

# Αλγόριθμοι εξόρυξης πληροφορίας

Εργασία 08: Ειρήνη Λυγκώνη

ΣΕΜΦΕ ΕΜΠ

## 1° ΣΤΑΔΙΟ: Προεπεξεργασία δεδομένων

Το training set αποτελείται από 39 numeric attributes των 2528 instances. Αφαιρούμε με την εντολή Remove τις:

- attribute38 (όλα τα instances είναι 0)
- attribute36 (όλα τα instances είναι 0 εκτός από 15)
- attribute34 (όλα τα instances είναι 0 εκτός από 1)
- Attribute selection (10 attributes)

ΕΠΑΛΗΘΕΥΣΗ: Σύγκριση αλγορίθμων για τα training sets με attribute34 και attribute343638 αντίστοιχα και αυτό με τα 10 attributes . Καλύτερο το δεύτερο.

2° ΣΤΑΔΙΟ: Αφαιρώ τα ίδια attributes και στο quiz set.

3° ΣΤΑΔΙΟ: Δοκιμάζω διάφορους αλγορίθμους επιλέγοντας πάντα εκείνους με το μικρότερο σφάλμα training set και το μεγαλύτερο ποσοστό επιτυχίας για το quiz set.

## ΣΥΝΔΥΑΣΜΟΙ ΑΛΓΟΡΙΘΜΩΝ

- **Bagging->RegressionByDiscretization->Reptree**
- **RandomSubspace->Bagging-> RegressionBy Discretization-> Reptree**
- **Bagging ->RandomSubspace->Reptree**
- **RegressionByDiscretization->J48**
- **Bagging->RepTree**
- **RandomSubspace-> RegressionByDiscretization ->J48**
- **RandomSubspace-> RegressionByDiscretization -> J48graft**

**SUCCESS RATE:????**