

Enhancing Data Science Teaching through Competition-based Learning with Kaggle

Pengfei Fan, Nikesh Bajaj, Jesús Requena Carrión



In the rapidly evolving landscape of Data Science education, the Competition-based Learning approach has emerged as a dynamic pedagogical strategy. This research delves into the efficacy of utilising Kaggle, a renowned data science competition platform, to enhance teaching quality, student engagement, and learning outcomes. By synthesising theoretical frameworks and empirical insights, the study aims to unravel the impact of this innovative approach. Through a comprehensive exploration of the alignment with literature, meticulous research questions, a mixed-methods methodology, and insightful results, this project seeks to contribute valuable perspectives to the evolving realm of data-driven education.

Objective

- 1) Investigate and optimise the processes of dataset preparation and model development for baseline and sample solutions within Kaggle competitions.
- 2) Create an innovative GPTenabled tool that can objectively assess students' performance in Kaggle competitions.
- 3) Employ a mixed-methods evaluation approach to comprehensively assess the impact of competition-based learning on teaching quality, student engagement, and learning outcomes.
- 4) Improve teaching practices in Data Science through the integration of competition-based learning strategies.

Research Questions

- 1) How does the competition-based learning approach using the Kaggle platform enhance student engagement and motivation in the context of Data Science?
- 2) What are the learning outcomes and skill development facilitated by the competition-based learning approach in teaching Data Science?
- 3) How does the competition-based learning approach align with relevant literature, models for learning, and best practices in the field of Data Science?
- 4) How does the competition-based learning approach address the transnational learning and teaching context in the field of Data Science?
- 5) To what extent does the competition-based learning approach contribute to pedagogic development in teaching Data Science?

Methods & Results

- A. Building the Kaggle Competition
 - 1) Dataset Preparation
 - 2) Model Development

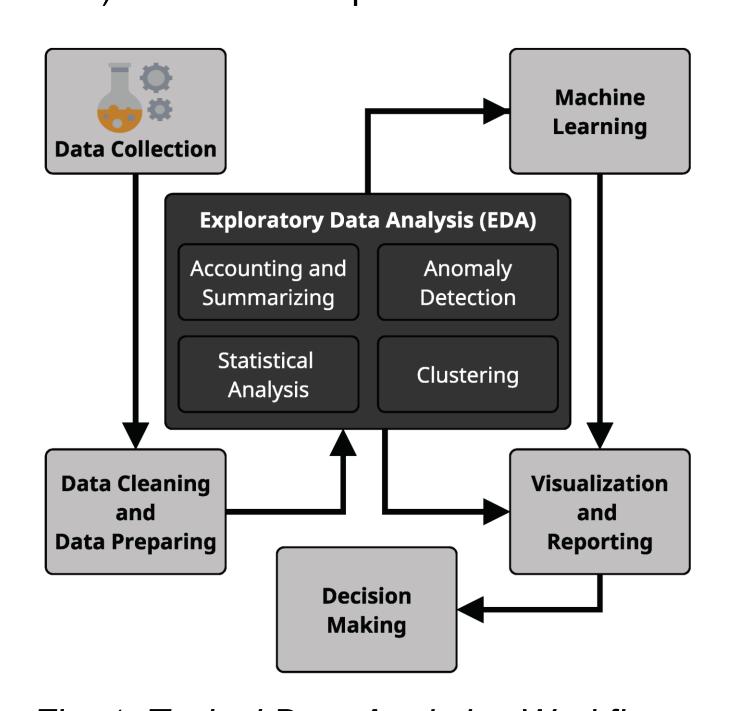


Fig. 1. Typical Data Analytics Workflow.

- B. Mixed-Methods Research Design
 - 1) Qualitative Method
 - 2) Quantitative Method

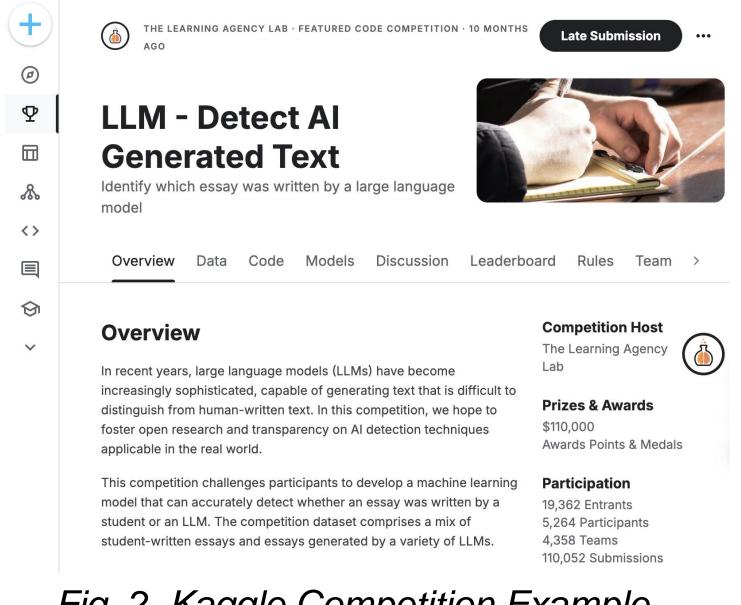


Fig. 2. Kaggle Competition Example.

To what extent do you believe the skills acquired are transferable to real-world scenarios?

Not transferable at all Neutral Highly transferable

Moderately transferable

Fig. 3. Survey Question Example.

C. GPT-enabled Assessment Tool

Code
Score
Entries Record
Leaderboard
Ranking
Report
Responses to
Survey
...
Personalised
Assessment

Fig. 4. GPT-enabled Assessment Generation

Conclusions

This project transforms Data
Science education through Kagglebased competition learning,
preparing students for real-world
challenges. It evaluates student
engagement, learning outcomes,
and quality using a mixed-methods
approach and introduces a GPT tool
for objective assessment,
addressing the limitations of
traditional subjective evaluations.