**Ερώτημα 1.Γ)**

Εδώ μας ζητείται να προβλέψουμε την παραγωγή ενέργειας από μη ανανεώσιμες πήγες για κάθε στιγμή της ημέρας χρησιμοποιώντας ένα LSTM παλινδρόμησή.

**ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ**

Αρχικά θεωρούμε ως μη ανανεώσιμες πήγες τον άνθρακα και το φυσικό αέριο, επειδή όμως ο άνθρακας έχει μέση τιμή παραγωγής 14.24 Μεγαβάτ ενώ το φυσικό αέριο έχει μέση τιμή 8286.76 Μεγαβάτ τα προσθέτουμε και φτιάχνουμε μια νέα στήλη με «fossil» οπού περιέχει το άθροισμα των 2 πηγών.

Τα δεδομένα μας θα έχουν την μορφή: X- (m\_samples, x\_time\_steps, n\_features),

Y-(m\_samples, y\_time\_steps,1).

Το πλήθος των time\_steps στο Y καθορίζεται από το πόσες χρονικές στιγμές της ημέρας θέλουμε να προβλέψουμε και με βάση αυτό επιλέγουμε και το πλήθος των time\_steps στο Χ, το οποίο πρέπει να είναι ίσο ή και μεγαλύτερο. Επειδή εμείς θέλουμε να προβλέπουμε την τιμή για την επόμενη χρονική στιγμή θα έχουμε: Υ(~300k, 1).

**FEATURE SELECTION**

Αρχικά επειδή έχουμε 17 features θα φτιάξουμε τον πίνακα αυτοσυσχέτισης χρησιμοποιώντας την απόσταση **spearman** καθώς αυτή δουλεύει καλυτέρα με μη γραμμικές συσχετίσεις.

Τα χαρακτηριστικά που έχουν τιμή στον πίνακα μικρότερη του 0.1 θα τα αφαιρέσουμε χωρίς να χάσουμε πληροφορία.

A picture containing graphical user interface

Description automatically generated

Παρατηρούμε πως οι πήγες: geothermal, biomass, biogas και other έχουν σχεδόν μηδενική συσχέτιση πράγμα λογικό γιατί αυτές οι πήγες παράγουν και πολύ λιγότερη ενέργεια αλλά και είναι σταθερές σε μόνιμο επίπεδο. Βλέπουμε επίσης ότι την μεγαλύτερη συσχέτιση την έχει η ζήτηση με 0.44

**TIME DIMENSION**

Chart, line chart

Description automatically generatedChart, line chart

Description automatically generated

Σαν είσοδο στο δίκτυο μας θα έχουμε όλες τις πήγες καθώς και τη ζήτηση για κάθε χρονική στιγμή συν μια χρονική παράμετρο που θα αναφέρουμε παρακάτω και για έξοδο την στήλη fossil τις επόμενες χρονικές στιγμές.

Η παραγωγή από μη ανανεώσιμες πήγες ενέργειας εξαρτάται κυρίως από τη ζήτηση σε κάθε χρονική στιγμή. Η ζήτηση ακολουθεί 2 μοτίβα, ένα για τη διάρκεια της μέρας και ένα για τη διάρκεια του χρόνου. Δηλαδή παίζει σημαντικό ρολό αν προβλέπουμε μέρες του Ιανουάριου σε σχέση με μέρες του Μάιου και για ποιες στιγμές της μέρας προβλέπουμε και αυτό το μοτίβο δεν το έχουμε στα δεδομένα.

Για αυτό θα πάρουμε το index των δεδομένων μας και θα εφαρμόσουμε μετασχηματισμό sin και cos με περίοδο 1 ημέρα και 1 χρόνο.

Χρησιμοποιούμε sin και cos γιατί θέλουμε κάθε σημείο να έχει μοναδική τιμή και για κάθε περίοδο οι ημιτροπικές συναρτήσεις έχουν 2 ιδίες τιμές.

Αυτό δίνει μια μοναδική τιμή για κάθε χρονική στιγμή της ημέρας και κάθε ημέρας στο χρόνο.

**NORMALIZATION**

Επειδή οι τιμές των δεδομένων κυμαίνονται από -30 μέχρι +70k θα κανονικοποιήσουμε τα δεδομενα μας με το μετασχηματισμο: Z = (X – mean) / Std για κάθε στηλη κρατωντας τις 2 μεταβλητες για την target στηλη μας ώστε να αντιστρέψουμε τον μετασχηματισμό μετά τις προβλέψεις.

**MODEL SELECTION**

Το μοντέλο που θα χρησιμοποιήσουμε αποτελείται από 4 layers

LSTM -> Dropout -> Dense -> Output.

Για loss χρησιμοποιούμε το Mean Squarred Error ενώ για metric το RMSE και για optimizer το Adam.

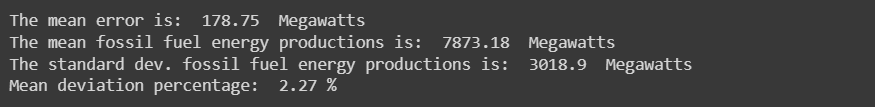
Για να εκπαιδεύσουμε το μοντέλο μας χρησιμοποιούμε EarlyStopping προκειμένου να κρατήσουμε τα καλύτερα αποτελέσματα.

Μέσα σε λίγα epochs το μοντέλο προσεγγίζει RMSE = 0.0498 στο validation dataset. Ενώ στο testing RMSE = 0.0455

Chart, line chart

Description automatically generated4 τυχαίες χρονικές περίοδοι 100-στιγμων από το test dataset.

Από τα αποτελέσματα που παίρνουμε έχουμε ότι

Η μέση τιμή του λάθους σε τιμές MW είναι 178MW, αυτό αποτελεί 2% της τυπικής απόκλισης που έχουν οι σταθμοί παραγωγής από ορυκτά καύσιμα τιμή που είναι αρκετά αποδεκτή.

Ο κώδικας υπάρχει εδώ: <https://colab.research.google.com/drive/1S-05T9LjXab1YasXV-8G03ClVnBVD918?usp=sharing>