A picture containing text, sign

Description automatically generated

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΕΙΔΙΚΕΥΣΗ: “ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ”

Wa. Po. – Water Potability

(Analytics as a Service)

ΕΥΦΥΗ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

 Επιβλέπων Καθηγητής: Δημοσθένης Κυριαζής

Ιωάννα Κανδή – Κωνσταντίνος Μαυρογιώργος

ΜΕ2136 – ΜΕ2144

{[ioannakandi, kostismvg}@gmail.com](mailto:ioannakandi@gmail.com)

Ιούνιος 2022

Πίνακας Περιεχομένων

[Κατάλογος Εικόνων 1](#_Toc106364424)

[Κατάλογος Πινάκων 1](#_Toc106364425)

[1. Εισαγωγή 2](#_Toc106364426)

[2. Υλοποίηση Συστήματος 3](#_Toc106364427)

[2.1. Χρησιμοποιούμενες Τεχνολογίες 3](#_Toc106364428)

[2.2. Σύνολο Δεδομένων 3](#_Toc106364429)

[2.3. Ανάλυση Δεδομένων 4](#_Toc106364430)

[2.4. Υλοποιημένες Υπηρεσίες 4](#_Toc106364431)

[3. Εγχειρίδιο Χρήσης 5](#_Toc106364432)

[4. Συμπεράσματα 6](#_Toc106364433)

# Κατάλογος Εικόνων

[Εικόνα 1: Σιγμοειδής συνάρτηση 4](#_Toc98767834)

[Εικόνα 2: Συντελεστές παλινδρόμησης 4](#_Toc98767835)

[Εικόνα 3: Απεικόνιση σιγμοειδούς συνάρτησης 4](#_Toc98767836)

[Εικόνα 4 Διαγραμματική απεικόνιση γραμμικής & λογιστικής παλινδρόμησης 5](#_Toc98767837)

# Κατάλογος Πινάκων

[Πίνακας 1: Πρότυπο πίνακα για υλοποιημένες υπηρεσίες 4](#_Toc106364434)

[Πίνακας 1: Υπηρεσία predictPotability 4](#_Toc106364435)

# Εισαγωγή

Η εκπόνηση εργασιών στα πλαίσια των μαθημάτων της Μεταπτυχιακής Εξειδίκευσης Σπουδών με τίτλο «Προηγμένα Πληροφοριακά Συστήματα» ήταν απαραίτητη για την μέγιστη δυνατή εξοικείωση των φοιτητών με το αντικείμενο του Προγράμματος. Η παρούσα εργασία αφορά το μάθημα εαρινού εξαμήνου με τίτλο «Ευφυή Πληροφοριακά Συστήματα και Τεχνητή Νοημοσύνη», και πραγματεύεται περιληπτικά τη δημιουργία μιας containerized εφαρμογής που αξιοποιεί την βιβλιοθήκη Keras για την πραγματοποίηση πρόβλεψης σχετικά με τη ποσιμότητα δειγμάτων νερού. Στο κείμενο που ακολουθεί περιγράφεται το σύνολο δεδομένων που αξιοποιήθηκε, οι τεχνολογίες που εφαρμόστηκαν και ο τρόπος με τον οποίο λειτουργεί η εφαρμογή . Τέλος, παρατίθενται τα συμπεράσματα που προκύπτουν από την εκπόνηση της εργασίας αυτής.

# Υλοποίηση Συστήματος

## Χρησιμοποιούμενες Τεχνολογίες

Η web εφαρμογή αποτελείται από το front-end, το οποίο αποτελεί την διεπαφή του χρήστη, και το back-end που είναι όλες οι υπηρεσίες που έχει πρόσβαση ο χρήστης από την διεπαφή. Για το front-end χρησιμοποιήθηκαν οι ακόλουθες τεχνολογίες:

* HTML5 για την οργάνωση των επιμέρους τμημάτων από τα οποία αποτελείται η διεπαφή και την προσθήκη κανόνων στις φόρμες υποβολής του χρήστη για διασφάλιση της εγκυρότητας αυτών πριν σταλούν στο back-end.
* Bootstrap 4: για την μορφοποίηση των παραπάνω τμημάτων ούτος ώστε να είναι ευδιάκριτα και ελκυστικά προς τον χρήστη. Το Bootstrap είναι ένα framework το οποίο αποτελείται από HTML, CSS και JavaScript. Ουσιαστικά, παρέχει στους προγραμματιστές έτοιμες κλάσεις τις οποίες μπορούν να χρησιμοποιήσουν έτσι ώστε να υλοποιήσουν γρήγορα μία πολύ ελκυστική διεπαφή χρήστη.
* CSS για την μορφοποίηση των παραπάνω τμημάτων ούτος ώστε να είναι ευδιάκριτα και ελκυστικά προς τον χρήστη.
* JavaScript για την προσθήκη του στοιχείου της δυναμικότητας στο front-end, διατήρηση cookies και γενικά διαχείριση διάφορων τιμών στο front-end.
* jQuery, η οποία αποτελεί μία βιβλιοθήκη JavaScript και ουσιαστικά παίρνει πολλές κοινές εργασίες που απαιτούν πολλές γραμμές κώδικα JavaScript για να ολοκληρωθούν και τις «τυλίγει» σε μεθόδους που μπορούν να χρησιμοποιηθούν με μία γραμμή κώδικα. Μέσω της jQuery γίνονται και οι κλήσεις Ajax στα διάφορα API με τα οποία επικοινωνεί η εφαρμογή, όπως επίσης και η εφαρμογή τεχνικών form validation, αξιοποιώντας την αντίστοιχη βιβλιοθήκη (jQuery Validation Plugin).

Για το back-end χρησιμοποιήθηκαν οι ακόλουθες τεχνολογίες:

• Python και ειδικότερα η βιβλιοθήκη Keras για την πραγματοποίηση των προβλέψεων.

## Σύνολο Δεδομένων

Στα πλαίσια της συγκεκριμένης εργασίας επιλέχθηκε ο τομέας του περιβάλλοντος και η ενασχόληση με ένα αντίστοιχο πρόβλημα. Ειδικότερα, ανακτήθηκε από το [Kaggle](https://www.kaggle.com/datasets/adityakadiwal/water-potability) ένα σύνολο δεδομένων το οποίο περιέχει μετρήσεις που πραγματοποιήθηκαν σε διαφορετικά δείγματα νερού και τη διερεύνηση σχετικά με τη ποσιμότητά τους. Τα χαρακτηριστικά που περιέχει το συγκεκριμένο σύνολο δεδομένων είναι τα ακόλουθα:

* pH: pH του δείγματος (0 έως 14).
* Σκληρότητα: Η ικανότητα του νερού να καθιζάνει σαπούνι σε mg/L.
* Στερεά: Ολικά διαλυμένα στερεά σε ppm.
* Χλωραμίνες: Ποσότητα χλωραμινών σε ppm.
* Θειικά: Ποσότητα θειικών αλάτων διαλυμένη σε mg/L.
* Αγωγιμότητα: Ηλεκτρική αγωγιμότητα του νερού σε μS/cm.
* Οργανικός Άνθρακας: Ποσότητα οργανικού άνθρακα σε ppm.
* Τριαλομεθάνια: Ποσότητα Τριαλομεθανίων σε μg/L.
* Θολότητα: Μέτρηση της ιδιότητας εκπομπής φωτός του νερού σε NTU.
* Ποσιμότητά: Υποδεικνύει εάν το νερό είναι ασφαλές για ανθρώπινη κατανάλωση. Πόσιμο -1 και Μη πόσιμο -0

## Ανάλυση Δεδομένων

Στόχος της εργασίας είναι η πρόβλεψη του χαρακτηριστικού «Ποσιμότητα» για ένα δείγμα νερού το οποίο δίνεται από τον χρήστη της εφαρμογής. Για την πραγματοποίηση της πρόβλεψης πραγματοποιήθηκαν συγκεκριμένα βήματα, ενώ αξιοποιήθηκε η βιβλιοθήκη Keras. Αρχικά, τυχών ελλιπείς τιμές που περιέχονται στο σύνολο δεδομένων αντικαταστάθηκαν με τον μέσο όρο της αντίστοιχης στήλης, ενώ οι διπλότυπες εγγραφές αφαιρέθηκαν. Εν συνεχεία, αναπτύχθηκε ένα νευρωνικό δίκτυο τριών (3) επιπέδων. Το πρώτο επίπεδο περιείχε δώδεκα (12) νευρώνες με συνάρτηση ενεργοποίησης την «ReLU», το δεύτερο περιείχε πέντε (5) νευρώνες με συνάρτηση ενεργοποίησης την «ReLU», ενώ το τρίτο περιείχε ένα (1) νευρώνα με συνάρτηση ενεργοποίησης την «Sigmoid». Έπειτα, δημιουργήθηκε το μοντέλο μηχανικής μάθησης το οποίο πλέον μπορούσε να αξιοποιηθεί για τη πραγματοποίηση προβλέψεων σε τυχαία δείγματα νερού.

## Υλοποιημένες Υπηρεσίες

Στη συγκεκριμένη ενότητα περιγράφονται οι υλοποιημένες υπηρεσίες ιστού σε μορφή πινάκων. Οι πίνακες έχουν την ακόλουθη μορφή:

|  |  |
| --- | --- |
| Όνομα Υπηρεσίας | |
| Περιγραφή | Περιγραφή της λειτουργίας της υπηρεσίας |
| Endpoint URL | http://localhost:5000/... |
| HTTP μέθοδος | Δυνατές τιμές: GET, POST, PUT, DELETE |
| Παράμετροι | Οι παράμετροι που χρειάζεται η υπηρεσία (αν υπάρχουν) |
| Επιστροφή | Τι επιστρέφει η υπηρεσία στον χρήστη |

Πίνακας 1: Πρότυπο πίνακα για υλοποιημένες υπηρεσίες

|  |  |
| --- | --- |
| predictPotability | |
| Περιγραφή | Αξιοποιείται για τη πραγματοποίηση της πρόβλεψης μέσω της Keras σχετικά με το αν ένα δείγμα νερού είναι πόσιμο |
| Endpoint URL | http://localhost:5000/ predictPotability |
| HTTP μέθοδος | GET |
| Παράμετροι | ph: το pH του δείγματος  hardness: η σκληρότητα του δείγματος  solids: η ποσότητα διαλυμένων στερεών στο δείγμα  chloramines: η ποσότητα χλωραμινών στο δείγμα  sulfate: η ποσότητα θειικών αλάτων στο δείγμα  conductivity: η αγωγιμότητα του δείγματος  organic\_carbon: η ποσότητα οργανικού άνθρακα στο δείγμα  trihalomethanes: η ποσότητα τριαλομεθανίων στο δείγμα  turbidity: η θολότητα του δείγματος |
| Επιστροφή | Επιστρέφει 0 αν το δείγμα νερού δεν είναι πόσιμο και 1 αν το δείγμα νερού είναι πόσιμο |

Πίνακας 1: Υπηρεσία predictPotability

# Εγχειρίδιο Χρήσης

Εφόσον εκκινήσει η εφαρμογή, αυτή είναι διαθέσιμη στη διεύθυνση <http://localhost:5000>, όπως φαίνεται παρακάτω.

Ο χρήστης έχει τη δυνατότητα να συμπληρώσει τη διαθέσιμη φόρμα και να ανακτήσει την αντίστοιχη πρόβλεψη για το δείγμα νερού που έχει στη διάθεσή του. Ύστερα από την ανάλυση του δείγματος επιστρέφεται ανάλογο μήνυμα που ενημερώνει τον χρήστη αν το νερό είναι πόσιμο (ΧΧΧ) ή όχι (ΧΧΧ).

Στους χρήστες παρέχονται επίσης πληροφορίες που ανακτώνται από εξωτερικά APIs και αφορούν την ποιότητα νερού σε επιλεχθείσες περιοχές (ΧΧΧ), όπως επίσης και σχετικά tweets (ΧΧΧ).

Επίσης, αξίζει να σημειωθεί ότι ο κώδικας της εφαρμογής είναι ανεβασμένος και σε ιδιωτικό repository στο GitHub, στη διεύθυνση <https://github.com/ioannakandi/waterPotability>. Για να έχει κανείς πρόσβαση θα πρέπει να επικοινωνήσει σε ένα από τα emails που αναγράφονται στο εξώφυλλο της εργασίας, έτσι ώστε να προστεθεί ως contributor. Τέλος, η εφαρμογή είναι διαθέσιμη και ως Docker Container στη διεύθυνση XXX.

# Συμπεράσματα

Συνοψίζοντας, η παρούσα εργασία ήταν απαραίτητη για να επιτευχθεί μια πρώτη επαφή με σημαντικές έννοιες του μαθήματος. Πιο συγκεκριμένα, υλοποιήθηκε μία containerized εφαρμογή που αξιοποιεί την βιβλιοθήκη Keras για την πραγματοποίηση πρόβλεψης σχετικά με τη ποσιμότητα δειγμάτων νερού. Για τη δημιουργία του προβλεπτικού μοντέλου δημιουργήθηκε νευρωνικό δίκτυο τριών (3) επιπέδων και οι συναρτήσεις ενεργοποίησης «ReLU» και «Sigmoid». Ωστόσο, μία μελλοντική βελτίωση της παρούσας εργασίας θα ήταν η εφαρμογή περισσοτέρων αλγόριθμων μηχανικής μάθησης, έτσι ώστε να βελτιωθεί η ακρίβεια του παραγόμενου μοντέλου.