JOEL - JOHANA - LUCIE - MANUEL

LOAN APPROVAL PREDICTION USING MACHINE LEARNING

MISSION

How can we automate the loan approval system?

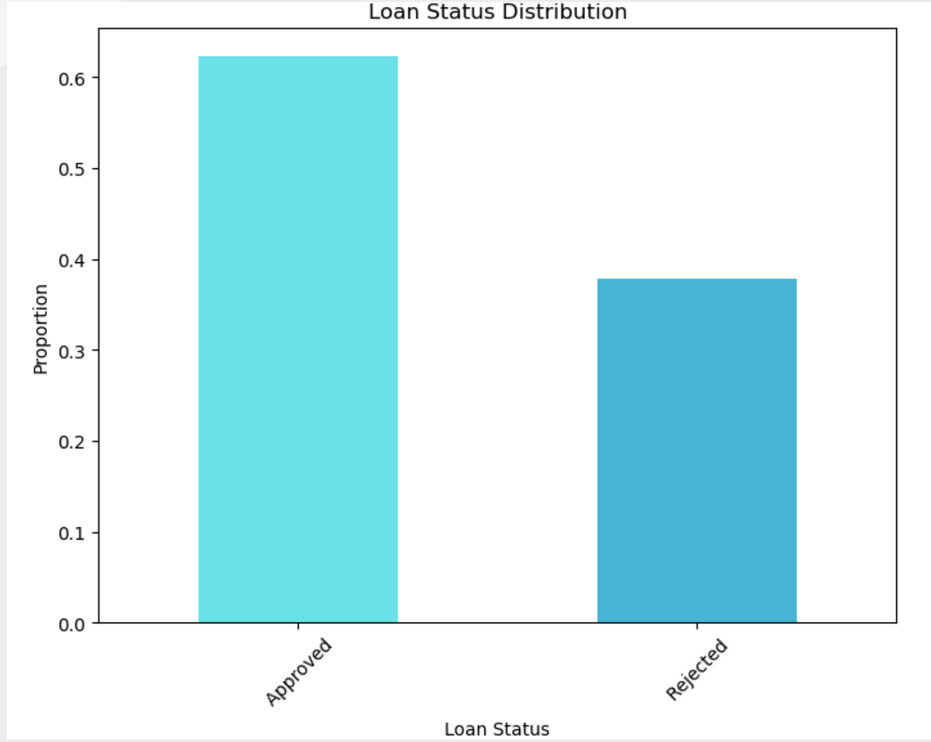
Our goal is to use a machine learning model to avoid human error and improve customer experience by reducing loan approval waiting time.
We will select the best performing model, with a focus on reducing false positive to avoid lending to risky applicants.

DATA PREPARATION

The project aims to develop a machine learning model to predict loan approval based on key financial factors such as :

- Credit score
- Income yearly
- Assets value (luxury/residential/commercial)
- Education
- Employment
- Loan amount and loan term.

kaggle



FEATURE ENGINEERING

Dummification

```
loan_id                    int64
 no_of_dependents          int64
 education                 object
 self_employed             object
 income_annum              int64
 loan_amount               int64
 loan_term                 int64
 cibil_score               int64
 residential_assets_value  int64
 commercial_assets_value   int64
 luxury_assets_value       int64
 bank_asset_value          int64
 loan_status               object
```

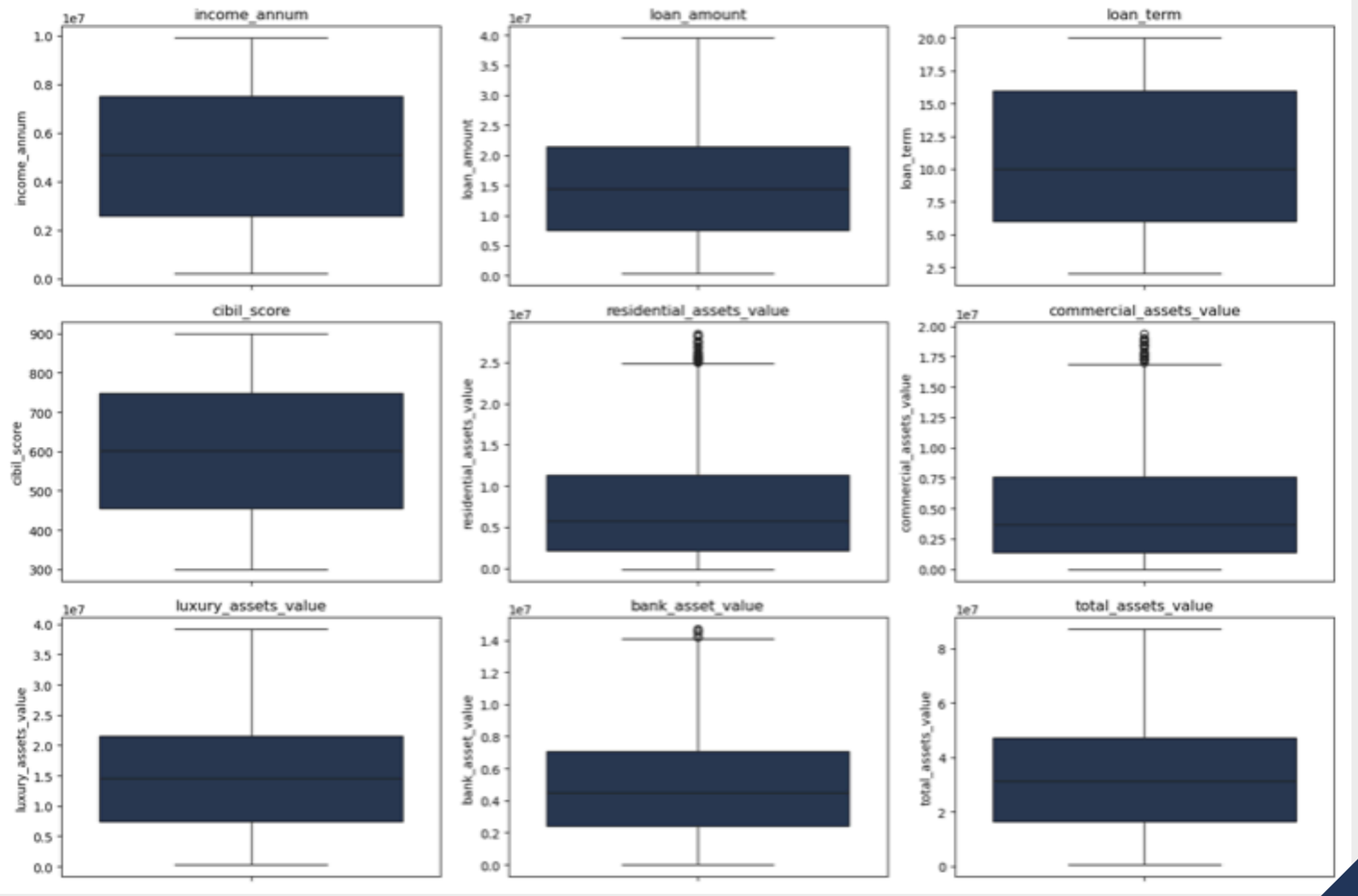
	loan_id	no_of_dependents	education	self_employed	income_annum	loan_amount	loan_term
0	1	2	Graduate	No	9600000	29900000	12
1	2	0	Not Graduate	Yes	4100000	12200000	8
2	3	3	Graduate	No	9100000	29700000	20
3	4	3	Graduate	No	8200000	30700000	8
4	5	5	Not Graduate	Yes	9800000	24200000	20

	education_Graduate	education_Not Graduate	self_employed_No	self_employed_Yes	no_of_dependents
2488	1.0	0.0	0.0	1.0	2
3703	0.0	1.0	0.0	1.0	1
3347	0.0	1.0	1.0	0.0	2
3487	0.0	1.0	1.0	0.0	5
3957	0.0	1.0	1.0	0.0	0

FEATURE ENGINEERING

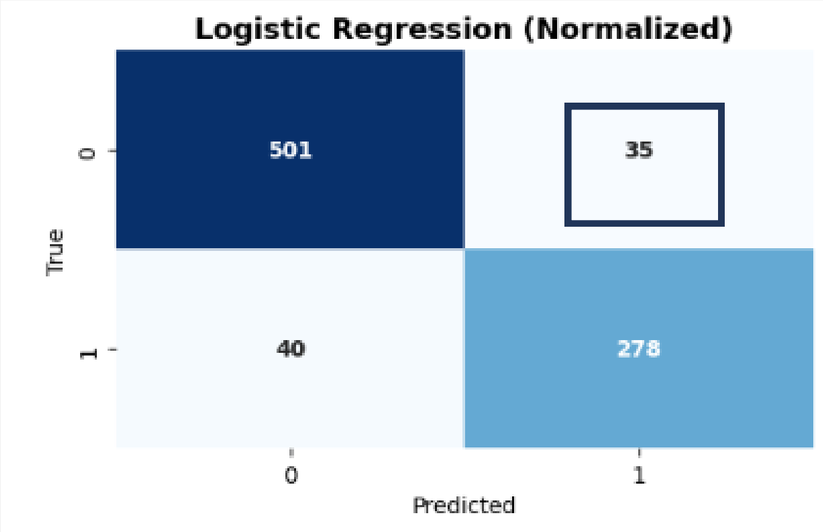
Normalisation versus Standardisation

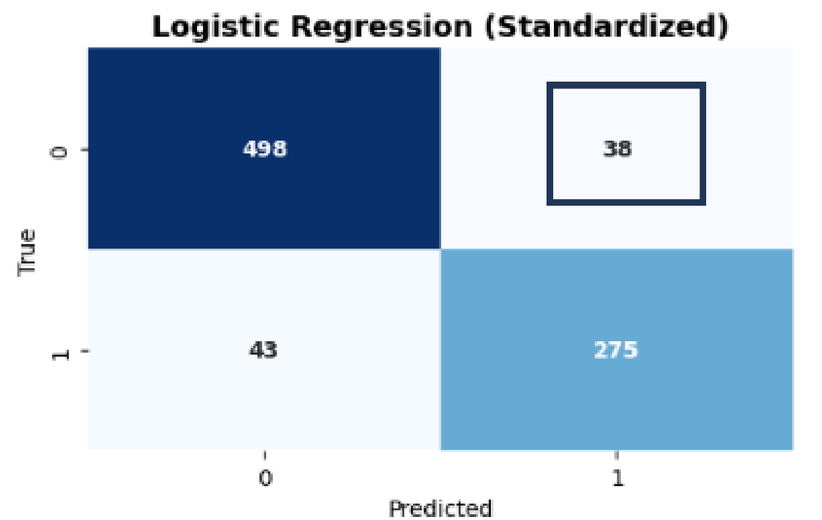
- We have values in the test set that are outside these values' ranges in the train set
- We have outliers in some of the columns

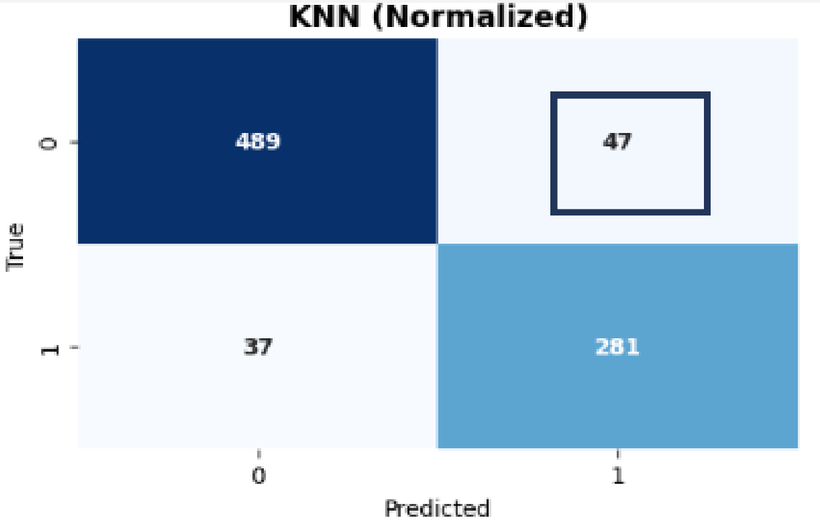


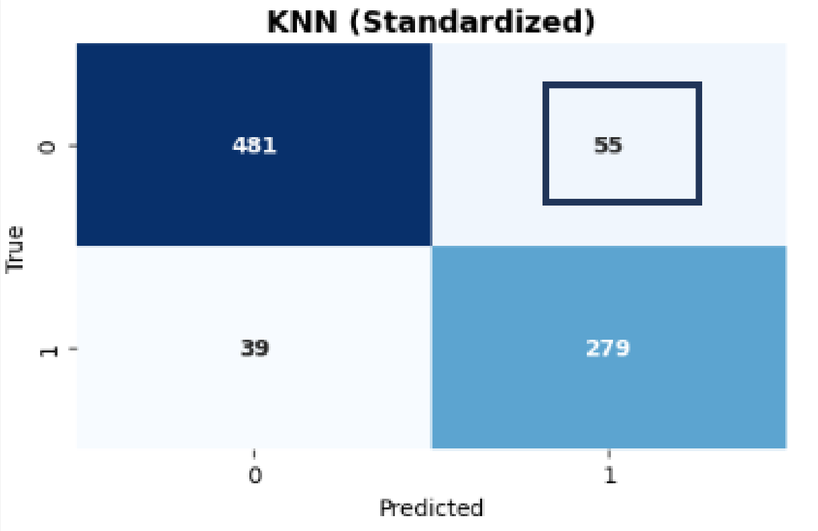
FEATURE ENGINEERING

Normalisation versus Standardisation









MODEL BUILDING

Types of Models experimented

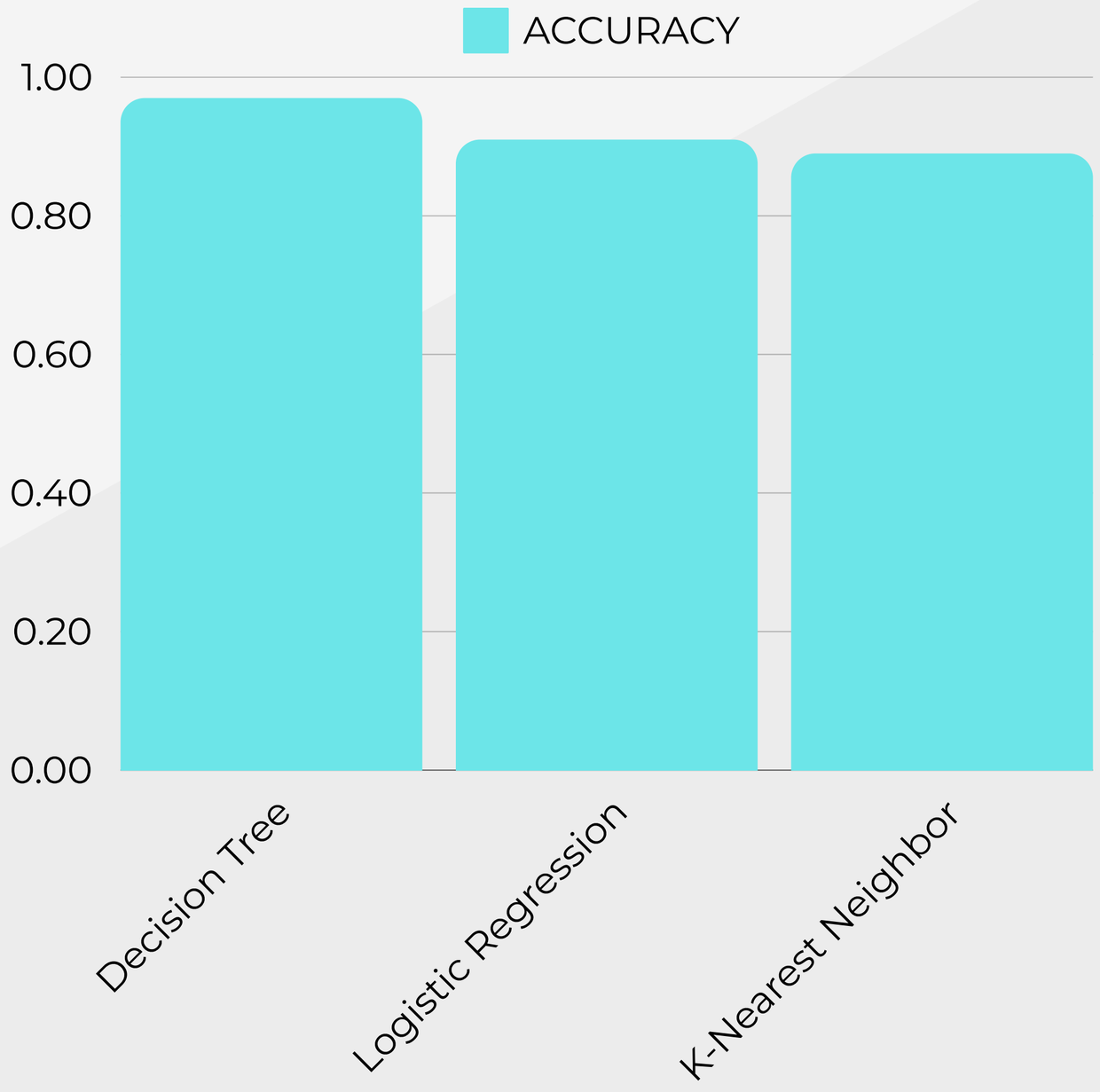
Decision tree

KNN

Logistic regression

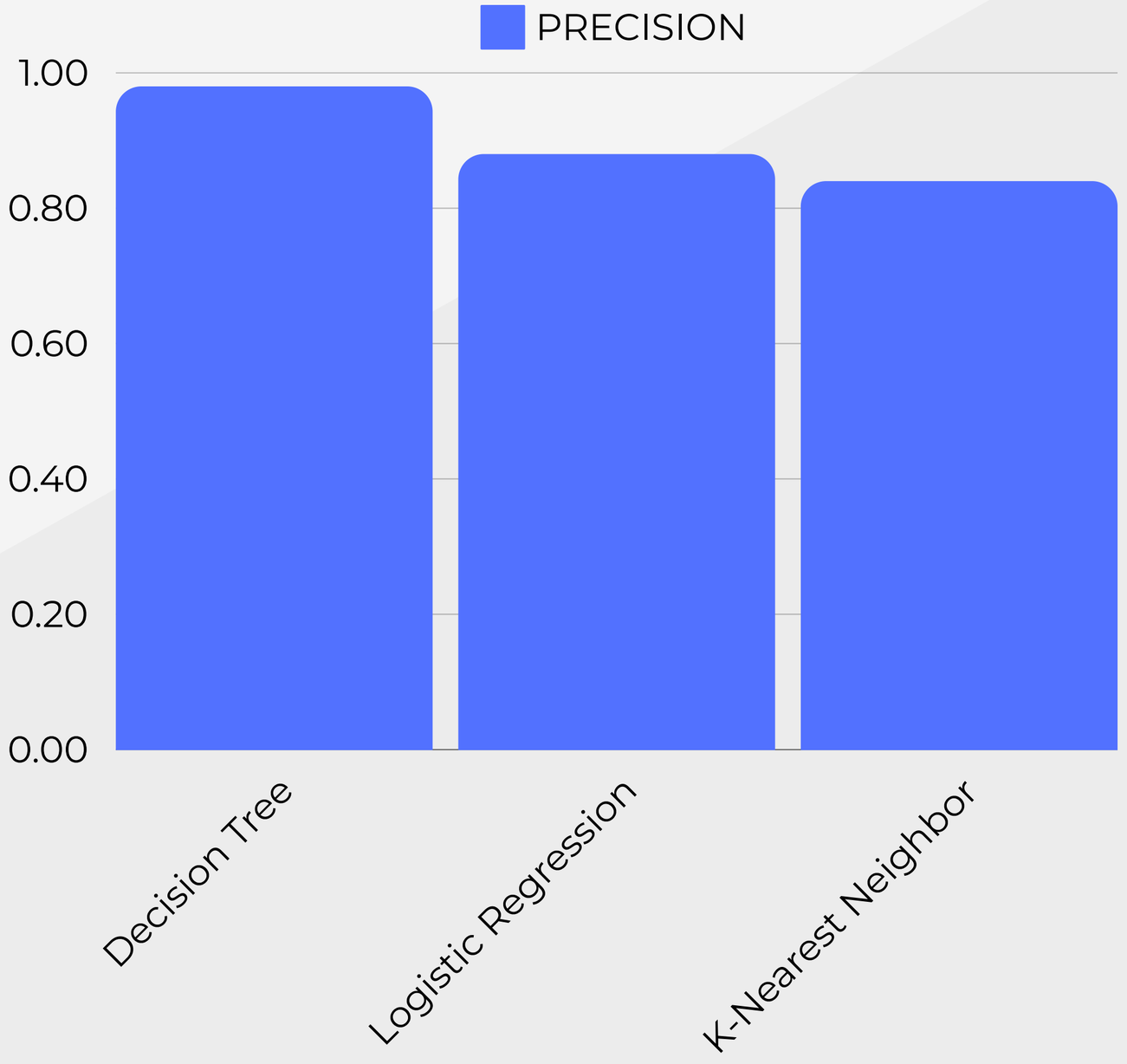
MODEL EVALUATION

COMPARISON OF MODEL PERFORMANCE



MODEL EVALUATION

COMPARISON OF MODEL PERFORMANCE



HYPERPARAMETER TUNING AND MODEL OPTIMIZATION FOR DECISION TREE

GridSearch

```
param_grid = {
    'max_depth': [3, 5, 10, None],
    'min_samples_split': [2, 10, 20],
    'min_samples_leaf': [1, 5, 10],
    'criterion': ['gini', 'entropy']
}
```

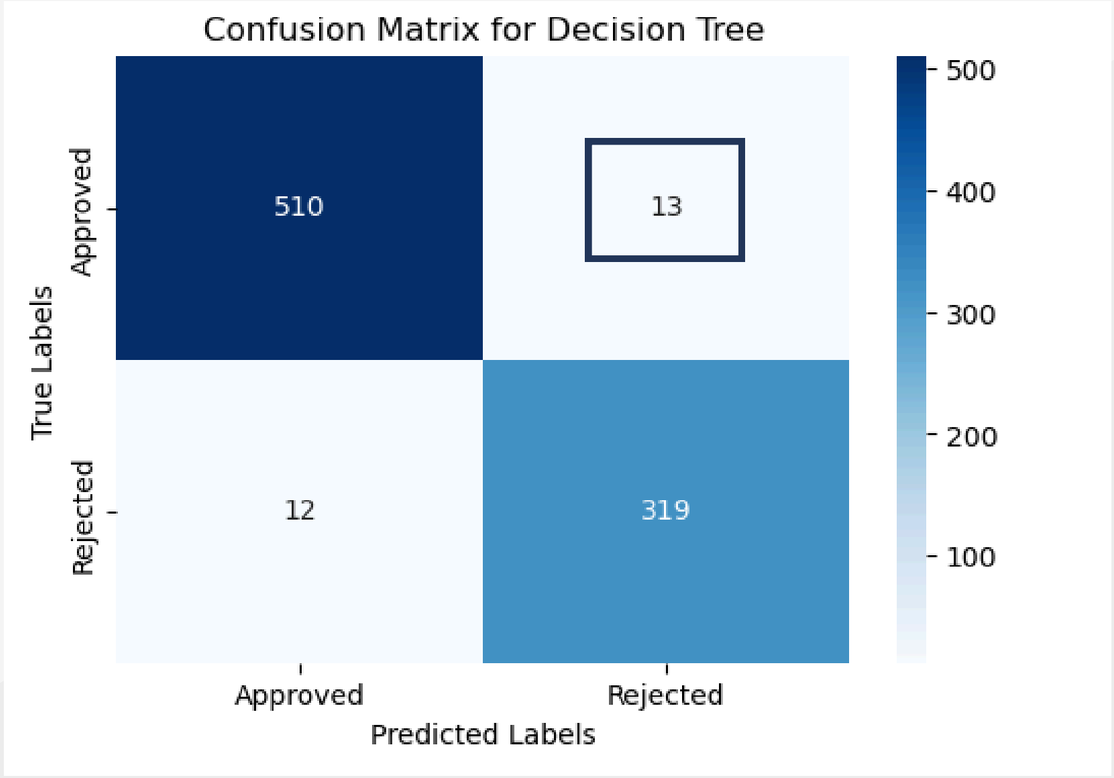
Best Hyperparameters

- **Max_depth:** none
- **Min_samples_leaf:** 1
- **Min_samples_split:** 10
- **Criterion:** entropy

Best Score: 0.97

No improvement, same score compared to initial accuracy score (0.97)

CONFUSION MATRIX - DECISION TREE



CHECKING FOR MODEL OVERFITTING

ACCURACY OF THE SPLIT

Training set Accuracy: 0.9646

Test set Accuracy: 0.9649

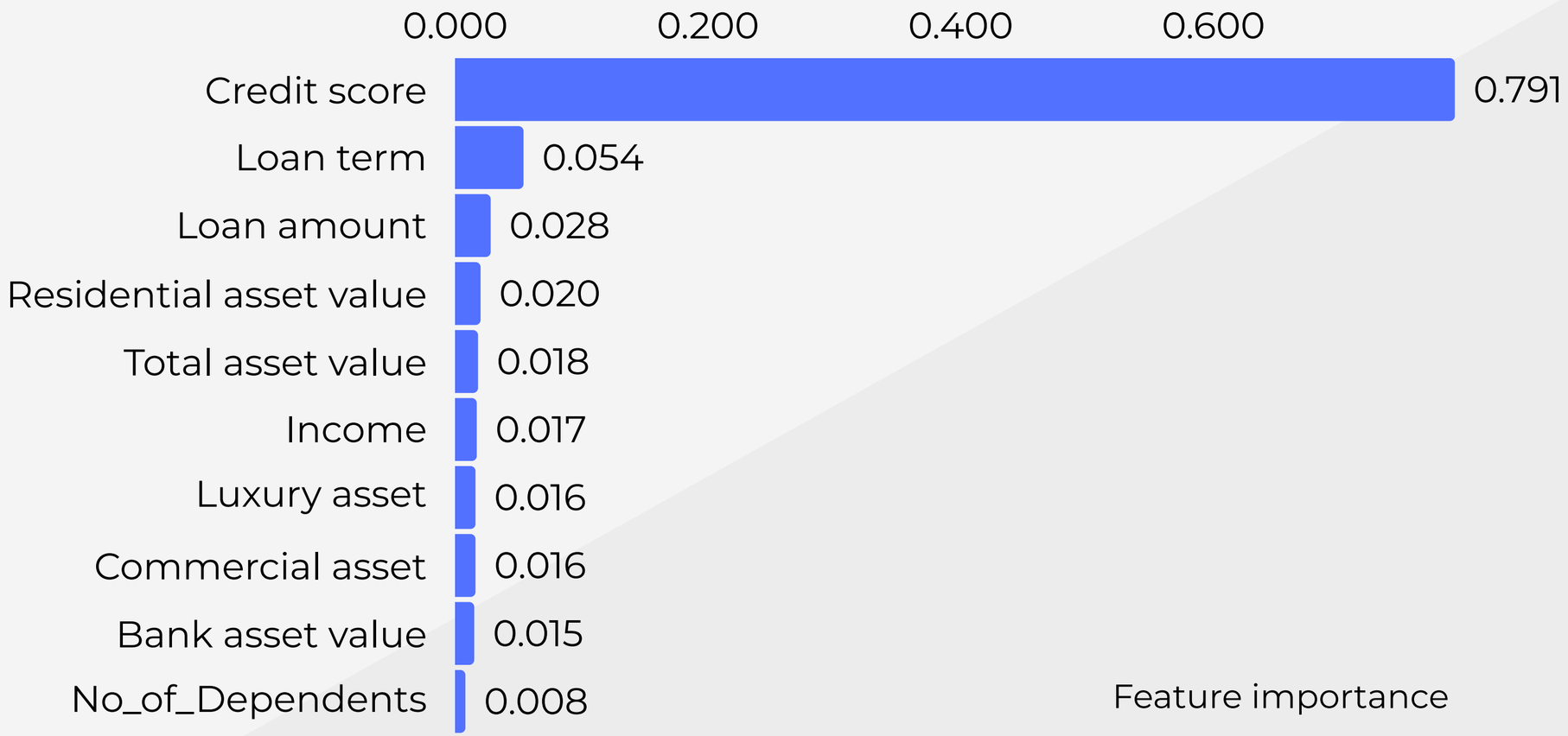
ERROR CHECKING

Training Error: 0.0354

Test Error: 0.0351

Our model shows no over-fitting or under-fitting

FEATURE IMPORTANCE



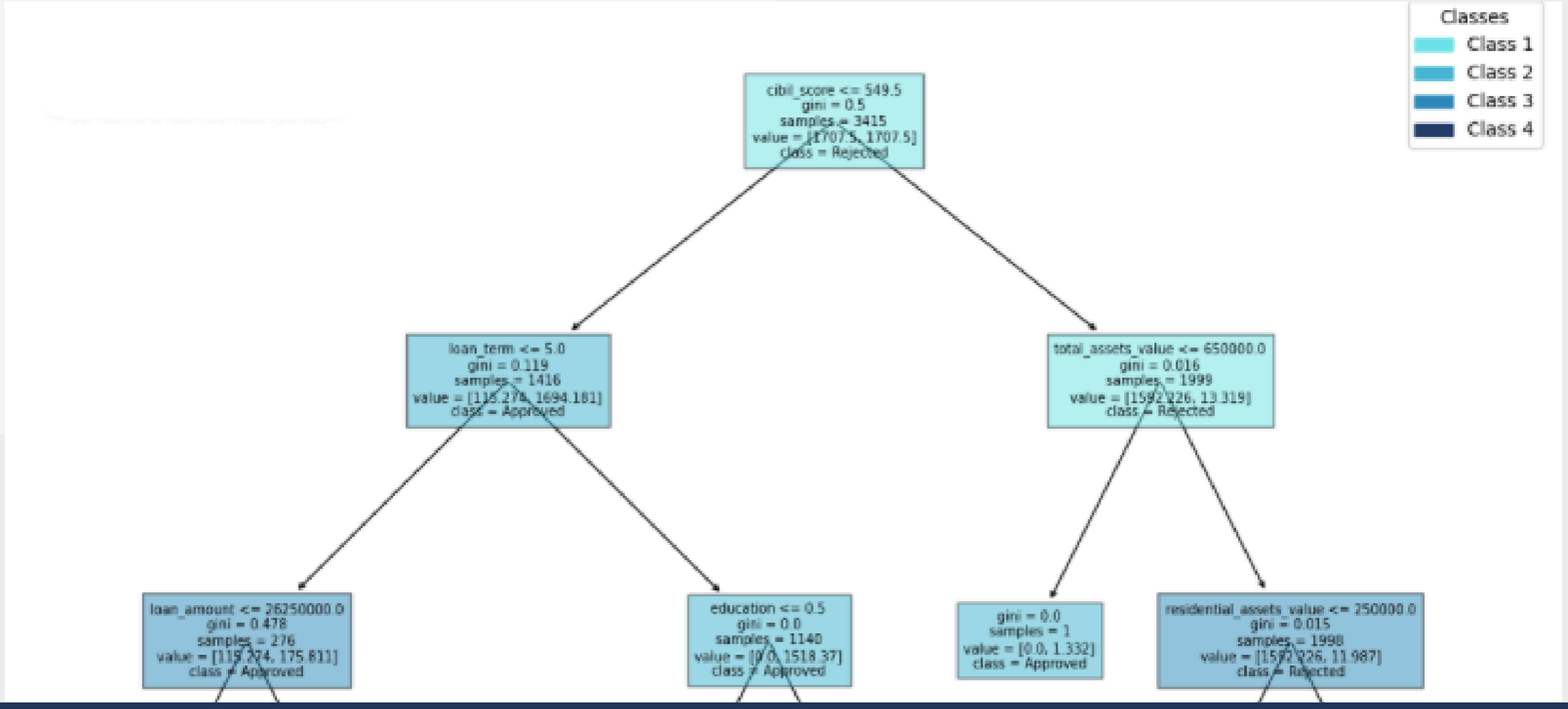
KEY FINDINGS AND INSIGHTS

Best Performing Models:

Decision Tree consistently achieved the best accuracy (~97%) and impressive precision compared to the other models.

Feature Importance:

Credit Score is the most influential feature, indicating its critical role in determining loan approvals.



REAL-WORLD APPLICATION AND IMPACT

Application:

. Implementation in financial institutions to automate and enhance the loan approval process, allowing for more efficient assessments of applicants.

Potential Impact:

· Reducing the time for processing loan approvals
. Increasing accuracy
. Better risk management and increased customer satisfaction

Ethical Considerations:

· Identifying and addressing potential biases in data to ensure fair loan approvals while complying with regulatory standards in lending practices.

CHALLENGES AND LEARNINGS

Challenges:

- Comprehensive evaluation
- Choosing the right models and tuning them for optimal performance.

Solution:

- Used various performance metrics and confusion matrices,
- Experimenting with different algorithms and preprocessing techniques to identify the best-fit model.

Learning:

Trying different machine learning algorithms is crucial to discovering the most suitable model for the data.

FUTURE WORK AND IMPROVEMENTS

- 5 Cs of credit analysis are:
  Character, **Capacity**, **Capital**, **Collateral**, and **Conditions.**

- Improving the model by:
  - ignoring less important information.
  - analysing additional information (debt-to-income ratio, prevailing interest rates).

- Calculating the cost of potential False Positive depending on the loan amount.

THANK YOU

Unlock the power of machine learning to transform tomorrow, today.

JOEL - JOHANA - LUCIE - MANUEL