

Final Project

Επισυνάπτεται το dataset Walmart cleaned, από το [Kaggle](#). Το εν λόγω σύνολο δεδομένων περιέχει διαθέσιμα στοιχεία πωλήσεων για 45 καταστήματα της Walmart. Αυτά είναι τα ιστορικά δεδομένα που καλύπτουν τις πωλήσεις από το 2010-02-05 έως το 2012-11-01 για 45 καταστήματα.

Σκοπός της εργασίας είναι να μελετηθούν τα δεδομένα και να γίνει κατάλληλη επεξεργασία τους, και στη συνέχεια αυτά να χρησιμοποιηθούν για την εκπαίδευση αλγορίθμων Μηχανικής Μάθησης με στόχο την πρόβλεψη των εβδομαδιαίων πωλήσεων. Η μεταβλητή-στόχος στο dataset είναι το Next week.

Κάθε ένας με βάση $AM \bmod 45 + 1$, θα επιλέξει ένα από τα καταστήματα που είναι διαθέσιμα στο dataset.

Έπειτα θα πρέπει να προβλέψετε τις εβδομαδιαίες πωλήσεις για τρία departments (ελεύθερη επιλογή).

Θα πρέπει να συμπληρωθεί η next week, για κάθε ένα department που θα επιλέξετε, με τις πωλήσεις της επόμενης εβδομάδας του συγκεκριμένου department για όλο το dataset. (Οι πρώτες γραμμές συμπληρώθηκαν ενδεικτικά)

Τρόπος υπολογισμού

Next week για παράδειγμα της εβδομάδας 05-02-10:

- To Next week για τη γραμμή 0 υπολογίζεται ως το weekly_sales της επόμενης εβδομάδας (12-02-10) για το Department 1, δηλαδή 46039.49
- To Next week για τη γραμμή 1 υπολογίζεται ως το weekly_sales της επόμενης εβδομάδας (12-02-10) για το Department 26, δηλαδή 10050.92
- To Next week για τη γραμμή 1 υπολογίζεται ως το weekly_sales της επόμενης εβδομάδας (12-02-10) για το Department 17, δηλαδή 13403.66

Στο dataset υπάρχουν και οι στήλες MarkDown 1 με 5, οι οποίες έχουν να κάνουν με προωθητικές ενέργειες. Προαιρετικά τις χρησιμοποιείτε και σε αυτήν την περίπτωση συσχετίστε τα αποτελέσματα με και χωρίς αυτές.

Να περιγράψετε τη διαδικασία που ακολουθήσατε μέχρι και αυτό το σημείο μαζί με τα αποτελέσματα (data preprocessing, model training, inference), καθώς επίσης και να παρουσιάσετε συνοπτικά (π.χ. με απλή ή πιο προχωρημένη στατιστική δηλαδή MAPE και MAE) τα συμπεράσματά σας όσον αφορά τις εβδομαδιαίες πωλήσεις.

Συνοπτικά, τα απαιτούμενα για την εργασία είναι:

1. Ανάλυση/διαχείριση δεδομένων
2. Εκπαίδευση/σύγκριση μοντέλων (output Y: Next week) (ο χωρισμός του dataset να γίνει με μία χρονιά train, μια χρονιά test και την τελευταία validation)
3. Πρόβλεψη Y (Next week)
4. Παρουσίαση διαδικασίας, πειραμάτων, αποτελεσμάτων, συμπερασμάτων

Παραδοτέα:

- Κώδικας(/ες)
- Αρχείο αναφοράς (report)
- Τελικό dataset (csv) μετά από οποιαδήποτε επεξεργασία έχει υποστεί το αρχικό

Για οποιαδήποτε άλλη πληροφορία, μπορείτε να αναρτήσετε τις απορίες και τα σχόλιά σας στο forum του μαθήματος