμένης, τύπου συμβολοσειράς, είτε προστίθεται μια νέα δεσμευμένη στήλη τύπου συμβολοσειράς.

One to Many (PMML) : Μετατρέπει όλες τις πιθανές τιμές σε μια επιλεγμένη στήλη η καθεμία σε μια νέα στήλη. Η τιμή ορίζεται ως το όνομα της νέας στήλης, οι τιμές των κελιών σε αυτήν τη στήλη είναι είτε 1, εάν αυτή η σειρά περιέχει αυτήν την πιθανή τιμή, είτε 0, εάν όχι. Ο κόμβος προσαρτά όσες στήλες είναι δυνατές τιμές που ορίζονται για τις επιλεγμένες στήλες. Εάν μια σειρά περιέχει μια τιμή που λείπει σε μια επιλεγμένη στήλη, όλες οι αντίστοιχες νέες στήλες περιέχουν την τιμή 0. Για την αποφυγή διπλών ονομάτων στηλών με πανομοιότυπες πιθανές τιμές σε διαφορετικές επιλεγμένες στήλες, το όνομα της στήλης που δημιουργείται περιλαμβάνει το αρχικό όνομα της στήλης σε αυτήν την περίπτωση. Το παράθυρο διαλόγου του κόμβου σας επιτρέπει μόνο να επιλέξετε στήλες με ονομαστικές τιμές. Εάν δεν εμφανίζεται κανένα όνομα στήλης στο παράθυρο διαλόγου, αλλά ο πίνακας εισαγωγής περιέχει ονομαστικές στήλες, μπορείτε να χρησιμοποιήσετε το Domain Calculator node και να συνδέσετε την έξοδο του σε αυτόν τον κόμβο.

Rule Set Editor : Χρησιμοποιώντας τους καθορισμένους κανόνες, δημιουργεί ένα PMML Rule Set model. Εφαρμόζει επίσης τους κανόνες στον πίνακα εισόδου. Αυτός ο κόμβος παίρνει μια λίστα κανόνων που ορίζονται από το χρήστη και προσπαθεί να τους αντιστοιχίσει σε κάθε σειρά στον πίνακα εισαγωγής. Εάν ένας κανόνας ταιριάζει, η τιμή του αποτελέσματος του προστίθεται σε μια νέα στήλη. Ο πρώτος κανόνας αντιστοίχισης κατά σειρά ορισμού καθορίζει το αποτέλεσμα.

Rule Set Predictor : Αυτός ο κόμβος παίρνει έναν πίνακα δεδομένων και ένα μοντέλο PMML για Rule Sets, π.χ δημιουργήθηκε από τον κόμβο επεξεργασίας συνόλου κανόνων PMML και προβλέπει μια νέα στήλη σύμφωνα με τους κανόνες.

Rule Set to Table : Μετατρέπει τα PMML Rule Set models σε πίνακα που περιέχει τους κανόνες. Αυτός ο πίνακας είναι κατάλληλος για τον κόμβο Rule

Engine ως είσοδο κανόνα. Χρησιμοποιήστε περιπτώσεις για αυτόν τον κόμβο: Μετατρέψτε το μοντέλο PMML σε συνηθισμένο πίνακα σε

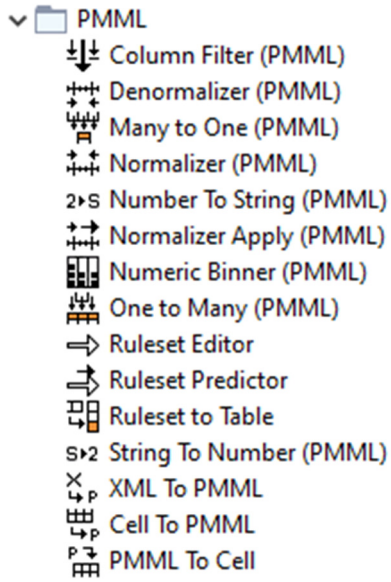
- Σύνδεση διαφορετικών πινάκων κανόνων
- Φιλτράρει ορισμένα αποτελέσματα
- Επανεπιπέδωση / ανα-κωδικοποίηση των αποτελεσμάτων

String To Number (PMML) : Μετατρέπει τις συμβολοσειρές από μια στήλη σε αριθμούς. Εάν ο κόμβος αποτύχει να αναλύσει μια συμβολοσειρά, θα δημιουργήσει ένα προειδοποιητικό μήνυμα κελιού και παραρτήματος που λείπει στην κονσόλα KNIME με λεπτομερείς πληροφορίες.

XML To PMML : Αυτός ο κόμβος διαβάζει μια στήλη τιμών XML και εξάγει μια στήλη τιμών PMML.

Cell To PMML : Μετατρέπει το κελί PMML στην πρώτη σειρά στη θύρα PMML.

PMML To Cell : Μετατρέπει τη θύρα PMML σε πίνακα που περιέχει το κελί PMML.[47]



ΕΙΚΟΝΑ 47. PMML Menu

7.2.23 Scoring

Scorer : Συγκρίνει δύο στήλες με βάση τα ζεύγη τιμών χαρακτηριστικών τους και εμφανίζει τον πίνακα σύγχυσης, δηλ. πόσες σειρές του χαρακτηριστικού και της ταξινόμησής τους ταιριάζουν. Επιπλέον, είναι δυνατό να επισημάνετε κελιά αυτού του πίνακα για να προσδιορίσετε τις υποκείμενες σειρές. Το παράθυρο διαλόγου σάς επιτρέπει να επιλέξετε δύο στήλες για σύγκριση· οι τιμές από την πρώτη επιλεγμένη στήλη αντιπροσωπεύονται στη σύγχυση τις σειρές του πίνακα και τις τιμές από τη δεύτερη στήλη από τις στήλες του πίνακα σύγχυσης. Η έξοδος του κόμβου είναι η μήτρα σύγχυσης με τον αριθμό των αντιστοιχιών σε κάθε κελί. Επιπλέον, η δεύτερη θύρα εξόδου αναφέρει μια σειρά από στατιστικά στοιχεία ακρίβειας όπως True-Positives, False-Positives, True-Negatives, False-Negatives, Recall, Precision, Sensitivity, Specificity, F-measure, as well as the overall accuracy and Cohen's kappa .

ΚΕΦΑΛΑΙΟ 8
ΔΗΜΙΟΥΡΓΙΑ ΠΑΡΑΔΕΙΓΜΑΤΩΝ ΡΟΗΣ ΔΕΔΟΜΕΝΩΝ ΣΤΟ ΠΡΟΓΡΑΜΜΑ KNIME

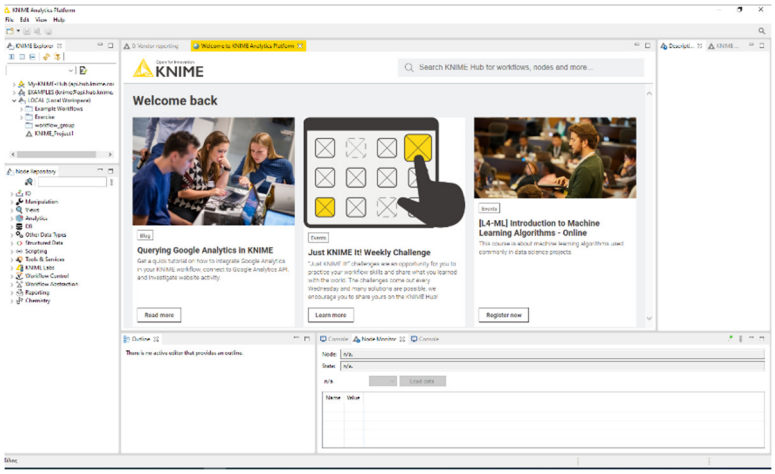
Εισαγωγή

Σε αυτό το κεφάλαιο θα δημιουργήσουμε και να αναλύσουμε κάποια παραδείγματα, για το πώς λειτουργεί το πρόγραμμα KNIME.

8.1 Παράδειγμα 1ο (Δημιουργία ροής δεδομένων με αρχείο Excel)

8.1.1 Αρχικό περιβάλλον KNIME

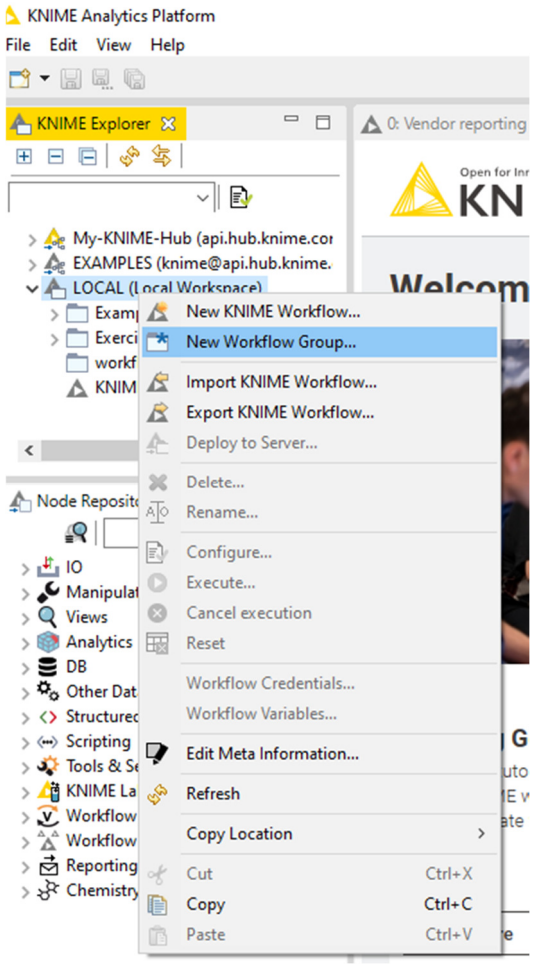
Σαν πρώτο παράδειγμα θα δείξουμε μία απλή ροή αναλύοντας κάθε βήμα. Στην παρακάτω εικόνα βλέπουμε την αρχική εμφάνιση του περιβάλλοντος KNIME.[48]



ΕΙΚΟΝΑ 48. Περιβάλλον KNIME

8.1.2 Δημιουργία Νέας Ροής Εργασίας

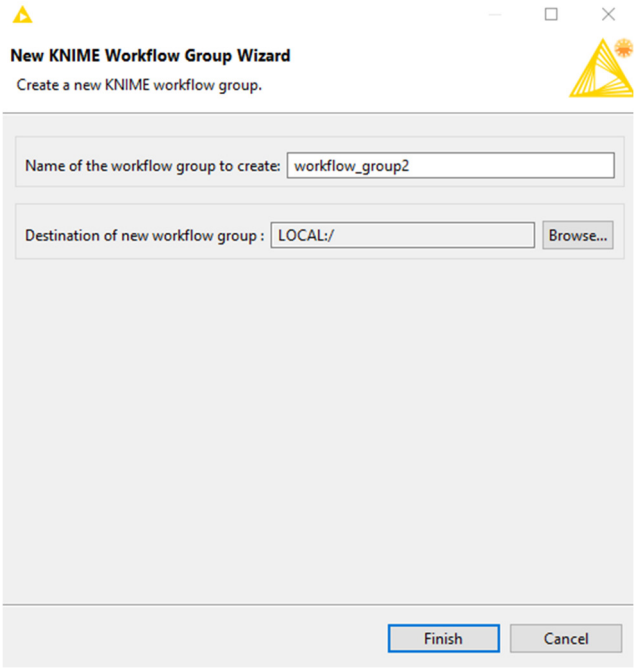
Για να δημιουργήσουμε μία νέα ροή, πάμε πάνω αριστερά στο KNIME Explorer – Local Workspace και κάνουμε δεξί κλικ πάνω στο LOCAL και επιλέγουμε New Workflow Group.[49]



ΕΙΚΟΝΑ 49. Δημιουργία νέας Ροής

8.1.3 Επιλογή ονόματος φακέλου

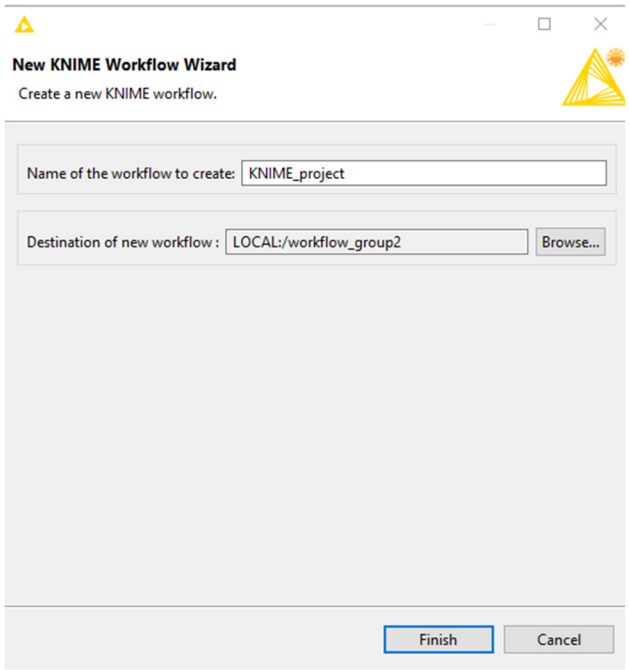
Επιλέγουμε το όνομα φακέλου για τις ροές εργασίας που θα δημιουργήσουμε μέσα στο φάκελο και πατάμε FINISH.[50]



ΕΙΚΟΝΑ 50. Όνομα φακέλου αποθήκευσης παραδείγματος

8.1.4 Επιλογή ονόματος εργασίας

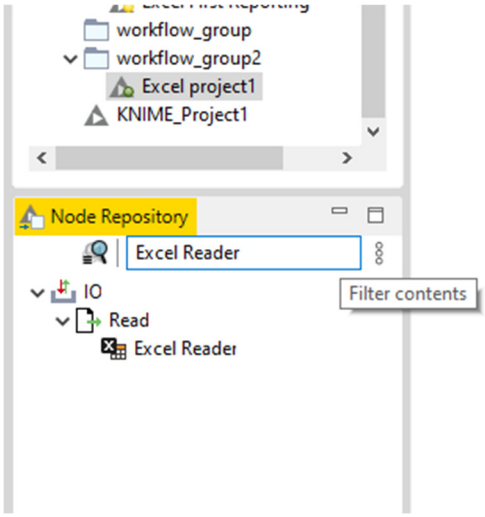
Επιλέγουμε το όνομα της εργασίας που θα δημιουργήσουμε και πατάμε FINISH.[51]



ΕΙΚΟΝΑ 51. Επιλογή ονόματος εργασίας

8.1.5 Δημιουργία Excel Workflow

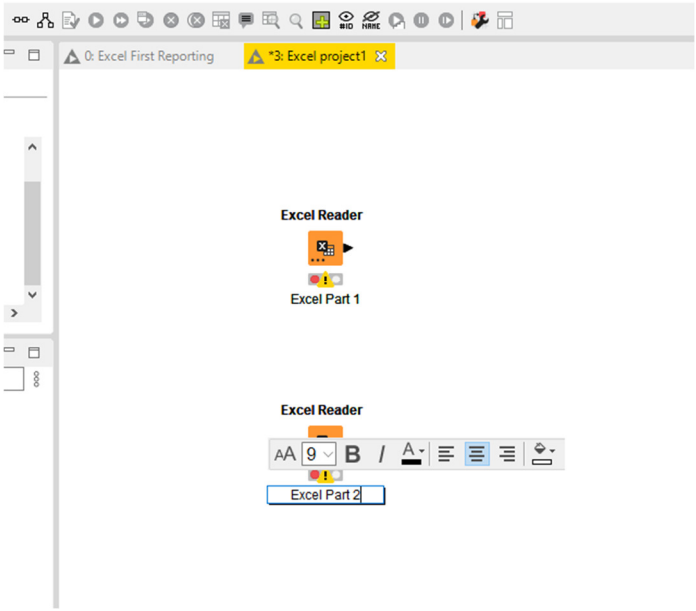
Ξεκινώντας, θα δημιουργήσουμε μια εργασία για να αντλήσουμε πληροφορίες από δύο Excel και να τα ενώσεις σε 1. Αρχικά για να προσθέσουμε το Excel Reader. Για να μην ψάχνουμε πάμε στην αναζήτηση και πληκτρολογούμε Excel Reader για να μας το εμφάνιση.[52]



ΕΙΚΟΝΑ 52. Excel Reader

8.1.6 Excel Reader

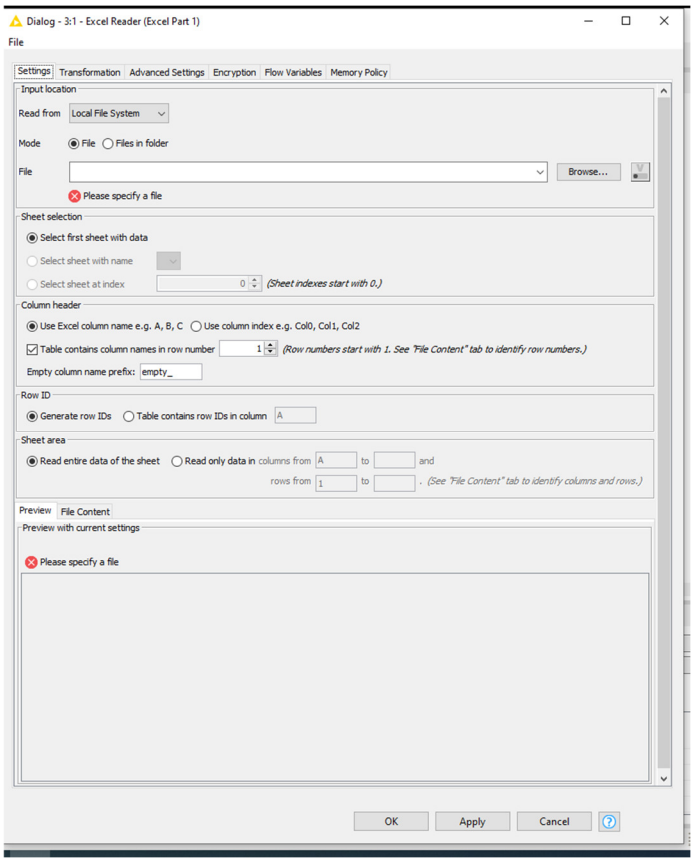
Πατάμε διπλό κλικ πάνω στο Excel Reader για να το προσθέσουμε μέσα στην ροή εργασίας. Προσθέσαμε το Excel Reader για να μπορέσουμε να περάσουμε μέσα σε αυτό τον πίνακα από τον οποίο θα λάβουμε τις πληροφορίες.[53]



EIKONA 53. Excel Reader Column

8.1.7 Πίνακας Excel Reader

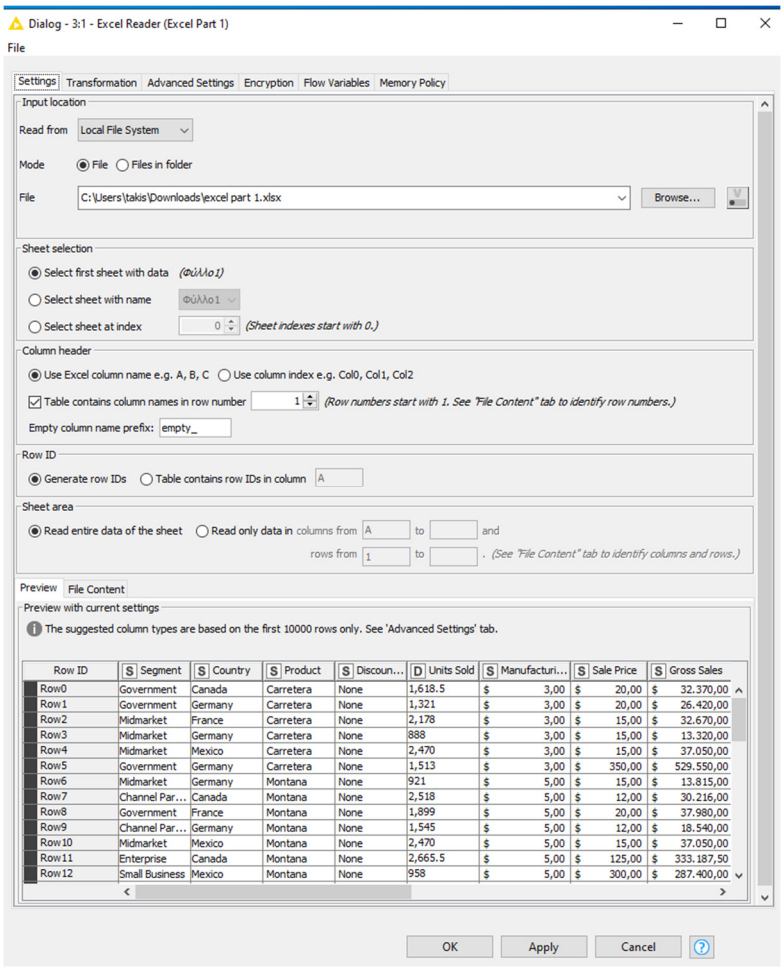
Πατάμε διπλό κλικ πάνω του για να μας εμφανίσει τον παρακάτω πίνακα , πάνω αριστερά έχει διάφορες επιλογές για να προσθέσουμε το αρχείο. Έπειτα στο File πάμε browse και διαλέγουμε το Excel που θέλουμε.[54]



ΕΙΚΟΝΑ 54. Πίνακας Excel Reader

8.1.8 Preview του αρχείου.

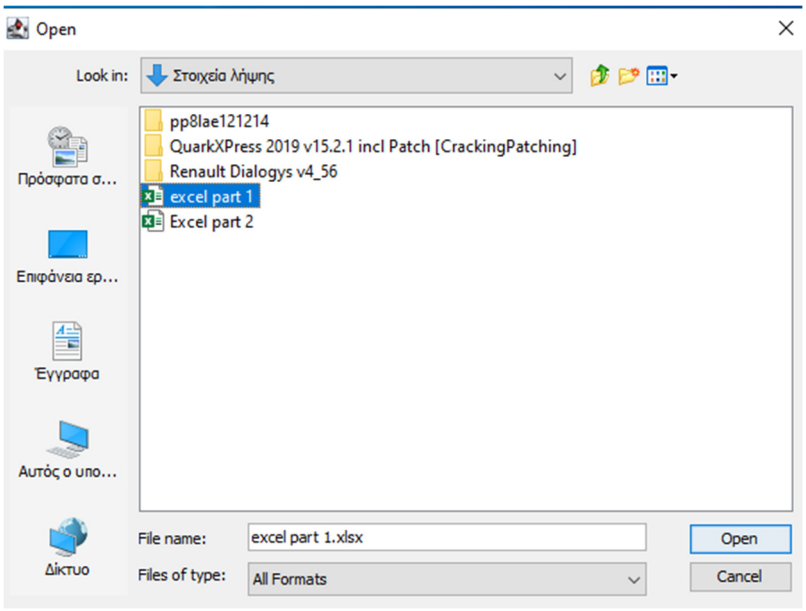
Στο κάτω μέρος του KNIME εμφανίζονται τα αρχεία του φακέλου που επιλέξατε και πατάτε OK. Αν εμφανιστεί ο πίνακας κανονικά τότε έχουμε βάλει το σωστό τύπο αρχείου και επιπλέον θα βγάλουμε ένα σωστό αποτέλεσμα.[55]



ΕΙΚΟΝΑ 55. Preview αρχείων

8.1.9 Επιλογή αρχείου excel .

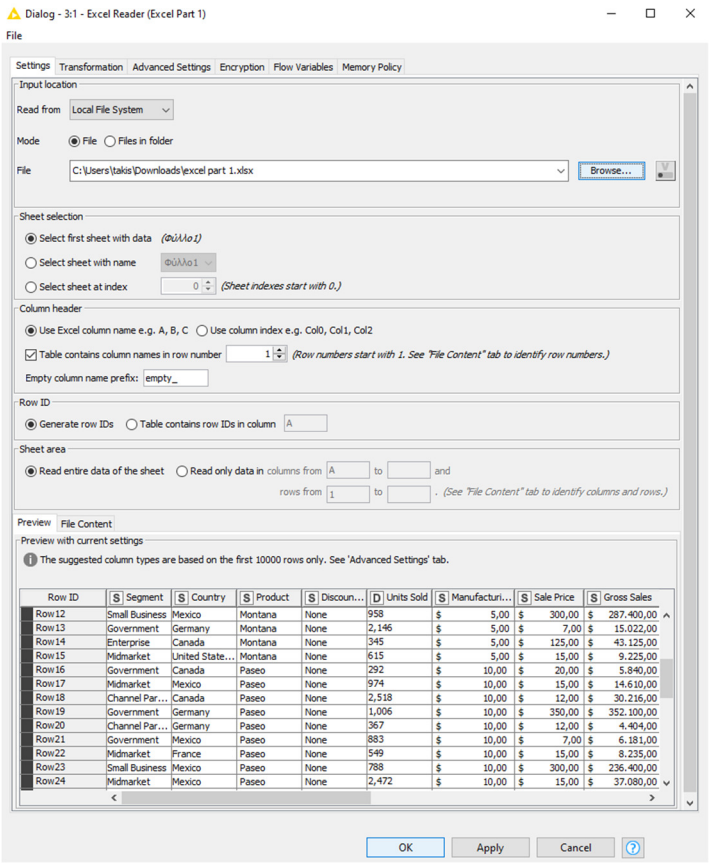
Σε αυτό το σημείο επιλέγουμε το αρχείο excel που θέλουμε να πάρουμε πληροφορίες, και πατάμε Open. [56]



ΕΙΚΟΝΑ 56. Αρχείο Excel

8.1.10 Settings Excel Reader

Εφόσον έχουμε προσθέσει το αρχείο στον πίνακα, θα εμφανιστεί στο κάτω μέρος τα αρχεία που έχει μέσα το Excel και πατάμε οκ.[57]

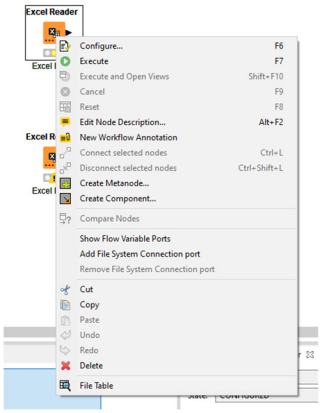


EIKONA 57. Settings Excel Reader

8.1.11 Execute Excel Reader

Μετά την προσθήκη του πίνακα , πατάμε δεξί κλικ πάνω στον κόμβο και Execute(F7).

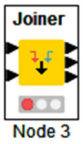
Την ίδια διαδικασία ακολουθάμε και για τον δεύτερο κόμβο.[58]



ΕΙΚΟΝΑ 58. Execute Reader

8.1.12 Joiner

Με τον ίδιο τρόπο που προσθέσαμε το Excel Reader προσθέτουμε το Joiner. Node Repository - Manipulation – Column – Split &Combine – Joiner. Άλλος ένας τρόπος είναι να γράψουμε στην αναζήτηση Joiner και θα σας το εμφανίσει.[59]



ΕΙΚΟΝΑ 59. Joiner Column

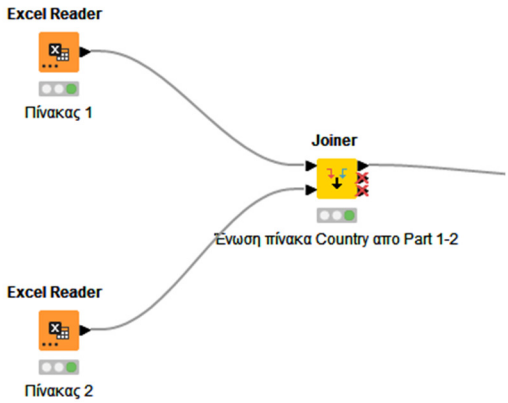
8.1.13 Green Light

Εφόσον έχουμε προσθέσει τους πίνακες και πατήσουμε execute, πρέπει να ανάψει το πράσινο λαμπάκι, αλλιώς κάτι είναι λάθος και πρέπει να διορθωθεί, το πράσινο λαμπάκι είναι η ένδειξη ότι η διαδικασία που ακολουθούμε είναι σωστή. Αν έχουμε κάνει κάποιο λάθος θα εμφανίσει πορτοκαλί λαμπάκι με !, το οποίο αν σύρουμε τον κέρσορα πάνω στον σύμβολο αυτό, αναφέρει που βρίσκεται το λάθος. [60]

ΕΙΚΟΝΑ 60. Green Light

8.1.14 Ένωση Πινάκων 1 και 2

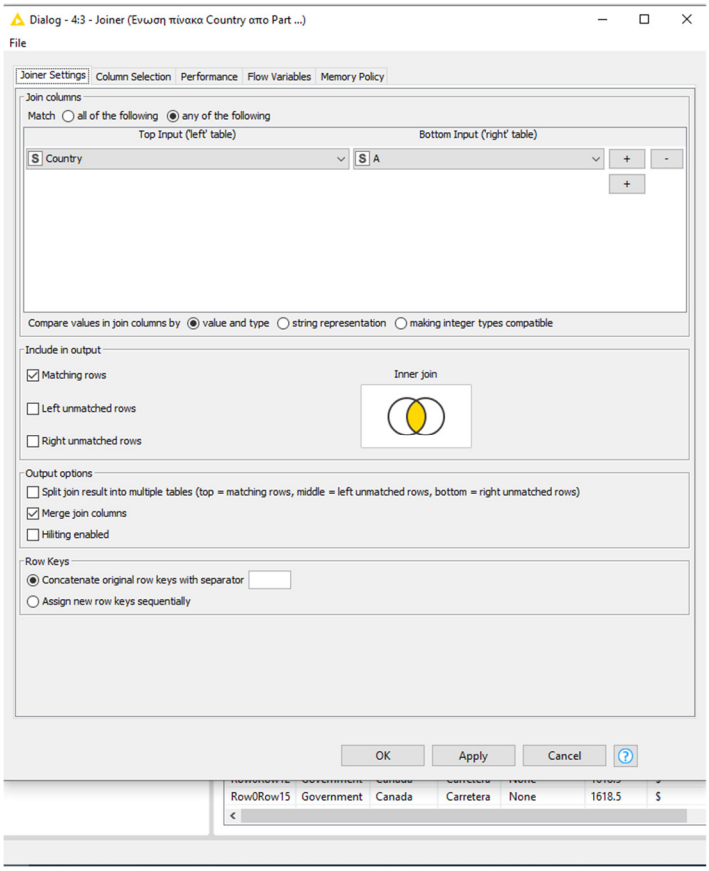
Ενώνουμε τους πίνακες 1-2 στον κόμβο Joiner. Με το ποντίκι σύρουμε από τον πίνακα 1 και 2 πάνω στον κόμβο ένωσης όπως φαίνεται στην παρακάτω εικόνα.[61]



ΕΙΚΟΝΑ 61. Joiner Column

8.1.15 Joiner Settings

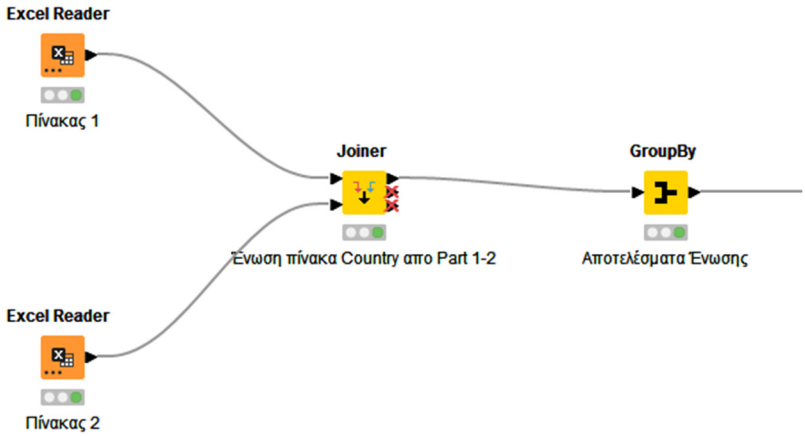
Πατάμε διπλό κλικ πάνω στον κόμβο Joiner και προσθέτουμε τους πίνακες που θέλουμε να ενώσουμε, πατάμε οκ και F7(Execute). Μπορούμε να προσθέσουμε από 1 έως και όλες τις στήλες. [62]



EIKONA 62. Joiner Settings

8.1.16 GroupBy Column

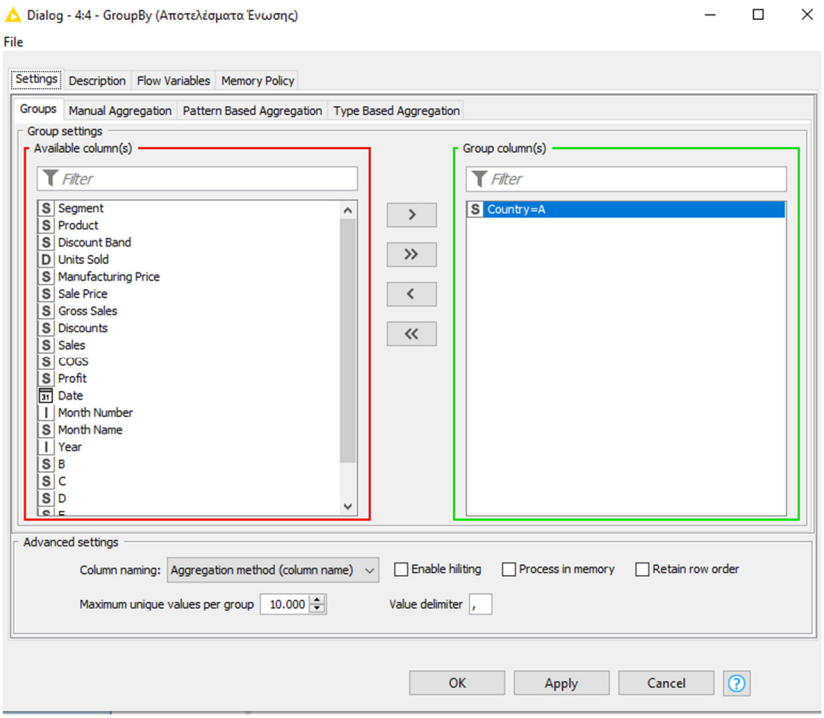
Στην συνέχεια προσθέτουμε τον κόμβο GroupBy τον ενώνουμε με τον κόμβο Joiner, για να ομαδοποιήσουμε τους πίνακες που προσθέσαμε στον κόμβο Joiner.[63]



EIKONA 63. GroupBy Column

8.1.17 GroupBy Setting

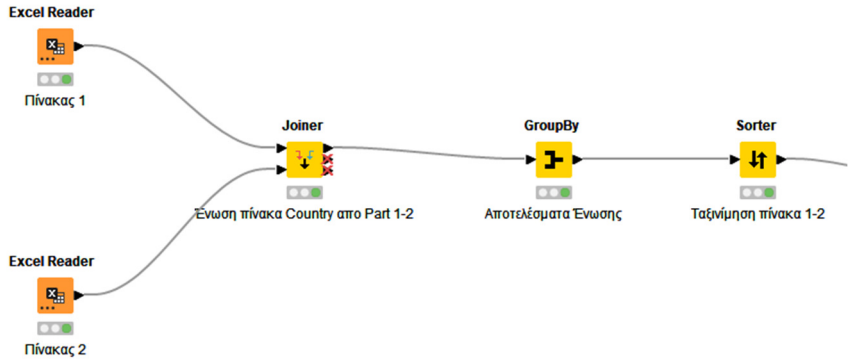
Επιλέγουμε την ομάδα που θέλουμε το αποτέλεσμα και πατάμε οκ, και στην συνέχεια F(7) Execute.[64]



EIKONA 64. GroupBy Settings

8.1.18 Κόμβος Sorter

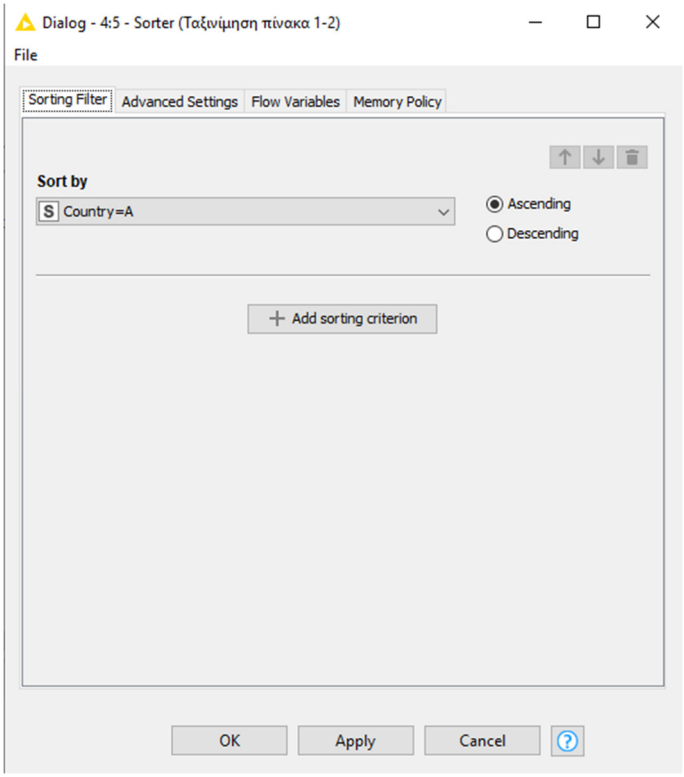
Προσθέτουμε τον κόμβο Sorter για να ταξινομεί τους πίνακες που έχουμε προσθέσει.[65]



ΕΙΚΟΝΑ 65. Sorter Column

8.1.19 Sorter Settings

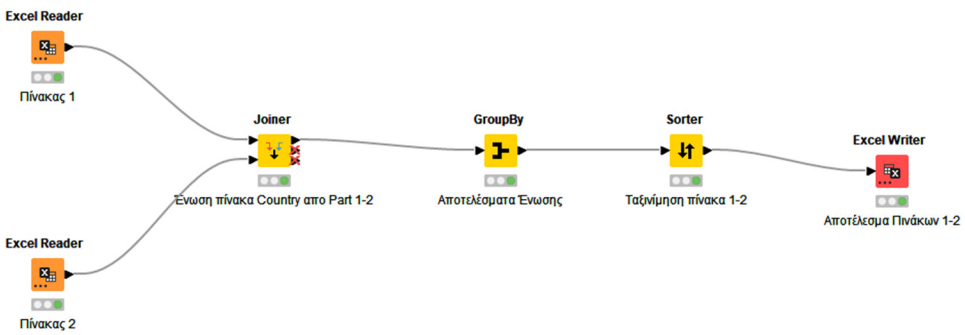
Επιλέγουμε τους πίνακες που θέλουμε να κάνουμε ταξινόμηση, μπορούμε να προσθέσουμε και με ποια κριτήρια. Πατάμε OK και F7(Execute).[66]



EIKONA 66. Sorter Settings

8.1.20 Excel Writer Column

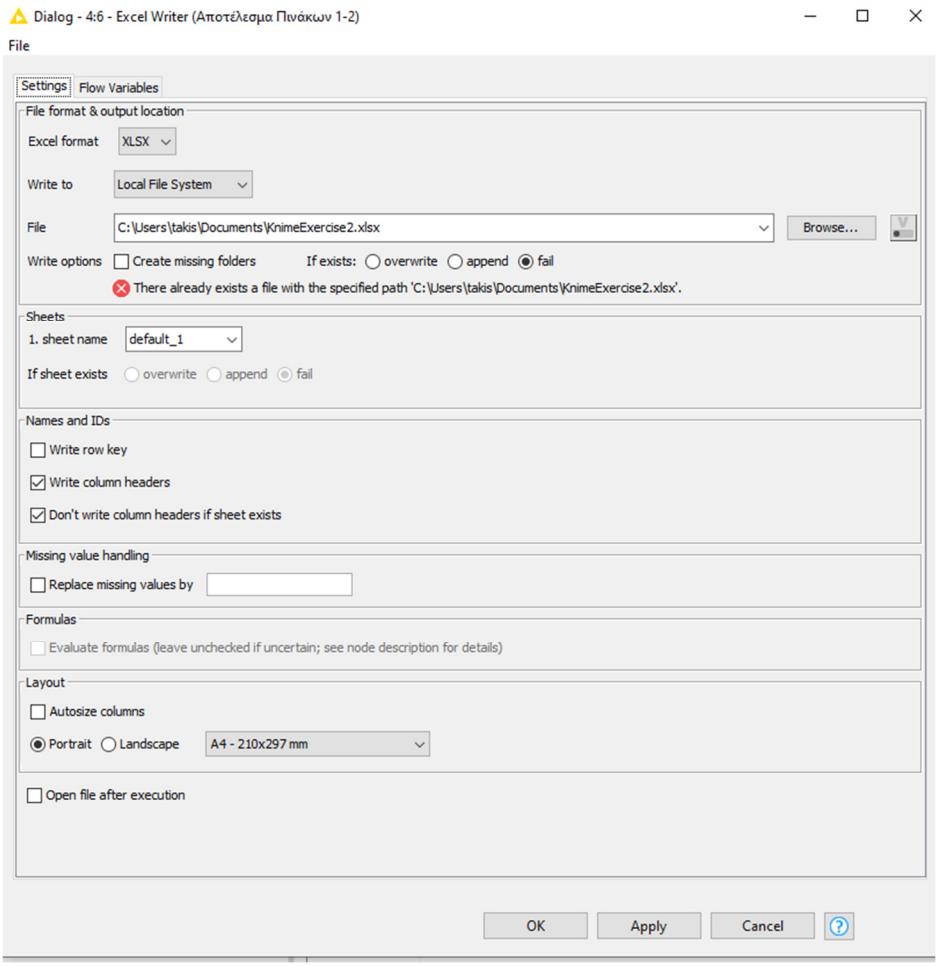
Προσθέτουμε το Excel Writer και το ενώνουμε με τον Sorter. Αυτός ο κόμβος δίνει το αποτέλεσμα των στηλών που επιλέξαμε. Θα εμφανίσει μια νέα πληροφορία με βάση τα στοιχεία των στηλών που προσθέσαμε στους προηγούμενους κόμβους.[67]



EIKONA 67. Column Excel Writer

8.1.21 Excel Writer Settings

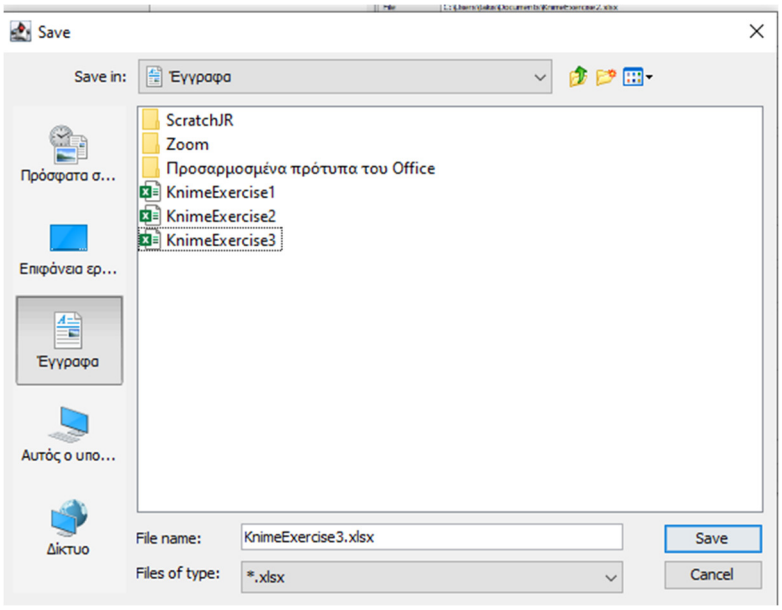
Μέσα στις ρυθμίσεις του Excel Writer μπορούμε να αποθηκεύσουμε όπου θέλουμε εμείς το αρχείο και διάφορες άλλες ρυθμίσεις όπως το μέγεθος του χαρτιού κτλ. Πατάμε OK και F7(Execute).[68]



ΕΙΚΟΝΑ 68. Excel Writer Settings

8.1.22 Τοποθεσία αποθήκευσης αποτελέσματος

Το αποτέλεσμα θα είναι σε αρχείο Excel , επιλέγουμε το όνομα που θέλουμε και Save.[69]



ΕΙΚΟΝΑ 69. Τοποθεσία αποθήκευσης αποτελέσματος

Συμπεράσματα : Σε αυτό το παράδειγμα ενώνουμε 2 πίνακες με σκοπό την ομαδοποίηση και σύγκριση των στηλών τις επιλογής μας. Για πίνακες Excel, χρησιμοποιούμε το Excel Reader για να προσθέσουμε των πίνακα. Το Joiner θα το χρησιμοποιήσουμε μόνο αν θέλουμε να ενώσουμε 2 πίνακες. Στην συνέχεια θα προσθέσουμε το GroupBy για να ομαδοποιήσουμε τις σειρές των πινάκων που θέλουμε. Έπειτα θα προσθέσουμε το Sorter για να ταξινομήσουμε τις σειρές που επιλέξαμε από τους δύο πίνακες με βάση τα κριτήρια μας, και τέλος θα προσθέσουμε το Excel Writer για να πάρουμε το αποτέλεσμα τις ομαδοποίησης των πινάκων.

Ερωτήματα :

1) Τι αποτέλεσμα θα μας εμφανίσει αυτό το παράδειγμα ;

- Μας εμφανίζει το αποτέλεσμα της ένωση των στοιχείων που επιλέξαμε, δίνοντας μία νέα πληροφορία την οποία δεν γνωρίζαμε.

2) Που μπορούμε να βρούμε ένα CSV Αρχείο ;

- Μπορούμε να το δημιουργήσουμε ή ποιο εύκολος τρόπος είναι να το κατεβάσουμε από το αυθεντικό site “ KNIME”. Έχει ένα έτοιμο αρχείο το οποίο είναι ένα δείγμα για να μπορούμε να το χρησιμοποιούμε για να καταλάβουμε πως λειτουργεί το πρόγραμμα.

3) Αν σε κάποιον πίνακα είχαμε ελλιπείς στοιχεία τι θα έπρεπε να κάνουμε ;

- Εάν σε κάποιο πίνακα που θα χρησιμοποιήσουμε υπάρχουν κενά σημεία (ελλιπείς στοιχεία) τότε θα προσθέσουμε τον κόμβο “Missing Value”.

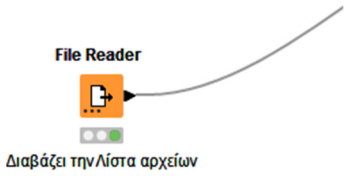
4) Μπορώ να έχω ότι αποτέλεσμα θέλω και να ενώσω πολλούς πίνακες μαζί ;

- Μπορούμε να ενώσουμε 2 πίνακες και να πάρουμε μια κοινή πληροφορία. Δεν μπορούμε να ενώσουμε σειρές η γραμμές αν δεν είναι η ίδια μεταβλητή.

8.2 παράδειγμα 2ο (Δημιουργία εξόρυξης δεδομένων με αρχείο pdf / word κ.τ.λ.- Data Mining Example with file reader)

8.2.1 File Reader

Αρχικά θα ξεκινήσουμε με το File Reader για να προσθέσουμε τα αρχεία που θέλουμε να αντλήσουμε πληροφορίες.[70]



ΕΙΚΟΝΑ 70. File Reader

8.2.2 Missing Value

Προσθέτουμε το Missing Value μόνο εάν έχουμε μεταβλητές που λείπουν στο αρχείο που προσθέσαμε.[71]

ΕΙΚΟΝΑ 71. Missing Value

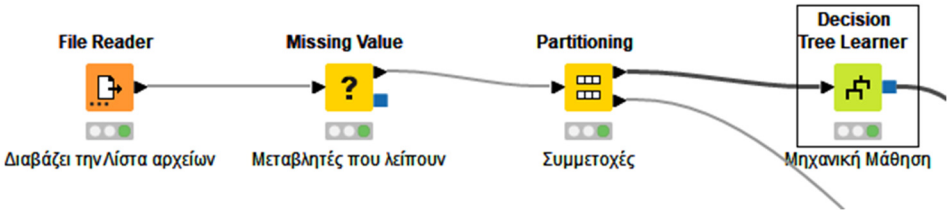
8.2.3 Partitioning

Τον κόμβο Partitioning τον προσθέτουμε για να επιλέξουμε από που θέλουμε να αντλήσουμε στοιχεία.[72]

ΕΙΚΟΝΑ 72. Partitioning

8.2.4 Decision Tree Learner

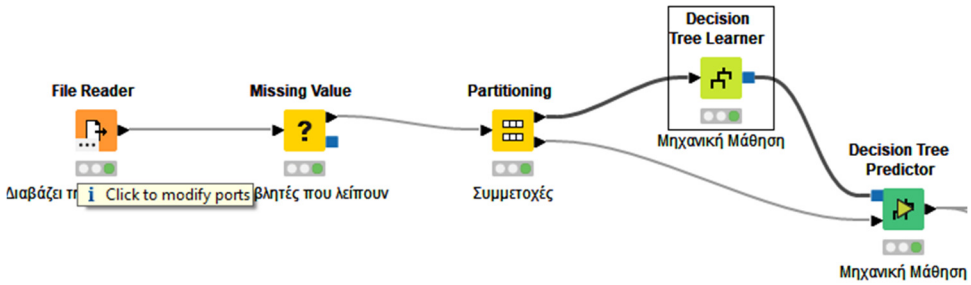
Αυτός ο κόμβος είναι ένα δέντρο απόφασης και χρησιμοποιείται για την ταξινόμηση στην κύρια μνήμη. [73]



ΕΙΚΟΝΑ 73. Decision Tree Learner

8.2.5 Decision Tree Predictor

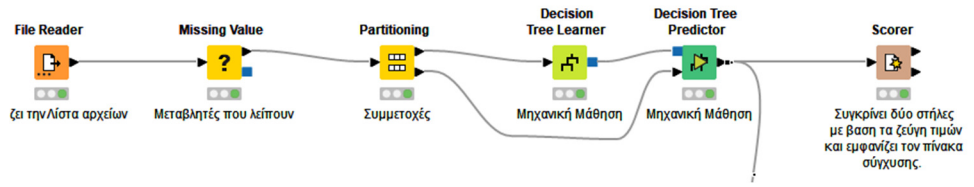
Αυτός ο κόμβος χρησιμοποιείται για την πρόβλεψη αποφάσεων αντλώντας πληροφορίες από την κύρια μνήμη.[74]



ΕΙΚΟΝΑ 74. Decision Tree Predictor

8.2.6 Scorer

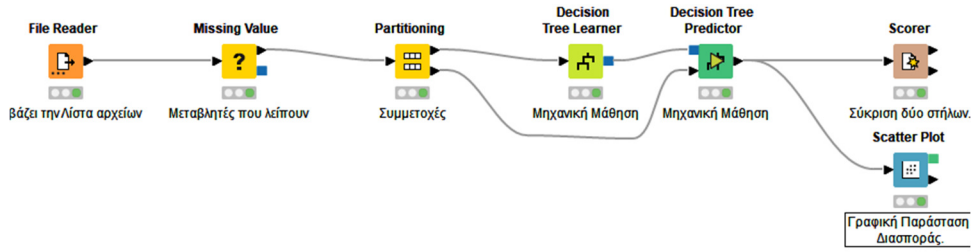
Αυτή η μεταβλητή συγκρίνει δύο στήλες με βάση τα ζεύγη τιμών και εμφανίζει τον πίνακα σύγχυσης.[75]



ΕΙΚΟΝΑ 75. Scorer

8.2.7 Scatter Plot

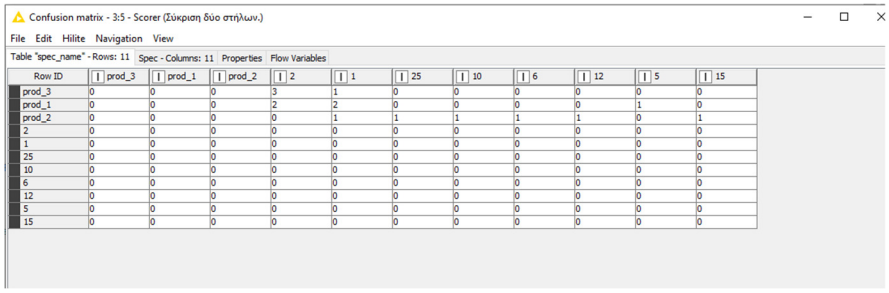
Αυτός ο κόμβος είναι μια γραφική παράσταση διασποράς, όπου χρησιμοποιεί μια βιβλιοθήκη γραφημάτων που βασίζεται σε JavaScript. Μας εμφανίζει ένα δια δραστικό γράφημα.[76]



ΕΙΚΟΝΑ 76. Scatter Plot

8.2.8 (Scorer) Confusion Matrix

Το Scorer μας εμφανίζει 2 ειδών αποτελέσματα, ένα από αυτά είναι το Confusion Matrix, το οποίο συγχέει τα στοιχεία των δύο πινάκων.[77]

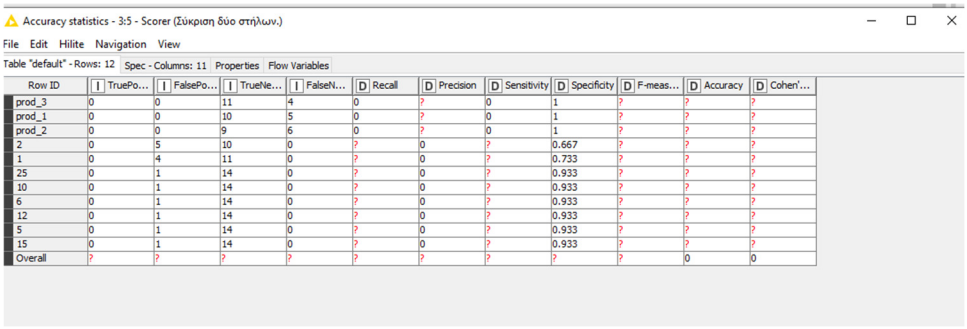


Row ID	prod_3	prod_1	prod_2	2	1	25	10	6	12	5	15
prod_3	0	0	0	3	1	0	0	0	0	0	0
prod_1	0	0	0	2	2	0	0	0	0	1	0
prod_2	0	0	0	0	1	1	1	1	1	0	1
2	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0

EIKONA 77. Confusion Matrix

8.2.9 (Scorer) Accuracy Statistics

Το δεύτερο scorer εμφανίζει το Accuracy Statistics δηλαδή τα ακριβή στατιστικά των δύο πινάκων.[78]

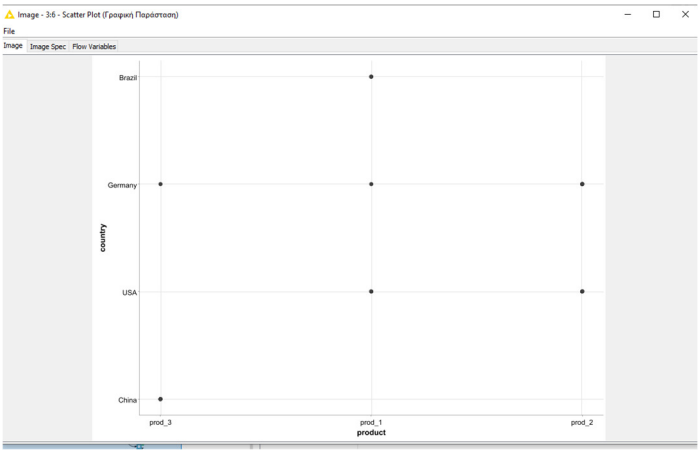


Row ID	TruePo...	FalsePo...	TrueNe...	FalseN...	Recall	Precision	Sensitivity	Specificity	F-meas...	Accuracy	Cohen'...
prod_3	0	0	11	4	0	?	0	1	?	?	?
prod_1	0	0	10	5	0	?	0	1	?	?	?
prod_2	0	0	9	6	0	?	0	1	?	?	?
2	0	5	10	0	?	0	?	0.667	?	?	?
1	0	4	11	0	?	0	?	0.733	?	?	?
25	0	1	14	0	?	0	?	0.933	?	?	?
10	0	1	14	0	?	0	?	0.933	?	?	?
6	0	1	14	0	?	0	?	0.933	?	?	?
12	0	1	14	0	?	0	?	0.933	?	?	?
5	0	1	14	0	?	0	?	0.933	?	?	?
15	0	1	14	0	?	0	?	0.933	?	?	?
Overall	?	?	?	?	?	?	?	?	?	0	0

EIKONA 78. Accuracy Statistics

8.2.10 (Scatter Plot) Image

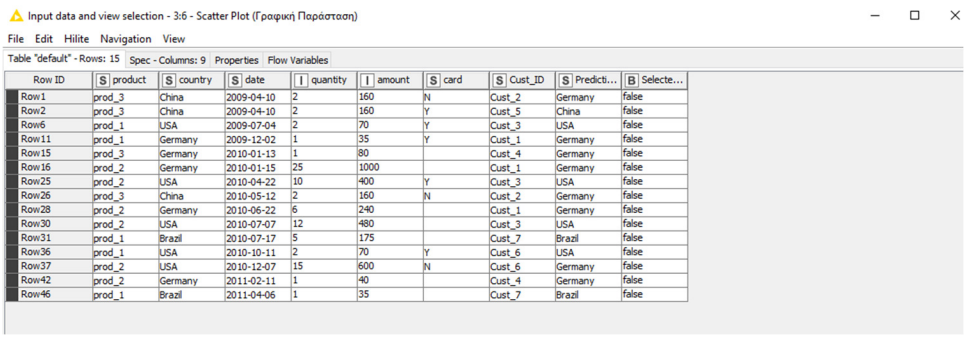
Στην παρακάτω εικόνα δείχνει την εικόνα από τα στοιχεία που προσθέσαμε.[79]



EIKONA 79. Image Scatter Plot

8.2.11 (Scatter Plot) Input Data View

Ο παρακάτω πίνακας δείχνει τα δεδομένα που έχουμε προσθέσει.[80]



Row ID	product	country	date	quantity	amount	card	Cust_ID	Predicti...	Selecte...
Row1	prod_3	China	2009-04-10	2	160	N	Cust_2	Germany	false
Row2	prod_3	China	2009-04-10	2	160	Y	Cust_5	China	false
Row6	prod_1	USA	2009-07-04	2	70	Y	Cust_3	USA	false
Row11	prod_1	Germany	2009-12-02	1	35	Y	Cust_1	Germany	false
Row15	prod_3	Germany	2010-01-13	1	80		Cust_4	Germany	false
Row16	prod_2	Germany	2010-01-15	25	1000		Cust_1	Germany	false
Row25	prod_2	USA	2010-04-22	10	400	Y	Cust_3	USA	false
Row26	prod_3	China	2010-05-12	2	160	N	Cust_2	Germany	false
Row28	prod_2	Germany	2010-06-22	6	240		Cust_1	Germany	false
Row30	prod_2	USA	2010-07-07	12	480		Cust_3	USA	false
Row31	prod_1	Brazil	2010-07-17	5	175		Cust_7	Brazil	false
Row36	prod_1	USA	2010-10-11	2	70	Y	Cust_6	USA	false
Row37	prod_2	USA	2010-12-07	15	600	N	Cust_6	Germany	false
Row42	prod_2	Germany	2011-02-11	1	40		Cust_4	Germany	false
Row46	prod_1	Brazil	2011-04-06	1	35		Cust_7	Brazil	false

EIKONA 80. Input Data View

Ερωτήματα :

1. Μόνο από συγκεκριμένα αρχεία μπορούμε να αντλήσουμε πληροφορίες ; αν όχι σε ποια ;

- Πληροφορίες μπορούμε να πάρουμε από πολλά διαφορετικά αρχεία, κάποια από αυτά είναι : Excel , Word , Εικόνες , Αρχεία συγκεκριμένου τύπου , Google URL , Twitter URL κ.τ.λ.

2. Είναι αναγκαίο να ακολουθήσουμε κάποια συγκεκριμένη ροή ;

- Η ροή που θα ακολουθήσουμε εξαρτάται από το είδος πληροφορίας που θέλουμε να έχουμε ως αποτέλεσμα. Υπάρχουν πολλοί κόμβοι και αρκετοί συνδυασμοί από τους οποίους μπορούμε να χρησιμοποιήσουμε για να πάρουμε το αποτέλεσμα που θέλουμε.

3. Αν λείπουν μεταβλητές, μπορούμε μέσα από το KNIME να τα συμπληρώσουμε ή να τα παραβλέψουμε;

- Εάν από κάποιο πίνακα λείπουν κάποια στοιχεία , υπάρχει η επιλογή να τα συμπληρώσει αυτόματα ή ακόμα και να τα παραβλέψει .

4. Πως να μάθεις να κάνεις την κατάλληλη ροή ανάλυσης δεδομένων;

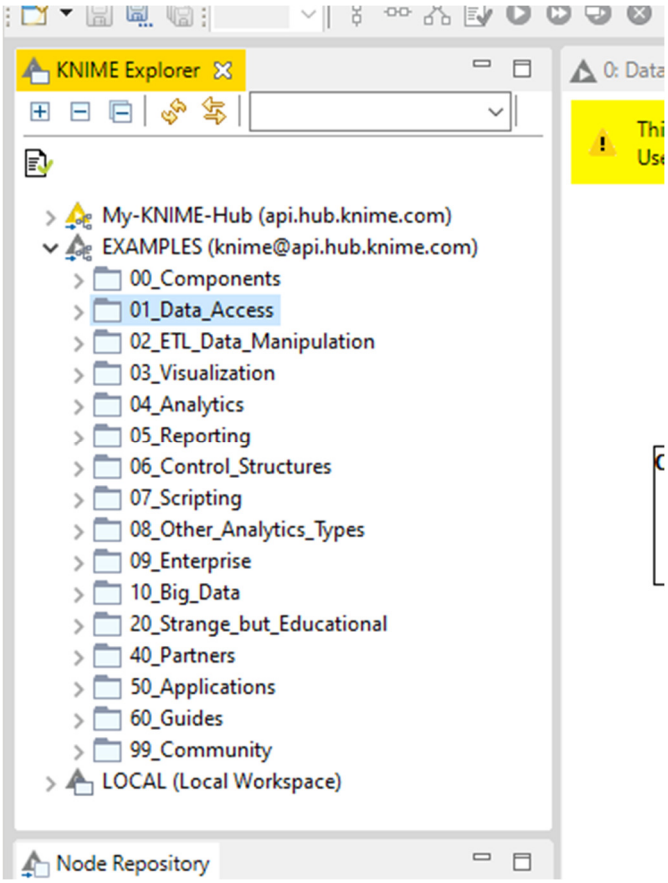
- Υπάρχουν έτοιμα παραδείγματα από τα οποία μπορείς να μάθεις και να καταλάβεις ποια είναι η σωστή ροή. Επιπλέον δίνει πολλούς και διάφορους κόμβους για να κατάληξης στο κατάλληλο αποτέλεσμα .

8.3 AUTOML

Για να δημιουργήσουμε ένα παράδειγμα αυτόματης Ανάλυσης δεδομένων, χρησιμοποιούμε διαφορετικούς κόμβους. Υπάρχουν έτοιμα παραδείγματα μέσα στην πλατφόρμα του KNIME.

8.3.1 KNIME AUTOML Example Guide

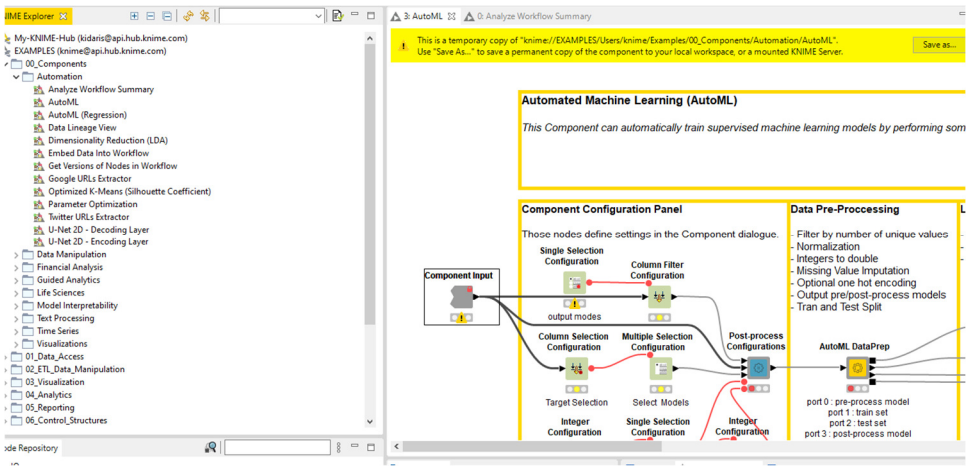
Πάνω αριστερά υπάρχει το παράθυρο KNIME Explorer, πάμε στην επιλογή Examples και εμφανίζει αρκετά και διαφορετικά παραδείγματα τα οποία μπορούμε να ακολουθήσουμε και να φτιάξουμε το δικό μας παράδειγμα.[81]



ΕΙΚΟΝΑ 81. AutoML

8.3.2 AUTOML Examples

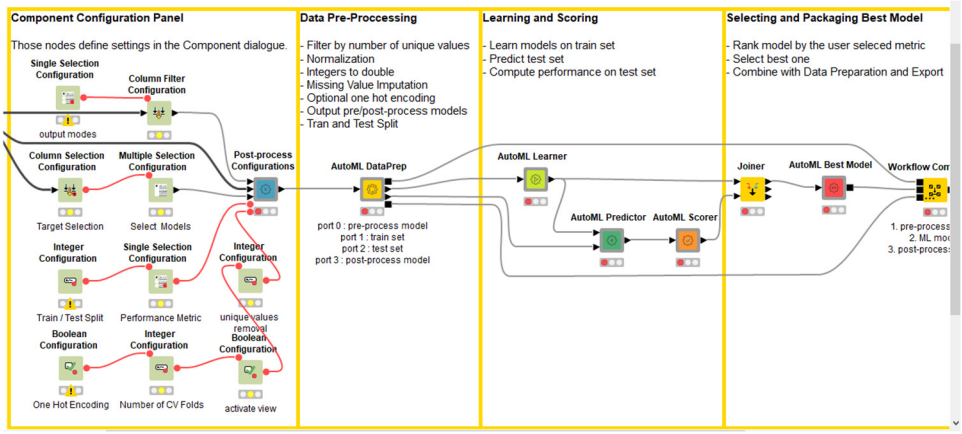
Για να βρούμε το παράδειγμα AUTOML : Examples – 00_Components – Automation – AutoML και πατάμε διπλό κλικ να το ανοίξουμε.[82]



EIKONA 82. AutoML Example

8.3.3 AutoML Workflows

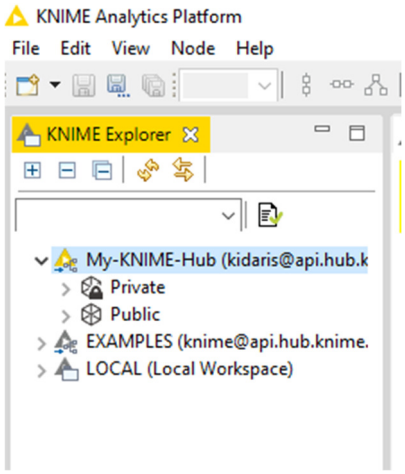
Στο παράδειγμα αυτό εμφανίζει όλες τις πιθανές επιλογές που έχουμε σε κάθε στάδιο της ροής δεδομένων. Υπάρχουν 4 στάδια Component Configuration Panel – Data Pre-Processing- Learning and Scoring- Selecting and Packaging Best Model .[83]



EIKONA 83. AutoML Workflow

8.4 My-KNIME-Hub

Σε αυτή την επιλογή μπορούμε να ανεβάσουμε ή να κατεβάσουμε μια ροή δεδομένων . Δημιουργούμε ένα λογαριασμό για να μπορούμε να μεταφέρουμε ροές στο δικό μας περιβάλλον KNIME. Υπάρχει και η επιλογή Private ή Public στο ενδεχόμενο που θελήσετε να κοινοποιήσετε στην πλατφόρμα του KNIME ή ακόμα και να κατεβάσετε μια ροή στο περιβάλλον KNIME.[84]



ΕΙΚΟΝΑ 84. Knime-Hub

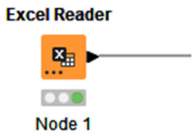
8.5 AutoML workflow , Αυτόματη ροή δεδομένων με κόμβο Excel

Σκοπός : Το AutoML " αυτόματη μηχανική μάθηση", αυτή η ροή δημιουργεί μια πλήρως αυτοματοποιημένη εφαρμογή βασισμένη στον ιστό για την επιλογή, την εκπαίδευση, τη δοκιμή και τη βελτιστοποίηση ενός αριθμού μοντέλων μηχανικής εκμάθησης. Η ροή εργασιών σχεδιάστηκε για επιχειρηματικούς αναλυτές ώστε να δημιουργούν εύκολες λύσεις πρόβλεψης αναλυτικών στοιχείων. Κάθε ένα στοιχείο θα δημιουργεί μια ιστοσελίδα με την οποία ο επιχειρηματικός αναλυτής θα μπορεί να αλληλεπιδράσεις.

Ξεκινώντας, όπως ανάφερα και στην προηγούμενη παράγραφο η κανονική ροή του AutoML δημιουργείται στον ιστότοπο του KNIME αλλά είναι για επιχειρηματικούς αναλυτές. Εμείς θα δημιουργήσουμε παρόμοια ροή αλλά στην πλατφόρμα του KNIME.

8.5.1 Excel Reader

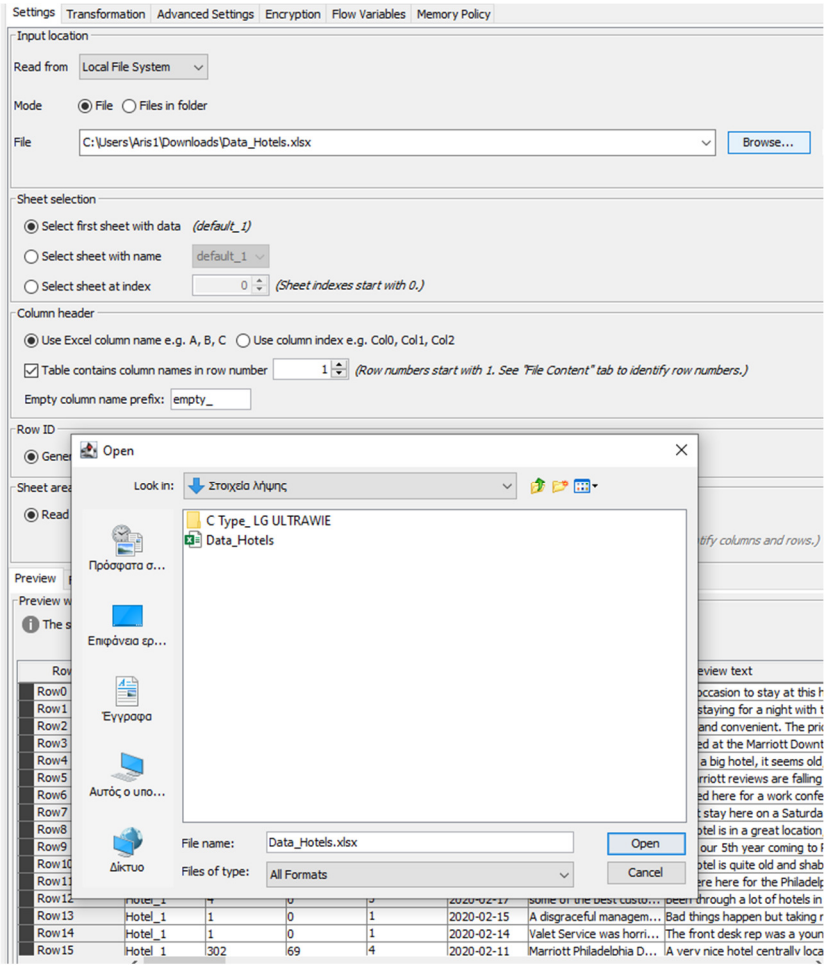
Αρχικά θα χρησιμοποιήσουμε το Excel Reader για να προσθέσουμε τις πληροφορίες που θέλουμε να επεξεργαστούμε στο πρόγραμμα KNIME.[85]



EIKONA 85. Excel Reader

8.5.2 Προσθήκη Αρχείων δεδομένων

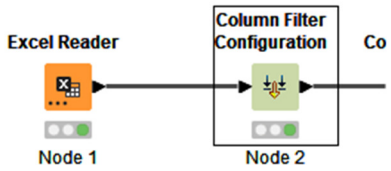
Κάνοντας διπλό κλικ στον κόμβο Excel Reader εμφανίζετε ο παρακάτω πίνακας επιλέγουμε το αρχείο και άνοιγμα, αφότου ανοίξει πατάμε οκ και F7 ή Εκτέλεση.[86]



ΕΙΚΟΝΑ 86. Προσθήκη Αρχείων

8.5.3 Column Filter Configuration

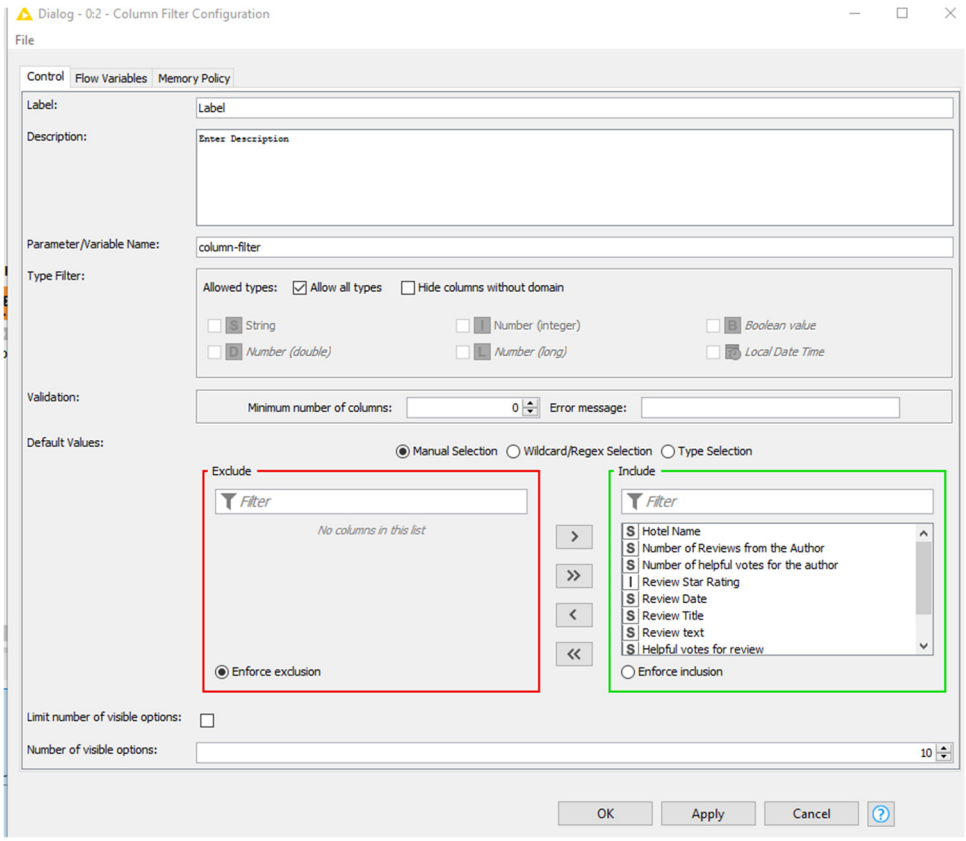
Στην συνέχεια θα προσθέσουμε τον κόμβο Column Filter Configuration , για να επιλέξουμε από ποια στοιχεία του κόμβου θέλουμε να αντλήσουμε πληροφορίες.[87]



ΕΙΚΟΝΑ 87. Column Filter

8.5.4 Column Filter settings

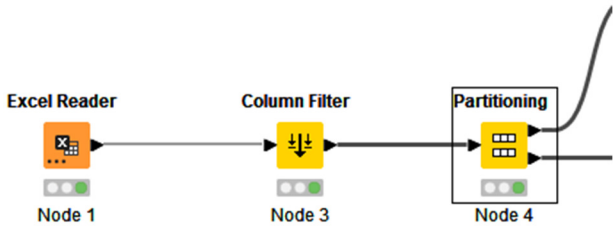
Αφού ανοίξουμε τον κόμβο, αφήνουμε στο Include τους πίνακες που θέλουμε να πάρουμε πληροφορίες και πατάμε OK.[88]



EIKONA 88. Filter Setting

8.5.5 Partitioning

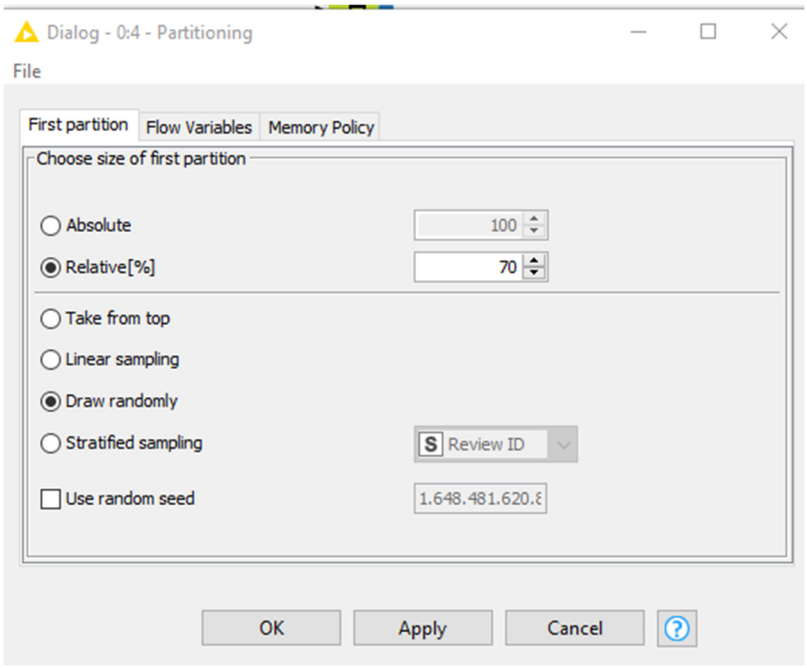
Στην συνέχεια χρησιμοποιούμε τον κόμβο Partitioning για να διασπάσουμε τον πίνακα σε δύο σημεία.[89]



EIKONA 89. Partitioning

8.5.6 Partitioning Settings

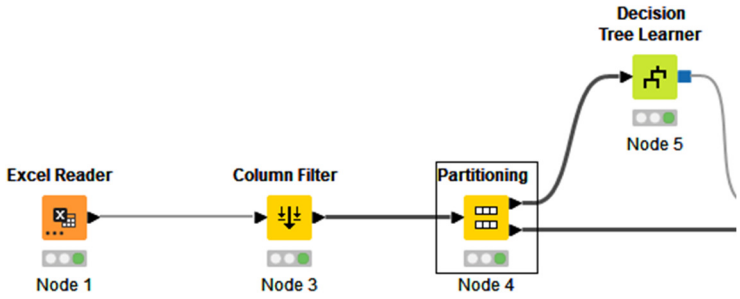
Σε αυτό το σημείο απλά πατάμε οκ και Εκτέλεση. [90]



EIKONA 90. Partitioning Settings

8.5.7 Decision Tree Learner

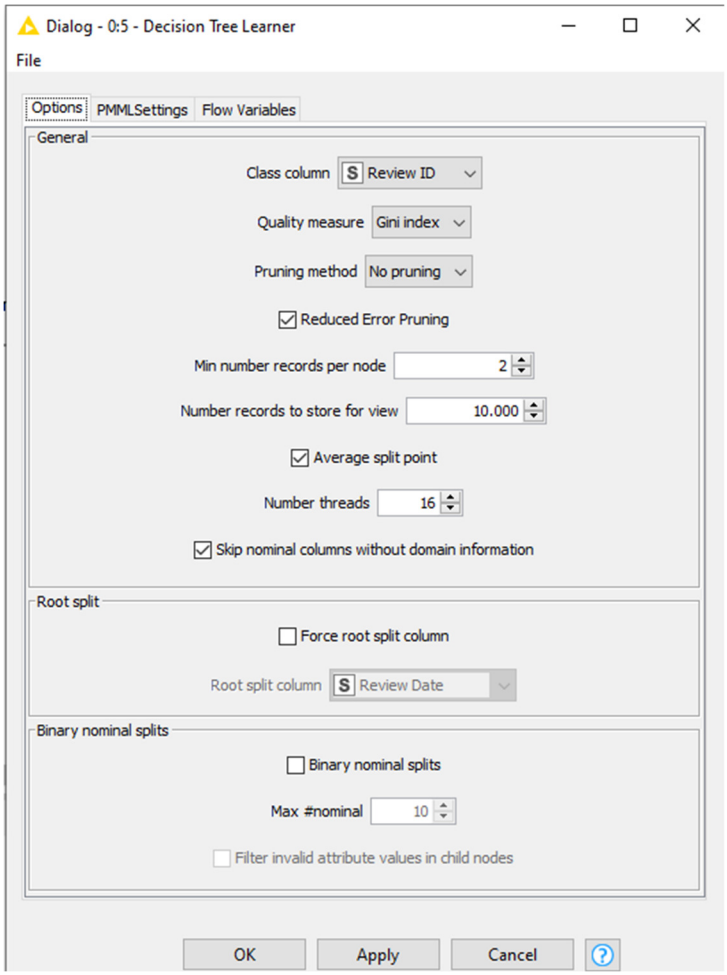
Έπειτα θα χρησιμοποιήσουμε το δέντρο αποφάσεων για την ταξινόμηση στην κύρια μνήμη. [91]



EIKONA 91. Decision Tree Learner

8.5.8 Decision Tree Learner Settings

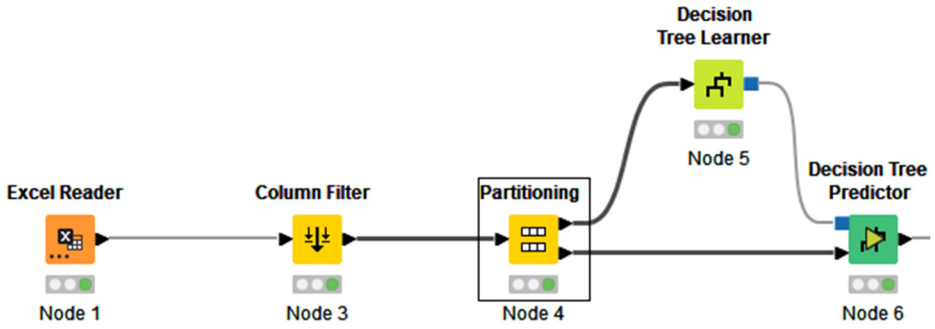
Κάνουμε διπλό κλικ πάνω στον κόμβο, πατάμε OK και Εκτέλεση.[92]



ΕΙΚΟΝΑ 92. Decision Tree Settings

8.5.9 Decision Tree Predictor

Αυτός ο κόμβος χρησιμοποιεί ένα υπάρχων δέντρο αποφάσεων για να προβλέψει την τιμή κλάσης για νέο μοτίβο.[93]



EIKONA 93. Decision Tree Predictor

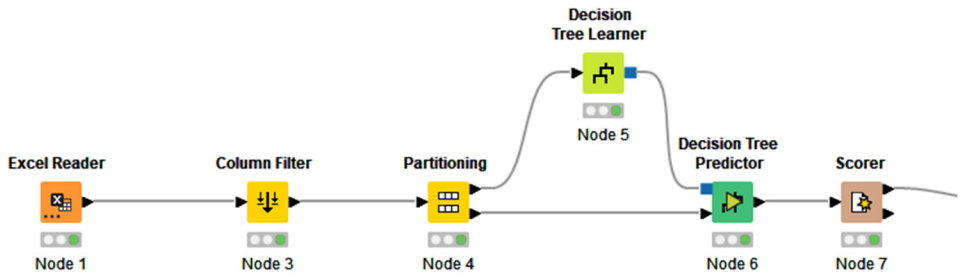
8.5.10 Decision Tree Predictor Settings

Πατάμε διπλό κλικ πάνω στον κόμβο και OK.[94]

EIKONA 94. Decision Tree Predictor Settings

8.5.11 Scorer

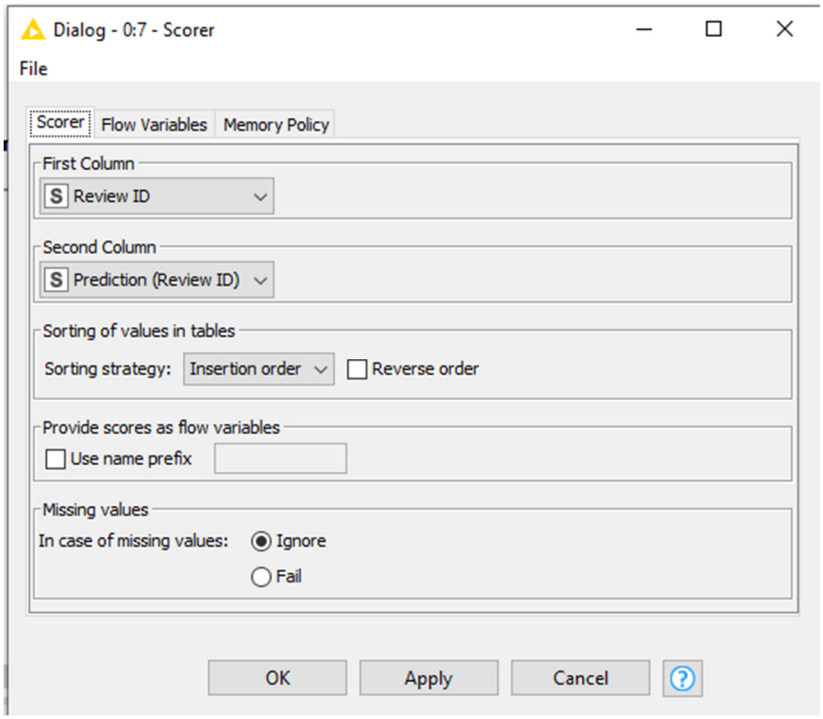
Σε αυτή την φάση προσθέτουμε τον κόμβο Scorer , για να συγκρίνουμε δύο στήλες που θέλουμε να πάρουμε πληροφορίες.[95]



EIKONA 95. Scorer

8.5.12 Scorer Settings

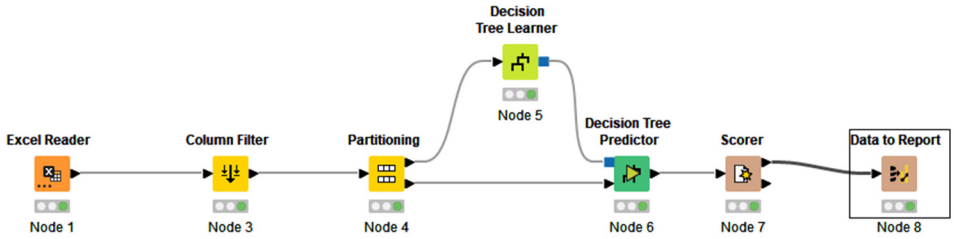
Γενικός πίνακας αλλαγών Scorer.[96]



EIKONA 96. Scorer Settings

8.5.13 Data to Report

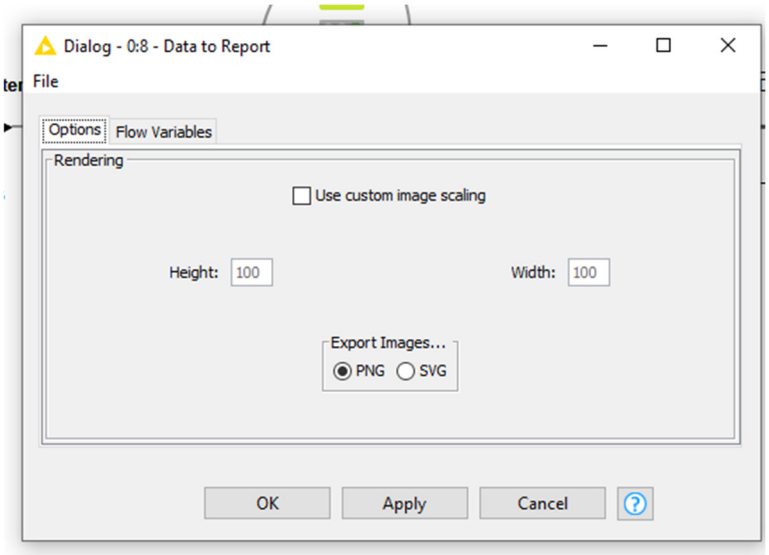
Χρησιμοποιούμε το Data to Report για να εισαχθούν τα εισερχόμενα δεδομένα στο πρόγραμμα και να μας βγάλει το αποτέλεσμα.[97]



EIKONA 97. Data Report

8.5.14 Data to Report Settings

Πατάμε διπλό κλικ πάνω στον κόμβο , αφού ανοίξουν οι ρυθμίσεις απλά πατάμε OK και εκτέλεση.[98]



ΕΙΚΟΝΑ 98. Data Report Settings

8.6 AUTOML Example II , Αυτόματη ροή δεδομένων με κόμβο CSV Reader

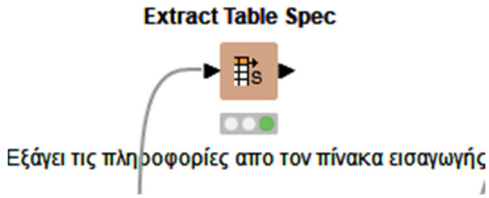
8.6.1 CSV Reader

Προσθέτουμε τον κόμβο CSV για να διαβάσει τον πίνακα που έχουμε προσθέσει.[99]

ΕΙΚΟΝΑ 99. Κόμβος CSV

8.6.2 Extract Table Spec

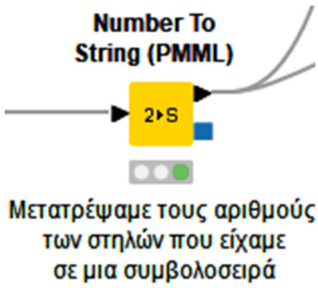
Εξάγει τις πληροφορίες από τον πίνακα εισαγωγής.[100]



ΕΙΚΟΝΑ 100. Extract Table

8.6.3 Number to String(PMML)

Μετατρέπει τους αριθμούς των στηλών που έχουμε σε μία συμβολοσειρά.[101]



ΕΙΚΟΝΑ 101. Number To String

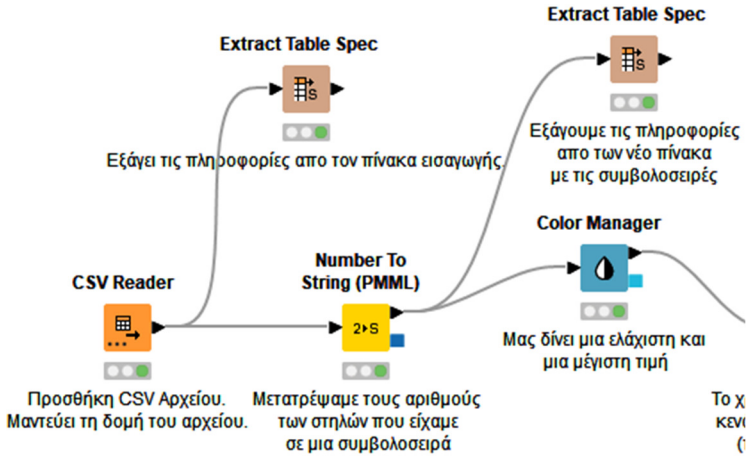
8.6.4 Color Manager

Μας δίνει μία ελάχιστη και μία μέγιστη τιμή.[102]

ΕΙΚΟΝΑ 102. Color Manager

8.6.5 Second Extract Table Spec

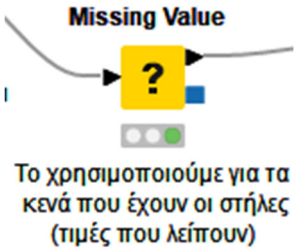
Εξάγει τις πληροφορίες από τον νέο πίνακα με τις συμβολοσειρές.[103]



ΕΙΚΟΝΑ 103. Second Extract Table

8.6.6 Missing Value

Το χρησιμοποιούμε για τα κενά που έχουν οι στήλες, δηλαδή τιμές που λείπουν.[104]



ΕΙΚΟΝΑ 104. Missing Value

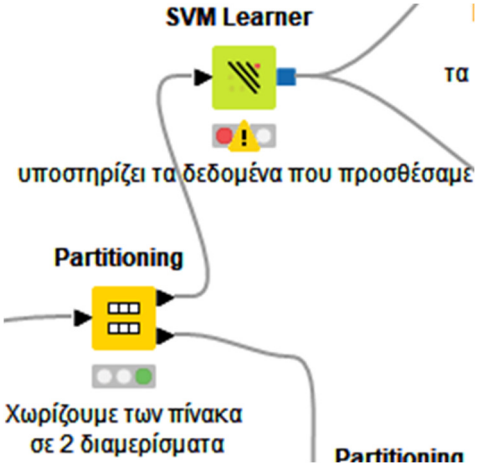
8.6.7 Partitioning

Χωρίζουμε των πίνακα σε 2 διαμερίσματα.[105]

ΕΙΚΟΝΑ 105. Partitioning

8.6.8 SVM Learner

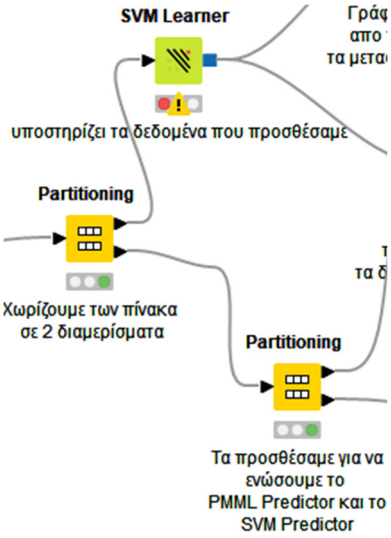
Υποστηρίζει τα δεδομένα που προσθέσαμε.[106]



ΕΙΚΟΝΑ 106. SVM Learner

8.6.9 Second Partitioning

Προσθέσαμε τον δεύτερο κόμβο Partitioning για να ενώσουμε το PMML Predictor και το SVM Predictor.[107]



ΕΙΚΟΝΑ 107. Second Partitioning

8.6.10 SVM Predictor

Προβλέπει την έξοδο για τα δεδομένα που προσθέσαμε.[108]

ΕΙΚΟΝΑ 108. SVM Predictor

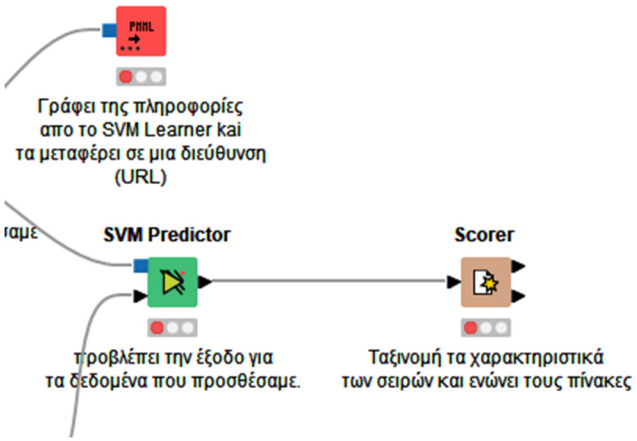
8.6.11 PMML Writer

Γράφει τις πληροφορίες από το SVM Leaner και τα μεταφέρει σε μία διεύθυνση URL.[109]

ΕΙΚΟΝΑ 109. PMML Writer

8.6.12 Scorer για ένωση των πινάκων και ταξινόμηση

Αυτός ο κόμβος ταξινομεί τα χαρακτηριστικά των σειρών και ενώνει τους πίνακες.[110]



ΕΙΚΟΝΑ 110. Scorer για ένωση πινάκων και ταξινόμηση

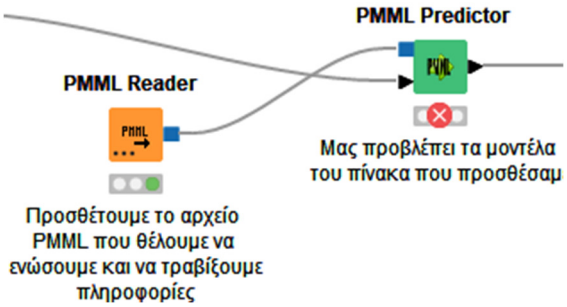
8.6.13 PMML Predictor

Μας προβλέπει τα μοντέλα του πίνακα που προσθέσαμε.[111]

ΕΙΚΟΝΑ 111. PMML Predictor

8.6.14 PMML Reader

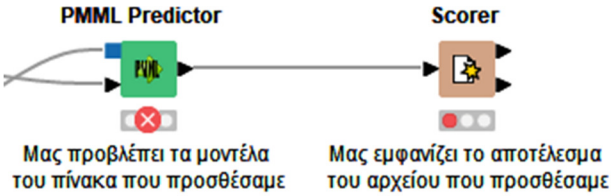
Προσθέτουμε το αρχείο PMML που θέλουμε να ενώσουμε και να τραβήξουμε πληροφορίες.[112]



ΕΙΚΟΝΑ 112. PMML Reader

8.6.15 Scorer

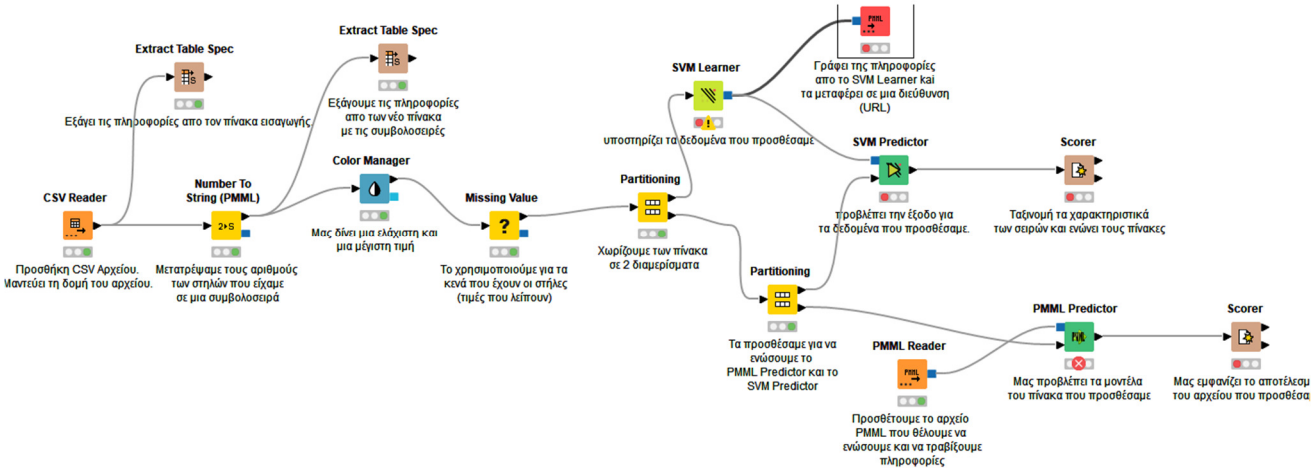
Μας εμφανίζει το αποτέλεσμα του αρχείου που προσθέσαμε.[113]



ΕΙΚΟΝΑ 113. Scorer για εμφάνιση του αποτελέσματος

8.6.16 Όλη η ροή δεδομένων

Όλες η ροή και οι ενώσεις από τους κόμβους που αναφέραμε αναλυτικά σε αυτή την ενότητα.[114]



ΕΙΚΟΝΑ 114. Ροή Ενότητας

ΚΕΦΑΛΑΙΟ **9**
Σύγκριση λογισμικών

ΔΙΑΓΡΑΜΜΑ ΠΙΝΑΚΩΝ

Πίνακας 1.

Γενικές πληροφορίες των λογισμικών WEKA-KNIME-RAPIDMINER.

	WEKA	KNIME	RapidMiner
Website	https://www.cs.waikato.ac.nz/ml/weka/	https://www.knime.com/	https://rapidminer.com/
Έτος Κυκλοφορίας	1992	2006	2006
Τελευταία έκδοση	3.8-3.9 / 2018	4.5.1 /20 Ιανουαρίου 2022	9.10.3 /16 Δεκεμβρίου 2021
Γλώσσα Προγραμματισμού	JAVA	JAVA	JAVA
Άδεια χρήσης	General Public Licence(GPL)	General Public Licence(GPL)	(AGPL)Small, Medium, Large . Οι εκδόσεις είναι ιδιόκτητες.
Λειτουργικά συστήματα	Cross platform	Cross platform	Cross platform
Διεπαφή χρήσης	➢ Μέτρια εμπειρία χρήσης ➢ Όχι αρκετά διαισθητικό περιβάλλον όσον αφορά τη δημιουργία της ροής εργασίας και την ένωση μεταξύ των κόμβων	➢ Εύκολη δη-επαφή (ένωση, αντικατάσταση κόμβων) ➢ Διαισθητικό περιβάλλον	➢ Μέτρια εμπειρία χρήσης ➢ Διαισθητικό περιβάλλον ➢ Μείωση πολυπλοκότητας

	➢ Καλή οργάνωση των στοιχείων του μενού	➢ Δημιουργία μετα-κόμβων για μείωση πολυπλοκότητας ➢ Ποιο σύγχρονη εμφάνιση	και αυτόματη διόρθωση λαθών ➢ Σχετικά καλή εμφάνιση των στοιχείων του μενού.
Τύπος	Περιέχει μία συλλογή από εργαλεία οπτικοποίησης και αλγορίθμους για την ανάλυση δεδομένων-προγνωστική μοντελοποίηση, Προ επεξεργασία δεδομένων.	Καθοδηγούμενη ανάλυση, Αναφορές επιχειρήσεων, Επιχειρηματική ευφυία, Εξόρυξη δεδομένων, Βαθιά μάθηση, Ανάλυση δεδομένων, Εξόρυξη κειμένου, Μεγάλα δεδομένα.	Επιστήμη δεδομένων, μηχανική μάθηση(ML), προγνωστική ανάλυση, (NLP) Επεξεργασία φυσικής γλώσσας, Εξηγήσιμη τεχνητή νοημοσύνη (XAI).

Πίνακας 2.

Πλεονεκτήματα των WELA-KNIME-RAPIDMINER

WEKA	KNIME	RAPIDMINER
ΠΛΕΟΝΕΚΤΗΜΑΤΑ	**ΠΛΕΟΝΕΚΤΗΜΑΤΑ**	**ΠΛΕΟΝΕΚΤΗΜΑΤΑ**
♦ Έχει χρησιμοποιηθεί από πολλούς επιστήμονες για την υλοποίηση	♦ Πιο εύχρηστο γραφιστικό περιβάλλον για τον μέσο χρήστη.	♦ Το RapidMiner είναι ένα Open-source software(Λογισμικό ανοικτού κώδικα)

των εργασιών τους.		που χρησιμοποιείται για την εξόρυξη δεδομένων και κειμένου για την επιστημονική και εμπορική χρήση.
♦ Περιέχει πολλές μεθόδους κατηγοριοποίησης , παλινδρόμησης , ανάλυση συστάδων και κανόνες συσχέτισης , όπου μπορείς να τα επεξεργαστείς όπως εσύ θες.	♦ Εύκολη σύνδεση κόμβων	♦ Είναι ένα λογισμικό αναγνωρισμένο σε παγκόσμια κλίμακα. Μεταξύ των χρηστών είναι οι εταιρίες Ford, Honda, Nokia και πολλές άλλες μεγάλες και μεσαίες.
♦ Μπορείς επιπλέων να τροποποιήσεις τον αλγόριθμο του αφού είναι λογισμικό ανοικτού κώδικα , αν ξέρεις δε πως να προγραμματίζεις και θέλεις να βελτιώσεις , τότε έχεις αυτή την επιλογή.	♦ Παρέχει περισσότερη πληροφορία μέσω έτοιμων παραδειγμάτων	♦ Είναι αρκετά εύκολο στην χρήση του, σε σύγκριση με άλλα παρόμοια προγράμματα.
♦ Η γλώσσα που έχει χρησιμοποιηθεί για την κατασκευή του είναι η Java, κάτι που το κάνει να εγκαθίσταται εύκολα σε	♦ Η τεκμηρίωση του κάθε κόμβου που εισάγεται στο Workflow, μας βοηθά να κατανοήσουμε ευκολότερα την εργασία που	♦ Επίσης έχει την δυνατότητα βαθμολόγησης του προγράμματος για να γνωρίζουν οι ερευνητές που υστερεί, τυχών κατασκευαστικά

πλατφόρμες υλικού και λογισμικού.	εκτελεί κάθε κόμβος.	λάθη και ελαττώματα.
♦ Υπάρχει μεγάλη ποικιλία βιβλιοθηκών για μηχανική μάθηση και εξόρυξης δεδομένων , απλά χρειάζεται να γραφτεί κώδικας. Διαθέτει όμως και το γραφικό του περιβάλλον στο οποίο δεν χρειάζονται γνώσεις προγραμματισμού .	♦ Προειδοποιήσεις, μηνύματα και χρώματα που εμφανίζονται απευθείας επάνω στον κόμβο που παρουσιάζει το πρόβλημα βοηθά στην κατανόηση και άμεση διόρθωση του προβλήματος.	♦ Βασίζεται στην γλώσσα προγραμματισμού Java
♦ Διαθέτει λογισμικό ανοικτού κώδικα(Freeware) , τον οποίο μπορείς να το επεξεργαστείς όπως θες εσύ.	♦ Η επιλογή των χαρακτηριστικών που θα λάβουν μέρος στην ανάλυση σε κάθε βήμα είναι πιο εύκολη καθώς γίνεται με βάση την ονομασία της στήλης.	♦ Περιλαμβάνεται μια εσωτερική xml αναπαράσταση ώστε να εξασφαλίζεται η τυποποιημένη μορφή ανταλλαγής εξόρυξης δεδομένων σε διάφορα πειράματα.
♦ Το WEKA διαθέτη 2 εκδόσεις, την stable η οποία απευθύνεται σε απλούς χρήστες και την Development	♦ Διαισθητικό περιβάλλον που έχει ως αποτέλεσμα καλύτερη εμπειρία χρήσης.	♦ Εξασφαλίζεται η αποτελεσματική διαχείριση των δεδομένων αφού υπάρχει δυνατότητα προβολής αυτών

στην οποία προορίζεται για τους προγραμματιστές , οι οποίοι θέλουν να την διορθώσουν και να την εξελίξουν.		σε πολλά επίπεδα.
♦ Έχει διάφορες επιλογές λειτουργικών συστημάτων , Windows, Mac OS X και Linux.	♦ Υποστηρίζει την ενσωμάτωση εκατοντάδων εργαλείων (Επεκτάσεις) extensions: Plugins, Modules.	♦ GUI, γραμμή εντολών Mode (λειτουργία batch) και Java API για την χρήση του από άλλα προγράμματα.
♦ Επιπλέον σου δίνει την επιλογή να καταλάβεις καλύτερα τον κώδικα, εφόσον σου προσφέρει οδηγίες και απαντήσεις για τυχών προβλήματα που αντιμετωπίζεις	♦ Δίνει τη δυνατότητα στον χρήστη να περιορίσει τις εγγραφές που θα χρησιμοποιηθούν για την εκπαίδευση του αλγορίθμου, ώστε η διαδικασία να εκτελεστεί πιο γρήγορα.	♦ Περιλαμβάνει μεγάλη ποικιλία από Plugins και Extensions
	♦ Έχει αρκετά μεγάλη χωρητικότητα σε σχέση με άλλα παρόμοια προγράμματα.	♦ Μια μεγάλη σειρά αναπαράστασης των δεδομένων με λεπτομερή διάσταση.
	♦ Έχει διάφορες επιλογές λειτουργικών συστημάτων , Windows, Mac OS X και Linux.	♦ Πλήρως αυτόματη διαδικασία όπου εμείς επιθυμούμε.

	♦ Διαθέτει λογισμικό ανοικτού κώδικα(Freeware) , τον οποίο μπορείς να το επεξεργαστείς όπως θες εσύ.	♦ Ολοκληρωμένο σεμινάριο παραδειγμάτων.
	♦ Βασίζεται στην γλώσσα προγραμματισμού Java	
	♦ Σε περίπτωση που δεν σας καλύπτουν όλες οι λειτουργίες, μπορείτε να προσθέσετε και επεκτάσεις ανοικτού κώδικα για επεξεργασία σύνθετων τύπων δεδομένων και την προσθήκη προηγμένων αλγορίθμων μηχανικής μάθησης.	

Πίνακας 3.

Μειονεκτήματα WEKA-KNIME-RAPIDMINER

WEKA	KNIME	RAPIDMINER
ΜΕΙΟΝΕΚΤΗΜΑΤΑ	**ΜΕΙΟΝΕΚΤΗΜΑΤΑ**	**ΜΕΙΟΝΕΚΤΗΜΑΤΑ**
♦ Δεν επιτρέπει την ενσωμάτωση	♦ Λιγότεροι αλγόριθμοι παλινδρόμησης	♦ Το RapidMiner δίνει αυτόματες λύσεις και

άλλων εργαλείων.		συμβουλές στον χρήστη όταν η διαδικασία που κάνει είναι λανθασμένη, αλλά για να το αποκτήσεις υπάρχει μια ετήσια συνδρομή.
♦ Δύσκολη επιλογή χαρακτηριστικών που θα λάβουν μέρος στην ανάλυση σε κάθε βήμα, καθώς γίνεται με βάση τον δείκτη όχι την ονομασία της στήλης (στον κόμβο Remove όταν γίνεται χειρωνακτικά)	♦ Περιορίζονται σε εργασίες ταξινόμησης, αφού επιτρέπουν μόνο ονομαστικές μεταβλητές ως μεταβλητές στόχους	♦ Μειωμένη ικανότητα καταμερισμού
♦ Εύχρηστο γραφιστικό περιβάλλον το οποίο χρήζει κάποιας βελτίωσης	♦ Η γραμμική παλινδρόμηση, το Random Forest και το δέντρο αποφάσεων προσφέρουν λιγότερες δυνατότητες παραμετροποίησης.	♦ Προ απαιτούμενη γνώση βάσεων δεδομένων
♦ Δεν είναι προφανές ποιοι κόμβοι πρέπει να χρησιμοποιηθούν και σε ποιο σημείο και με ποιους κόμβους	♦ Υστερεί στις μεθόδους επιλογής χαρακτηριστικών	♦ Δεν διαθέτει μεγάλο αποθηκευτικό χώρο, σε σύγκριση με άλλα παρόμοια προγράμματα.

μπορούν να συνδεθούν		
	♦ Σημαντικός περιορισμός η μετατροπή της μεταβλητής Creator, από ονομαστική σε αριθμητική με τιμές 0,1.	

Στους παραπάνω πίνακες αναφέρουμε αναλυτικά τα τεχνικά χαρακτηριστικά των λογισμικών πακέτων WEKA/KNIME/RAPIDMINER και στην συνέχεια προσθέσαμε τα πλεονεκτήματα/μειονεκτήματα αυτών των προγραμμάτων.

Επιγραμματικά μπορούμε να αναφέρουμε ότι για το κάθε ένα από τα λογισμικά προγράμματα που χρησιμοποιήσαμε έχουν κάποια συγκεκριμένα χαρακτηριστικά και μπορούμε να γράψουμε ότι :

- Όλα χρησιμοποιούν την ίδια γλώσσα προγραμματισμού JAVA.

- Έχουν το ίδιο λειτουργικό σύστημα Cross Platform.

- Μέσα στην πλατφόρμα έχουν αρκετά παραδείγματα για την κατανόηση και την γρήγορη εξοικείωση τους.

- Όλα έχουν την επιλογή ‘ λογισμικό ανοικτού κώδικα’ όπου αν ξέρουν και θέλουν να πειραματιστούν πάνω σε αυτό.

- Και στα 3 προγράμματα που ασχοληθήκαμε έχουν την επιλογή AutoML δηλαδή αυτόματη μηχανική μάθηση, είναι μια προσθήκη για την διευκόλυνση των χρηστών χωρίς τη χρήση κάποιου κώδικα.

Αναλύοντας το κάθε ένα ξεχωριστά και συγκρίνοντας τα προγράμματα καταλήξαμε στα εξής συμπεράσματα :

- Κάθε λογισμικό έχει διαφορετικό τρόπο χρήσης, δηλαδή διαφορετικό μενού μέσα στο πρόγραμμα (εικονίδια, ροή δεδομένων κ.τ.λ) .

- Κάποια από αυτά είναι δωρεάν και αν επιθυμείς περισσότερο χώρο ή ποιο εξιδεικευμένη έκδοση, τότε υπάρχει και η επιλογή συνδρομής. Υπάρχουν λογισμικά όπως το WEKA όπου είναι μόνο με συνδρομή αλλά προσφέρει αρκετές επιλογές.

- Στο KNIME δεν είναι απαραίτητη η γνώση προγραμματισμού εκτός και θέλεις να πειραματιστείς και να εξελίξεις το υπάρχων περιβάλλον. Στο RapidMiner είναι απαραίτητη η γνώση προγραμματισμού, ως αποτέλεσμα οι χρήστες που δεν ξέρουν προγραμματισμό θα δυσκολευτούν πολύ στο χειρισμό του ή ακόμα και να μην μπορούν να το χρησιμοποιήσουν. Από την άλλη το WEKA με την Developer έκδοση απευθύνεται κυρίως σε προγραμματιστές, παρόλα αυτά έχει και την επιλογή stable για τους απλούς χρήστες.

- Το KNIME και το RapidMiner δίνουν ενσωματωμένες λύσεις μέσα στο πρόγραμμα για την διευκόλυνση του χρήστη, ενώ το WEKA δεν σου δίνει αυτή την επιλογή.

- Στις απλές εκδόσεις, το WEKA δεν επιτρέπει την ενσωμάτωση άλλων εργαλείων , το KNIME επιτρέπει την προσθήκη άλλων εργαλείων , όπως επίσης και το RapidMiner έχει την επιλογή προσθήκης άλλων εργαλείων.

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στις μέρες μας η μηχανική μάθηση έχει εισβάλει μέσα στην καθημερινότητα ως βασικό στοιχείο ταχύτητας και επεξεργασία των πληροφοριών. Αρκετοί άνθρωποι τη χρησιμοποιούν καθημερινά για την διεκπεραίωση και την διευκόλυνση των εργασιών τους. Μέσα από την μηχανική μάθηση δημιουργήθηκε η εξόρυξη δεδομένων , για λόγους εξοικονόμησης χρόνου και ταχύτερης αναζήτησης πληροφοριών.

Στην παρούσα πτυχιακή εργασία στοχεύσαμε στην μελέτη και την σύγκριση προγραμμάτων μηχανικής μάθησης και εξόρυξης δεδομένων. Η υλοποίηση της ξεκίνησε με σκοπό την ανάλυση της εξόρυξης δεδομένων και την μηχανική μάθησης, όπου αναφέρουμε αναλυτικά τον λόγο που εφευρέθηκε και που μας εξυπηρετούν. Βασικό συμπέρασμα είναι ότι μέσα από αυτή την διαδικασία ανακαλύπτουμε νέες πληροφορίες που βοηθά στην κατανόηση και την μείωση χρόνου που θα χρειαζόταν να δαπανήσουμε για να τις βρούμε. Η δημιουργία ανταγωνιστικών προγραμμάτων βοήθησε στην εξέλιξη και την πρόοδο τους, προσθέτοντας περισσότερες επιλογές στην ανάλυση και την εγκυρότητα της πληροφορίας.

ΑΞΙΟΛΟΓΗΣΗ

Αρχικά θα αναφέρω κάποια πράγματα για το KNIME μιας και έχω μια σχετικά πλήρη εικόνα. Ξεκινώντας ως ένας απλός χρήστης, χωρίς να έχω κάποια εμπειρία εφόσον δεν είχα ασχοληθεί ποτέ με αυτό. Η εγκατάσταση του ήταν αρκετά εύκολη, επίσης στην αριστερή πλευρά έχει αρκετά διαφορετικά παραδείγματα τα οποία μας βοηθούν στην κατανόηση και την λειτουργία του προγράμματος με διάφορους συνδυασμούς. Ωστόσο τα παραδείγματα που υπάρχουν μέσα στην πλατφόρμα έχουν όλους τους πιθανούς συνδυασμούς ενωμένους και αυτό το καθιστά αρκετά δυσνόητο. Ακριβός από κάτω έχει όλες της κόμβους που μπορούμε να χρησιμοποιήσουμε για την δημιουργία ροών δεδομένων, και πατώντας πάνω στην κάθε μια υπάρχει η περιγραφή τους, το οποίο βοηθά πολύ στην κατανόηση και στην εξοικονόμηση χρόνου. Ένα πολύ βοηθητικό στοιχείο είναι ότι προσθέτοντας τον κάθε κόμβο, μας εμφανίζει στην γραμμή εργαλείων πιθανούς κόμβους που συνδυάζονται με τον προηγούμενο (δηλαδή το επόμενο βήμα), οπότε κερδίζεις χρόνο.

Έχοντας μία γενική εικόνα από το πρόγραμμα WEKA, μπορούμε να πούμε ότι είναι αρκετά καλό και σου δίνει πολλές επιλογές πάνω στην επεξεργασία και την δημιουργία νέων πληροφοριών. Εάν κατεβάσεις την δωρεάν έκδοση, τότε έχεις πολλούς περιορισμούς σε σχέση με το KNIME όπως λίγος χώρος αποθήκευσης ροής δεδομένων.

Έχοντας μελετήσει και συλλέξει πληροφορίες για το RapidMiner, θα μπορούσε να πει κανείς ότι είναι μία πλατφόρμα επιστήμης δεδομένων, όπου επεξεργάζεται πληροφορίες και δημιουργεί μια νέα. Η δωρεάν έκδοση απευθύνεται περισσότερο σε απλούς χρήστες που δεν έχουν μεγάλες απαιτήσεις μια και δεν έχει όλες τις δυνατότητες ενεργοποιημένες. Η πλατφόρμα είναι σχετικά εύκολη για τους απλούς χρήστες που δεν ψάχνουν εξιδεικευμένα πράγματα.

Οι περισσότεροι απλοί χρήστες χρησιμοποιούν το KNIME μιας και έχει πολύ περισσότερες επιλογές από το WEKA & RapidMiner στην απλή έκδοση, δεν έχει περιορισμούς και έχει δωρεάν επεκτάσεις σε περίπτωση που της χρειαστείς.

ΒΙΒΛΙΟΓΡΑΦΙΑ

https://el.wikipedia.org/wiki/Εξόρυξη_δεδομένων

https://el.wikipedia.org/wiki/Συστήμα_ενδοεπιχειρησιακού_σχεδιασμού

https://el.wikipedia.org/wiki/Εξόρυξη_δεδομένων

https://en.wikipedia.org/wiki/Text_mining

https://el.wikipedia.org/wiki/Weka_(μηχανική_μάθηση)

https://www.cs.waikato.ac.nz/ml/weka/book.html

https://en.wikipedia.org/wiki/KNIME

https://en.wikipedia.org/wiki/KNIME

https://waikato.github.io/weka-wiki/downloading_weka/

https://www.researchgate.net/figure/WEKA-GUI-Chooser-2-HISTORY-OF-WEKA-The-first-release-of-WEKA-was-brought-in-the-market_fig1_266593066

https://www.google.gr/search?q=rapidminer+release+date&sxsrf

https://www.knime.com/knime-open-source-story

https://rapidminer.com/why-rapidminer/

https://www.weka.io/company/about-us/

https://www.gartner.com/reviews/market/data-science-machine-learning-platforms/vendor/knime/product/knime-analytics-platform/likes-dislikes

https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/weka-gui-learn-machine-learning/

https://repository.kallipos.gr/handle/11419/1227

https://www.ebooks4greeks.gr/epixeirhmatikh-eyfyia-kai-eksoryksh-dedomenwn

https://people.iee.ihu.gr/~kdiamant/MachineLearning/MachineLearningLesson01.pdf

https://repository.kallipos.gr/handle/11419/3382

https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf

https://www.knime.com/sites/default/files/2021-03/KNIME%20Beginner%27s%20Luck%204.3_20210219_sample.pdf

https://docs.rapidminer.com/downloads/DataMiningForTheMasses.pdf

https://docs.rapidminer.com/latest/studio/operators/rapidminer-studio-operator-reference.pdf

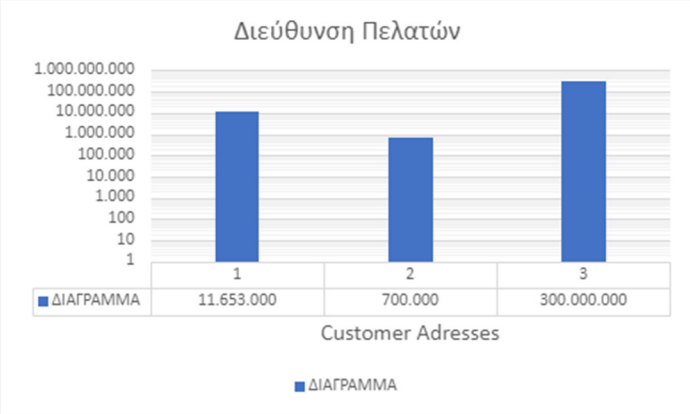
https://www.csc.com.gr/machine-learning-/

ΠΑΡΑΡΤΗΜΑ

ΔΙΑΓΡΑΜΜΑ ΠΙΝΑΚΩΝ

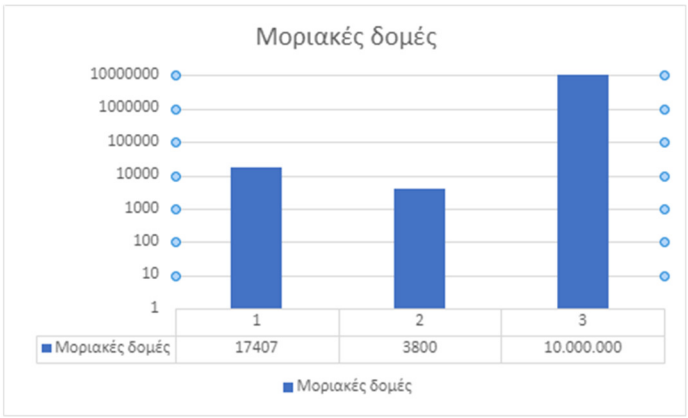
Διάγραμμα 1.

Πόσες Εγγραφές υπάρχουν στο κάθε πρόγραμμα WEKA-KNIME-RAPIDMINER



Διάγραμμα 2.

Μοριακές δομές των WEKA-KNIME-RAPIDMINER



Διάγραμμα 3.

Ενεργοί Χρήστες των Προγραμμάτων WEKA-KNIME-RAPIDMINER


Ενεργοί Χρήστες
1000000
100000
10000
1000
100
10
1
1
2
3
Ενεργοί Χρήστες
584,252
400.000
15.000
Ενεργοί Χρήστες

Πνευματικά δικαιώματα

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1988 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα εργασία αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον.

Ιωάννου Αριστείδης ΑΜ: 15102

Φαζάκης Νικόλαος ΑΜ: 15327

Έτος ολοκλήρωσης πτυχιακής εργασίας Έτος 2022- 2023