



2^η Προγραμματιστική Εργασία στο μάθημα «Τεχνητή Νοημοσύνη»

Η παρούσα εργασία εκπονήθηκε ατομικά.

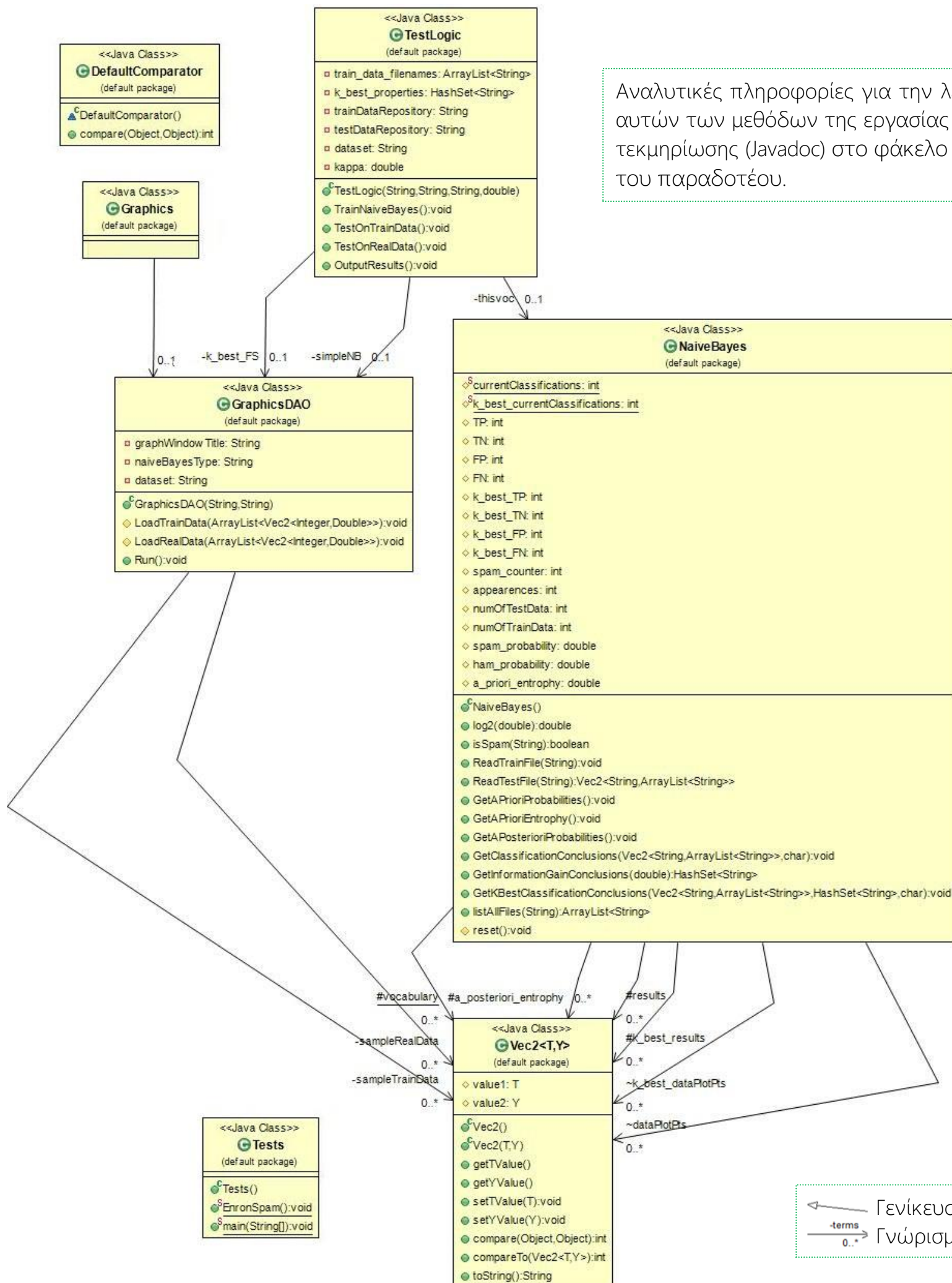
Ονοματεπώνυμο	Αριθμός Μητρώου	E-mail επικοινωνίας
Γεώργιος Μοσχόβης	3150113	p3150113@aueb.gr

Χρήσιμες πληροφορίες για την εργασία:

- ✓ Αναλυτικές πληροφορίες για τη λειτουργικότητα της εργασίας θα βρείτε στα αρχεία τεκμηρίωσης (Javadoc).
- ✓ Όνομα εκτελέσιμου αρχείου: **Tests.java**, στο φάκελο src/main/java (ή Tests.class, στο φάκελο target\classes).
- ✓ Λόγω εξαρτήσεων με τις βιβλιοθήκες Apache.math3 και org.jfree παραδίδεται IntelliJ Idea Maven Project.

Διάγραμμα κλάσεων σε UML

Για την υλοποίηση του αλγορίθμου αδαούς ταξινομητή Bayes (Naïve Bayes Classifier), χρησιμοποιήθηκαν οι ακόλουθες κλάσεις, καθώς και τα αρχεία text, των δεδομένων εκπαίδευσης και επικύρωσης, που προέρχονται από το σύνολο δεδομένων Enron-Spam με τα οποία διασυνδέονται:



Αποτέλεσμα εκτέλεσης

Κατόπιν επιτυχούς εκτέλεσης του αλγορίθμου, θα πρέπει να εμφανίζεται η ακόλουθη οθόνη. Για την λήψη του παρουσιασθέντος στιγμιοτύπου, έχει λάβει χώρα εκτέλεση με χρήση του IDE IntelliJ Idea Community, η ίδια διαδικασία μπορεί όμως να πραγματοποιηθεί μέσω οποιουδήποτε IDE υποστηρίζει Maven Projects με εξαρτήσεις (Maven Dependencies). Το IDE IntelliJ Idea επιτρέπει τα αρχεία να βρίσκονται σε οποιονδήποτε κατάλογο του project, ακριβή αντίγραφά τους έχουν περιληφθεί στους φακέλους `src/data_subdirectory` όπου ταυτόχρονα βρίσκονται τα αρχεία κώδικα και έχει οριστεί σχετικό μονοπάτι ανεύρεσής τους. Βεβαιωθείτε ότι έχετε τα αρχεία σε κατάλληλο φάκελο.

The screenshot shows the IntelliJ IDEA IDE with the `Tests.java` file open. The `main` method is highlighted, which calls `EnronSpam()`. The Run console at the bottom displays the output of the program, including accuracy and precision metrics for the Enron-Spam dataset. Two performance graphs are overlaid on the right side of the image.

Simple Naive Bayes Performance

Accuracy Value

Percentage of Data Used (%)

Train Data Accuracy (red dashed line), Test Data Accuracy (blue dashed line)

k-Best Feature Selection Naive Bayes Performance

Accuracy Value

Percentage of Data Used (%)

Train Data Accuracy (red dashed line), Test Data Accuracy (blue dashed line)

Run: Tests

"C:\Program Files\Java\jdk1.8.0_141\bin\java.exe" ...

Example: Enron-Spam

./src/enron_train_data

DIMENSIONALITY REDUCTION

INITIAL: 130255 REMAINING: 117230

Right predictions: 24619, 89% and wrong predictions: 3086

Precision: 0.9813958165728077

Recall: 0.7708530805687204

F1: 0.8634754910635286

*** USING FEATURE SELECTION ***

Right k-best predictions: 25165, 91% and wrong predictions: 2540

Precision: 0.9008238276299113

Recall: 0.89826224328594

F1: 0.8995412118335706

./src/enron_test_data

Right predictions: 4895, 81% and wrong predictions: 1115

Precision: 0.9950277859023107

Recall: 0.756

F1: 0.8591993938628615

*** USING FEATURE SELECTION ***

Right k-best predictions: 5333, 89% and wrong predictions: 667

Precision: 0.9327161887559269

Recall: 0.918

F1: 0.9252995856198902

All files are up-to-date (a minute ago)

20:5 CRLF UTF-8