Data Science & Machine Learning Portfolio Specification

Student Name: Qadeer Hussain (Student ID: C00270632)

Introduction

This portfolio specification showcases key information about myself including current education and work experience. I will also be linking my GitHub and LinkedIn. This portfolio contains my skill set and Data Science and Machine Learning project proposals, to each of which I have attached a brief description. These projects will be using either/both text and image data. I will be using multiple sites to get existing datasets for the proposals below. Algorithms will also be used in these project proposals.

Project Proposals

Traffic Management System: The aim of this project is to optimize traffic flow by reducing congestion and re-routing vehicles, and enhancing safety by ensuring safe movement of people and vehicles on the road. The system will predict traffic flow during rush hours such as commuting to work or school. It will also predict accidents by using existing datasets by cautioning vehicles to slow down or rerouting them on a safer path. Image detection can be utilized for this project. I will be getting the datasets from websites such as https://www.kaggle.com/, https://paperswithcode.com/ and https://data.gov.ie/.

There are many different types of algorithms that could be used for this proposal such as Linear regression which is a supervised machine learning algorithm, Logistic Regression another supervised machine learning algorithm and K-Means Clustering, a Unsupervised machine learning algorithm.

Sign language translator: This project will be using live camera feed to interpret sign language used by people with a hearing impairment. The translation will be displayed as words or sentences on screen. For example if the hand gesture is " food" and "hot" is detected, the system will output the sentence as "The food is hot". Using existing datasets of hand signs to train the system. I will be using datasets from websites such as https://www.kaggle.com/. An example algorithm that could be used here would be LSTM(Long Short-Term Memory) which is a supervised ML algorithm.

Global Economy, Population: The aim of this project is to use existing data to predict countries GDP growth, and population. This project will create accurate forecasts for individual countries. The predictions will be visualized through graphs. I will be using datasets from websites such as https://data.worldbank.org/, https://data.gov.ie/ and https://www.kaggle.com/. An example algorithm for this project could be Linear regression.

Meal suggester: The aim of this project is to create a personalized meal suggester. The user has different preferences such as dietary restrictions, health goals: weight loss and muscle gain. The system will also recommend meals aligned with the users preferences such as eating high carb or high protein meals. It

will consider dietary requirements like vegan, vegetarian or non-vegetarian. System will also provide the detailed recipes and the time frame it will take to cook the meals and calorie intake depending on the meal. I will be using datasets from websites such as https://www.kaggle.com/. A suitable machine learning algorithms for this could be Linear regression and Clustering algorithms.

Sudoku solver: The aim of this project is to create a sudoku solver that allows users to both manually enter each number on the grid or by using image detection to automatically recognize the grid. The Sudoku solver will analyse each grid row by row and fill in the blank missing numbers and getting the possible entries. An example machine learn algorithm could be Decision Trees which is a supervised algorithms.

Car registration detector: The aim of this project is to extract car registration numbers using image detection such as computer vision. Vehicles that are parked in shops or supermarket. This system can be used for parking management and monitoring overstays.

Sales Prediction: The aim of this project is to predict sales based on seasonal patterns and time of year. This system will do this by analysing existing datasets. It will identify high sale periods such as increased sales during the holidays, such as Christmas and Easter. This will help business optimize inventory management. I will be using datasets from websites such as https://archive.ics.uci.edu/ and https://www.kaggle.com/. Linear regression which is a supervised machine learning algorithm could be very suitable for this project.

Technologies

Through the duration of my course and work placement during year 3, I have gained proficiency in several programming languages, frameworks, I am currently learning algorithms/techniques for machine learning. Some of which I intend to utilize in my projects. Below is a list of all technologies I have encountered and will be learning:

- Java
- HTML
- CSS
- JavaScript (JS)
- Python
- Jupyter Notebook
- Go
- GD Script
- Angular
- Java Server Pages(JSP)
- SQL database
- Git, GitHub
- Machine learning algorithms/techniques:
    - Linear Regression
    - Logistic Regression
    - K- Means Clustering
    - LSTM(Long Short-Term Memory)

- Decision trees
- Clustering
- Computer Vision

Machine Spec

- Operating System: Windows 11 and Linux distro POP OS
- Ram: 16GB Ram
- Storage: 1TB SSD
- CPU: AMD Ryzen 5 5600H with integrated graphics
- GPU: Nvidia GeForce RTX 3050

References

Banerjee, S., n.d. *World Population Dataset*. [Online]
Available at: https://www.kaggle.com/datasets/iamsouravbanerjee/world-population-dataset
[Accessed September 2024].

Chen, D., n.d. *Online Retail*. [Online]
Available at: https://archive.ics.uci.edu/dataset/352/online+retail
[Accessed September 2024].

Data Gov, n.d. *Data.gov.ie*. [Online]
Available at:
https://data.gov.ie/dataset?q=population&tags=population&sort=score+desc%2C+metadata_created+desc
[Accessed September 2024].

Data Gov, n.d. *Data.gov.ie*. [Online]
Available at: https://data.gov.ie/dataset/?tags=traffic&theme=Transport&license_id=CC-BY-4.0&api=true
[Accessed September 2024].

Fedesoriano, n.d. *Traffic Prediction Dataset*. [Online]
Available at: https://www.kaggle.com/datasets/fedesoriano/traffic-prediction-dataset/data
[Accessed September 2024].

Geeks for Geeks, 2024. *Computer Vision Tutorial*. [Online]
Available at: https://www.geeksforgeeks.org/computer-vision/
[Accessed September 2024].

Geeks for Geeks, 2024. *Decision Tree in Machine Learning*. [Online]
Available at: https://www.geeksforgeeks.org/decision-tree-introduction-example/
[Accessed September 2024].

Geeks for Geeks, 2024. *Geeks for Geeks*. [Online]
Available at: https://www.geeksforgeeks.org/clustering-in-machine-learning/
[Accessed 2024 September].

Geeks for Geeks, 2024. *K means Clustering – Introduction*. [Online]
Available at: https://www.geeksforgeeks.org/k-means-clustering-introduction/
[Accessed September 2024].

Geeks for Geeks, 2024. *Linear Regression in Machine learning*. [Online]
Available at: https://www.geeksforgeeks.org/ml-linear-regression/
[Accessed September 2024].

Geeks for Geeks, 2024. *Logistic Regression in Machine Learning*. [Online]
Available at: https://www.geeksforgeeks.org/understanding-logistic-regression/
[Accessed September 2024].

Geeks for Geeks, 2024. *What is LSTM – Long Short Term Memory?*. [Online]
Available at: https://www.geeksforgeeks.org/deep-learning-introduction-to-long-short-term-memory/
[Accessed September 2024].

Majumder, S. L. a. B. P., n.d. *Food.com Recipes and Interactions*. [Online]
Available at: https://www.kaggle.com/datasets/shuyangli94/food-com-recipes-and-user-interactions
[Accessed September 2024].

paperswithcode, n.d. *Traffic Prediction*. [Online]
Available at: https://paperswithcode.com/task/traffic-prediction
[Accessed September 2024].

Shikha, 2021. *Computer Vision to Detect License Number Plate*. [Online]
Available at: https://www.analyticsvidhya.com/blog/2021/12/computer-vision-to-detect-license-number-plate/
[Accessed September 2024].

Tecperson, n.d. *Sign Language MNIST*. [Online]
Available at: https://www.kaggle.com/datasets/datamunge/sign-language-mnist
[Accessed September 2024].

World Bank Group, n.d. [Online]
Available at: https://data.worldbank.org/indicator/SP.POP.TOTL
[Accessed September 2024].