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

Classification Analysis Report

Abstract

Purpose: This study applies classification models to predict university ranking categories.

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

Key Results: The model evaluation relies on **accuracy, precision, recall, and F1-score**. Among all models, the **Random Forest Classifier achieved the highest accuracy**.

Conclusion: The classification models identified key factors influencing university rankings. **Hyper parameter tuning and feature selection improved model performance**.

1. Introduction

1.1 Problem Statement

The goal of this project is to classify universities into different ranking categories (**Top 100, 100-500, 500+**).

1.2 Dataset

The dataset used is the **2024 QS World University Rankings**. It includes various metrics such as **academic reputation, employer reputation, faculty-student ratio, and research impact**.

1.3 Objective

The primary objective is to develop a **classification model** to categorize universities based on ranking.

2. Methodology

2.1 Data Pre-processing

- Removed **missing values**.
- Encoded **categorical variables**.
- Standardized **numerical features**.

2.2 Exploratory Data Analysis (EDA)

- **Correlation heat map** to identify relationships between features.
- **Class distribution visualization**.

2.3 Model Building

- **Models Used:**
  - Logistic Regression
  - Decision Tree
  - Random Forest

2.4 Model Evaluation

- **Evaluation Metrics:**
  - **Accuracy**
  - **Precision**
  - **Recall**
  - **F1-score**

2.5 Hyper parameter Optimization

- **GridSearchCV** was used to optimize model performance.

2.6 Feature Selection

- **Recursive Feature Elimination (RFE)** was applied to select the most relevant features.

UNIVERSITY OF WOLVERHAMPTON

3. Conclusion

3.1 Key Findings

- **Random Forest** achieved the best classification accuracy.
- **Feature selection improved efficiency**.

3.2 Final Model

- The **Random Forest Classifier** was the most effective model after hyper parameter tuning.

3.3 Challenges

- **Complexity in feature selection**.

3.4 Future Work

- Experiment with **advanced ensemble models**.
- Test **additional feature engineering techniques**.

4. Discussion

4.1 Model Performance

- **Random Forest outperformed other classification models**.

4.2 Impact of Hyper parameter Tuning and Feature Selection

- **Tuning improved model accuracy**, and **feature selection reduced complexity** while maintaining performance.

4.3 Limitations

- The dataset had **missing values** that required pre-processing.

4.4 Future Research Suggestions

- **Expanding dataset scope**.
- **Implementing deep learning techniques**.


UNIVERSITY OF
WOLVERHAMPTON