Student Success Prediction

TAA - Tópicos de Aprendizagem Automática 24/25

André Alves 113962
Bruno Tavares 113372
Francisco Pinto 113763

Table of contents

01 **Overview**
Introduction and Reflexion of the Project

02 **Objectives**
Goals and State of Art

03 **Model Selection**
Chosen models and evaluation

04 **Results and Analysis**
Comparison of results between models

05 **Conclusion**
Final Thoughts

01 Overview

This project represents our first significant experience with machine learning, focusing on the prediction of academic outcomes for higher education students.

Overview

Dataset: 4,425 Portuguese students with 35 attributes (demographic, socioeconomic, and academic)

Goal: to early identify students at risk of dropping out of school.

Data Classes

- 49.92% (2209) Graduated
- 32.11% (1421) Dropout
- 17.94% (794) Enrolled

02

Objectives

Goals and State of Art

Objectives

Detection

Transform reactive approaches into proactive ones – early detection of academic risk

ML Analysis

Exploratory analysis, comparison of ML algorithms, identification of influential factors

State of art

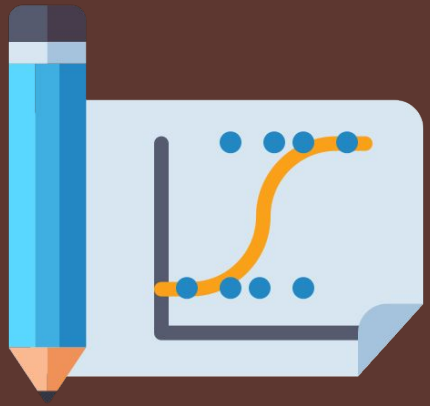
Evolution from statistical methods to deep learning (78–91% accuracy)

03

Model Selection

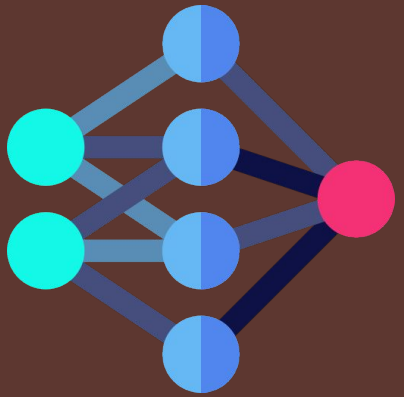
Chosen models and evaluation

Model Selection



Logistic Regression

Random Forest



Neural Network

metrics: precision, recall, F1-score and accuracy

04

Results And Analysis

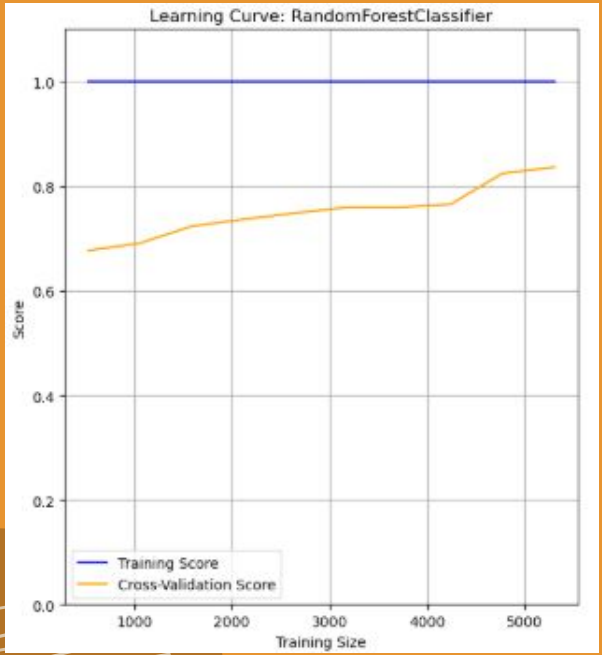
Comparison of results between models

Results (Logistic Regression)

Metric	Train	Test
Accuracy	0.7764	0.7597
F1 Score	0.7764	0.7609
Precision	0.7648	
Recall	0.7597	

- Consistency
- No overfitting

Results (Random Forest)



Metric	Train	Test
Accuracy	1.0000	0.8250
F1 Score	1.0000	0.8254
Precision	0.8287	
Recall	0.8250	

- Best absolut performance
- Overfitting

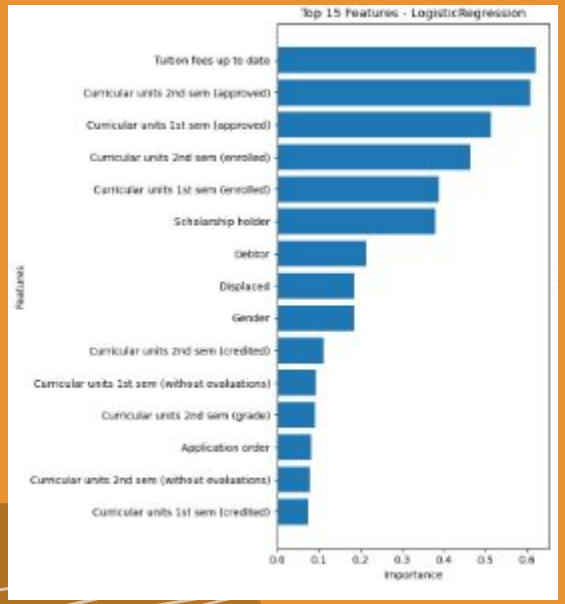
Results (Neural Networks)

Metric	Train	Test
Accuracy	0.9871	0.7697
F1 Score	0.9871	0.7698
Precision	0.7700	
Recall	0.7697	

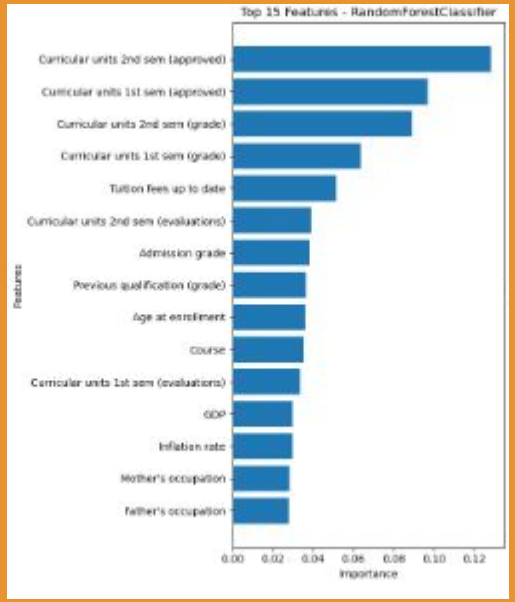
- Mid performance
- Overfitting

TOP 15 FEATURES

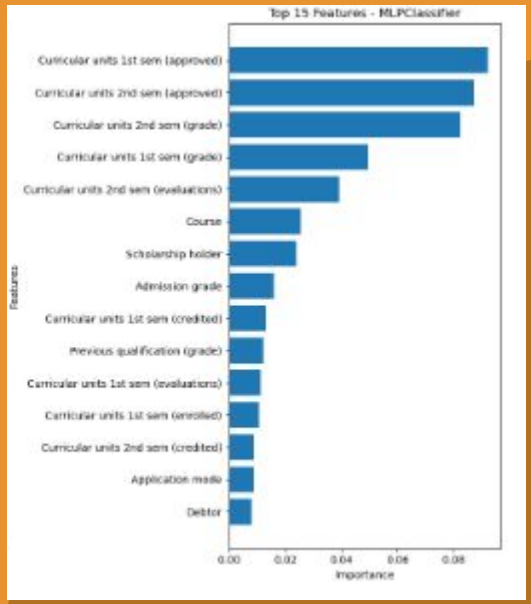
LOGISTIC REGRESSION



RANDOM FOREST



NEURAL NETWORKS



Results (Top 10 Features)

Metric	Train	Test
Accuracy	0.7559	0.7446
F1 Score	0.7554	0.7445
Precision	0.7448	
Recall	0.7446	

Logistic Regression

Metric	Train	Test
Accuracy	0.7930	0.7300
F1 Score	0.7933	0.7317
Precision	0.7384	
Recall	0.7300	

Neural Networks

Metric	Train	Test
Accuracy	1.0000	0.7964
F1 Score	1.0000	0.7975
Precision	0.8017	
Recall	0.7964	

Random Forest

Results (Tuning Hyper-parameters)

Hyperparameter	Values Tested
solver	['lbfgs', 'saga']
max_iter	[5000]
C	[0.01, 0.1, 1, 10]
class_weight	[None, 'balanced']
penalty	['l2']

Metric	Train	Test
Accuracy	0.7768	0.7617
F1 Score	0.7768	0.7628
Precision	0.7664	
Recall	0.7617	

Logistic Regression

Hyperparameter	Values Tested
n_estimators	[100, 150]
max_depth	[10, 20, None]
min_samples_split	[2, 5]
min_samples_leaf	[1, 2, 4]

Metric	Train	Test
Accuracy	1.0000	0.8321
F1 Score	1.0000	0.8324
Precision	0.8353	
Recall	0.8321	

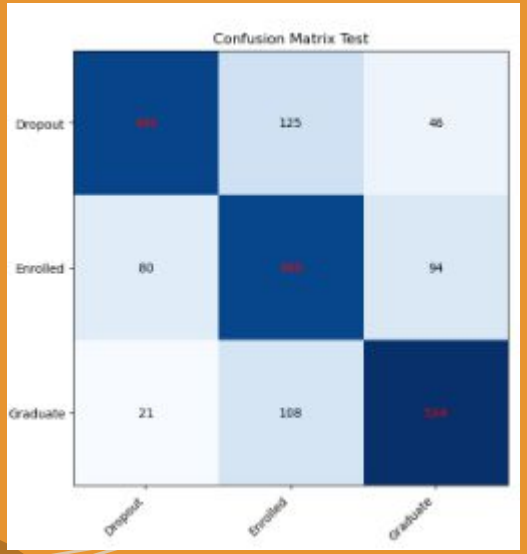
Random Forest

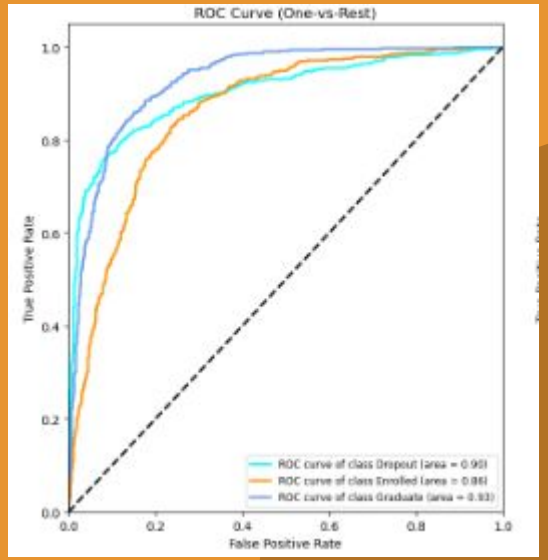
Hyperparameter	Values Tested
hidden_layer_sizes	[(50,), (100,), (50,30)]
activation	['relu', 'tanh']
alpha	[0.0001, 0.001]
learning_rate_init	[0.001, 0.01]

Metric	Train	Test
Accuracy	0.9646	0.7853
F1 Score	0.9646	0.7857
Precision	0.7865	
Recall	0.7853	

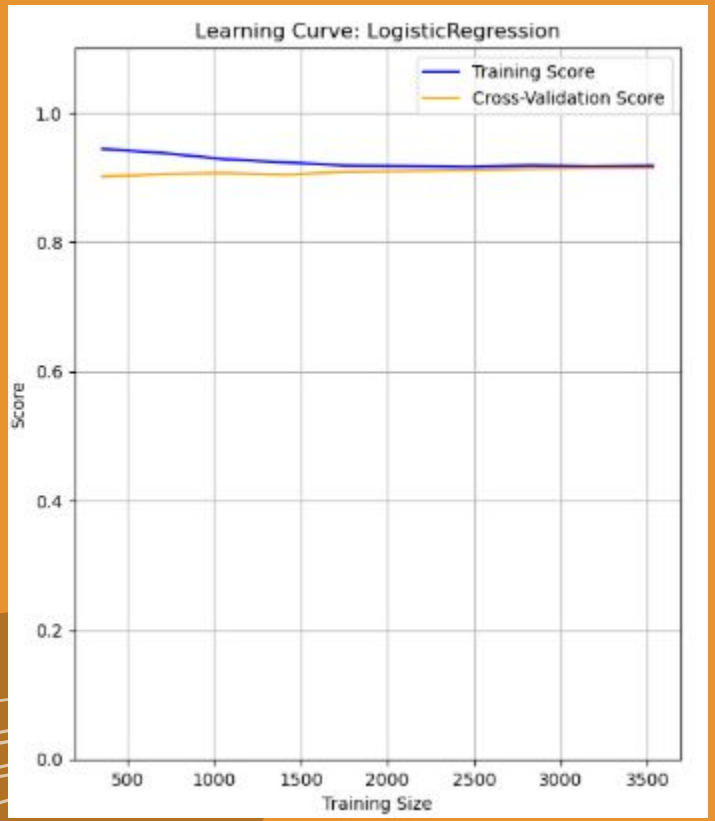
Neural Networks

Results (Tuning Hyper-parameters)





Results (Binary System)



Metric	Logistic Regression	Random Forest	MLP Classifier
Train			
Accuracy	0.9196	1.0000	1.0000
F1 Score	0.9196	1.0000	1.0000
Test			
Accuracy	0.9095	0.9061	0.8959
F1 Score	0.9094	0.9059	0.8959
Precision	0.9117	0.9090	0.8962
Recall	0.9095	0.9061	0.8959

05

Conclusion