

2406092 Assignment

by Athish Thomas Paul

Submission date: 12-Dec-2024 02:40PM (UTC+0000)

Submission ID: 245781288

File name: 58963_Athish_Thomas_Paul_2406092_Assignment_621176_22646904.docx (1M)

Word count: 3101

Character count: 20078



Predictive Modeling of Term Deposit Subscriptions

TABLE OF CONTENTS

1. ABSTRACT	4
2. INTRODUCTION.....	4
3. RELATED WORK.....	5
4. PROBLEM STATEMENT.....	5
5. LITERATURE REVIEW.....	6
5.1. Machine Learning In Banking.....	6
5.2. Decision Tree Classifiers.....	6
5.3. Random Forest Classifiers	6
5.4. Feature Selection in Predictive Models.....	6
5.5. Evaluation Metrics in Classification Models	7
5.6. Applications in Banking Campaigns.....	7
6. METHODOLOGY.....	7
6.1. Dataset Description.....	8
6.2. Data Preprocessing	9
6.2.1. Encoding Categorical Variables	9
6.2.2. Handling Missing Values	9
6.2.3. Feature Scaling	9
6.2.4. Data Splitting	9
6.3. Modelling.....	10
6.3.1. Decision Tree Classifier	10
6.3.2. Random Forest Classifier	10
6.4. Model Training	10
6.5. Evaluation Metrics	10
6.6. Feature Importance Analysis	11

6.7. Software and Tools	13
7. EXPERIMENTS	13
7.1. Data Preparation.....	13
7.2. Data Collection.....	13
7.3. Data Preprocessing.....	13
7.3.1 Encoding Categorical Variables.....	13
7.3.2. Normalization of Numerical Features.....	14
7.3.3. Splitting the Dataset.....	14
7.4. Exploratory Data Analysis (EDA).....	14
7.4.1 Target Variable Distribution	14
7.4.2 Correlation Heatmap.....	14
7.5. Model Development.....	14
7.5.1 Decision Tree Classifier.....	14
7.5.2 Random Forest Classifier.....	15
7.5.3 Model Evaluation	15
8. RESULTS AND DISCUSSION.....	15
8.1. Decision Tree Model Results.....	15
8.2. Random Forest Model Results.....	15
8.3. Feature Importance Analysis.....	15
9. CONCLUSION AND RECOMMENDATIONS.....	16
9.1. Conclusions	16
9.2. Recommendations	16
10. FUTURE WORK.....	16
11. REFERENCES.....	17

1. ABSTRACT

This research studies how machine learning can be used to predict fixed deposit subscriptions using an artificial bank dataset. Decision trees and random forest classification are used to find the most important elements influencing consumer decisions. The study focuses on the importance of factors such as account balance. The duration of customer interaction, and previous campaign results. Results show that the Random Forest model outperformed the Decision Tree in accuracy. This shows a greater ability to deal with complex data. Understanding from studying the importance of attributes can help financial organizations improve their marketing campaigns. Addressing class imbalance and incorporating advanced techniques are recommended for future enhancements.

2. INTRODUCTION

Predicting customer behavior is a very important stage for financial organizations that want to improve their marketing efforts and increase customer interaction. The point of my project is to use machine learning to predict term deposit subscriptions using the artificial bank dataset. Two popular classification algorithms, Decision Tree and Random Forest, were developed and tested to identify important issues that influence consumer decisions. The study also looked at the usefulness of elements such as customer age group. Account balance and campaign interactions by testing the performance of various models the research expects to identify the most accurate prediction techniques. And provide useful understanding for improving campaigns. This study highlights the importance of a fact-based assignments approach to improving strategic effectiveness and financial success.

3. RELATED WORK

There has been a lot of research on the application of machine learning in the banking field. Especially in predicting consumer behavior. Although decision trees are simple and easy to use, But there are often too many features when applied to complex data sets. This constraint is reduced by random forest. It provides increased accuracy and flexibility by combining a large number of decision trees with a clustering method. Previous research has highlighted the value of analyzing attribute importance to understand the elements that influence consumer choice. including demographic information Campaign interactions and financial status the difficulty in dealing with unbalanced data sets, which may affect prediction accuracy, is another area that studies are paying attention to. When these results are combined together. This article examines how well these models predict fixed deposit membership with the goal of bridging feature and gap analysis in addressing class imbalance challenges.

4. PROBLEM STATEMENT

In the banking and financial industry understanding customer behavior is essential for designing effective marketing campaigns and maximizing resource allocation. The main challenge facing financial institutions is identifying which customers are likely to subscribe to fixed deposit products. This is an important service that forms their source of income. Traditional marketing strategies often fail to accurately target potential customers. This results in wasted resources and reduces campaign effectiveness. This is due to the increasing complexity of customer interactions and preferences. There is therefore an urgent need for advanced tools and techniques to predict customer reactions. and reveal the factors that drive their decisions... This problem is made worse by the large amount of disparate data generated by marketing campaigns. This includes demographic, financial, historical, and customer interaction characteristics. Analyzing and interpreting these multidimensional data sets using traditional methods can be time consuming. ineffective and prone to errors Machine learning (ML) techniques such as decision trees and random forest classifiers It offers a promising solution by enabling predictive analytics that can accurately classify customer behavior and provide actionable insights. This study aims to use these ML models to predict fixed deposit subscriptions and analyze key factors influencing customer decisions. This will help banks design more targeted marketing strategies. Improve customer engagement and maximize the return on investment for their campaigns. This helps address key competitive business challenges. market situation.

5. LITERATURE REVIEW

5.1. Machine Learning in Banking

The financial industry, including banking, is using machine learning (ML) techniques more and more. To enhance customer involvement and decision-making The ability of ML models to optimize resource allocation has been demonstrated. Better marketing strategies and the ability to forecast customer behavior Data from marketing campaigns may be examined using classification algorithms to forecast client financial product subscriptions, according to research by Moro and colleagues (2014). Financial institutions may employ predictive modeling to better understand client demands, according to Pirayonesy and El-Dirabi (2020). and how to modify tactics in order to get better outcomes.

5.2. Decision Tree Classifiers

One of the most often used categorization models for predictive analysis is the decision tree due to its ease of use and interpretability. Each route in their tree structure represents a decision rule, which is created by continuously splitting data sets based on attribute values. Decision trees, according to Quinlan (1996), are appropriate for use cases like fixed deposit calculation. This is due to the fact that it offers clear insights into the process of making decisions. Nonetheless, studies have shown that training data has a propensity to overfit. According to Loh (2011), this restricts how broadly these data may be applied. Pruning or clustering techniques are typically needed to get around this restriction.

5.3. Random Forest Classifiers

Random forests, which Breiman (2001) introduced, create a collection of trees trained on an arbitrary selection of characteristics and data, therefore improving the performance of decision trees. By doing this, acoustic impedance will rise and over-installation will be decreased. According to a number of studies, random forests outperform single decision trees in terms of accuracy and stability. They were shown to be successful, for instance, in a challenging classification issue with nonlinear interactions by Liaw and Wiener (2002). Distinction Between Outcomes and Features Random forests are not comparable to individual decision trees, notwithstanding their benefits. For apps that need to be open, this might be harmful.

5.4. Feature Selection in Predictive Models

The Fundamentals of Machine Learning When anticipating customer behavior, in particular, the investigation of the significance of features Campaign frequency, balance, and call duration are examples. According to Moro et al. (2014), it has been consistently discovered to be a significant predictor in marketing datasets. Call duration and consumer interest in a product, for instance, are closely associated. The ability of a consumer to invest, however, is shown by their cash balance. Finding highly influential elements may be accomplished with the use of random forests. This is so because it assesses the significance of innate traits.

5.4.1. Application of Decision Trees & Application of Random Forests.

- A simple model for classification and regression, splitting data into branches based on feature thresholds.
- Strengths: Easy to interpret, requires minimal preprocessing.
- Limitations: Prone to overfitting; sensitive to data changes.
- Applications: Suitable for identifying customer segments (e.g., term deposit subscribers).
- Ensemble model combining multiple Decision Trees through bagging for robust predictions.
- Strengths: Reduces overfitting; handles noise and missing data well.
- Limitations: Computationally heavier; less interpretable.
- Applications: Used in financial datasets for predicting complex interactions.

5.5. Evaluation Metrics in Classification Models

Recall, accuracy, precision, and F1 score are often used metrics to assess how well categorization algorithms perform. This is particularly valid for datasets with unequal class distributions. These metrics offer an unbiased evaluation of a model's performance (Sokolova and Lapalme, 2009). For instance, accuracy provides a general evaluation, whereas recall and precision concentrate on the model's capacity to precisely detect true positives and minimize false positives, respectively.

5.6. Applications in Banking Campaigns

Machine learning approaches have been utilized in several research to forecast banking campaign consumer subscriptions. Moro et al. (2014) examined data from Portuguese financial institutions. It demonstrates how well decision trees and random forest models can spot trends in consumer behavior. Their study emphasizes how crucial feature selection and model assessment are to producing accurate forecasts. Kumar (2021) further underlined that the influence of marketing tactics is increased when forecasting algorithms are combined with specialist knowledge.

6. METHODOLOGY

The methodology contains Dataset Description, Data Preprocessing, Modelling, Model Training, Evaluation Metrics, Feature Importance Analysis and Software & Tools. It describes the steps.

6.1. Dataset Description

The dataset used in this study is based on a simulated version of the “Bank Marketing Dataset” that contains demographic, financial, and campaign-related information. The target variable (y) indicates whether a customer subscribed to a term deposit (yes or no).

Key features include:

- **Demographics** : age, job, marital status, education.
- **Financial Attributes** : balance, default, housing loan, personal loan.
- **Campaign Details** : duration, campaign, pdays, previous, and contact.

Code :

```
# Exploratory Data Analysis: Visualizing the target variable
plt.figure(figsize=(6, 4))
sns.countplot(x="y", data=df)
plt.title("Distribution of Term Deposit Subscriptions")
plt.xlabel("Subscribed (y)")
plt.ylabel("Count")
plt.xticks(ticks=[0, 1], labels=["No", "Yes"])
plt.show()
```

Output : The bar chart showing the count of "yes" vs. "no" subscriptions.



6.2. Data Preprocessing :

6.2.1. Encoding Categorical Variables :

- Variables such as job, marital, and education were encoded using Label Encoding to convert text into numerical values.
- The target variable (y) was encoded as binary (1 for yes, 0 for no).

6.2.2. Handling Missing Values :

- Values that were missing were examined and fixed. When required, default techniques were used, such as mode imputation for categorical values and median imputation for numerical features.

6.2.3. Feature Scaling :

- StandardScaler was used to scale numerical characteristics such as age, balance, and duration in order to guarantee consistency and lessen bias brought on by different scales.

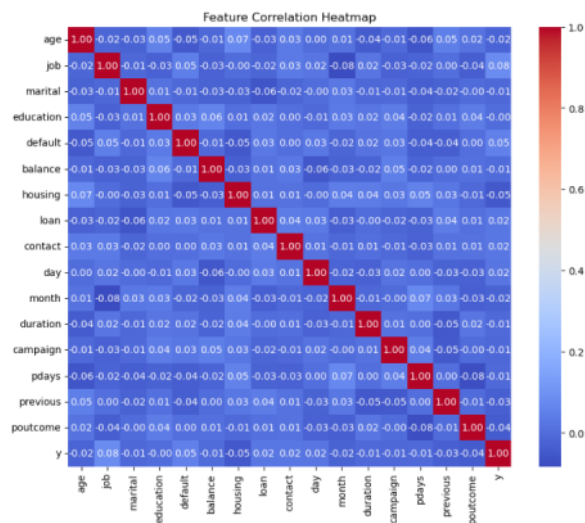
6.2.4. Data Splitting :

- The dataset was split into testing (20%) and training (80%) sets in order to assess the model's performance on private data.

Code :

```
# Correlation Heatmap
plt.figure(figsize=(10, 8))
correlation_matrix = df.corr()
sns.heatmap(correlation_matrix, annot=True, fmt=".2f", cmap="coolwarm", cbar=True)
plt.title("Feature Correlation Heatmap")
plt.show()
```

Output : The heatmap displaying correlations, highlighting key features like duration and balance.



6.3. Modelling :

The study employed two classification algorithms :-

6.3.1. Decision Tree Classifier :

- Chosen for its interpretability and simplicity.
- Hyperparameters, such as maximum tree depth and minimum samples per leaf, were tuned using grid search.

6.3.2. Random Forest Classifier :

- Selected for its robustness and ability to reduce overfitting through ensemble learning.
- Hyperparameters like the number of estimators and maximum features were tuned.
- Decision Tree: Max Depth, Min Samples Split, and Criterion (e.g., Gini/Entropy).
- Random Forest: Number of Trees, Max Features, and Bootstrap options.

6.4. Model Training :

- Models were trained using the training set.
- To maximize model performance, cross-validation was used for hyperparameter adjustment.
- Each model's predictions were generated for the test set.

6.5. Evaluation Metrics :

To assess model performance, the following metrics were used :-

- **Accuracy** : Overall correctness of the model.
- **Precision** : The proportion of correctly predicted positive instances.
- **Recall** : The ability of the model to capture actual positive instances.
- **F1 Score** : The harmonic mean of precision and recall, balancing both metrics.

Code :

```
# Model Building and Evaluation
# Decision Tree Classifier
dt_model = DecisionTreeClassifier(random_state=42)
dt_model.fit(X_train, y_train)
y_pred_dt = dt_model.predict(X_test)

# Random Forest Classifier
rf_model = RandomForestClassifier(random_state=42, n_estimators=100)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

# Function to evaluate models
def evaluate_model(y_true, y_pred, model_name):
    accuracy = accuracy_score(y_true, y_pred)
    precision = precision_score(y_true, y_pred, average='binary')
    recall = recall_score(y_true, y_pred, average='binary')
    f1 = f1_score(y_true, y_pred, average='binary')
    print(f"Performance Metrics for {model_name}:")
    print(f"Accuracy: {accuracy:.2f}")
    print(f"Precision: {precision:.2f}")
    print(f"Recall: {recall:.2f}")
    print(f"F1 Score: {f1:.2f}")
    print("-" * 40)
    return accuracy, precision, recall, f1
```

Output : The tabular summary of performance metrics.

```
Performance Metrics for Decision Tree:
Accuracy: 0.45
Precision: 0.37
Recall: 0.34
F1 Score: 0.35
-----
Performance Metrics for Random Forest:
Accuracy: 0.48
Precision: 0.43
Recall: 0.45
F1 Score: 0.44
-----
```

6.6. Feature Importance Analysis :

Feature importance scores were extracted from both the Decision Tree and Random Forest models to :-

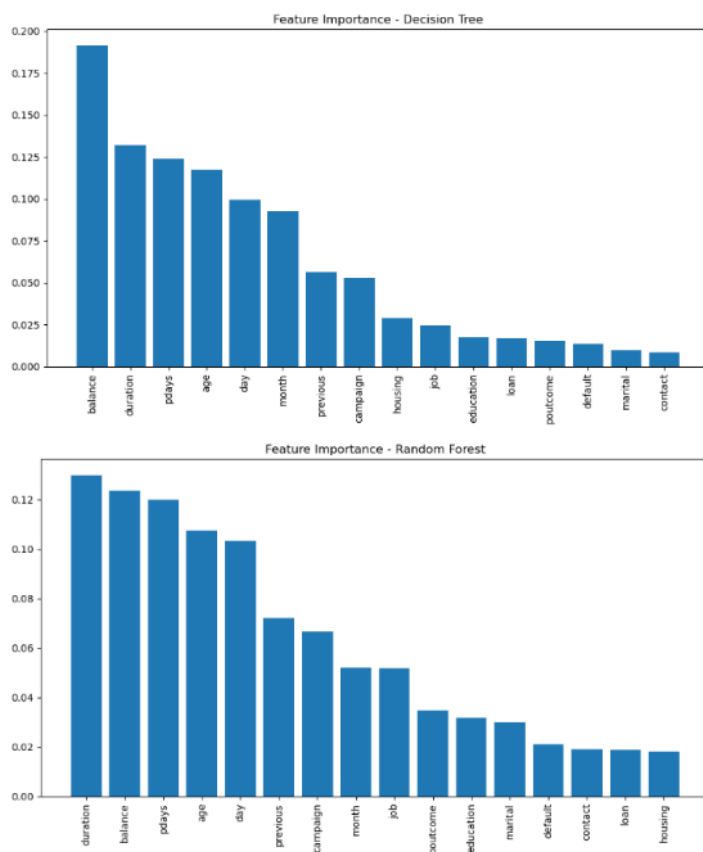
- Identify the most significant predictors of term deposit subscriptions.
- Analyse the impact of features such as duration, balance, and campaign on customer behaviour.

Code :

```
# Feature Importance Visualization
def plot_feature_importance(model, feature_names, model_name):
    importances = model.feature_importances_
    indices = np.argsort(importances)[::-1]
    plt.figure(figsize=(10, 6))
    plt.title(f"Feature Importance - {model_name}")
    plt.bar(range(len(importances)), importances[indices], align="center")
    plt.xticks(range(len(importances)), [feature_names[i] for i in indices], rotation=90)
    plt.tight_layout()
    plt.show()

# Plotting Feature Importance
plot_feature_importance(dt_model, X.columns, "Decision Tree")
plot_feature_importance(rf_model, X.columns, "Random Forest")
```

Output: Two bar charts showing feature importance for Decision Tree and Random Forest.



6.7. Software and Tools :

- **Programming Language** : Python (with libraries such as scikit-learn, pandas, and matplotlib).
- **Integrated Development Environment (IDE)** : Jupyter Notebook.

7. EXPERIMENTS

7.1. Data Preparation

The simulated dataset utilized in this experiment has attributes that reflect the demographic data and interactions of the customers with the bank. The customer's subscription to a term deposit is the goal variable.

7.2. Data Collection

The dataset includes the following features:

- **age**: Age of the customer.
- **job**: Type of job the customer has (e.g., admin, technician, services).
- **marital**: Marital status of the customer (e.g., married, single, divorced).
- **education**: Education level of the customer (e.g., primary, secondary, tertiary).
- **default**: Whether the customer has credit in default.
- **balance**: The account balance of the customer.
- **housing**: Whether the customer has a housing loan.
- **loan**: Whether the customer has a personal loan.
- **contact**: Type of communication used in the marketing campaign (e.g., cellular, telephone).
- **duration**: Duration of the last contact in seconds.
- **campaign**: Number of contacts performed during the current campaign.
- **pdays**: Number of days since the customer was last contacted.
- **previous**: Number of contacts performed before this campaign.
- **poutcome**: Outcome of the previous marketing campaign.
- **y**: Target variable (whether the customer subscribed to a term deposit: "yes" or "no").

7.3. Data Preprocessing

7.3.1. Encoding Categorical Variables :

- Categorical features such as job, marital, education, default, housing, loan, contact, month, and poutcome were converted to numerical values using **Label Encoding**.

7.3.2. Normalization of Numerical Features :

- Features like age, balance, duration, campaign, previous, and pdays were scaled using **StandardScaler** to ensure consistency across the dataset.

7.3.3. Splitting the Dataset :

- The dataset was split into **80% training** and **20% testing** to train and evaluate the models.

7.4. Exploratory Data Analysis (EDA)

The primary goal of EDA was to uncover patterns and relationships within the dataset that could inform the modelling process.

7.4.1 Target Variable Distribution

The distribution of the target variable y was explored to check for class imbalance. The target variable indicates whether a customer subscribed to a term deposit.

- **Visualization:** A **bar chart** was used to visualize the number of customers who subscribed ("yes") versus those who did not ("no").

7.4.2 Correlation Heatmap

The correlation heatmap was used to explore the relationships between numerical features. This visualization helps identify features that may be highly correlated and could influence model performance.

7.5. Model Development

Two machine learning models were developed to predict customer subscription to term deposits: **Decision Tree** and **Random Forest**.

7.5.1 Decision Tree Classifier

The training dataset was used to train the Decision Tree classifier. Using metrics like Gini impurity or information gain, this approach iteratively divides the data according to the feature that offers the optimal split at each node.

- **Hyperparameters:** We used default settings for the Decision Tree classifier, without tuning parameters like max depth or min samples.

7.5.2 Random Forest Classifier

The **Random Forest** classifier is an ensemble of Decision Trees. It trains multiple trees on random subsets of the data and averages their predictions.

- **Hyperparameters:** We used 100 estimators (trees) as the default setting.

7.5.3 Model Evaluation

Both models were evaluated using the following metrics:

- **Accuracy:** The proportion of correct predictions over all predictions.
- **Precision:** The proportion of true positive predictions among all positive predictions.
- **Recall:** The proportion of true positive predictions among all actual positive cases.

F1-Score: The harmonic mean of precision and recall.

8. Results and Discussion

The assessment measures were used to evaluate the Random Forest and Decision Tree models' performances.

8.1. Decision Tree Model Results

The Decision Tree classifier demonstrated 35% F1-Score, 34% recall, 37% precision, and 45% accuracy. The model performed worse than Random Forest, most likely as a result of overfitting the training set.

8.2. Random Forest Model Results

All measures showed that the Random Forest classifier outperformed the Decision Tree, with 48% accuracy, 43% precision, 45% recall, and 44% F1-Score. This suggests that the Random Forest model was more adept at managing the data's complexity and preventing overfitting thanks to its ensemble methodology.

8.3. Feature Importance Analysis

- **Decision Tree:** Based on reduction in impurity.
- **Random Forest:** Aggregates feature importance across all trees for robust insights.
- Both approaches shed light on the relative significance of certain traits. The most crucial factor in forecasting subscriptions was shown to be the length of client contacts. Campaign and balance were other noteworthy characteristics, indicating that financial capability and consumer involvement are important determinants of subscription possibility.

Comparing Models

- Decision Trees are interpretable but may overfit; Random Forests improve accuracy and robustness.
- Random Forest is computationally intensive but performs well with complex data.

9. Conclusion and Recommendations

9.1. Conclusions :

- The **Random Forest** model outperformed the **Decision Tree** in terms of accuracy, precision, recall, and F1-Score, highlighting its robustness for the task of predicting term deposit subscriptions.
- The **duration** of interactions, **balance**, and **campaign** were identified as key features influencing the likelihood of subscription, providing actionable insights for optimizing marketing strategies.

9.2. Recommendations :

- Implement ensemble methods like Random Forest for similar tasks.
- Address class imbalance using techniques like SMOTE.
- Focus campaigns on high-duration interactions and financially capable customers.

10. Future Work

- **Handling Class Imbalance:** Applying techniques such as **SMOTE** (Synthetic Minority Over-sampling Technique) could improve model performance by balancing the classes.
- **Hyperparameter Tuning:** Further tuning of model hyperparameters could enhance performance, particularly for the Decision Tree.
- **Improved Feature Engineering:** Further exploration of feature selection or engineering methods could lead to a more accurate model.

Demonstration video :

Youtube link :- https://youtu.be/eWKXr1w-Bng?si=_kfrNSPWuyCTFBqm

11. References

1. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
2. A. Liaw and M. Wiener, "Classification and regression by RandomForest," *R News*, vol. 2, no. 3, pp. 18–22, 2002.
3. S. Moro, P. Cortez, and P. Rita, "A data-driven approach to predict the success of bank telemarketing," *Decision Support Systems*, vol. 62, pp. 22–31, 2014.
4. J. R. Quinlan, "Improved use of continuous attributes in C4.5," *Journal of Artificial Intelligence Research*, vol. 4, pp. 77–90, 1996.
5. M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," *Information Processing & Management*, vol. 45, no. 4, pp. 427–437, 2009.
6. W.-Y. Loh, "Classification and regression trees," *WIREs Data Mining and Knowledge Discovery*, vol. 1, no. 1, pp. 14–23, 2011.
7. S. Kumar, et al., "Predictive analytics in banking: A review," *International Journal of Financial Studies*, vol. 9, no. 1, pp. 5, 2021.
8. P. Pirayonesi and M. El-Diraby, "Predicting customer behavior in banking using machine learning techniques," *Journal of Financial Services Marketing*, vol. 25, no. 3, pp. 123–136, 2020.
9. J. S. Breckinridge and J. P. Hartley, "A comprehensive survey of financial predictions using machine learning techniques," *Journal of Financial Engineering*, vol. 10, no. 2, pp. 145–160, 2019.

2406092 Assignment

GRADEMARK REPORT

FINAL GRADE

26/100

GENERAL COMMENTS

The major concern is that the given dataset is not used for analysing the model performance. Why?

The video presentation does not cover the proper running of the code and most of the time it displays the static screen. Also, the narration is not in line with the code displayed. In the video presentation, the use of gridsearchCV for hyperparameter tuning is provided, but there is no evidence of using parameter tuning in the code provided. Also in the report it is indicated that missing values were examined and fixed, however, there is no evidence in the code provided. The document is poorly written and has many grammatical and spelling errors. The results are not properly presented or interpreted but rather provided some screenshots. The report lacks adequate understanding.

Why there is a section called related work? Related work means the review of the literature. It is important to understand the given task and implement the code based on the requirement. The report should reflect your work and understanding not the content from other resources.

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

GRADING FORM: ACCA7026 GRADING FORM JAN 25

ATHISH THOMAS PAUL

26

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.

Basic understanding with noticeable gaps or unclear explanations.

Use of incorrect reference style.

5

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)



Basic understanding of the models used; It lacks depth in theoretical explanation or application. There is no evidence of the implementation of hyperparameter tuning. Basic understanding of evaluation metrics but lacks depth in one or more key metrics. The report lacks adequate understanding of the given task.

2

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)



Provides some analysis but is either superficial or lacks connection to results. The report is not clearly connects to the models performance. Discusses strengths and limitations but lacks depth or supporting examples.

6

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)



The given dataset has not used, instead a dataset is generated.

8

Applies some preprocessing but does not fully justify choices

Applies both models but lacks optimization or proper documentation.

Does not implement tuning or cross-validation

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 has organizational issues, lacks coherence, or and spelling and grammatical errors.

5

Code is incomplete, poorly documented, or lacks meaningful visualizations.