

**SCHOOL OF TECHNOLOGY**

**INDUSTRIAL DIGITAL TALENT PROGRAMME**

FINAL HACKATHON

Task 5: Presentations of Tasks’ Results

The motivation behind this task is to present Final Hackathon results and obtain feedback from the panelists on

areas that need improvement

Instruction

i) Submit word **processed document** and **presentation slides** with results of all the tasks (1-4) on or before 22nd /11/2022 at 4pm.

ii) Present the slides during the Final Hackathon Presentation Session that will be held on 24th November 2022

Requirement

i Compile a word processes document that summarizes the results of Tasks 1-4 ii Compile presentation slides with results of Tasks 1-4.

# Defining the problem

* The project's goal is to leverage time series analysis to predict energy consumption in 10-minute windows for the city of Tétouan in Morocco.
* Given the strong dependency on non-renewable sources (64%), forecasting energy consumption could help the stakeholders better manage purchases and stock. On top of that, Morocco’s plan is to   
  reduce energy imports by increasing production from renewable sources. It’s common knowledge   
  that sources like wind and solar present the risk of not being available all year round. Understanding the energy needs of the country, starting with a medium-sized city, could be a step further in planning these resources.
* The distribution network is powered by 3 zone stations, Quads, Smir and Boussafou. The 3 zone stations power 3 different areas of the city, this is why we have three potential target variables.
* The dataset is located in <http://archive.ics.uci.edu/ml/datasets/Power+consumption+of+Tetouan+city>

# Loading and cleaning the dataset

Table

Description automatically generated with medium confidence

Graphical user interface, table

Description automatically generated with medium confidence

Graphical user interface

Description automatically generated

Graphical user interface, text, application

Description automatically generatedGraphical user interface, text, application

Description automatically generatedTable

Description automatically generatedGraphical user interface, text

Description automatically generatedTable

Description automatically generated with low confidenceData VisualizationGraphical user interface

Description automatically generated with medium confidenceChart, histogram

Description automatically generatedChart, line chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, line chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, box and whisker chart

Description automatically generatedChart, scatter chart

Description automatically generated

# Data Transformation

Table

Description automatically generated

Table

Description automatically generated

# Selecting the predictor and the output variable

Table

Description automatically generated

# Splitting the dataset into training and testing data

Table

Description automatically generated

# Model building

## Linear Regression

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generated

## Decision Tree

The maximum depth was set to 6.

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text

Description automatically generated

## Random Forest

Graphical user interface, text, application, email

Description automatically generatedGraphical user interface, text, application

Description automatically generated

## k-Nearest Neighbors

Graphical user interface, text, application, email

Description automatically generatedText

Description automatically generated

## Support Vector Machine

The Linear SVR was picked because SVR itself took too long to train. Upon finding out why, we realized that SVR scales poorly with larger datasets

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text

Description automatically generated

# Evaluating and selection of the models

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Coefficient of determination** | **Mean absolute error** | **Mean squared error** | **Root mean squared error** |
| Linear Regression | 0.645513 | 3400.205 | 18281112 | 4275.642 |
| Decision tree | 0.895157 | 1717 | 5406812 | 2325.255 |
| Random Forest | 0.903543 | 1640.9 | 4974350 | 2230.325 |
| k-Nearest Neighbours | 0.91521 | 1372.777 | 4372661 | 2091.091 |
| Support Vector Machine | 0.559441 | 3740.107 | 22719885 | 4766.538 |

The algorithm that we feel is the best-performing is the k-nearest neighbour since the decision trees and the random forest algorithm both have a high risk of overfitting.

The Multiple Linear regression and the Support Vector Machine perform the poorest in terms of the coefficient of determination and the closeness in which values are predicted.

# Exporting the model

Joblib was used to dump the model so that it could be reused to make predictions without restarting a Jupyter kernel.

Graphical user interface, text, application

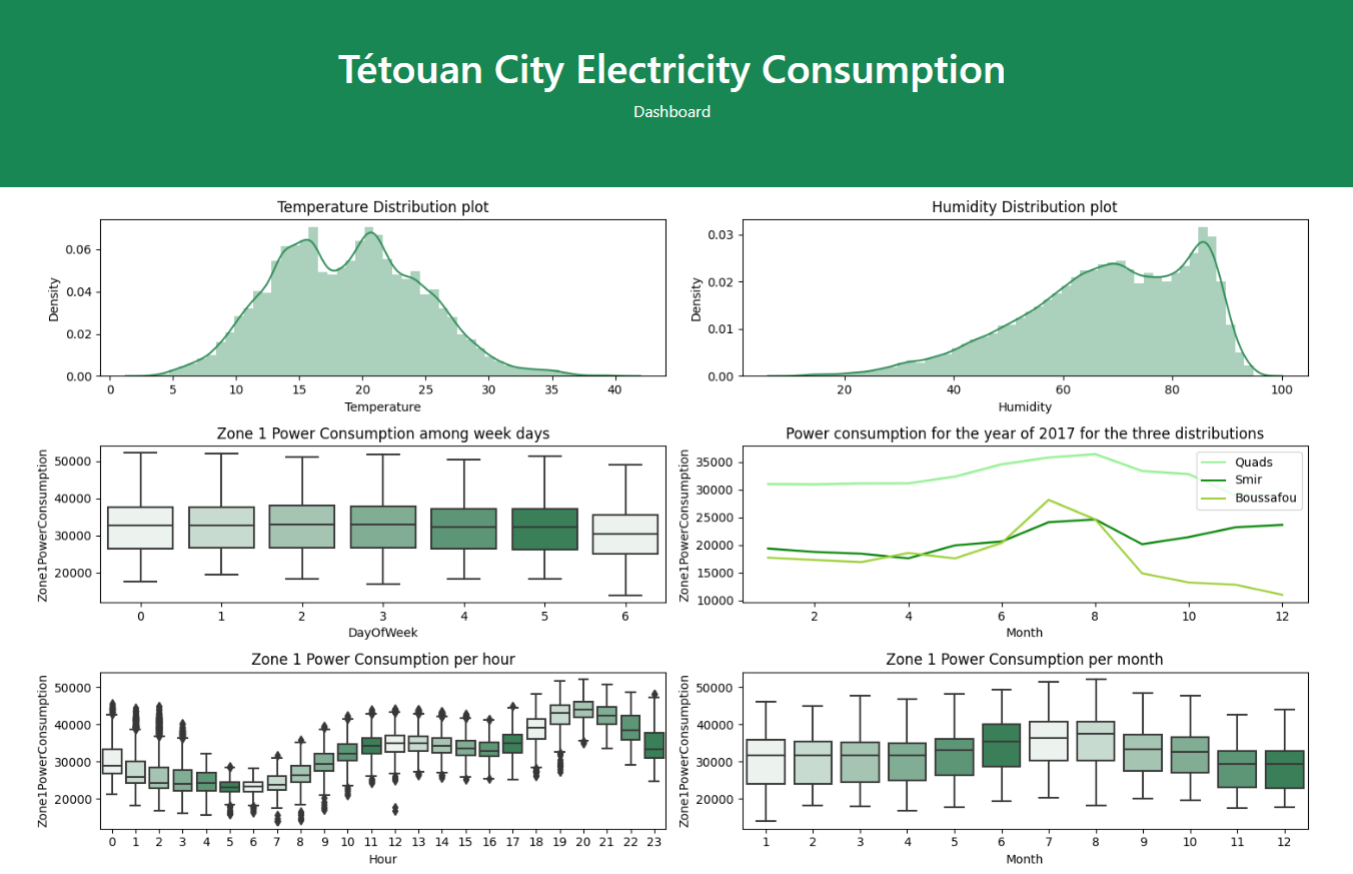
Description automatically generated

# Deploying the machine learning model

The following are screenshots for our machine learning application which predicts electricity consumption in Tétouan City located in Morocco. The application was built using Flask and utilizes the k-nearest neighbour regression model in predictions.

The homepage consists of elaborate charts including distribution plots, box plots and line plots. These plots describe distribution of temperature and humidity, Zone 1 (Quads) power consumption among weekdays, Power consumption for the year of 2017 for the three zones (Quads, Smir and Boussafou), Zone 1 (Quads) Power Consumption per hour, and Zone 1 (Quads) Power Consumption per month.

The application also consists of a form which collects information from the user and predicts electricity consumption based on attributes such as temperature, humidity, wind speed, general diffuse flows and diffuse flows. The selected model with a r2 score of 0.9152101236 predicts the electricity consumption which is then displayed at the bottom.

Graphical user interface, text, application

Description automatically generated