Data Science Principles with Applications on Educational Data GROUP PROJECT PROPOSAL

Exploring Key Factors in Educational Data: A Case Study on University Rankings

Name: 張育丞,   張宇青,   周姿妤,   黃詳諺,   Duong Van Nhat Quang

Sid: 113598043,   113598032,   113598090,   113598091,   113998411

Date: 04/16/2025

National Taipei University of Technology
Computer Science and Information Engineering

OUTLINE

01
**Introduction
(About Dataset)**

02
**Introduction
(Motivation)**

03
**Introduction
(Objectives)**

04
Dataset Preview

05
Related Work

06
**Data Processing
and Analysis Flow**

07
Expected Results

06
Timeline

Introduction - About Dataset

- World University Ranking - focuses on university rankings across different regions.

- Datasets are sourced from Times Higher Education, Academic Ranking of World Universities, and Center for World University Rankings.

- Data collected during 2014–2015 for consistency.

Introduction - Motivation

- Global university rankings significantly impact higher education policies and strategies.

- Understanding key indicators is crucial for academic institutions aiming to enhance competitiveness and influence.

- Educational data provide a rich foundation for exploring patterns and proposing actionable insights.

Introduction - Objectives

1. Identify and analyze key educational indicators influencing global university rankings.

2. Integrate multiple features (citations, patents, publications) to explore combined effects.

3. Visualize relationships among indicators to enhance interpretability.

4. Provide data-driven analysis to assist in understanding university competitiveness.

Dataset Preview

- Dataset from Kaggle: "World University Ranking."

- Selected 14 headers for analysis:

  (List in two columns)

- World Rank
- Institution
- Country
- Quality of Faculty
- Influence
- Broad Impact
- Score

- National Rank
- Quality of Education
- Alumni Employment
- Publications
- Citations
- Patents
- Year

Related Work - Research Topics & Methods

Topic	Research Question	Method
Citations vs. World Rank	Does more citations lead to better rank?	Correlation + Linear Regression
Alumni Employment vs. Score	Does better employment lead to higher score?	Correlation + Regression
Key Indicators Analysis	Which indicator influences rank most?	Relation Weight Analysis
Country Impact	Do specific countries rank higher at same score?	Group Analysis
Patents vs. Academic Impact	Which type of university ranks higher?	Comparative Analysis
University Type Classification	Classify by citations and patents	Clustering

Data Processing and Analysis Flow

1. Data Preprocessing

   ○ Clean missing or inconsistent data.

   ○ Normalize features for comparability.

2. Exploratory Data Analysis (EDA)

   ○ Generate correlation matrix.

   ○ Visualize feature distributions and relationships.

3. Feature Importance Analysis

   ○ Apply Decision Trees and Random Forests.

   ○ Conduct Principal Component Analysis (PCA).

4. Predictive Modeling

   ○ Build regression models (Linear Regression, Decision Tree Regression).

5. Visualization

   ○ Correlation matrix, feature importance, regression results.

Expected Results

- Identify key factors significantly impacting university rankings.

- Build predictive models based on educational indicators.

- Visualize insights into relationships among indicators.

- Provide actionable suggestions for academic institutions.

Timeline

Week	Task
10	Familiarize with dataset, literature review, initial data cleaning.
11	EDA and correlation matrix visualization.
12	Feature importance analysis (Decision Trees, Random Forests).
13	Predictive modeling (Regression Analysis).
14	Visualization and results review.
15	Project review.
16	Project report and presentation preparation.
17	Final report submission and presentation.

Thank You

If you have any questions, please feel free to contact us by email.

t { 113598043, 113598032, 113598090, 113598091, 113998411 } @ntut.org.tw