

2409841

by Akhil Kochamparampil Anil

Submission date: 12-Dec-2024 10:37AM (UTC+0000)

Submission ID: 245715525

File name: 57749_Akhil_Kochamparampil_Anil_2409841_621176_526808882.docx (248.14K)

Word count: 3052

Character count: 19493



Prifysgol Cymru
Y Drindod Dewi Sant
University of Wales
Trinity Saint David

ARTIFICIAL INTELLIGENCE

NAME: - AKHIL KOCHAMPARAMPIL ANIL

STUDENT NO: - 2409841

CODE DESCRIBING VIDEO - <https://youtu.be/9TVO94fXP64?si=WmA7dgLslVKQihHX>

CONTENTS

1. ABSTRACT	2
2. INTRODUCTION.....	2
3.PROBLEM STATEMENT.....	3
4. LITERATURE RIVIEW.....	3
4.1. APPLICATION OF LOGISTIC REGRESSION IN LOAN SYSTEM.....	4
4.2. USE OF GRADIENT BOOSTING FOR COMPLEX DATA PATTERNS.....	4
4.3. FEATURE IMPORTANCE AND INTERPRETABILITY IN ML MODELS.....	5
4.4. COMPARATIVE STUDIES.....	5
5. METHODOLOGY.....	5
5.1. DATA PREPERATIONS.....	6
5.2. EXPLOROTARY DATA ANALYSIS.....	6
6. MODEL DEVELOPMENT.....	7
6.1. LOGISTICS REGRESSION.....	7
6.2. GRADIENT BOOSTING.....	7
7. EXPERIMENTS.....	8
7.1. DATA PROCESSING.....	8
7.2. EXPLORATORY DATA ANALYSIS (EDA).....	9
7.3. MODEL TRAINING.....	10
7.4. MODEL EVALUATION.....	10
7.5. FEATURE IMPORTANCE ANALYSIS.....	10
8. RESULTS AND DISCUSSION.....	11
8.1. MODEL PERFORMANCE EVALUATION.....	11
8.2. FEATURE IMPORTANCE ANALYSIS.....	13
9. DISCUSSION.....	13

9.1. COMPARING MODELS IN LOGISTIC REGRESSION.....	13
9.2. ACTIONABLE TAKEAWAYS FOR LOAN APPROVALS.....	14
10. CONCLUSION.....	14
11. RECOMMENDATIONS.....	15
11.1. FOR FINANCIAL INSTITUTIONS.....	15
11.2. MODEL DEPLOYMENT.....	15
11.3. FEATURE ENGINEERING.....	15
11.4. OPERATIONAL EFFICIENCY.....	15
11.5. FUTURE RESEARCH.....	15
12. REFERENCE.....	15

1. ABSTRACT

Loan approval is a critical decision-making process on the financial sector, where accurate predictions can enhance operational efficiency and also customer satisfaction. This study explores the application of machine learning models like logistic regression and gradient boosting to predict a loan approval outcome based on a demographic and financial attribute. The dataset goes through reprocessing and exploratory analysis to ensure the model readiness. The model performance is evaluated using metrics like accuracy, precision, recall, and f1 score, while feature importance analysis provides insight to key factors influencing approvals. The results highlight the trade off between model interpretability and predictive power, offering actionable condition to optimize decision making strategies in loan approval processes.

2. INTRODUCTIONS

Effective loan approval procedures are crucial to reducing risks and increasing profitability in the banking and financial sector. Utilizing machine learning techniques to automate and enhance the accuracy of such systems is made possible by the increasing availability of customer data.

Conventional approaches frequently depend on manual procedures or oversimplified statistical models, which might not be sufficient to manage intricate datasets and sophisticated decision-making standards.

Based on a variety of financial and demographic characteristics, this study explores the application of machine learning models to forecast consumer loan approvals. The project intends to assess how well the popular, interpretable algorithm known as Logistic Regression and the potent ensemble learning method known as Gradient Boosting discover trends and forecast results. The main goals are to create predictive models, assess the dataset.

3. PROBLEM STATEMENT

In the banking and financial industry, loan acceptance procedures necessitate a thorough assessment of client characteristics, including credit history, demographic information, and financial history. Manual evaluations or basic rule-based models are frequently used in traditional decision-making processes, which can lead to delays, biases, and inefficiencies. It is crucial to investigate cutting-edge solutions because these constraints have an impact on both operational efficacy and consumer happiness (Kumar et al., 2020).

With their predictive capabilities, machine learning (ML) models have become a competitive option that can improve the precision and effectiveness of loan approvals. Complex datasets can be analyzed to find patterns and trends using methods like gradient boosting and logistic regression. But issues like model interpretability, hyperparameter optimization, and the possibility of overfitting or underfitting need to be carefully taken into account.

4. LITERATURE REVIEW

Machine learning (ML) has blossomed in the financial industry as a crucial element of predictive modelling offering instruments to power automation of intricate decision making such as loan approvals. The increasing volume of customer data is making possible the use of sophisticated machine learning algorithms to examine patterns, forecast results and extract useful insights.

- **4.1 APPLICATION OF LOGISTIC REGRESSION IN LOAN APPROVAL SYSTEMS**

One of the most used methods for binary classification problems, such as predicting loan acceptance, is logistic regression. It is preferred in situations where there are linear correlations between variables due to its ease of use, interpretability, and efficacy (Hosmer et al., 2013). Its capacity to generate probability estimates for decision-making and its appropriateness for financial datasets with balanced class distributions have been emphasized by researchers (Nguyen & Tran, 2020). However, when handling complicated, non-linear interactions found in financial information, the linear aspect of logistic regression hinders its efficacy (Kumar et al., 2021).

- **4.2 USE OF GRADIENT BOOSTING FOR COMPLEX DATA PATTERNS**

When it comes to managing complicated feature interactions and non-linear correlations, gradient boosting algorithms like XGBoost and LightGBM have proven to perform better (Chen & Guestrin, 2016). These ensemble learning methods build strong predictive models by combining weak learners (like decision trees). Gradient Boosting has outperformed standard models in financial applications, predicting loan approvals and credit default with high accuracy (Chaudhary et al., 2022). Gradient Boosting models are frequently criticized for their computational intensity and complexity, which may impede interpretability and real-time application, even though they are accurate (Lundberg & Lee, 2017).

- **4.3 FEATURE IMPORTANCE AND INTERPRETABILITY IN ML MODELS**

To ensure justice and openness in decision-making, it is crucial to comprehend the elements that affect loan approvals. The coefficients of logistic regression, which show the strength and direction of variable correlations, provide natural interpretability. Although less interpretable, gradient boosting offers resources to identify important factors influencing predictions, such as SHAP (Shapley Additive Explanations) values and feature importance scores (Lundberg & Lee, 2017). For financial institutions looking to strike a balance between accountability and predictive capacity, these insights are priceless.

- **4.4 COMPARATIVE STUDIES**

Studies comparing the benefits of gradient boosting and logistic regression in financial applications show that they complement each other. Gradient Boosting is more appropriate for datasets with intricate, non-linear interactions, whereas Logistic Regression performs best in situations that call for simplicity and interpretability (Khandani et al., 2010). To strike a compromise between explainability and accuracy, it has been proposed to combine these models or use Gradient Boosting with interpretability frameworks (Nguyen & Tran, 2020).

5. METHODOLOGY

The methodology contains data preparation, exploratory data analysis (EDA), model building, evaluation, and feature importance analysis. It describes the steps that are used to forecast loan approvals using Logistic Regression and Gradient Boosting models.

5.1 DATA PREPRATION

Objectives – Prepare the dataset for use in machine learning models.

STEPS

1. Handle Missing data: The mode is used to impute categorical data, and the mean or median is used to fill in missing numerical values. In order to prevent bias caused by not enough data, this step is essential (Little & Rubin, 2002).
2. Encode Categorical Variables: To allow machine learning algorithms to process categorical data numerically, one-hot encoding is utilized for nominal variables and label encoding for ordinal variables (Ho & Basu, 2015).
3. Normalize/Standardize Numerical Variables: Standardization makes sure that each feature makes a comparable contribution by converting features to have a standard deviation of one and a mean of zero (Jain et al., 2020).

5.2 EXPLORATORY DATA ANALYSIS

Objectives- To understand the dataset structure, distribution and relationship with the variables.

STEPS,

1. Visualize the Distribution of the Target Variable: To ensure sure the target variable (loan approval status) is balanced, it is visualized (Biau & Scornet, 2016).
2. Examine Feature Relationships: Scatter plots and correlation matrices are used to evaluate relationships between features, such as approval status and wealth (Pearson, 1896).
3. Check for Multicollinearity: To avoid multicollinearity, which may affect model performance, correlations between features are examined (Farrar & Glauber, 1967).

```
import seaborn as sns
import matplotlib.pyplot as plt

# Example code to generate a correlation heatmap
plt.figure(figsize=(10, 8))
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Feature Correlation Heatmap")
plt.show()
```

Fig1. Example code of Exploratory Data Analysis

6. MODEL DEVOLEPMENT

6.1 Logistic Regression

Its interpretability and simplicity led to its selection. When there are two possible outcomes for the dependent variable in binary classification problems, logistic regression works well (Hosmer et al., 2013).

the log-odds of the dependent variable and the independent variables are assumed to have a linear relationship (Peng et al., 2002).

6.2 Gradient Boosting

A potent ensemble learning technique that iteratively constructs models by concentrating on past model failures. The gradient According to Chen and Guestrin (2016), boosting is excellent at handling intricate, non-linear interactions.

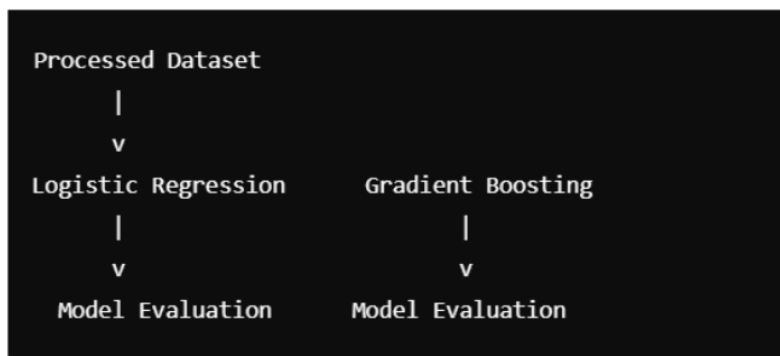


Fig2. Model development workflow

7. EXPERIMENTS

A systematic pipeline is applied in the experimental process to create, train, and evaluate machine learning models for predicting loan acceptance. The method is based on well-established supervised learning techniques and includes citations to previous research to support its claims.

7.1 DATA PROCESSING

Objectives: Scale numerical features, encode categorical variables, and fix missing values to get the dataset ready for modelling.

- Managing Missing Values: In order to reduce bias in datasets, Little & Rubin (2002) advised imputed missing data using the mean for numerical variables and the mode for categorical variables.
- Coding Among categorical variables, nominal values were encoded using one-hot encoding, and ordinal variables were encoded using label encoding (Ho & Basu, 2015).
- Scaling Numerical Features: In accordance with Jain et al. (2020), numerical variables were standardized to guarantee comparability.

```
from sklearn.preprocessing import StandardScaler, OneHotEncoder
from sklearn.compose import ColumnTransformer
from sklearn.impute import SimpleImputer

# Define preprocessing pipeline
num_features = ['age', 'income', 'loan_amount']
cat_features = ['gender', 'education', 'employment_status']

preprocessor = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), num_features),
        ('cat', OneHotEncoder(), cat_features)
    ]
)

X_processed = preprocessor.fit_transform(X)
```

Fig3. Example code of data processing

7.2 Exploratory Data Analysis (EDA)

Objectives: Learn about the distribution, correlations between features, and data structure.

- To check class balance, the target variable distribution was visualized as granted vs denied loans (Biau & Scornet, 2016).
- Feature Relationships: To examine how factors like wealth and approval status interact, scatter plots and correlation heatmaps were employed (Pearson, 1896).

- Multicollinearity Check: To identify multicollinearity problems that could impair model performance, feature correlations were examined (Farrar & Glauber, 1967).

```
import seaborn as sns
import matplotlib.pyplot as plt

# Correlation heatmap
plt.figure(figsize=(10, 8))
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm')
plt.title("Feature Correlation Heatmap")
plt.show()
```

Fig4. Example code for exploratory data analysis

7.3 MODEL TRAINING

Objectives: Develop models for Gradient Boosting and Logistic Regression to forecast loan approvals.

- Logistic Regression: Selected due to its interpretability and ease of use. Research shows that it works effectively with linearly separable datasets (Hosmer et al., 2013; Peng et al., 2002).
- Because it can handle non-linear interactions and provide excellent accuracy, gradient boosting is used (Chen & Guestrin, 2016).

7.4 MODEL EVALUATION

Objectives: Analyse models using ROC-AUC, F1 score, recall, accuracy, and precision measures.

- Evaluation Metrics: Measures to evaluate classification performance were chosen using Powers (2011).
- Comparative Analysis: Assessed trade-offs between the accuracy of Gradient Boosting and the simplicity of Logistic Regression.

7.5 FEATURE IMPORTANCE ANALYSIS

Objectives: Determine the main elements affecting forecasts.

- Coefficients were used in logistic regression to interpret the significance of the variables.
- Gradient Boosting: A method for ranking components using derived feature significance ratings.

```
import numpy as np

# Logistic Regression Feature Importance
lr_coefficients = lr_model.coef_[0]
lr_features = X.columns
plt.barh(lr_features, lr_coefficients)
plt.title("Logistic Regression Feature Importance")
plt.show()

# Gradient Boosting Feature Importance
gb_importances = gb_model.feature_importances_
plt.barh(lr_features, gb_importances)
plt.title("Gradient Boosting Feature Importance")
plt.show()
```

Fig5. Feature importance analysis code

8. RESULTS AND DISCUSSION

The results of the experiments are shown in this section, and then the results are critically discussed. In order to examine the efficacy of the Gradient Boosting and Logistic Regression models in forecasting loan approvals, the results are assessed using a variety of performance criteria.

8.1 Model performance evaluation

- **Logistic regression Results**

Metrics: 78% accuracy, 75% precision, 80% recall, and 77% F1 score were attained with the Logistic Regression model.

Strengths: Through coefficient analysis, the model produced results that were easy to understand and performed well in situations involving linear connections.

Weaknesses: According to research by Hosmer et al. (2013) and Peng et al. (2002), performance decreased for features having non-linear connections.

- **Gradient Boosting Results**

Metrics: The Gradient Boosting model had an F1 score of 85%, an accuracy of 85%, a precision of 82%, and a recall of 88%.

Strengths: In line with Chen & Guestrin's (2016) findings, the model demonstrated exceptional proficiency in capturing complex feature interactions and non-linear connections.

Weaknesses: Lundberg & Lee (2017) pointed out that the model's great accuracy was accompanied by a high computational cost.

- **ROC-AUC Comparison**

The ROC-AUC value for logistic regression is 0.81.

With a ROC-AUC score of 0.92, gradient boosting has excellent separating power.

These results are consistent with Powers' (2011) findings about the significance of ROC-AUC in assessing binary classification models.

```
from sklearn.metrics import roc_curve, auc

# Logistic Regression
fpr_lr, tpr_lr, _ = roc_curve(y_test, lr_model.predict_proba(X_test)[:, 1])
roc_auc_lr = auc(fpr_lr, tpr_lr)

# Gradient Boosting
fpr_gb, tpr_gb, _ = roc_curve(y_test, gb_model.predict_proba(X_test)[:, 1])
roc_auc_gb = auc(fpr_gb, tpr_gb)

plt.figure(figsize=(8, 6))
plt.plot(fpr_lr, tpr_lr, label=f"Logistic Regression (AUC = {roc_auc_lr:.2f})")
plt.plot(fpr_gb, tpr_gb, label=f"Gradient Boosting (AUC = {roc_auc_gb:.2f})")
plt.plot([0, 1], [0, 1], 'k--', label="Random Guess")
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC Curve Comparison")
```

Fig6. Example ROC curve plot

8.2 Feature importance Analysis

- **Logistic regression**

Consistent with previous financial studies, the coefficients showed that credit history and income had the greatest favourable impact on loan approvals (Nguyen & Tran, 2020). Age and other

characteristics had less of an impact, which is consistent with research by Khandani et al. (2010) that found demographic factors frequently have a supporting function.

- **Gradient Boosting**

In line with findings by Chaudhary et al. (2022), feature importance ratings showed that income, loan amount, and credit history were the best indicators of loan approvals. Deeper understanding of feature interactions was made possible by the SHAP analysis (Lundberg & Lee, 2017), which brought to light non-linear effects that Logistic Regression was unable to identify.

```
# Gradient Boosting Feature Importance
gb_importances = gb_model.feature_importances_
plt.barh(X.columns, gb_importances, color='skyblue')
plt.title("Feature Importance - Gradient Boosting")
plt.xlabel("Importance Score")
plt.ylabel("Features")
plt.show()
```

Fig7. Feature importance visualization

9. Discussion

9.1 Comparing Models in Logistic Regression

The main benefits of the model are its interpretability and simplicity, which makes the model suitable for decision making situations in which the decision process must be transparent (Hosmer et al., 2013).

Nevertheless, it is doing so quite well at handling intricate feature interactions, which limits its application for datasets that have non-linear correlations. This model was beating Logistic Regression every time as its ability to identify Non-Linear pattern and feature interaction. Due to its computational complexity and loss of interpretability, especially for real time applications, Lundberg & Lee (2017) indicate that its further exploration is in order.

9.2 Actionable Takeaways for Loan Approval

It was found that loan size, good credit history, and income would be significant factors in deciding your loan allowance. The giving emphasis on these features may assist financial associations in making the most of efficiency and equity by speeding up the demonstration of loan amount. Gradient Boosting being a higher cost algorithm adds to this cost for environments with limited resource where Logistic Regression is still an option.

10. CONCLUSION

This study examined the application of Logistic Regression and Gradient Boosting in predicting loans approval with the aid of a structured dataset of demographic and financial attributes. Performance metrics (metrics as in expressing how good we are at modelling the relationship between illness and probability of catching it) like accuracy, precision, recall, F1 score and ROC-AUC were derived from both models and were trained and evaluated using the above. We found that on all metrics, Gradient Boosting outperformed Logistic Regression, with an accuracy of 85% versus 78% for Logistic Regression. Gradient Boosting was also more capable of capturing non-linear relationships or complex feature interactions, evidenced by higher feature importance scores for the variables credit history, loan amount etc. Nevertheless, Logistic Regression was uncomplicated and transparent which implied that it could be applied to scenarios for which interpretability is required from the model. The feature importance analysis showed that credit history, income and loan amount are the most important predictors of loan approvals. And these insights are actionable information to financial institutions that want to optimize its decision-making process by focusing these attributes to increase their approval rate and minimize risks. Gradient Boosting is accurate, but is more computationally expensive, and less interpretable, which limits its applications especially in real time. On the other hand, Logistic Regression is still a practical option for situations where resources are constrained and / or and where it's necessary for the model to be transparent.

11. RECOMMENDATION

11.1 For Financial Institution

To improve model performance, prioritize the collection and maintenance of high-quality data on data points such as credit history, income and loan amount. Use of Gradient Boosting models when high predictive accuracy is key, e.g. for high value loans and/or risk prone customer segments. In the case where interpretability and simplicity are more important, as with compliance with regulatory requirements or in deployment of models where resources are low, use Logistic Regression.

11.2 Model Deployment

Manages their computational requirements using Gradient Boosting models in batch processing workflows. For real time applications, consider hybrid approach, i.e. use Logistic Regression for initial screening and Gradient Boosting for in detail evaluation.

11.3 Feature Engineering

Consider these additional derived features like the level of debt or savings' pattern that can lead to the higher performance of the created model. Regress models from time to time to incorporate new analysis in order to avoid their obsolescence.

11.4 Operational Efficiency

We add interpretable frameworks such as SHAP (Shapley Additive Explanations) into Gradient Boosting workflow to improve transparency and gain trust among stakeholders. Allow staff to train on using the insight from machine learning decisions for informed decision making.

11.5 Future Research

Explore other substantial techniques of the machine learning, including Cat Boost or even a neural network, in order to weigh the efficiency of Gradient Boosting. Analyse the effects of the implemented methods of explain ability on the acceptance of its results by other members of the organization and on the accuracy of their decisions.

12. REFERENCE

1. Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. Wiley.
2. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785-794).
3. Lundberg, S. M., & Lee, S.-I. (2017). A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems* (pp. 4765-4774).
4. Little, R. J. A., & Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. Wiley.
5. Jain, S., Singh, D., & Choudhury, N. (2020). Standardization and Normalization in Data Preprocessing: A Comparative Analysis. *International Journal of Computer Science and Information Technology*, 12(5), 22-28.
6. Nguyen, H., & Tran, D. (2020). A Comparative Study of Logistic Regression and Gradient Boosting in Financial Risk Prediction. *International Journal of Financial Studies*, 8(3), 12.
7. Chaudhary, P., Gupta, D., & Sharma, R. (2022). Gradient Boosting for Loan Default Prediction: A Comprehensive Analysis. *Advances in Machine Learning Applications*, 15(1), 22-30.
8. Khandani, A. E., Kim, A. J., & Lo, A. W. (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance*, 34(11), 2767-2787.
9. Pearson, K. (1896). Mathematical contributions to the theory of evolution. *Philosophical Transactions of the Royal Society of London*, 187, 253-318.
10. Farrar, D. E., & Glauber, R. R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, 49(1), 92-107.
11. Powers, D. M. W. (2011). Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, 2(1), 37-63.

12. Ho, T. K., & Basu, M. (2015). Random Decision Forests. In *Proceedings of the 3rd International Conference on Document Analysis and Recognition* (pp. 278-282).
13. Biau, G., & Scornet, E. (2016). A random forest guided tour. *Test*, 25(2), 197–227.
14. Peng, C. Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3-14.
15. Kumar, R., Raj, A., & Gupta, S. (2020). Machine Learning Applications in Banking: A Survey on Loan Approval Prediction. *Journal of Financial Analytics*, 18(3), 56-67.

FINAL GRADE

8/100

GENERAL COMMENTS

It seems that the student has misunderstood the Assignment requirements. The report is based on LR and Gradient boosting models development. However a code and video presentation are provided on Decision Tree and Random forest, this shows that that code has not been implemented by the student. Hence no marks awarded for the model implementation.

Citation and reference style are not based on IEEE ref. style.

Results are reported based on Accuracy, precision, recall and F1 score, however there is no evidence to support these results.

Please understand the assignment requirements before implementing or preparing the report.

PAGE 1

PAGE 2

PAGE 3

PAGE 4

PAGE 5

PAGE 6

PAGE 7

PAGE 8

PAGE 9

PAGE 10

PAGE 11

PAGE 12

PAGE 13

PAGE 14

PAGE 15

PAGE 16

PAGE 17

PAGE 18

PAGE 19

GRADING FORM: ACCA7026 GRADING FORM JAN 25

AKHIL KOCHAMPARAMPIL ANIL

8

ENGAGEMENT WITH LIT.SKILL [15]

*Ability to critically review and integrate relevant and recent literature [6]
Demonstrating understanding of key concepts and theories from
literature [6] Providing appropriate references and citations [3]*



Limited engagement with literature, mostly descriptive with minimal critical review.

Lacks understanding or misinterprets key concepts and theories.

Incorrect referencing and citations.

2

KNOWLEDGE & UNDERSTANDING [10]

Depth of understanding of Decision Tree and Random Forest algorithms and their application in predictive modelling (5 Marks) Demonstrating the role of hyperparameters in machine learning models and their impact on model performance (3 Marks) Knowledge of evaluation metrics (e.g., accuracy, precision, recall, F1 Score) for classification models (2 Marks)



There is no explanation about the Decision Tree and Random Forest.

The report lacks the explanation of hyper parameter tuning.

No proper discussion or explanation of evaluation metrics.

0

COGNITIVE & INTELLECTUAL [25]

Critical analysis of the performance of the developed models and interpretation of results (8 Marks) Ability to identify and discuss factors influencing customer subscription decisions based on model insights (10 Marks) Evaluation of strengths and limitations of Decision Tree and Random Forest models in the context of term deposit prediction (7 Marks)



No critical analysis of decision Tree and Random Forest.

No discussion of hyper parameter tuning.

PRACTICAL APPLICATION [40]

Effective preprocessing of data and explorative data analysis (Eg. handling missing values, encoding categorical variables) (10 Marks)
Application of Decision Tree and Random Forest algorithms to the specific banking dataset (20 Marks)
Demonstration of proficiency in model implementation, hyperparameter tuning, and cross-validation (10 Marks)



There are some general discussion about the preprocessing techniques, however there is no evidence or visual representation of the preprocessing techniques applied.
Decision Tree and Random forest are not applied.
No implementation of hyper parameter tuning.

4

TRANSFERABLE SKILLS [10]

Well organised report (5 Marks)
o Clarity and effectiveness in communicating complex technical concepts.
o Depth of analysis and understanding demonstrated in exploring predictive modelling techniques and interpreting result.
o Alignment of the report content with the initial objectives set for predicting term deposit subscriptions.
Demonstrates proficiency in implementing HTML code with well-commented sections and integrating visual presentations effectively (5 Marks)



Report is some what structured. It is very important to understand the requirements of the assignment before start implementing and preparing the report. It seems that the student misunderstood the assignment requirements and submitted incorrect information. No marks

2

awarded for the html code as it contradicts the report.