UNIVERSITY OF WOLVERHAMPTON

HERALD COLLEGE KATHMANDU | ing GROUP

5CS037

Concepts and Technologies of AI

Final Assignment – Predicting University Rankings and Scores Using Machine Learning

Name: Lam Pasang Lama

Student ID: 2408830

Lecturer: Siman Giri

Tutor: Bibek Khanal

UNIVERSITY OF WOLVERHAMPTON

Regression Analysis Report

Abstract

Purpose: This study aims to develop a **regression model** to predict university scores.

Approach: The research uses the **2024 QS World University Rankings** dataset. The methodology includes **Exploratory Data Analysis (EDA), training regression models (Linear Regression and Decision Tree Regression), hyper parameter tuning, and feature selection**.

Key Results: The model evaluation relies on **R^2 and Mean Squared Error (MSE)**. **Decision Tree Regression outperformed Linear Regression in prediction accuracy**.

Conclusion: Feature selection and hyper parameter tuning enhanced model performance, with **Decision Tree Regression providing the best results**.

---

1. Introduction

1.1 Problem Statement

The objective is to **predict the overall university score** using various ranking metrics.

1.2 Dataset

The dataset used is the **2024 QS World University Rankings**, containing **numerical metrics** that affect a university's ranking.

1.3 Objective

The goal is to develop a **predictive regression model** to estimate university scores.

---

2. Methodology

2.1 Data Preprocessing

- Removed **irrelevant columns**.
- Handled **missing values and outliers**.

- Standardized **numerical features**.

2.2 Exploratory Data Analysis (EDA)

- **Scatter plots and histograms** were used to analyse data distribution.
- **Correlation analysis** identified important relationships between features.

2.3 Model Building

- **Models Used:**
  - Linear Regression
  - Decision Tree Regression

2.4 Model Evaluation

- **Evaluation Metrics:**
  - **R^2 Score**
  - **Mean Squared Error (MSE)**

2.5 Hyper parameter Optimization

- **GridSearchCV** was used to fine-tune model parameters.

2.6 Feature Selection

- **Recursive Feature Elimination (RFE)** identified the most relevant features.

---

3. Conclusion

3.1 Key Findings

- **Decision Tree Regression** outperformed **Linear Regression** in predictive accuracy.
- **Feature selection improved interpretability**.

3.2 Final Model

- **Decision Tree Regression** was the best-performing model.

3.3 Challenges

- **Handling multicollinearity** in the dataset.

3.4 Future Work

- Implement **advanced regression models** such as **Random Forest Regression**.
- Test **feature engineering techniques**.

---

4. Discussion

4.1 Model Performance

- **Decision Tree Regression produced the most accurate predictions**.

4.2 Impact of Hyper parameter Tuning and Feature Selection

- **Hyper parameter tuning improved accuracy**, and **feature selection simplified the model**.

4.3 Limitations

- The dataset had a **limited size**, which could impact generalizability.

4.4 Future Research Suggestions

- **Expanding dataset scope**.
- **Experimenting with deep learning models**.