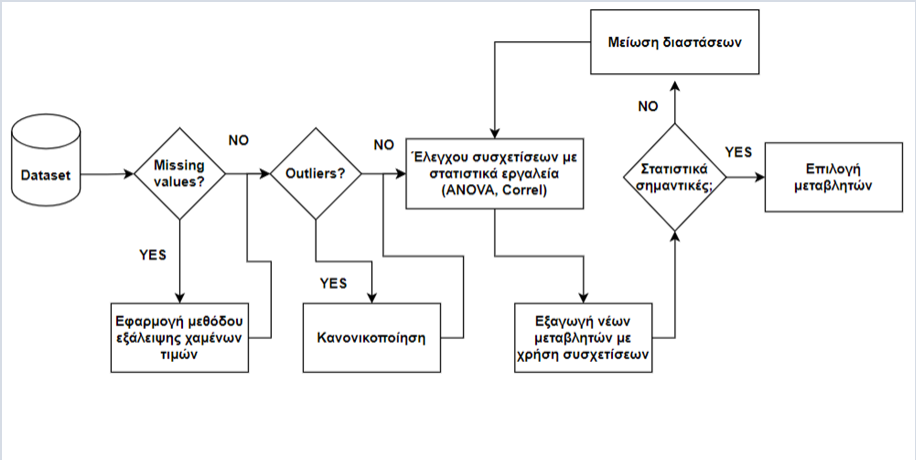




Περίληψη: Στόχος της παρούσας εργασίας είναι η παράθεση του θεωρητικού υποβάθρου για την υλοποίηση ευφών υπολογιστικών συστημάτων, που υπάγονται στο πλαίσιο της Τεχνητής Νοημοσύνης. Το συγκεκριμένα συστήματα, βασίζονται στην δια λειτουργικότητα επιμέρους συστημάτων Μηχανικής Μάθησης. Με βάση αυτή την παραδοχή, η εργασία εστιάζεται στην παρουσίαση των τεχνικών υλοποίησης, των κρίσιμων παραμέτρων και των πεδίων εφαρμογής τριών βασικών αλγορίθμων επιτηρούμενης Μηχανικής Μάθησης. Συγκεκριμένα, οι αλγόριθμοι που επιλέγονται είναι τα δέντρα αποφάσεων, ο ταξινομητής κατά Naïve Bayes και ο αλγόριθμος των κοντινότερων γειτόνων (k-NN). Για κάθε έναν από τους παραπάνω αλγορίθμους παρατίθεται και ένα εκτενές παράδειγμα, με στόχο την πλήρη αποσαφήνιση του τρόπου υλοποίησής τους. Επιπλέον, έμφαση δίνεται και στο στάδιο προ – επεξεργασίας των δεδομένων, το οποίο θεωρείται ιδιαίτερα σημαντικό για την αποτελεσματική υλοποίηση των αλγορίθμων. Τέλος, δίνεται το σύνολο των μετρικών ελέγχου της αποτελεσματικότητας των αλγορίθμων ταξινόμησης.

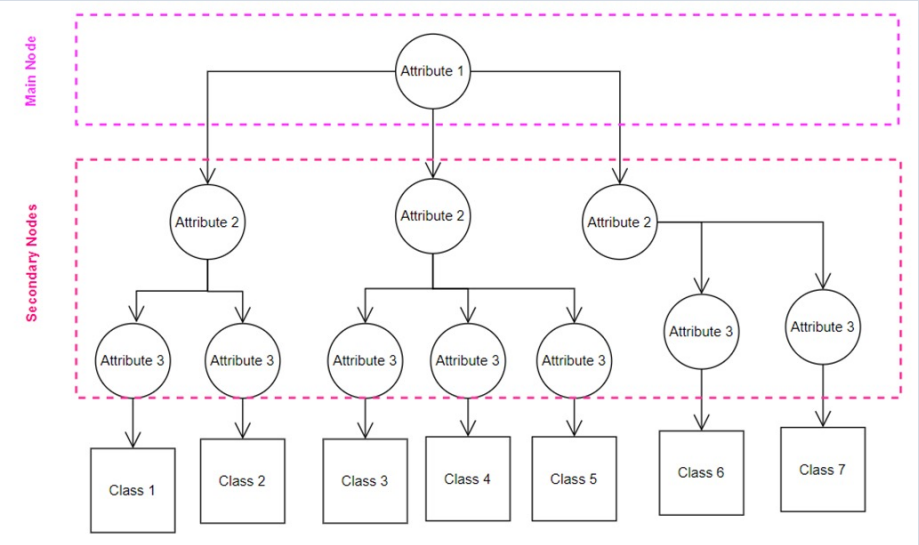
Data preprocessing



Evaluation metrics

Ονομασία μετρικής ελέγχου	Συνάρτηση υπολογισμού
Precision (P)	$Precision = \frac{TP}{(TP + FP)}$
Recall (R)	$Recall = \frac{TP}{(TP + FN)}$
F - Score	$F - Score = \frac{2 * Precision * Recall}{Precision + Recall}$
Ακρίβεια - Accuracy	$Accuracy = \frac{(TP + TN)}{(TP + FP + TN + FN)}$
Εξειδίκευση - Specificity	$Specificity = \frac{TN}{(FP + TN)}$
Περιοχή κάτω από την καμπύλη – Area under Curve - AUC	$AUC: FPR = 1 - Specificity = \frac{FP}{(FP + TN)}$
FPR: False Positive Ratio	

Γενική δομή δέντρων αποφάσεων



- Απλή μέθοδος
- Βασικό πρόβλημα η υπερ – εστίαση
- Μέθοδος κέρδους πληροφορίας - εντροπίας για κλάδεμα δέντρων

Βασική σχέση υπολογισμού Naïve Bayes

$$P(C_k|V_i) = \frac{P(C_k) * \prod(V_i, C_k)}{P(V_i)}$$

Κατάλληλη μέθοδος για επεξεργασία φυσικού κειμένου και γλώσσας

- Εναλλακτικές προσεγγίσεις:
- Multinomial Naïve Bayes classifiers
 - Bernoulli – Bayes
 - Gaussian Naïve Bayes

Υπολογισμός αποστάσεων πλησιέστερων γειτόνων: Ευκλείδεια απόσταση

$$d(i,j) = \sqrt{\frac{(x^i - y^i)^2 + (x^j - y^j)^2}{2}}$$