|  |
| --- |
| **Διδακτική Ενότητα 1:** Εισαγωγή |
| Η παρούσα ενότητα έχει ως στόχο να επισημάνει και να αναδείξει τις διαφορές που παρουσιάζονται στην προσέγγιση της ανάλυσης δεδομένων ανάμεσα σε λογισμικά τύπου spreadsheet και την εφαρμογή KNIME. Οι δύο αυτές πλατφόρμες διαφέρουν σημαντικά σε θεωρητικό και πρακτικό επίπεδο, καθώς ακολουθούν διαφορετικές προσεγγίσεις και παρέχουν διαφορετικές λειτουργίες και δυνατότητες ανάλυσης δεδομένων.Αρχικά, θα παρουσιασθεί με ακρίβεια και λεπτομέρεια ο τρόπος εγκατάστασης της εφαρμογής KNIMΕ και θα αναλυθεί το περιβάλλον εργασίας της πλατφόρμας.Θα παρουσιασθούν τα ακόλουθα:* **Οι έννοιες Node, Node creation/configuration/execution:** Επεξηγείται η λειτουργία και ο ρόλος των κόμβων (nodes) στο περιβάλλον του KNIME. Αναλύονται οι διαδικασίες δημιουργίας, ρύθμισης και εκτέλεσης των κόμβων, καθώς και ο τρόπος λειτουργίας τους μέσω της επικοινωνίας και της αλληλεπίδρασης μεταξύ τους.
* **Η έννοια του WorkFlow:** Εξηγείται η έννοια και η σημασία του WorkFlow ως της βασικής μονάδας οργάνωσης και εκτέλεσης εργασιών στο περιβάλλον του KNIME. Παρουσιάζονται οι διαδικασίες εισαγωγής και εξαγωγής των WorkFlows, καθώς και ο τρόπος διαχείρισης και οργάνωσής τους.
* **Εισαγωγή σχολίων στο WorkFlow (Annotations & Comments):** Αναλύεται ο τρόπος εισαγωγής σχολίων και σημειώσεων στο περιβάλλον του WorkFlow του KNIME. Εξηγούνται οι διαδικασίες δημιουργίας σχολίων και η χρησιμότητά τους για την τεκμηρίωση, την οργάνωση και την κατανόηση του WorkFlow

Στη συνέχεια, θα παρουσιασθεί η διαδικασία δημιουργίας μιας πρώτης ροής εργασίας καθώς και οι επιπλέον δυνατότητας μέσω τις εγκατάστασης επεκτάσεων . |
| **Ώρες**2 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 2:** Καθαρισμός δεδομένων  |
| Η παρούσα θεματική ενότητα αναλύει τον τρόπο αντιμετώπισης των πρωτογενών δεδομένων που σε πολλές περιπτώσεις δεν είναι σε επιθυμητή μορφή για επεξεργασία. Συχνά, αυτά τα δεδομένα περιέχουν ελλιπή στοιχεία που επηρεάζουν την ακρίβεια και την αξιοπιστία της επεξεργασίας τους, ή βρίσκονται σε μη πρόσφορη μορφή για ανάλυση. Στο πλαίσιο αυτής της ενότητας, παρουσιάζονται τα σχετικά "nodes" που προσφέρει το KNIME για τον καθαρισμό των δεδομένων.* **Παρουσίαση των κόμβων(node)s για ανάγνωση δεδομένων:** Εξετάζονται οι κόμβοι που παρέχονται από το KNIME για την ανάγνωση και αποθήκευση των δεδομένων από τοπικά αρχεία, βασεις δεδομένων και από το cloud
* Παρουσιάζονται διάφοροι κόμβοι που είναι διαθέσιμοι στο περιβάλλον του KNIME για τον καθαρισμό των δεδομένων, όπως τα "raw filter" , "column filter" και “missing value”. Αυτοί οι κόμβοι επιτρέπουν την απομάκρυνση ή την επιλογή συγκεκριμένων στοιχείων από τα δεδομένα, με σκοπό τη βελτίωση της ποιότητας και την εξαγωγή της απαραίτητης πληροφορίας.
* Επιπλέον, παρέχονται αναλυτικά παραδείγματα για τον καθαρισμό των δεδομένων και την πρωτογενή επεξεργασία τους. Αυτά τα παραδείγματα αναδεικνύουν τη χρήση των διαθέσιμων κόμβων και των τεχνικών καθαρισμού δεδομένων μέσα στο περιβάλλον του KNIME.
* Τέλος, παρέχονται ασκήσεις για την ενίσχυση των γνώσεων των εκπαιδευομένων στον τομέα του καθαρισμού δεδομένων και της πρωτογενούς επεξεργασίας, καθώς και για την εξάσκησή τους στη χρήση των κόμβων και των τεχνικών που παρουσιάζονται.
 |
| **Ώρες**2 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 3:** Πρωτογενής επεξεργασία |
| Η βελτίωση της ποιότητας των δεδομένων συνεχίζεται με την πρωτογενή τους επεξεργασία. Ο σκοπός είναι να εξαχθούν από τα πρωτογενή δεδομένα χρήσιμες ποσότητες που θα συμβάλλουν στην επίλυση του υπό διερεύνηση προβλήματος.* Παρουσιάζονται διάφοροι κόμβοι όπως ο “string manipulation”, “rule engine”, “math formula” , column expression”, “cell splitter” κ.α. Αυτοί οι κόμβοι επιτρέπουν την εξαγωγή πρόσθετης πληροφορίας από τα πρωτογενή δεδομένα.
* Επιπλέον, παρέχονται αναλυτικά παραδείγματα για τον καθαρισμό των δεδομένων και την πρωτογενή επεξεργασία τους. Αυτά τα παραδείγματα αναδεικνύουν τη χρήση των διαθέσιμων κόμβων και των τεχνικών καθαρισμού δεδομένων μέσα στο περιβάλλον του KNIME.
* Τέλος, παρέχονται ασκήσεις για την ενίσχυση των γνώσεων των εκπαιδευομένων στον τομέα του καθαρισμού δεδομένων και της πρωτογενούς επεξεργασίας, καθώς και για την εξάσκησή τους στη χρήση των κόμβων και των τεχνικών που παρουσιάζονται.
 |
| **Ώρες**2 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 4:** Επεξεργασία Δεδομένων |
| Στο πλαίσιο αυτής της ενότητας, παρουσιάζονται τεχνικές που αφορούν την επεξεργασία δεδομένων που προέρχονται από πολλαπλές πηγές, με στόχο τη συσχέτισή τους, την ομαδοποίησή τους και τη διευκόλυνση της οπτικοποίησής τους και της περαιτέρω ανάλυσής τους.Παρουσιάζονται οι έννοιες:* **Ομαδοποίησης δεδομένων (Grouping)**: αφορά την κατηγοριοποίηση των δεδομένων βάσει ενός κριτηρίου ή μιας μεταβλητής. Αυτή η τεχνική επιτρέπει τη δημιουργία ομάδων δεδομένων που έχουν κοινά χαρακτηριστικά, προκειμένου να πραγματοποιηθεί συγκεκριμένη ανάλυση ή οπτικοποίηση σε επίπεδο ομάδων.
* **Συνάθροιση δεδομένων (Aggregation):** αφορά τη συγκέντρωση δεδομένων με βάση μια ή περισσότερες μεταβλητές, συνήθως υπολογίζοντας μια συνοπτική μέτρηση, όπως ο μέσος όρος, ο αριθμητικός πίνακας ή η μέγιστη/ελάχιστη τιμή.
* **Συνένωση δεδομένων (Concatenation):** αφορά τον συνδυασμό δύο ή περισσότερων πηγών δεδομένων, διατηρώντας τη δομή και το περιεχόμενό τους, προκειμένου να δημιουργηθεί ένα μεγαλύτερο σύνολο δεδομένων.
* Δ**ιαχωρισμός δεδομένων (Splitting):** αφορά τον διαχωρισμό των δεδομένων σε μικρότερα τμήματα, και τη μαθηματική επεξεργασία δεδομένων, η οποία περιλαμβάνει την εκτέλεση μαθηματικών πράξεων και τον υπολογισμό αποτελεσμάτων βάσει αριθμητικών αλγορίθμων.
* **Μετασχηματισμός δεδομένων (Pivoting):** αφορά τη μετατροπή της δομής των δεδομένων από μια μορφή σε μια διαφορετική μορφή. Αυτή η τεχνική επιτρέπει την αλλαγή της κατάταξης των μεταβλητών ή των ετικετών, προκειμένου να δημιουργηθούν νέες προβολές ή να διευκολυνθεί η ανάλυση των δεδομένων.

Επιπλέον, παρέχονται:* **Αναλυτικά παραδείγματα** για την καλύτερη κατανόηση και εφαρμογή των παρουσιαζόμενων τεχνικών.
* **Ασκήσεις** που επιτρέπουν στον εκπαιδευόμενο να εξασκηθεί και να εφαρμόσει τις προαναφερθείσες τεχνικές με σκοπό την ενίσχυση της κατανόησης και των δεξιοτήτων του στον τομέα της ανάλυσης δεδομένων.
 |
| **Ώρες**4 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 5:** Οπτικοποίηση Δεδομένων |
| Η οπτικοποίηση δεδομένων αποτελεί ένα σημαντικό βήμα στη διαδικασία ανάλυσης δεδομένων. Στη συγκεκριμένη ενότητα, θα γίνει παρουσίαση των κόμβων (nodes) που χρησιμοποιούνται στο KNIME για την οπτικοποίηση των δεδομένων, ενώ θα δοθούν αναλυτικά παραδείγματα για την καλύτερη κατανόηση των διαδικασιών.Η ενότητα καλύπτει διάφορες έννοιες σχετικές με την οπτικοποίηση των δεδομένων, περιλαμβάνοντας:* **Διάγραμμα διασποράς (Scatter plot):** Μια γραφική αναπαράσταση που απεικονίζει τις σχέσεις μεταξύ δύο μεταβλητών. Χρησιμοποιείται για να εξετάσουμε την κατανομή και τη συσχέτιση των δεδομένων.
* **Διάγραμμα γραμμών (Line plot):** Ένα γραφικό που απεικονίζει την εξέλιξη μιας μεταβλητής με την πάροδο του χρόνου ή κάποιας άλλης παραμέτρου. Χρησιμοποιείται για να εντοπίσουμε τάσεις και μοτίβα στα δεδομένα.
* **Ραβδόγραμμα (Bar chart):** Ένα γραφικό που απεικονίζει τις τιμές μιας μεταβλητής με χρήση ράβδων. Χρησιμοποιείται για τη σύγκριση κατηγοριών και την ανάδειξη διαφορών μεταξύ τους.
* **Ιστόγραμμα (Histogram):** Ένα γραφικό που απεικονίζει την κατανομή μιας μεταβλητής. Χρησιμοποιείται για να εντοπίσουμε τις συχνότητες και τα μοτίβα στα δεδομένα.

Επιπλέον, η ενότητα περιλαμβάνει:* **Αναλυτικά παραδείγματα** για την εφαρμογή των παραπάνω τεχνικών, προσφέροντας μια πλήρη εικόνα της διαδικασίας οπτικοποίησης δεδομένων.
* **Ασκήσεις** που συμβάλουν στην ενίσχυση των γνώσεων και των δεξιοτήτων του εκπαιδευόμενου στον τομέα της οπτικοποίησης δεδομένων.
 |
| **Ώρες**2 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 6:** Παρουσίαση Αποτελεσμάτων (Εκθέσεις ή Αναφορές) |
| Αυτή η ενότητα θα παρουσιάσει νέο τρόπο δημιουργίας αναφορών στο πλαίσιο της ανάλυσης δεδομένων με τη χρήση του KNIME, που παρέχεται από το έκδοση 5.0 και μετέπειτα. Θα αναλυθεί η δημιουργία components με απότερω στόχο την δημιουργία αναφορώνΗ ενότητα αυτή περιλαμβάνει: * **Αναλυτικά παραδείγματα** στη δημιουργία αναφορών μέσω της πλατφόρμας KNIME και αποτελούν πρακτικές εφαρμογές των παραπάνω τεχνικών σε πραγματικά δεδομένα.
* **Ασκήσεις**, με τις οποίες οι επιμορφούμενες/-οι θα αποκτήσουν την απαραίτητη εξοικείωση στη δημιουργία αναφορών.
 |
| **Ώρες**2 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καϊσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **2** | Εξ αποστάσεως, Σύγχρονη τηλεκπαίδευση | **Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 7:** Χρονικά δεδομένα |
| Στο πλαίσιο αυτής της ενότητας, θα διερευνηθούν οι τεχνικές και μέθοδοι επεξεργασίας δεδομένων ημερομηνίας και ώρας. Θα επικεντρωθούμε στην εξαγωγή επιπρόσθετων πληροφοριών από αυτά τα δεδομένα, όπως, για παράδειγμα, την αναγνώριση και τον προσδιορισμό του μήνα ή της ημέρας από μία συγκεκριμένη ημερομηνία.Σε συνέχεια, θα ασχοληθούμε με τις τεχνικές επεξεργασίας χρονοσειρών. Ένας βασικός άξονας θα είναι ο υπολογισμός του κινούμενου μέσου όρου, μια τεχνική που χρησιμοποιείται για την ομαλοποίηση δεδομένων χρονοσειρών, μειώνοντας την τυχαία μεταβλητότητα και αποκαλύπτοντας υποκείμενες τάσεις.Επιπλέον, παρέχονται:* **Αναλυτικά παραδείγματα** για την καλύτερη κατανόηση και εφαρμογή των παρουσιαζόμενων τεχνικών.
* **Ασκήσεις** που επιτρέπουν στον εκπαιδευόμενο να εξασκηθεί και να εφαρμόσει τις προαναφερθείσες τεχνικές με σκοπό την ενίσχυση της κατανόησης και των δεξιοτήτων του στον τομέα της ανάλυσης δεδομένων.
 |
| **Ώρες**2 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καϊσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 8:** Μεταβλητές ροής δεδομένων (flow variables) και Components |
| Στην συγκεκριμένη ενότητα, θα ασχοληθούμε εκτενώς με τους διάφορους τύπους των μεταβλητών ροής δεδομένων (flow variables) στο περιβάλλον του KNIME. Θα εξετάσουμε σε βάθος τη χρησιμότητα και τον σημαντικό ρόλο τους στην αυτοματοποίηση των διαδικασιών μιας ροής εργασίας. Η παρουσίαση θα περιλαμβάνει την ανάλυση των διάφορων κόμβων που χρησιμοποιούνται για τη δημιουργία των flow variables, καθώς και τον τρόπο με τον οποίο αυτές οι μεταβλητές μπορούν να τροποποιήσουν ή να προσαρμόσουν τις ρυθμίσεις ενός κόμβου, προσδίδοντας μεγαλύτερη ευελιξία και δυναμικότητα στη ροή εργασίας.Επιπλέον, η ενότητα θα προχωρήσει στην ανάλυση του τρόπου δημιουργίας και ρύθμισης των components στο KNIME. Θα εξετάσουμε τις λειτουργίες και τις δυνατότητες των components, επικεντρώνοντας στη σημασία τους για την οργάνωση και αποτελεσματικότερη διαχείριση των ροών εργασίας. Θα εξηγήσουμε πώς οι components μπορούν να διαμορφωθούν και να παραμετροποιηθούν για να ανταποκρίνονται σε συγκεκριμένες ανάγκες επεξεργασίας δεδομένων, καθιστώντας την όλη διαδικασία πιο ολοκληρωμένη και ευέλικτη. Επιπλέον, παρέχονται:* **Αναλυτικά παραδείγματα** για την καλύτερη κατανόηση και εφαρμογή των παρουσιαζόμενων τεχνικών.
* **Ασκήσεις** που επιτρέπουν στον εκπαιδευόμενο να εξασκηθεί και να εφαρμόσει τις προαναφερθείσες τεχνικές με σκοπό την ενίσχυση της κατανόησης και των δεξιοτήτων του στον τομέα της ανάλυσης δεδομένων.
 |
| **Ώρες**3 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καϊσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 9:** Έλεγχος ροής |
| Στην παρούσα ενότητα, θα γίνει μια λεπτομερής και συνολική παρουσίαση των διαφόρων τρόπων διαχείρισης και ελέγχου της ροής εργασίας (workflow) στο περιβάλλον του KNIME. Θα εξεταστούν ενδελεχώς οι βρόγχοι (loops), οι οποίοι είναι ζωτικής σημασίας για την εκτέλεση επαναλαμβανόμενων διαδικασιών εντός της ροής εργασίας. Οι βρόγχοι επιτρέπουν την αυτόματη επανάληψη συγκεκριμένων εργασιών, βασιζόμενοι σε συγκεκριμένα κριτήρια ή συνθήκες, ενισχύοντας την αυτοματοποίηση και την ευελιξία του workflow.Παράλληλα, θα εξεταστούν επίσης οι κόμβοι **"IF Switch"** και **"Case Switch"**, που παίζουν κρίσιμο ρόλο στον έλεγχο της ροής δεδομένων. Ο κόμβος "IF Switch" χρησιμοποιείται για την εφαρμογή συνθηκών "αν-τότε" στη ροή εργασίας, διακλαδώνοντας τη ροή βάσει της εκπλήρωσης ή μη ενός συγκεκριμένου κριτηρίου. Αντίθετα, ο κόμβος **"Case Switch"** επιτρέπει τη διακλάδωση της ροής σε πολλαπλές κατευθύνσεις, ανάλογα με την εκπλήρωση πολλαπλών κριτηρίων ή συνθηκών. Και οι δύο αυτοί κόμβοι βοηθούν στη δημιουργία πιο πολύπλοκων και ευέλικτων ροών εργασίας, επιτρέποντας την πιο αποτελεσματική και εξατομικευμένη επεξεργασία δεδομένων.Επιπλέον, παρέχονται:* **Αναλυτικά παραδείγματα** για την καλύτερη κατανόηση και εφαρμογή των παρουσιαζόμενων τεχνικών.
* **Ασκήσεις** που επιτρέπουν στον εκπαιδευόμενο να εξασκηθεί και να εφαρμόσει τις προαναφερθείσες τεχνικές με σκοπό την ενίσχυση της κατανόησης και των δεξιοτήτων του στον τομέα της ανάλυσης δεδομένων.
 |
| **Ώρες**3 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καϊσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **Διδακτική Ενότητα 10:** Ανάλυση Δεδομένων |
| Η ανάλυση των δεδομένων αποτελεί μια διαδικασία που εφαρμόζει τεχνικές υπολογισμού στατιστικών μετρήσεων και εξόρυξης δεδομένων με σκοπό την ανάδειξη πληροφοριών και μοτίβων που ενδέχεται να είναι κρυμμένα σε μεγάλα σύνολα δεδομένων. Κατά την ανάλυση αυτή, αναπτύσσονται διάφορες τεχνικές με τη χρήση του KNIME, που περιλαμβάνουν τα ακόλουθα:* **Υπολογισμός στατιστικών μέτρων:** Η τεχνική αυτή αποσκοπεί στον υπολογισμό και την αξιολόγηση διάφορων στατιστικών μεγεθών, όπως ο μέσος όρος, η τυπική απόκλιση και οι συσχετίσεις μεταξύ μεταβλητών. Αυτά τα μέτρα μπορούν να παρέχουν πληροφορίες για τη διακύμανση και τις συσχετίσεις μεταξύ των μεταβλητών που εξετάζονται.
* **Παλινδρόμηση (Regression):** Η ανάλυση αυτή επιδιώκει την κατανόηση της σχέσης μεταξύ μιας εξαρτημένης μεταβλητής και μιας ή περισσότερων ανεξάρτητων μεταβλητών. Η παλινδρόμηση μπορεί να χρησιμοποιηθεί για πρόβλεψη και ανάλυση των συνεισφορών των μεταβλητών στην εξαρτημένη μεταβλητή.
* **Δένδρα Αποφάσεων:** Τα δένδρα αποφάσεων αναπαριστούν μια δομή αποφάσεων που βασίζεται σε ερωτήσεις και συνθήκες που οδηγούν σε συγκεκριμένες αποφάσεις. Χρησιμοποιούνται για την αναγνώριση προτύπων και την πρόβλεψη αποτελεσμάτων βάσει των χαρακτηριστικών που παρουσιάζονται.
* **Νευρωνικά Δίκτυα:** Τα νευρωνικά δίκτυα αποτελούν μια τεχνική που εμπνέεται από τη λειτουργία του ανθρώπινου εγκεφάλου. Αυτά τα μοντέλα αποτελούνται από διάφορα επίπεδα νευρώνων, τα οποία συνεργάζονται για την επεξεργασία των δεδομένων. Τα νευρωνικά δίκτυα χρησιμοποιούνται σε πολλές εφαρμογές, όπως η αναγνώριση προτύπων, η επεξεργασία φωνής και η τεχνητή νοημοσύνη.

Επιπλέον, η ενότητα αυτή περιλαμβάνει: * **Αναλυτικά παραδείγματα**, τα οποία αποτελούν πρακτικές εφαρμογές των παραπάνω τεχνικών σε πραγματικά δεδομένα. Αυτά τα παραδείγματα συνοδεύονται από λεπτομερείς περιγραφές και ερμηνείες για να βοηθήσουν στην κατανόηση και την εφαρμογή των τεχνικών.
* **Ασκήσεις**, οι οποίες προσφέρουν την ευκαιρία για ενδυνάμωση των γνώσεων και των δεξιοτήτων στην ανάλυση δεδομένων.
 |
| **Ώρες**5 | **Μέθοδος υλοποίησης** Εξ αποστάσεως, Ασύγχρονη τηλεκπαίδευση | **Διδάσκουσα/Διδάσκων****Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |
| **2** | Εξ αποστάσεως, Σύγχρονη τηλεκπαίδευση | **Καΐσας Ιωάννης****Καραφασούλης Κωνσταντίνος** |