Intelligent Decision Support System

Practical work 3

January 8, 2024

“Educational Success Predictor”

Group 2

Gustavo Rayo

Umberto Salviati

Giacomo Sanguin

UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

Abstract

A general idea:

- In this project, an Intelligent Decision Support System has been crafted to forecast student academic outcomes by amalgamating traditional features with expert opinions.
- The IDSS is designed as a web application, in which the user can predict whether a student is going to graduate or not
- The IDSS features customizable models, accommodating users with a choice between basic and advanced configurations. An approach that aspires to empower educational institutions with an effective tool for tailored student support and strategic academic planning.

Domain of application

- **Educational Sphere:** Our IDSS operates at the heart of education, focusing on predicting and optimizing student academic outcomes.
- **Data Driven Foundation:** Leveraging a rich dataset, including demographic info, academic history, and socio-economic factors, our IDSS employs predictive modeling for insights into students' academic trajectories.
- **Expert Opinion Integration:** incorporation of an expert opinion component, simulating a college interview scenario.
- **Stakeholders: t**argeting academic advisors, administrators, and educators, our IDSS offers a proactive approach to identify at-risk students early.

Main identified decisions

- Grant or deny an application
- Decide special guidance to accepted students more likely to drop out
- Grant scholarship
- Improve learning experience

Functional architecture of the IDSS prototype

Components:

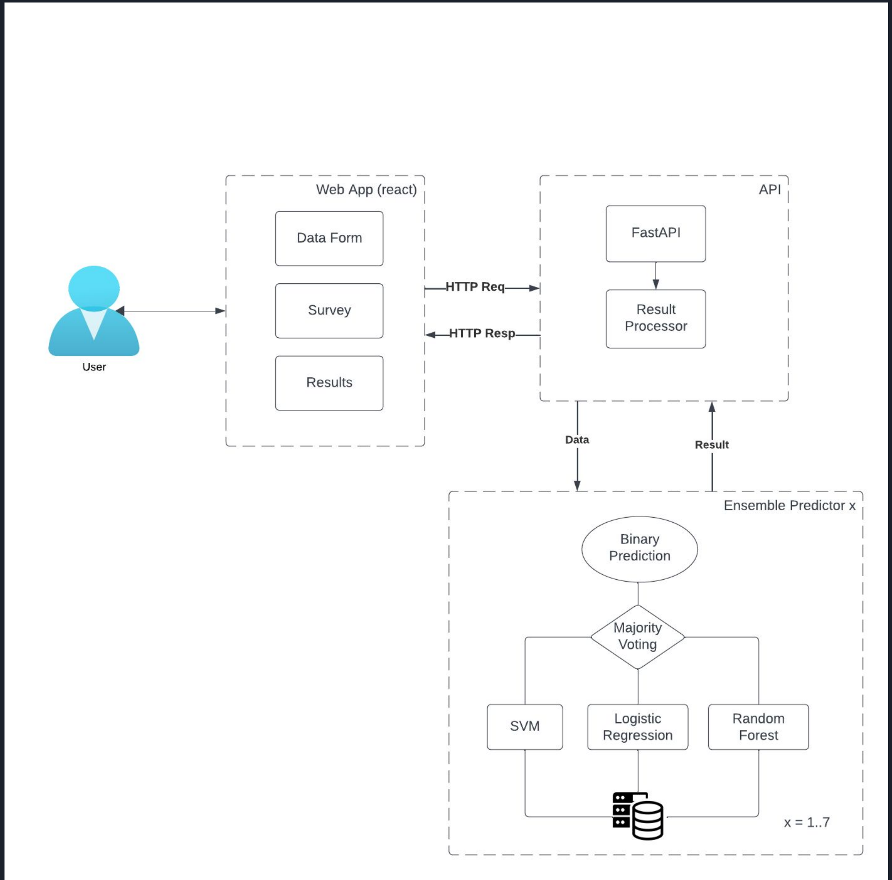
- Web Application

  React, Materia-kit-react
- API

  FastAPI
- Ensemble predictor

  SMV, Logistic Regression,

  Random Forest

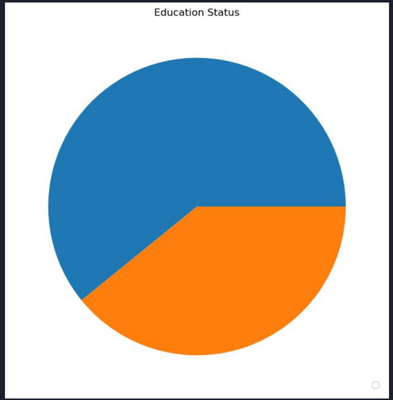


Data pre-processing

- Thorough Analysis with Python:
  - Used Matplotlib and Seaborn for comprehensive data visualization
- Data Filtering:
  - Focused on certainty by excluding students with unclear outcomes
  - Target variable: "graduate" or "dropout" status.
- Observed significant imbalance - more graduates than dropouts.
  - Explored various balancing techniques.
- SMOTE Oversampling:
  - outperform Kaggle past analysis
  - Achieved a 50% balance between graduates and dropouts, provides a fairer basis for analysis.

Outcome:

- Improved model performance through effective data pre-processing.
- Balanced dataset enhances predictive accuracy.



Data-driven IDSS model

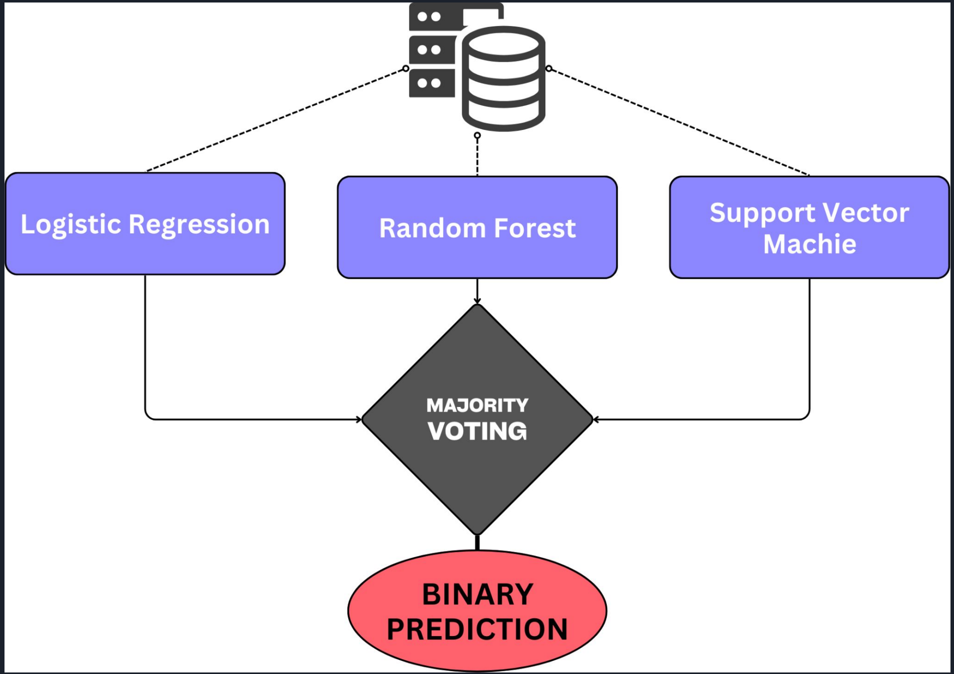
Model Exploration:

- Experimented with various algorithms for data-driven models:
    - Random Forest, Logistic Regression, SVM (multiple kernels), KNN,linear discriminant analysis and Gaussian Naive Bayes

Model Selection:

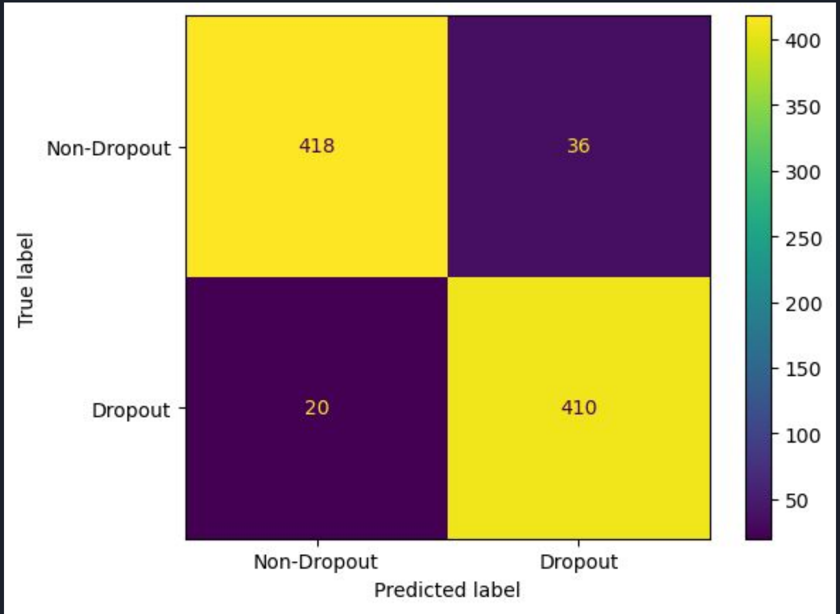
- Chose Three Models:
    - Random Forest
    - Logistic Regression
    - SVM

Ensemble Learning



Validation

- Model validation crucial for IDSS functionality.
  - Collaboration between data-driven models and expert-based component.
- Accuracy Scores:
  - Simple features: 0.654
  - Basic features: 0.673
  - Basic with academic features: 0.67
  - Basic with financial features: 0.76
  - Basic with full academic: 0.913
  - Basic with full academic and financial: 0.924
  - Full data model: 0.937



Post-prediction and Evaluation

Model Output Treatment:

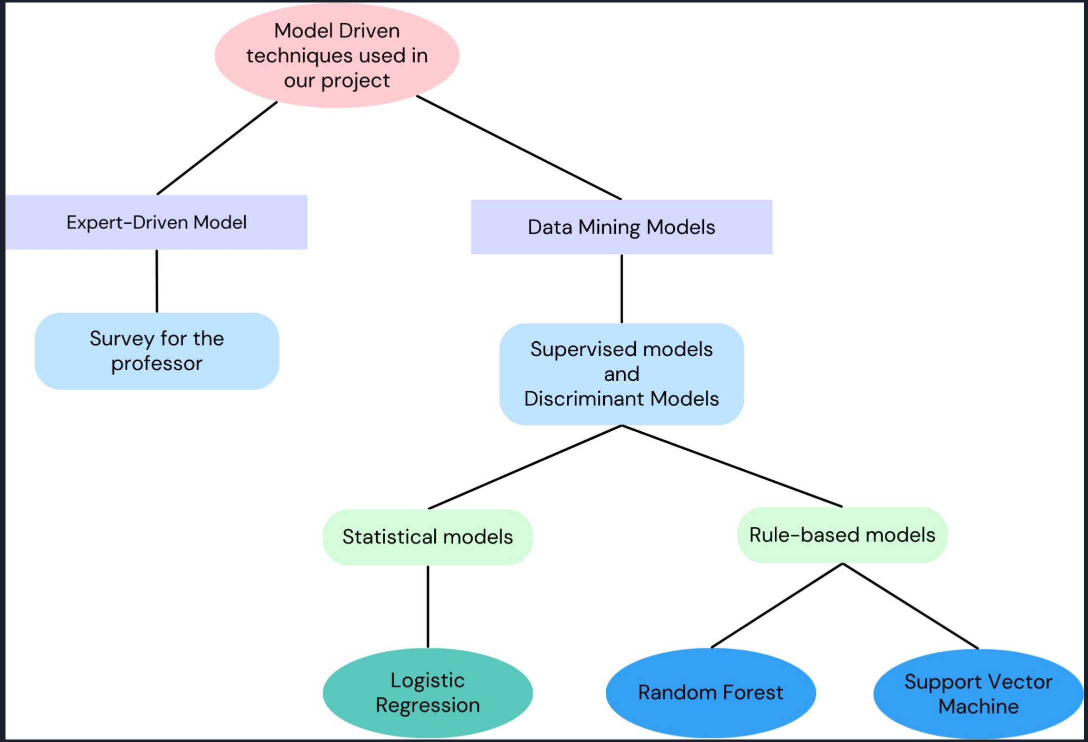
- Data-driven model outputs not independently treated.
- Weighted differently based on the accuracy of individual models.

Comparison with Kaggle previous analysis:

- Kaggle project focused on logistic regression, achieving around 92% accuracy.
- Our project, incorporating various models, slightly outperforms with a full model accuracy just below 94%.

Model-driven IDSS techniques

- Expert Opinion
- Data Mining Models
- Statistical models
- Rule Based Models
- Logistic Regression



Future work and improvements

- **Multi-Expert Evaluation**: Allow the selection of a student and evaluation by multiple experts.
- **User Feedback Mechanism:** Incorporation of a feedback method from the users where they could report issues in the predictions and areas of improvement.
- **Explainability Features:** Incorporate tools for explainability of the results. It could be implemented using techniques such as LIME or SHAP.

Gantt diagram and tasks planning

Educational Success Predictor

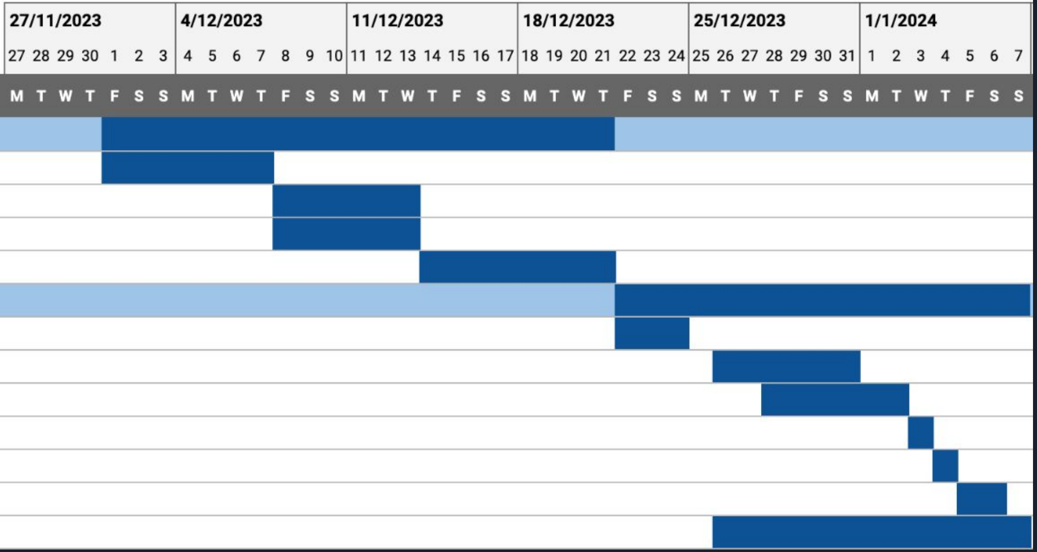
Group 2

Gustavo Rayo, Umberto Salviati, Giacomo Sanguin

Project Start: 1/12/2023

Display Week: 1

TASK	ASSIGNED	PROGRESS	START	DAYS	END
Phase 1		100%	**1/12/2023**		**21/12/2023**
Dataset Selection		100%	1/12/2023	7	7/12/2023
Defining Scope and Objectives		100%	8/12/2023	6	13/12/2023
Dataset Analysis		100%	8/12/2023	6	13/12/2023
Web App Framework Selection		100%	14/12/2023	8	21/12/2023
Phase 2		100%	**22/12/2023**		**7/1/2024**
Data Preparation		100%	22/12/2023	3	24/12/2023
Model Development		100%	26/12/2023	6	31/12/2023
Application development		100%	28/12/2023	6	2/1/2024
Expert opinion Integration		100%	3/1/2024	1	3/1/2024
Customization Options		100%	4/1/2024	1	4/1/2024
Testing and Deployment		100%	5/1/2024	2	6/1/2024
Report Documentation		100%	26/12/2023	13	7/1/2024



Thanks for the attention