

افنان الحجيلي

Smart Financial Payments System %28SFPS%29 afnan %282%29 %282%29.docx

 Midocean University

Document Details

Submission ID

trn:oid::2945:381712635

Submission Date

Apr 30, 2026, 7:15 PM GMT+3

Download Date

May 5, 2026, 11:17 AM GMT+3

File Name

Smart Financial Payments System (SFPS) afnan (2) (2).docx

File Size

1.9 MB

48 Pages

4,183 Words

24,970 Characters

21% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Match Groups

-  **72 Not Cited or Quoted 20%**
Matches with neither in-text citation nor quotation marks
-  **0 Missing Quotations 0%**
Matches that are still very similar to source material
-  **5 Missing Citation 1%**
Matches that have quotation marks, but no in-text citation
-  **0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 13%  Internet sources
- 14%  Publications
- 18%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

No suspicious text manipulations found.

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

Match Groups

- 72 Not Cited or Quoted 20%**
Matches with neither in-text citation nor quotation marks
- 0 Missing Quotations 0%**
Matches that are still very similar to source material
- 5 Missing Citation 1%**
Matches that have quotation marks, but no in-text citation
- 0 Cited and Quoted 0%**
Matches with in-text citation present, but no quotation marks

Top Sources

- 13% Internet sources
- 14% Publications
- 18% Submitted works (Student Papers)

Top Sources

The sources with the highest number of matches within the submission. Overlapping sources will not be displayed.

1	Internet	
	www.mdpi.com	1%
2	Submitted works	
	Midocean University on 2025-09-05	1%
3	Submitted works	
	University of Portsmouth on 2024-04-28	1%
4	Submitted works	
	Sheffield Hallam University on 2026-01-15	<1%
5	Internet	
	fastercapital.com	<1%
6	Submitted works	
	Middlesex University on 2026-01-01	<1%
7	Submitted works	
	University of Hertfordshire on 2025-01-05	<1%
8	Submitted works	
	University of Hertfordshire on 2026-03-02	<1%
9	Internet	
	bspa.ce.buid.ac.ae	<1%
10	Internet	
	open.metu.edu.tr	<1%

11	Internet	propulsiontechjournal.com	<1%
12	Submitted works	City University Malaysia on 2026-04-21	<1%
13	Publication	K. V. Sambasivarao, Anasuya Sesha Roopa Devi Bhima. "Artificial Intelligence, Co...	<1%
14	Submitted works	Universiti Malaysia Perlis on 2026-01-09	<1%
15	Submitted works	University of Hertfordshire on 2025-01-06	<1%
16	Internet	ir-library.ku.ac.ke	<1%
17	Submitted works	Carnegie Mellon University on 2023-11-06	<1%
18	Submitted works	Islington College,Nepal on 2025-12-17	<1%
19	Publication	Shiva Shankar Reddy, Mahesh Gadiraju, N. Meghana Preethi, V.V.R.Maheswara Ra...	<1%
20	Internet	blog.so8848.com	<1%
21	Internet	core.ac.uk	<1%
22	Internet	qspace.library.queensu.ca	<1%
23	Internet	www.coursehero.com	<1%
24	Submitted works	University of Bristol on 2013-05-20	<1%

25	Submitted works	University of Essex on 2026-04-24	<1%
26	Submitted works	University of Warwick on 2022-09-04	<1%
27	Internet	instadatahelp.com	<1%
28	Publication	Alrawili, Reem Farhan. "User Authentication Framework: Leveraging Vital Signs f...	<1%
29	Publication	B. G. Nagaraja, S. Kannadhasan. "Information and Communication Systems", CRC...	<1%
30	Internet	bmcpsy psychiatry.biomedcentral.com	<1%
31	Internet	cdr.lib.unc.edu	<1%
32	Internet	eprints.utar.edu.my	<1%
33	Internet	esource.dbs.ie	<1%
34	Internet	repository.aaup.edu	<1%
35	Internet	scholar.colorado.edu	<1%
36	Submitted works	Aston University on 2026-01-30	<1%
37	Publication	Chilomo, Innocent. "Using Machine Learning to Detect Customs Tariff Classificati...	<1%
38	Submitted works	Gisma University of Applied Sciences GmbH on 2026-03-27	<1%

39	Submitted works	University of West London on 2024-09-16	<1%
40	Submitted works	University of York on 2025-07-21	<1%
41	Internet	s3.eu-central-1.amazonaws.com	<1%
42	Submitted works	University of Bedfordshire on 2024-03-08	<1%
43	Internet	atrium.lib.uoguelph.ca	<1%
44	Internet	eprints.utm.my	<1%
45	Internet	stax.strath.ac.uk	<1%
46	Internet	viterbi-web.usc.edu	<1%
47	Submitted works	De LaSalle - College of Saint Benilde on 2025-12-05	<1%
48	Publication	Ervin Sejdić, Tiago H. Falk. "Signal Processing and Machine Learning for Biomedic...	<1%
49	Publication	Farah, A. "Assessment of FLUENT CFD Code as an Analysis Tool for SCW Applicatio...	<1%
50	Publication	Fernandes, Eduardo Rocha. "Integrar Inteligência Artiçial Para Melhorar a Capac...	<1%
51	Submitted works	Sheffield Hallam University on 2026-04-29	<1%
52	Submitted works	Staffordshire University on 2024-05-13	<1%

53	Internet	ijecs.in	<1%
54	Internet	ir.lib.nycu.edu.tw	<1%
55	Internet	scholar.sun.ac.za	<1%
56	Internet	spectrum.library.concordia.ca	<1%
57	Publication	Pinto, Gonalo Filipe Lucas Menino Rodrigues. "Sistema Inteligente de Previso e ...	<1%



Project Proposal

Smart Financial Payments System (SFPS)

Prepared By

Name: AFNAN ALHUJAILI

ID No.: 251002575

Submitted in fulfillment of the requirements for the degree Master of Science
in Artificial Intelligence

At

Collage of Informatics

Midocean University

UAE

Supervisor Name: Dr Khaled Eskaf

Collage of Informatics, Midocean University, Comoros

Email: AFNANNASSEERR@gmail.com

Acknowledgements

9 I would like to thank Dr. Khalid for his ongoing support, and I would like to thank my family—my parents, my sister, and my brothers and sisters—for all the support they have given me throughout my studies, both directly and indirectly. I would also like to thank my cats, Nooni, Lucky, and Lulu, for all the warm, tender, and unconditional love they have shown me.

Abstract

 40 In recent times, the concept of artificial intelligence and machine learning has evolved into many different approaches, helping us to develop and learn significantly and equipping us with information that makes our lives easier. Today, artificial intelligence exists to improve and simplify all aspects of life; for example, the financial system designed to assist students with payments, The study we are currently working on includes 1,200 students and their databases

making it easier for us to manage all financial obligations and the traditional processes involved in monthly tuition payments.

جدول المحتويات

1

	List of Abbreviations.....	7
23	List of Figures	Error! Bookmark not defined.
	List of Tables	Error! Bookmark not defined.
3	Chapter 1: Introduction.....	9
	1.1 Background and Context	9
	1.2 Problem Definition	10
	1.3 Research Motivation	10
	1.4 Research Objectives	11
	1.5 Research Contributions	11
	1.6 Scope and Limitations	11
44	1.7 Significance of the Study	12
	1.8 Thesis Organization	Error! Bookmark not defined.
22	Chapter 2: Background and Related Work.....	14
	2.1 Introduction	14
	2.2 Artificial Intelligence in Financial Analytics.....	14
	2.3 Supervised Learning and Classification Techniques	14
18	2.3.1 Logistic Regression	15
	2.3.2 Random Forest	16
	2.3.3 Gradient Boosting	17
	2.4 Data Preprocessing in Financial Datasets	18
38	2.5 Feature Engineering for Financial Behavior Analysis	19
	2.6 Model Evaluation Metrics	19
	2.7 Data Leakage in Machine Learning Models.....	20
	2.8 Machine Learning Tools and Implementation	20

4

3

2.9 Related Work	21
------------------------	----

2.10 Research Gap	22
-------------------------	----

Chapter 3: Proposed Approach.....	23
--	-----------

3.1 Introduction	23
------------------------	----

3.2 System Architecture	23
-------------------------------	----

8

3.3 Dataset Description	24
-------------------------------	----

3.4 Data Preprocessing	25
------------------------------	----

3.5 Feature Engineering.....	25
------------------------------	----

3.6 Rule-Based System Implementation	26
--	----

39

3.7 Model Training.....	26
-------------------------	----

3.8 Model Evaluation.....	27
---------------------------	----

3.9 System Workflow	27
---------------------------	----

21

3.10 System Output	28
--------------------------	----

Chapter 4: Experimental Results and Discussion	29
---	-----------

4.1 Introduction	29
------------------------	----

4.2 Experimental Environment and Setup.....	30
---	----

4.3 Evaluation Metrics and Performance Criteria.....	30
--	----

4.3.1 Primary Results	30
-----------------------------	----

4.4.2 Improved Results.....	31
-----------------------------	----

4.5 Interpretation of Results	41
-------------------------------------	----

4.6 Impact of Rule-Based System on Model Behavior	41
---	----

4.7 Model Comparison and Analysis.....	42
--	----

4.8 Critical Analysis and Discussion	42
--	----

4.9 Limitations and Generalization.....	43
---	----

54

4.10 Practical Implications.....	43
----------------------------------	----

Chapter 5: Conclusion and Future Work	44
--	-----------

16

5.1 Conclusion.....	44
---------------------	----

5.2 Key Findings	44
------------------------	----

5.3 Limitations of the Study.....	45
-----------------------------------	----

5.4 <i>Practical Implications</i>	45
5.5 <i>Future Work</i>	46
References	46

List of Abbreviations

- Acc.: Accuracy
- AI: Artificial Intelligence
- AUC: Area Under the Curve
- CV: Cross-Validation
- F1: F1-Score
- FN: False Negative
- FP: False Positive
- GB: Gradient Boosting
- GS: Grid Search
- LR: Logistic Regression
- ML: Machine Learning
- Prec.: Precision
- Rec.: Recall
- RF: Random Forest
- ROC: Receiver Operating Characteristic
- SFPS: Smart Financial Payments System
- TN: True Negative
- TP: True Positive

List of Figures

	Figure 2.2 Logistic Regression.....	19
	Figure 2.3 Random Forest	20
	Figure 2.4 Gradient Boosting	22
	Figure 2.5 Machine Learning Tools and Implementation	27
10	Figure 3.1 System architecture of the proposed Smart Financial Payments System	32
	Figure 3.2 Structure of the dataset showing the main components	34
29	Figure 3.3 Workflow of the proposed system	39
4	Figure 4.1 Accuracy comparison of models in the initial results	44
6	Figure 4.2 ROC curve for Logistic Regression (commitment prediction).....	44
	Figure 4.3 ROC curve for Random Forest (commitment prediction)	44
19	Figure 4.4 ROC curve for Gradient Boosting (commitment prediction).	44
6	Figure 4.5 ROC curve for Logistic Regression (exemption prediction).....	45
	Figure 4.6 ROC curve for Random Forest (exemption prediction)	45
19	Figure 4.7 ROC curve for Gradient Boosting (exemption prediction)	45
7	Figure 4.8 Confusion matrix for Random Forest (commitment prediction).....	46
	Figure 4.9 Confusion matrix for Gradient Boosting (exemption prediction)	47
28	Figure 4.10 Feature importance for exemption prediction using Gradient Boosting	48



Chapter 1: Introduction

1. Background and Context

Colleges, universities, and even schools rely on a traditional financial system for students, offering loans and allowing them to be paid back in manageable installments to accommodate students' circumstances—given the differences in socioeconomic backgrounds and financial situations among students—especially when students are unable to make payments. For this reason, traditional systems are often impractical; therefore, technological systems help and encourage students to commit

to monthly installments through rewards they can earn and discounts for meeting their obligations. Some traditional systems can become very complex and difficult to use; therefore, developing a technological system that minimizes human interaction makes the repayment process easier for both students and the institution.

1.2 Problem Definition

The problem we are facing is the lack of an integrated financial system that serves students, the administration, the finance department, and accountants; one that tracks students and their payment patterns, determines whether they have met their obligations, and identifies any delinquencies in the payment schedule. Everything shows that payments are due between the 25th of the current month and the 3rd of the following month, at which point a discount is applied. 30% discount on the installment. If the payment deadline is missed, the installment reverts to the base price. If the student pays all installments on time, they receive a waiver on the 29th and 30th installments. The current traditional system does not track any payment commitments or waivers; these are applied entirely manually.

1.3 Research Motivation

The most important decision behind conducting this study is to develop financial payment systems that minimize human error and facilitate access to payments, refunds, and eligibility for discounts at the click of a button—without the need to wait for human interaction or delayed responses. Most importantly, this will help us understand students' payment patterns and assess their overall financial situation.

1.4 Research Objectives

The main goal is to develop a comprehensive, smart financial payment system. I decided to make it a hybrid system that combines rule-based financial logic with predictive modeling techniques to analyze students' installment payment behavior, integrate these elements, and train the model to achieve the best possible results.

1.5 Research Contributions

This research contributes to the field of financial programming and data-driven decision-making. It proposes a hybrid system that integrates logic with predictive techniques and addresses the shortcomings of rule-based systems. The study incorporates the expected financial policies resulting from the system's design and explores how to transform installment-level data into characteristics that effectively represent repayment and improves the model and its interpretation, and most importantly,

integrates operational systems with analytical capabilities to advance the management of intelligent financial systems

1.6 Scope and Limitations

In this study, I focused on analyzing student repayment behavior using a dataset of 1,200 student records containing detailed information on the loan and installment amounts. The study addresses two main aspects: first, the application of financial rules regarding eligibility for discounts and exemptions; and second, predicting student behavior, although there are some variations due to missed payment deadlines.

1.7 Significance of the Study

Its importance lies in improving the financial payment system by integrating machine learning and artificial intelligence, thereby reducing human error—particularly in financial matters and when dealing with complex data—especially since it enables management to stay informed about the financial outlook regarding student payments and their compliance.

1.8 thesis organization

This thesis consists of five chapters, each of which addresses a specific issue.

8 Chapter 1 Research Background, Problem Definition, Objectives, and Contributions

Chapter 2 Theoretical Background, Review of Related Literature, and Data Analysis

Chapter 3 Proposed methodology, raw data processing, and feature engineering

50 Chapter 4 Results and detailed model performance

Chapter 5 Discussion and conclusion

Chapter 2: Background and Related Work

2.1 Introduction

Now I will present to you a selection of previous studies that I have reviewed and found to be relevant to my research, particularly those concerning financial forecasting, classification issues, and the most significant studies on model improvement.

2.2 Artificial Intelligence in Financial Analytics

The key benefits of applying artificial intelligence to financial analysis include data enhancement and improved decision-making. Machine learning is used to analyze large volumes of financial data, identify patterns and customer behavior, detect fraud, and assess credit risks. We frequently use models to estimate the likelihood of loan defaults or repayment behavior. When we use artificial intelligence, it helps us predict the future.

2.3 Supervised Learning and Classification Techniques

Supervised learning is one of the most widely used approaches in the field of machine learning, and classification techniques are extensively used in many fields, including financial analysis, where they are used to predict outcomes such as payment behavior or eligibility decisions.

Common classification models include logistic regression, random forests, and gradient boosting, each with different strengths depending on the nature of the data. For example, simpler models like logistic regression are easier to interpret but may struggle with complex patterns, while ensemble methods like random forests and gradient boosting are better at detecting nonlinear relationships. The performance of these models depends not only on the algorithm itself but also on the quality of the data and the selected features.

2.3.1 Logistic Regression

Logistic regression is one of the most important techniques in the study of ML

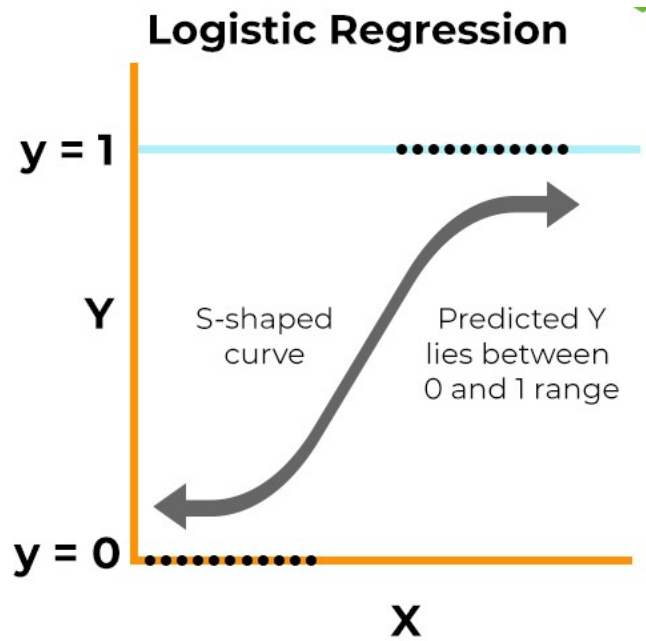


Figure 2.2 Logistic Regression

2.3.2 Random Forest

The Random Forest is a large-scale ensemble learning method for classification tasks

Random Forest

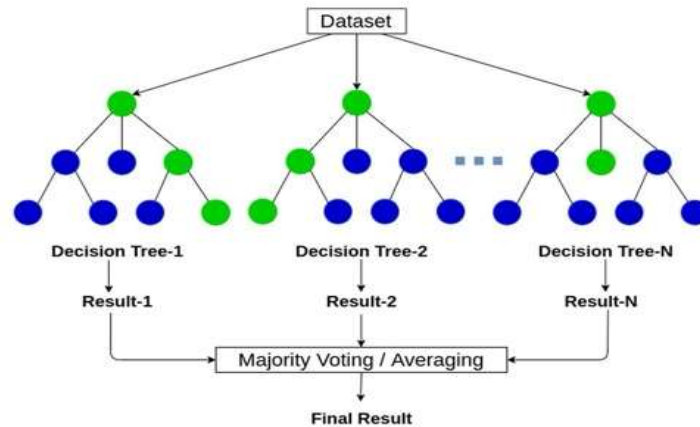


Figure 2.3 Random Forest

2.3.3 Gradient Boosting

Gradient Boosting is also a collective approach that sequentially combines weak models to form a strong predictive model.

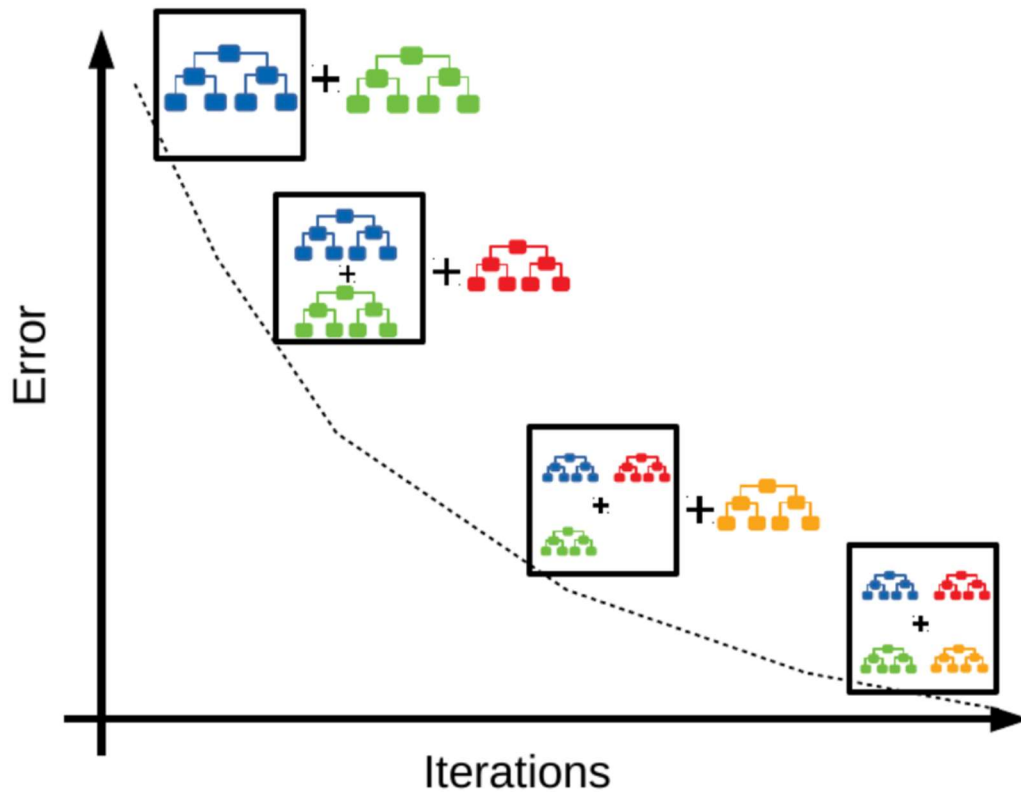


Figure 2.4 Gradient Boosting

2.4 Data Preprocessing in Financial Datasets

The most important data we have is the raw data, which is always incomplete and contains missing or incorrect values. It is now very important to clean the data, verify the quality of the input data, and ensure proper processing so that we don't end up with incorrect results.

2.5 Feature Engineering for Financial Behavior Analysis

Feature engineering plays a crucial role in improving the performance of machine learning models, better representing user behavior, and developing new features from existing data—such as aggregating transaction values, calculating specific events, and extracting temporal patterns—often focusing on payment history, frequency, and timing in groups that help identify relationships that are not apparent in the raw data.

2.6 Model Evaluation Metrics

To evaluate the performance of machine learning models—which are among the most fundamental models—the following common methodological metrics are used: accuracy, precision, recall, F1 score, and ROC curve. These metrics generally measure the model's validity. Relying on a single metric is not sufficient; therefore, we use a combination of metrics together.

2.7 Data Leakage in Machine Learning Models

Data leakage in financial forecasting tasks occurs when you obtain variables from information that reflects future outcomes or is linked to a target change; this may work but fails when generalized to new data.

2.8 Machine Learning Tools and Implementation

Model development relies heavily on machine learning techniques, and the use of these techniques greatly facilitates the development process and the creation of structured, repeatable workflows in the field of machine learning.

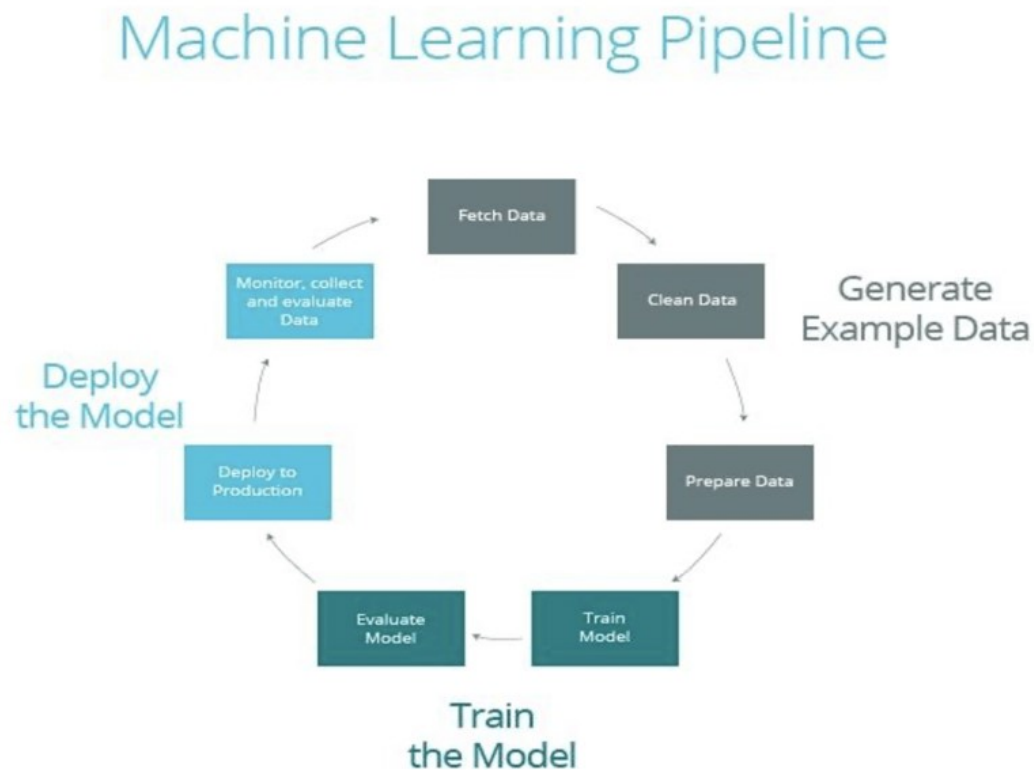


Figure 2.5 Machine Learning Tools and Implementation

2.9 Related Work

With a focus on improving decision-making and forecasting, many studies have applied and explored machine learning techniques to analyze and forecast financial behavior. In previous studies, **Random Forest and Gradient Boosting were used** for these tasks **to classify customers**, forecast loan repayments and student installment payments, and assess credit risk.

2.10 Research Gap

Despite the abundance of research on financial forecasting and behavioral analysis, there are only a few studies that have focused specifically on forecasting student financial behavior, particularly with regard to repayment compliance and eligibility for exemption. General financial studies have focused on credit risks, forecasting default on installments, and other related topics; however, learning methods have not been integrated with the systems. These methods have been applied to predefined groups, and some studies have yielded excellent results regarding the issue of data leakage and student repayment behavior after adjusting the set of features to ensure reliable predictions.

Chapter 3: Proposed Approach

3.1 Introduction

This chapter explains, step by step, the development of a smart financial payment system. The core concept involves integrating financial logic with machine learning models to analyze student loan repayment behavior and support better decision-making.

3.2 System Architecture

This section presents the architecture of the smart financial payments system, which is designed to analyze student loan repayment behavior and generate future predictions based on rules and machine learning models. It begins by inputting raw data, which then undergoes several processing stages until the final prediction output is produced. The first step involves receiving student financial data containing the loan amount, installment amounts, payment records, and other details. The data is then sent through preprocessing, where it is cleaned, formatted, and prepared. Next, features are engineered to extract relevant indicators from the raw data. These features are designed to capture important aspects of student repayment behavior, such as payment frequency, as this helps improve the models' ability to learn relevant patterns

The overall architecture of the system is shown in Figure 3.1.

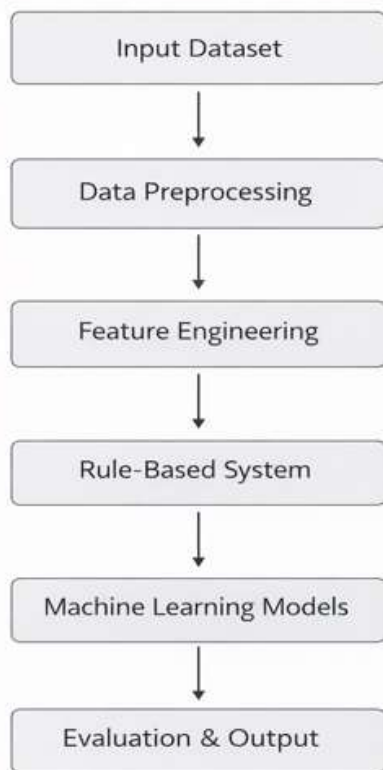


Figure 3.1 System architecture of the proposed Smart Financial Payments System

3.3 Dataset Description

The dataset consists of 1,200 student records representing installment payment behavior within the educational financial system. Each record includes basic information such as the total loan amount, the number of installments, and the value of each installment, as well as detailed data.

The dataset also includes information related to repayment compliance and eligibility for financial incentives such as tax and installment exemptions.

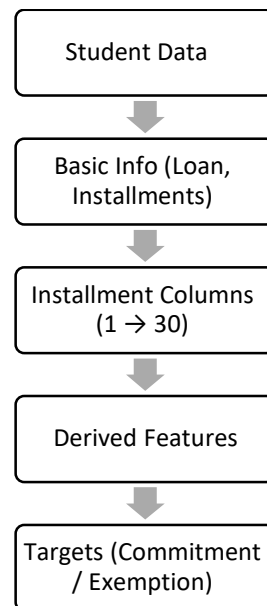


Figure 3.2 Structure of the dataset showing the main components

3.4 Data Preprocessing

Here, the dataset is prepared, cleaned, and preprocessed for machine learning modeling. This includes handling missing values, correcting data, and converting numerical fields—such as loan amounts, installment values, and total payments—into a numerical format to ensure accurate calculations.

3.5 Feature Engineering

I have incorporated feature engineering here to improve data quality and help the models better understand students' repayment behavior, rather than relying solely on the original variables. I created features based on financial data and payment history. These features help reflect the actual circumstances of students' repayment and financial well-being.

3.6 Rule-Based System Implementation

These rules are provided here to serve as a basis for analyzing student behavior and the conditions related to eligibility, payment, and installment exemptions, as they are among the most important guidelines we rely on. We have used simple criteria to verify the data and assign appropriate values.

3.7 Model Training

I used the prepared dataset to train a set of machine learning models to predict student payment behavior. I divided the data into training and

testing sets to evaluate the models on previously unseen data, then selected several models mentioned earlier because they offer different approaches. At this stage, the models learn patterns from the dataset and adapt to them.

3.8 Model Evaluation

Here, model performance gets checked through standard classification measures, showing how effectively each one works. What matters most shows up in accuracy, precision, recall, and the F1 score. Each importance distinct sides of prediction quality, like correct classifications or handling varied categories. Beyond that, separation power across classes draws from ROC AUC values. Relying on several indicators instead of just one gives a fuller, fairer picture of results. Because the testing dataset was used, performance metrics show how each model handles new cases. One after another, these outcomes help tell which approach predicts student repayment more consistently. By focusing here, differences among methods become clearer without relying on prior assumptions.

3.9 System Workflow

This section explains in detail the system's operational steps and how its components work, and illustrates how the system progresses step by step from the initial data to the final results in a systematic and clear manner.

The system workflow is shown in Figure 3.1

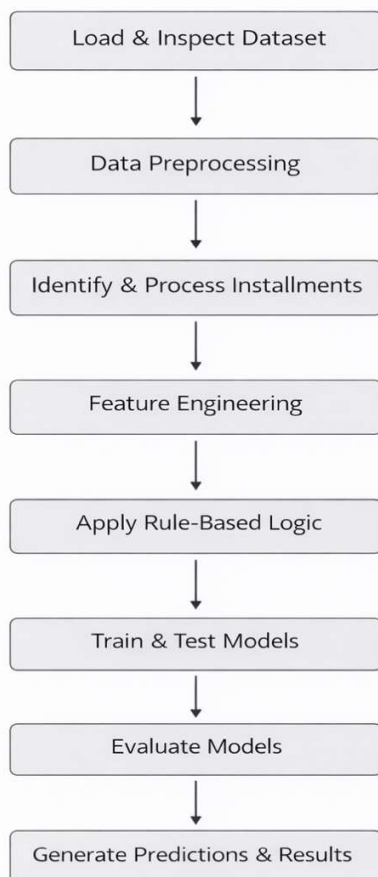


Figure 3.3 Workflow of the proposed system

3.10 System Output

Here, we address the final outputs of the system—the results generated after all models have been applied, machine learning has been performed, and model results have been obtained. Most importantly, the entire process runs within a single, user-friendly system.

Chapter 4: Experimental Results and Discussion

4.1 Introduction

This chapter presents the results of the system and the performance of machine learning models. The primary focus here is on predicting students' loan repayment compliance and their eligibility for installment exemptions. The evaluation was conducted using several models, and their performance was measured using various metrics to provide an understanding of the quality of performance across the models.

4.2 Experimental Environment and Setup

The system was developed using Python as the primary programming language, along with data analysis and machine learning libraries such as Pandas, NumPy, and Scikit-learn. The dataset consists of features and derived attributes that we prepared for training from student records.

4.3 Evaluation Metrics and Performance Criteria

This section presents the results of the machine learning models used in this study for both predictors—payment compliance and eligibility for exemption—and the results are divided into two parts: preliminary results, followed by improved results after feature engineering.

4.3.1 Primary Results

Here, I evaluated the performance of machine learning models using the original features, and the results were excellent across all metrics. The accuracy for logistic regression, random forest, and gradient boosting, as well as the precision, recall, F1 score, and ROC curve, all reached 1.0, which raised my concerns about the reliability of the data.

The initial results of the models are summarized in Figure 4.1

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Logistic Regression	1.00	1.00	1.00	1.00	1.00
Random Forest	1.00	1.00	1.00	1.00	1.00
Gradient Boosting	1.00	1.00	1.00	1.00	1.00

Figure 4.1 Accuracy comparison of models in the initial results

4.4.2 Improved Results

After reviewing the initial results, I had some concerns about their extremely high accuracy, so I refined the features and retrained the models using the updated dataset.

4.2 summarizes the refined model results.

All Targets Model Results

target	model	accuracy	precision	recall	f1	roc_auc
commitment_target	LogisticRegression	0.6833333333333333	0.6833333333333333	1.0	0.8118811881188119	0.5012034659820283
commitment_target	RandomForest	0.6833333333333333	0.6833333333333333	1.0	0.8118811881188119	0.516206675224647
commitment_target	GradientBoosting	0.6583333333333333	0.6752136752136753	0.9634146341463414	0.7939698492462312	0.488165917843389
exemption_target	GradientBoosting	0.8958333333333334	0.6575342465753424	1.0	0.7933884297520661	0.924587673611111
exemption_target	RandomForest	0.85	0.6111111111111112	0.6875	0.6470588235294118	0.9223090277777778
exemption_target	LogisticRegression	0.8	0.0	0.0	0.0	0.5545789930555557

ROC Curves

ROC Curves for Commitment

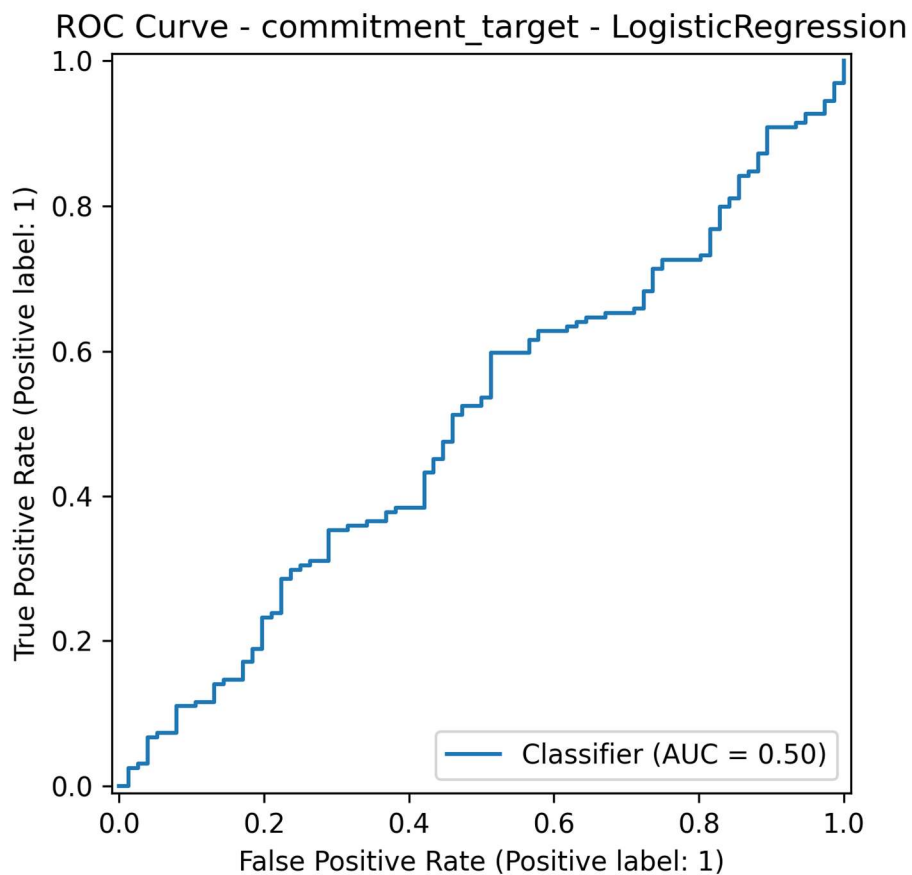


Figure 4.2: ROC curve for Logistic Regression (commitment prediction)

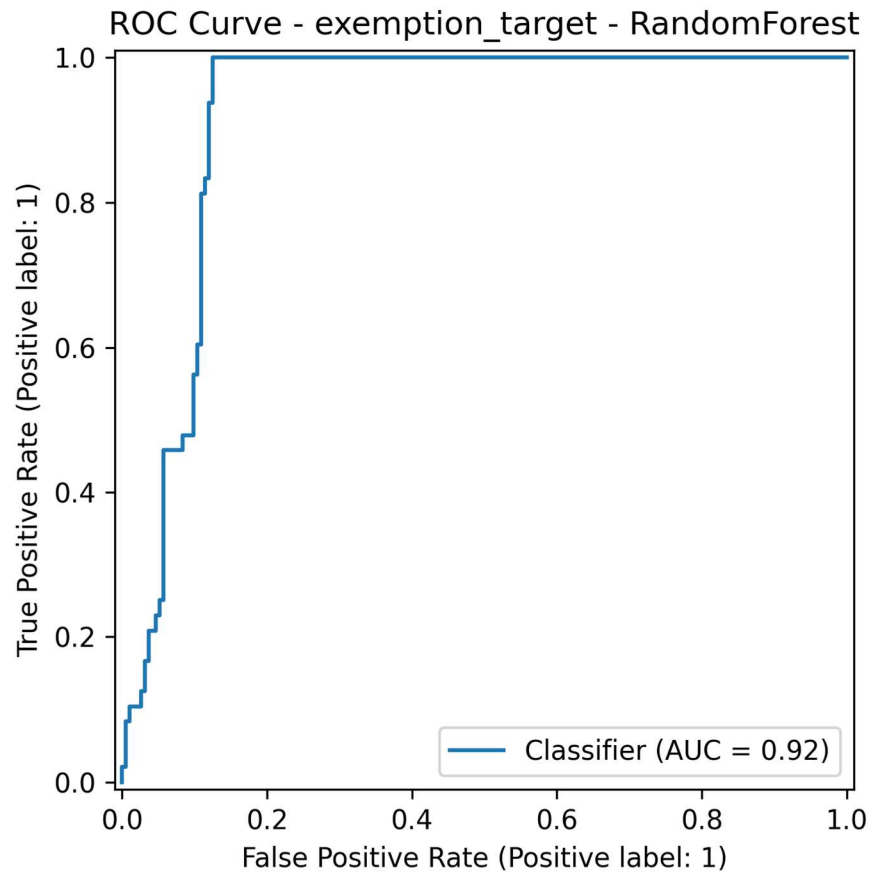


Figure 4.3: ROC curve for Random Forest (exemption prediction)

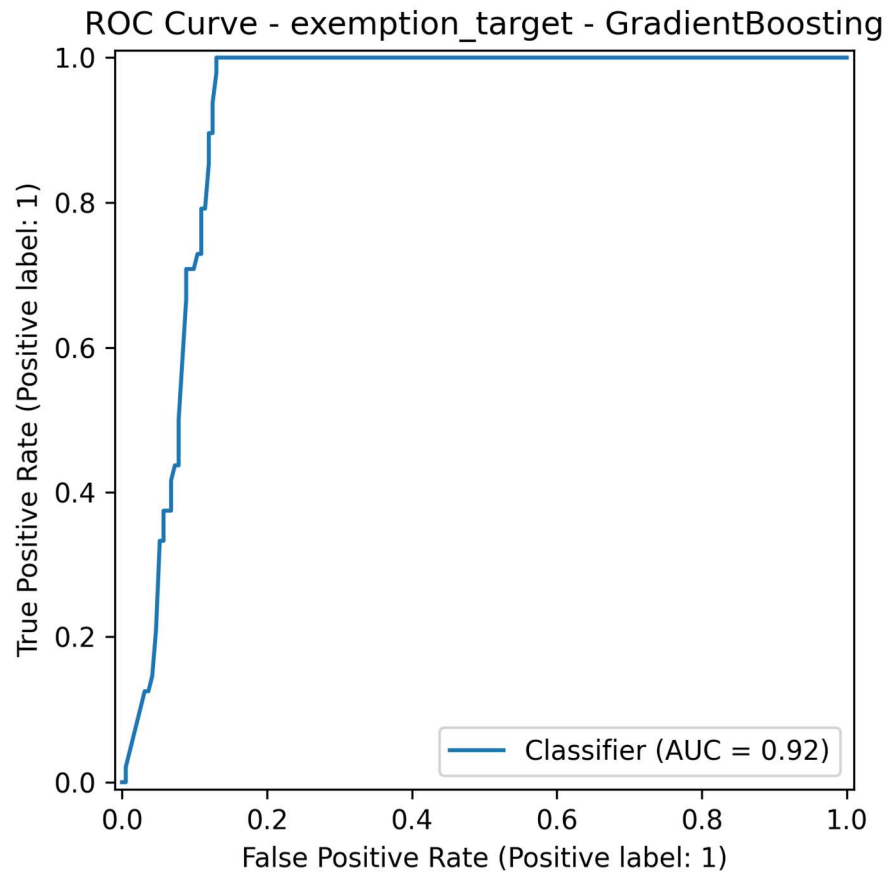


Figure 4.4: ROC curve for Gradient Boosting (exemption prediction)

ROC Curves for Exemption

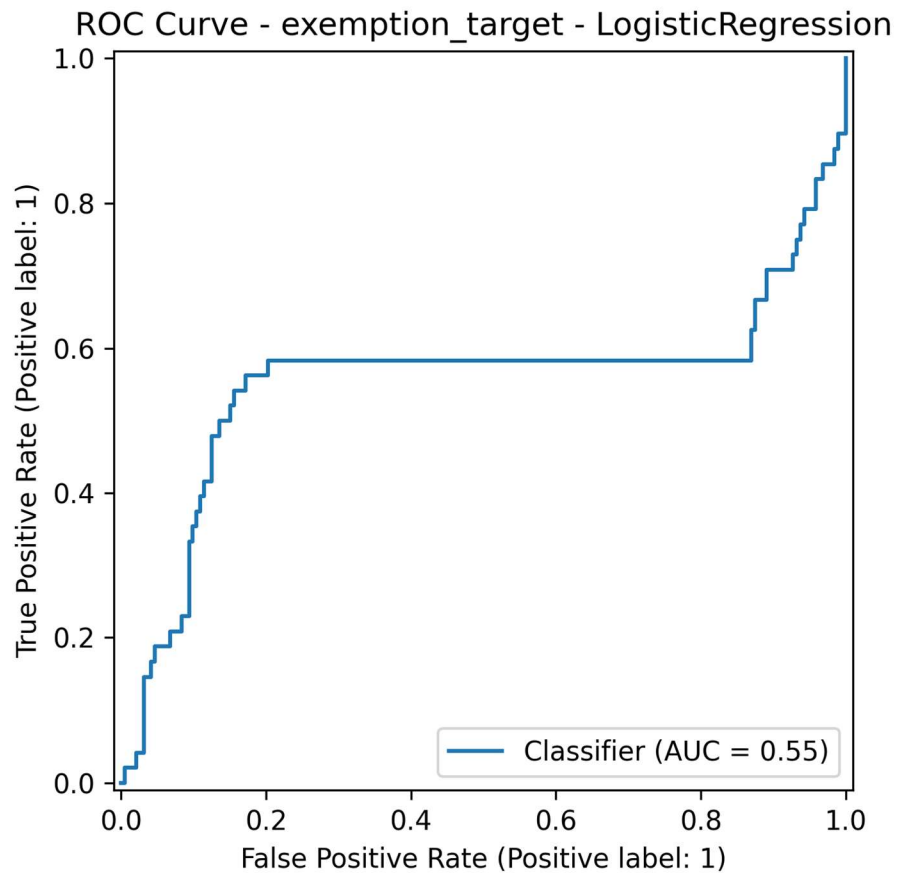


Figure 4.5: ROC curve for Logistic Regression (exemption prediction)

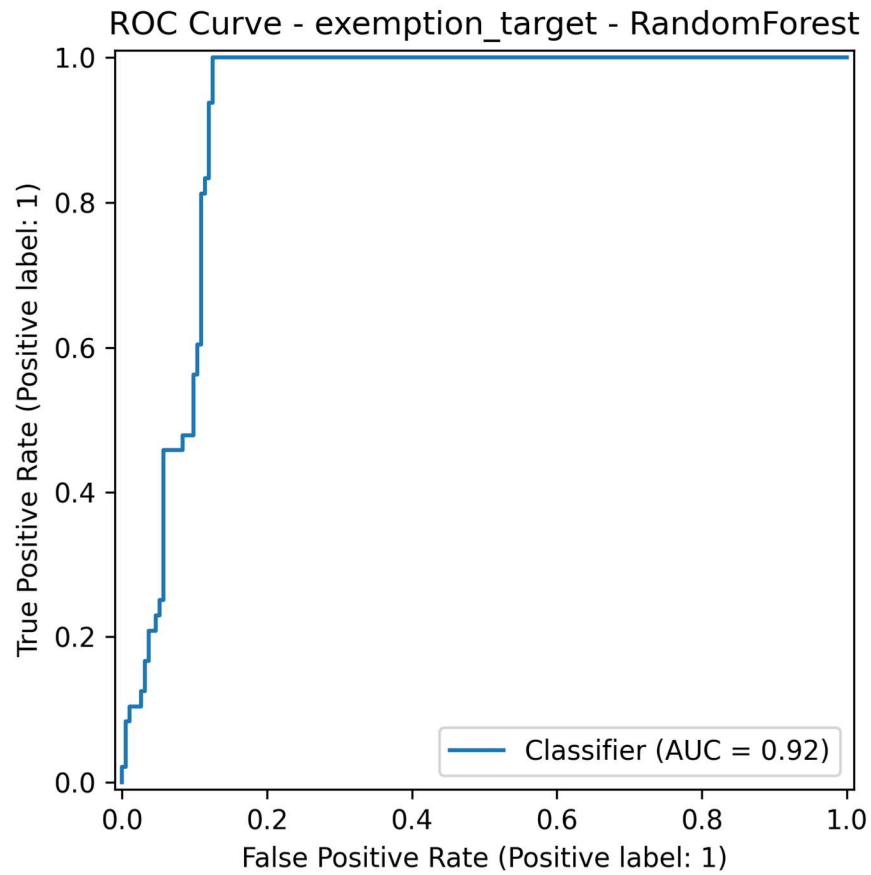


Figure 4.6: ROC curve for Random Forest (exemption prediction)

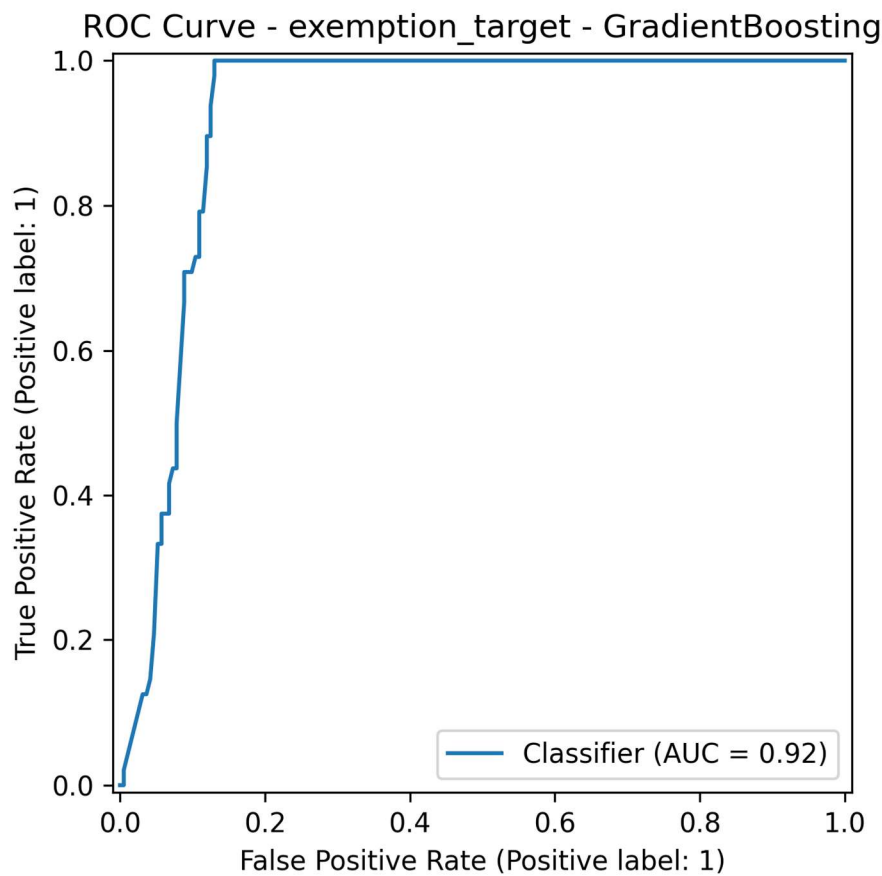


Figure 4.7: ROC curve for Gradient Boosting (exemption prediction)

Confusion Matrix

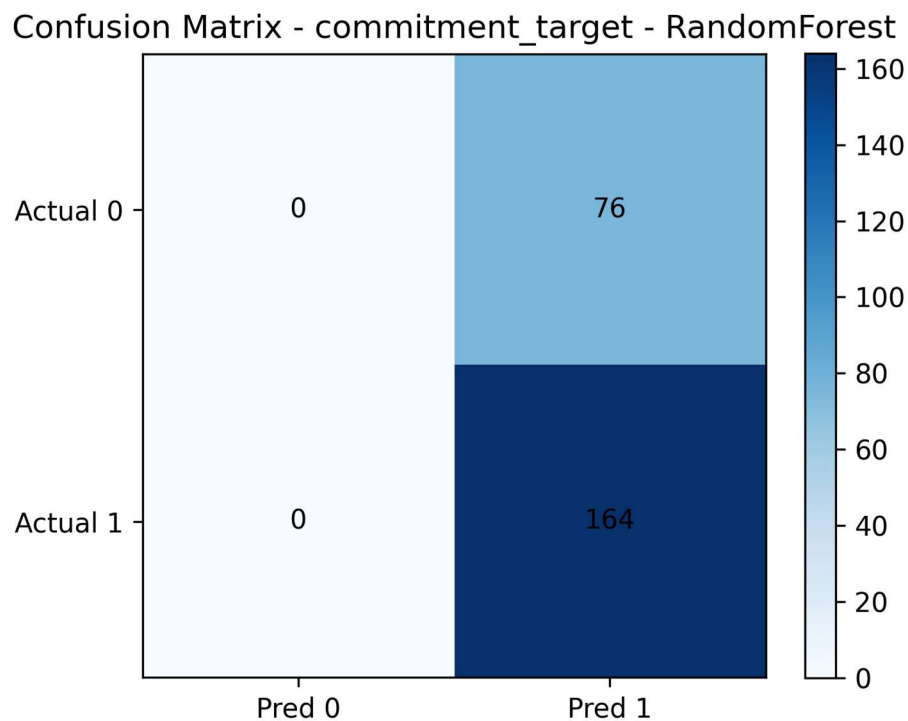


Figure 4.8: Confusion matrix for Random Forest (commitment prediction)

Confusion Matrix - commitment_target - GradientBoosting

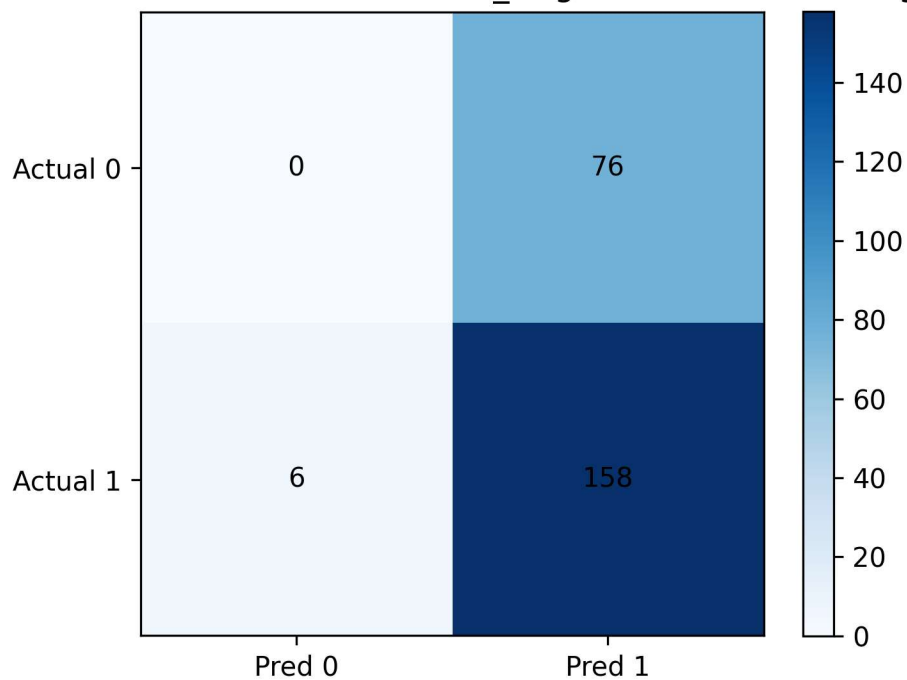


Figure 4.9: Confusion matrix for Gradient Boosting (exemption prediction)

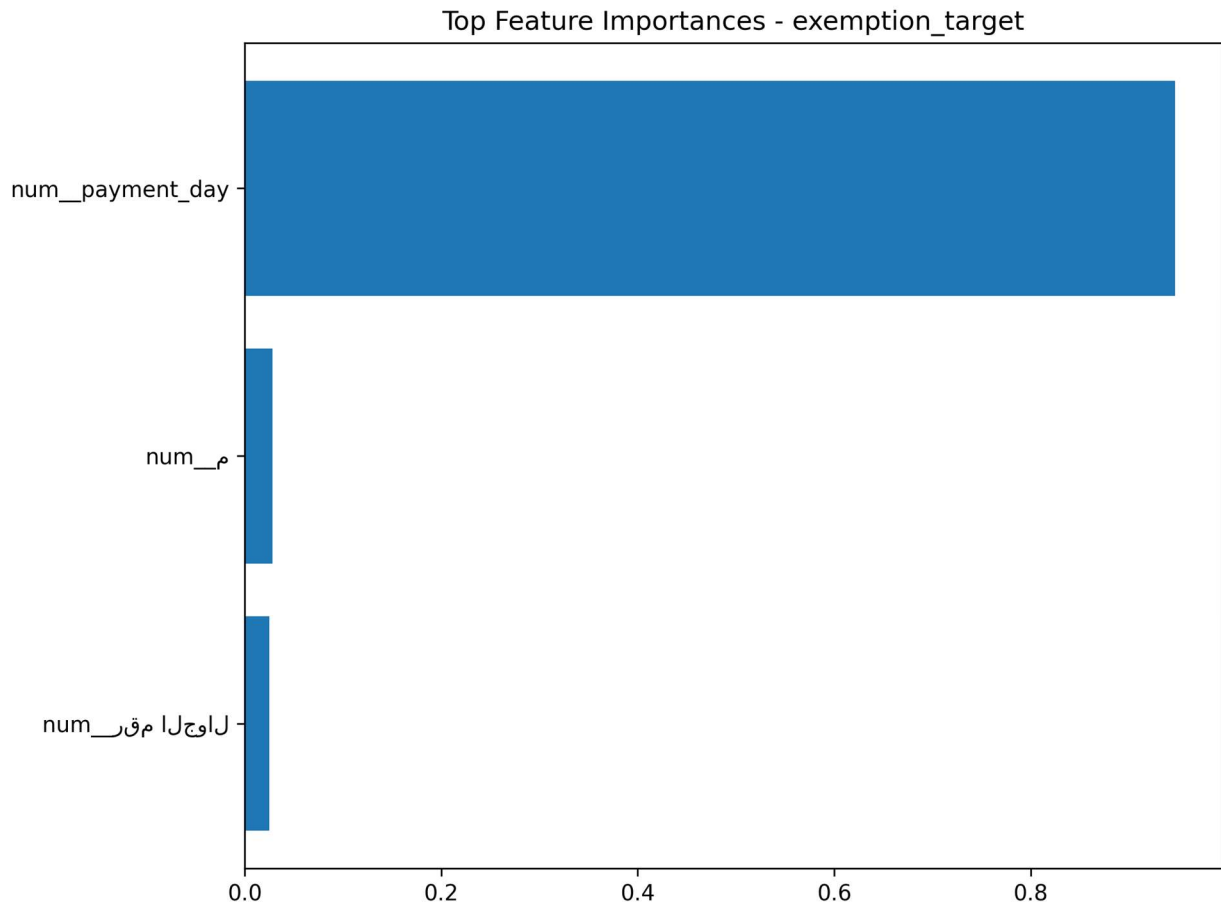


Figure 4.10: Feature importance for exemption prediction using Gradient Boosting

4.5 Interpretation of Results

In this section, I'll discuss the results from the first experiment. At first, I thought the results were more than perfect and achieved very high performance across all evaluation metrics. However, I realized that the results were not entirely accurate and were not at all reliable, especially in the financial forecasting tasks. This doubt about the validity of the results prompted me to start over from the beginning. I found that certain characteristics were closely linked to the variables used, which had made the forecasting process too easy. To address this, I improved the model by modifying and removing some of these variables.

4.6 Impact of Rule-Based System on Model Behavior

Here, I analyzed the impact of the rule-based system on the models' behavior. What I observed during the experiments was that I incorporated a rule-based logic to represent specific financial conditions, such as obtaining discounts and exemptions from installments. I noted that this had a greater impact on the forecast compared to the forecast based on obligations; in the results, the exemptions led to growth. the model performed better than the commitment-based model. The model

shows average results, which proves that the rules were not sufficient for this complex problem.

4.7 Model Comparison and Analysis

11 After comparing and analyzing the results of the performance of the various machine learning models used, I noticed that the results were not consistent across my prediction tasks. In terms of predictions, the performance of all models was similar; however, in terms of classification, there was a very clear difference. The best model was the gradient booster, followed by the random forest, and the weakest was the logistic regression

4.8 Critical Analysis and Discussion

My model is heavily influenced by how the features are prepared and selected. The initial results were very accurate, but the model was unreliable. After refining the features, the results became more realistic.

4.9 Practical Implications

The most significant implications of the proposed system and how to implement them in practice: I found that the developed models help analyze students' payment behavior and identify the cases most likely to qualify for exemption based on that behavior, thereby reducing the need for manual assessment. This study aids in analyzing behavior and making financial decisions, and benefits can be derived from integrating domain knowledge and data-driven models

Chapter 5: Conclusion and Future Work

5.1 Conclusion

I developed a **machine learning** system for this study to predict student compliance with loan repayments and their eligibility for exemptions by forecasting student behavior. Initially, I used the dataset to extract relevant data and added rule-based logic to model financial conditions such as discounts and installment exemptions which helped me explore several models and compare their performance before and after optimization Feature set.

5.2 Key Findings

The most important conclusion I reached was that the initial results were overly optimistic and completely unreliable, which led me to review the

dataset more carefully. After refinement, the prediction for payment compliance was the most difficult, but the prediction for exemption was the most accurate.

5.3 Limitations of the Study

The main limitation of the study was that the dataset did not fully capture all aspects of payment behavior, particularly regarding the ability to predict compliance, especially given the varying financial circumstances of students. For example, a student might be able to make the current month's payment on time and receive a discount, or they might miss the payment deadline and forfeit the discount. The most important thing for me was that the results be truly realistic and reliable

5.4 Practical Implications

Using the financial system helps students make timely payments, keeps them informed, allows them to repay their loans, and lets them know whether they are eligible for discounts or exemptions. Most importantly, it minimizes human errors that might otherwise occur during the process. The system relies on the quality of the data and the features used. In this

study, we can improve the efficiency of financial analysis by integrating the methods with domain expertise.

5.5 Future Work

In the future, I hope to enhance the system technically so that it interacts directly with students, the administration, and the accounting department. I also hope to incorporate deep learning to maximize its benefits, increase the system's flexibility, and make it more reliable and efficient.

References

[1] T. R. Hastie J. Tibshirani and Friedman, The Elements of Statistical Learning. Springer, 2009.

- [2] C. M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006.
- [3] I. Y. Goodfellow A, Bengio, and Deep learning, Courville. 2016, MIT Press [4] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [5] J. H. Annals of Statistics, 2001, Friedman, "Greedy Function Approximation: A Gradient Boosting Machine." [6] D. W. Hosmer, S. Lemeshow, and R. X. Sturdivant, Applied Logistic Regression. Wiley, 2013.
- [7] F. "Scikit-learn: Machine Learning in Python," by Pedregosa et al., published in 2011 in JMLR. [8] J. Han, M. J. Kamber and Pei, Data Mining: Concepts and Techniques. 2011.
- [9] S. B. Informatica, 2007, Kotsiantis, "Supervised Machine Learning: A Review." [10] A. Consumer Credit-Risk Models Using Machine Learning, Khandani et al., Journal of Banking & Finance, 2010. [11] G. Lessmann et al., "Benchmarking Classification Algorithms for Credit Scoring," European Journal of Operational Research, 2015.
- [12] R. S. K. and Baker Yacef, "Educational Data Mining," 2009.
- [13] S. Moro, P. Cortez, and P. Rita, Decision Support Systems, "A Data-Driven Approach to Predict the Success of Bank Telemarketing," 2014. [14] A. Bahnsen, D. Aouada, and B. Expert Systems with Applications, 2015, "Example-Dependent Cost-Sensitive Decision Trees," by Ottersten. [15] J. Brownlee, Machine Learning Mastery with Python, 2016.
- [16] X. Zhang and Y. Zhou, "Predicting Student Performance Using Machine Learning Techniques," IEEE Access, 2020.

[17] M. S. Islam et al., “Financial Prediction Using Machine Learning: A Survey,” IEEE Access, 2021.

[18] N. Gupta and S. Kumar, "Using Machine Learning to Predict Loan Defaults," International Journal of Advanced Computer Science, 2020. [19] H. Chen, R. Chiang, V. and MIS Quarterly, 2012, Storey, "Business Intelligence and Analytics: From Big Data to Big Impact." [20] A. Ng, “Machine Learning Yearning,” 2018.

[21] S. Python Machine Learning by Raschka, 2015. [22] D. Hand, H. P., Mannila, and Principles of Data Mining by Smyth, 2001.