ACADEMIC CITY UNIVERSITY COLLEGE
(Affiliated to University of Mines & Technology, Tarkwa and University of Cape Coast, Ghana)
CS4232/CE4232 : Machine Learning
END of SEMESTER EXAMINATION - 2023 / 2024

Question Paper

Allocated Time: 10 days | Maximum mark: 100 marks
Examination Date: March, 2024

Answer One (1) Question.

For your final year examination in Machine Learning, you are required to select **one of the following project topics** and develop a comprehensive solution incorporating the methodologies, technologies, and criteria specified. Your project will demonstrate your ability to apply data science and machine learning concepts and tools to solve real-world problems.

Question 1: Customer Segmentation for Marketing Strategies

Objective:

Develop a dashboard that segments a company's customers into distinct groups based on purchasing behaviour using Unsupervised Learning algorithms (K-means Clustering and Hierarchical Clustering). This segmentation can then be used to tailor marketing strategies.

Data:

A customer dataset from online retail or a similar domain, with attributes like purchase history, demographic information, browsing behaviour, etc.

Technologies:

a) **Model Training:** Apply clustering techniques such as *KMeans or Hierarchical Clustering* to segment customers.

b) **Analysis Tool:** Develop an interactive dashboard using *Streamlit* to visualise and analyse customer segments.

c) **Deployment:** Use *Docker* to containerise the dashboard for deployment.

d) **Project Management:** Manage the project workflow with *ZenML* and track experiments with *MLflow*.

Evaluation Criteria:

a) Effectiveness in identifying meaningful customer segments.

b) Application of dimensionality reduction techniques to visualise high-dimensional data.

c) Interactive and informative dashboard design.

d) Use of Docker for containerised deployment.

e) Workflow management with ZenML.

Question 2: News Article Categorization System

Objective:

Develop a system that automatically categorises news articles into predefined topics using Ensemble Machine Learning techniques.

Data:

A text dataset of news articles available on Kaggle or UCI, labelled categories like sports, politics, technology, etc.

Technologies:

a) **Model training:** Python, scikit-learn, NLTK or spaCy for NLP tasks.

b) **Deployment:** Flask or Streamlit for the web interface, Docker for containerisation, MLflow for tracking experiments.

c) **Project management:** Git, GitHub, ZenML.

Evaluation Criteria:

a) Accuracy of the categorisation model.

b) Use of text preprocessing and feature extraction techniques.

c) Efficient implementation of ensemble methods.

d) Deployment of a user-friendly web interface.

e) Experiment tracking with MLflow and project orchestration with ZenML.

Question 3: Sales Forecasting using Time Series Analysis

Objective:

Implement a system that forecasts future sales for a retail company using time series analysis and Neural Networks (RNN, LSTM).

Data Source:

A dataset containing historical sales data, possibly with additional features like holidays, promotions, and economic indicators, available on platforms like Kaggle or UCI.

Technologies:

a) **Model training:** Python, TensorFlow PyTorch, and pandas for data manipulation.

b) **Deployment:** Use Flask or Streamlit for creating a forecasting interface, Docker for deployment, and MLflow for model tracking.

c) **Project management:** Git, GitHub, and ZenML for workflow management.

Evaluation Criteria:

a) Accuracy and reliability of the forecast.

b) Application of RNN or LSTM networks for time series prediction.

c) Implementation of a user interface for displaying forecasts.

d) Use of Docker and MLflow for deployment and tracking.

e) Efficient workflow management with ZenML.

Good Luck
Ishaya, Jeremiah ayock