**Software for Network and**

**Services Design (SNS)**

**ELEC0088**

**Assignment Report**

by

**Candidate Number / Student Number:** Programme

**Candidate Number / Student Number:** Programme

**Candidate Number / Student Number:** Programme

**Candidate Number / Student Number:** Programme

**Candidate Number / Student Number:** Programme

**Candidate Number / Student Number:** Programme

29th March, 2024

Table of Contents

[Executive Summary 2](#_Toc161075141)

[1. Introduction 3](#_Toc161075142)

[2. Project Implementation 3](#_Toc161075143)

[*2.1 Data Collection 3*](#_Toc161075144)

[*2.2 Data Preparation 4*](#_Toc161075145)

[*2.3 Choosing a Machine Learning Algorithm 4*](#_Toc161075146)

[*2.4 Training the Model 4*](#_Toc161075147)

[*2.5 Evaluating the Model 4*](#_Toc161075148)

[*2.6 Parameter Tuning 5*](#_Toc161075149)

[*2.7 Making the Model Available via a Server. 5*](#_Toc161075150)

[*2.8 Making Predictions 5*](#_Toc161075151)

[3. Results and Conclusion 5](#_Toc161075152)

# **Executive Summary**

This report describes implementation of an Oracle chatbot for predicting the number of bicycles hires in the future in London. Prediction is done by using time series forecasting with machine learning. The Oracle runs in a server and can be reached by a client that interfaces via a terminal.

Project implementation included two main parts namely (i) Machine Learning part and (ii) Networking part. Altogether, eight (8) distinct steps were followed in execution of the project.

In the machine learning part, supervised learning with neural network was deployed with keras library. Linear regression, which is a supervised learning algorithm to predict the outcome of a continuous variable was used. The core task was to train a model to make it predict the number of bicycles hires in the future. Two datasets were used, which are weather dataset and bicycles hires dataset.

On the other side, in the networking part, socket interface was used to provide network functionality to send and receive data in the network. TCP socket that provides a reliable and connection-oriented service was chosen. The main task performed was to make the developed model available through a server. The client could request information from the server and get a response back.

The overall result was that the developed model could perform the task at hand with accuracy … loss ….. and … In addition, the model was accessible from a server, the client could request information from the server and get a prediction response back.

The project has been implemented collaboratively by six group members and a central repository code of GitHub was used to facilitate contributions by members. All codes have been implemented in Python programming language and a GitHub URL is included in this report.

# **Introduction**

Machine learning as a subfield of Artificial Intelligence gives computers ability to learn without being explicitly programmed. It is behind many sophisticated services in today’s world such as chatbots, predictive texts and suggestions to users when online.

In line with this, prediction of number of bicycles hires in a city is another area in everyday life that can benefit from the power of machine learning. This objective was therefore set to be implemented which can be imperative in improvement of bicycles hire services.

Predictions are done based on weather conditions. While recognizing that bicycles hires can be influenced by a number of other factors such as random events in the city, weather has constant relationship and therefore was the factor considered.

Machine learning can be Supervised, Unsupervised or Reinforcement. Supervised learning used labelled data, and is grouped into classification and regression. Regression is mostly used to predict data. Therefore, Supervised learning with neural network will be applied in this project namely “Prediction of number of bicycles hires in London”.

# **Project Implementation**

This section provides a description of steps followed in implementing the project. All data and codes involved are available at GitHub URL <https://github.com/Flowey0622/ELEC0088-SNS>.

The following eight (8) steps were executed:

## **Data Collection**

Relevant data in line with project objective was searched on the Internet. Much effort was employed in searching for and collecting data in order to secure good quality data because incorrect or outdated data can lead to wrong outcomes and irrelevant predictions.

Among the goals was also to have data large and diverse enough to cover sufficiently well the characteristics of the underlying task, so that the model can learn the task well. Data consistency was also key in the selection of datasets, there was much interest in having time series data.

Two datasets were collected:

1. London weather dataset (01/01/2000 – 17/02/2024)
2. Tfl daily cycles hires dataset (30/07/2010 – 31/01/2024)

## **Data Preparation**

At this stage, data was cleaned and pre-processed to make it ready for training. In such a process, a number of checks were conducted and subsequent improvements made where necessary. Main activities in data preparation were:

1. Merging the two collected datasets into a single dataset. Entries outside mutual dates of the two datasets were dropped and the resulting dataset comprised of entries in mutual dates only.
2. Checking datatypes of each parameter and converting to the right datatype for Machine Learning e.g. date parameter as the index of the dataset was converted from object type to datetime type
3. Checking percentage of missing data in each parameter. This was in order to identify the extent of null values in the dataset. The goal is not to pass entries with null values to training.
4. Dealing with columns with high percentage of null values. A value of 10% was chosen as cut off and columns with null percentage greater than 10% were removed from the dataset.
5. Filling up missing values. For the remaining columns, missing data were filled up. Previous day values were used in filling up because there is a high correlation of weather condition between consecutive days.

At this point, collected data was ready to be passed for next step.

## **Choosing a Machine Learning Algorithm**

A model relevant to the task at hand was identified. A machine learning model determines the output after running a machine learning algorithm on the collected data. Since the project is on projection, linear regression was chosen. Subsequently, keras library was selected.

## **Training the Model**

Collected and prepared data was passed to the machine learning model to find patterns and make predictions. The model learned from the data so that it can accomplish the set task.

## **Evaluating the Model**

Using testing data, an accurate measure of how the model will perform and its speed can be obtained.

## **Parameter Tuning**

The objective was to find out if the model’s accuracy could be improved by tuning the parameters present in the model, finding a particular value of the parameter where the accuracy would be the maximum.

## **Making the Model Available via a Server.**

## **Making Predictions**

In the end, the model was used to make predictions, being accessed from a server.

# **Results and Conclusion**