

ΘΕΜΑΤΑ ΠΤΥΧΙΑΚΩΝ ΕΡΓΑΣΙΩΝ

1 Πρόβλεψη Παραγωγής Ενέργειας Ανεμογεννήτριας

1.1 Σκοπός

Στην πτυχιακή αυτή θα αναλυθούν δεδομένα ενός έτους που παράχθηκαν από το σύστημα SCADA μιας ανεμογεννήτριας. Ο φοιτητής, μέσα από την εργασία αυτή, θα μάθει τις βασικές αρχές επεξεργασίας δεδομένων με εργαλεία μηχανικής μάθησης (και όχι μόνο), με σκοπό την εξαγωγή πληροφοριών για ένα πραγματικό πρόβλημα. Θα χρησιμοποιηθεί ένα ανοικτό πακέτο δεδομένων. Πιο συγκεκριμένα το πακέτο δεδομένων περιλαμβάνει μετρήσεις ανά 10 λεπτά σε:

- Πραγματική Παραγωγή ενέργειας (kW): Η πραγματική ενέργεια που παράγεται τη στιγμή της μέτρησης
- Ταχύτητα Ανέμου (m/s): Η ταχύτητα του ανέμου που καταγράφεται εκείνη τη στιγμή
- Κατεύθυνση ανέμου (°): Η κατεύθυνση του ανέμου σε μοίρες τη στιγμή της μέτρησης
- Ονομαστική Παραγωγή Ενέργειας: Η ενέργεια που δηλώνει ο κατασκευαστής ότι μπορεί να παραχθεί από τη γεννήτρια βάση των μετρούμενων συνθηκών

Για την πρόβλεψη, η στήλη της πραγματικής παραγωγής θα είναι η έξοδος του συστήματος, ενώ οι υπόλοιπες στήλες είσοδοι.

1.2 Μεθοδολογία υλοποίησης

Ο ενδιαφερόμενος φοιτητής μπορεί να επισκεφτεί την ιστοσελίδα Kaggle¹ που προσφέρεται το πακέτο δεδομένων. Εκεί θα βρει περισσότερες πληροφορίες για το πακέτο δεδομένων καθώς και κώδικα που προσφέρεται για την ανάλυση και την πρόβλεψη παραγωγής **στα αγγλικά**. Ο φοιτητής θα πρέπει στην εργασία να περιγράψει θεωρητικά την λειτουργία των ανεμογεννητριών, καθώς και του μοντέλου πρόβλεψης που επέλεξε (π.χ. Νευρωνικά Δίκτυα) μέσα από τη βιβλιογραφία που θα βρεί. Η δομή της εργασίας θα είναι:

- Εισαγωγή-Λειτουργία Ανεμογεννήτριας (Θεωρία)
- Μεθοδολογία
 - Περιγραφή δεδομένων
 - Περιγραφή του αλγόριθμου πρόβλεψης π.χ. Αρχιτεκτονική Νευρωνικού δικτύου (Στρώματα, νευρώνες, εποχές κτλπ)
 - Αποτελέσματα προσέγγισης (Απόλυτο σφάλμα, μέσο Απόλυτο σφάλμα, μέσο τετραγωνισμένο σφάλμα)
- Συμπεράσματα

Ο φοιτητής μπορεί να χρησιμοποιήσει οποιοδήποτε λογισμικό για την υλοποίηση των αλγόριθμων πρόβλεψης και την επεξεργασία δεδομένων (matlab, python, excel κτλπ). Συνιστάται η χρήση της γλώσσας Python καθώς θα βρει και έτοιμα κομμάτια κώδικα στον ιστότοπο το πακέτου δεδομένων. Εάν ο φοιτητής είναι εξοικειωμένος με την χρήση της αγγλικής, μπορεί να παρακολουθήσει το δωρεάν σεμινάριο για εκμάθηση Python στο netacad² ή στα ελληνικά από το mathesis³.

¹ <https://www.kaggle.com/datasets/berkerisen/wind-turbine-scada-dataset>

² <https://www.netacad.com/courses/programming/pcap-programming-essentials-python>

³ <https://mathesis.cup.gr/courses/course-v1:ComputerScience+CS1.1+21D/about>

2 Ανάλυση της ποιότητας του νερού και κατηγοριοποίηση με τη χρήση της Python

2.1 Σκοπός

Στην πτυχιακή αυτή θα αναλυθούν δεδομένα από ποιοτικό έλεγχο σε δεξαμενές νερού. Ο φοιτητής, μέσα από την εργασία αυτή, θα μάθει τις βασικές αρχές επεξεργασίας δεδομένων με εργαλεία μηχανικής μάθησης, με σκοπό την εξαγωγή πληροφοριών για ένα πραγματικό πρόβλημα. Πιο συγκεκριμένα θα υλοποιήσει αλγόριθμο κατηγοριοποίησης (classification) όπου ανάλογα τα μετρούμενα χαρακτηριστικά του νερού, ο αλγόριθμος θα εξετάζει αν αυτό είναι πόσιμο ή όχι (binary classification). Το πακέτο δεδομένων περιλαμβάνει 9 μετρήσεις αναφορικά με την ποιότητα του νερού (στήλες), από 3276 δεξαμενές (σειρές): Οι μετρήσεις περιλαμβάνουν:

- pH
- Σκληρότητα
- Επίπεδο ολικών διαλυμένων στερεών (TDS)
- Χημικές ουσίες (χλώριο)
- Συγκέντρωση θειικού
- Αγωγιμότητα
- Επίπεδο οργανικού άνθρακα (TOC)
- Τριαλομεθάνιο
- Καθαρότητα-διαφάνεια νερού (θολούρα)
- Ποσιμότητα

2.2 Μεθοδολογία υλοποίησης

Ο ενδιαφερόμενος φοιτητής μπορεί να επισκεφτεί την ιστοσελίδα Kaggle⁴ που προσφέρεται το πακέτο δεδομένων. Εκεί θα βρει περισσότερες πληροφορίες για το πακέτο δεδομένων καθώς και κώδικα που προσφέρεται για την ανάλυση και την κατηγοριοποίηση **στα αγγλικά**. Ο φοιτητής θα πρέπει στην εργασία να περιγράψει μέσα από βιβλιογραφία που θα βρει το πρόβλημα της ποσιμότητας του νερού⁵, των δεδομένων, καθώς και του μοντέλου πρόβλεψης που επέλεξε (π.χ. Νευρωνικά Δίκτυα). Η δομή της εργασίας θα είναι:

- Εισαγωγή: Καταλληλότητα-ποσιμότητα νερού
- Μεθοδολογία
 - Περιγραφή δεδομένων
 - Περιγραφή του αλγόριθμου
 - Αποτελέσματα κατηγοριοποίησης με μετρικά όπως η ακρίβεια (accuracy, precision, recall)
- Συμπεράσματα

Οι αλγόριθμοι θα πραγματοποιηθούν στη γλώσσα Python, που χρησιμοποιείται ευρέως για τέτοιου τύπου προβλήματα, με βιβλιοθήκες όπως η pandas, η tensorflow, η scikit-learn κ.α. για την ανάπτυξη αλγόριθμων και επεξεργασίας δεδομένων. Εάν ο φοιτητής είναι εξοικειωμένος με την χρήση της αγγλικής, μπορεί να παρακολουθήσει το δωρεάν σεμινάριο για εκμάθηση Python στο netacad⁶ ή στα ελληνικά από το mathesis⁷.

⁴ <https://www.kaggle.com/datasets/adityakadiwal/water-potability>

⁵

https://el.wikipedia.org/wiki/%CE%A0%CE%BF%CE%B9%CF%8C%CF%84%CE%B7%CF%84%CE%B1_%CF%84%CE%BF%CF%85_%CE%BD%CE%B5%CF%81%CE%BF%CF%8D

⁶ <https://www.netacad.com/courses/programming/pcap-programming-essentials-python>

⁷ <https://mathesis.cup.gr/courses/course-v1:ComputerScience+CS1.1+21D/about>