**PRCP-1000-ProtugeseBank**

|  |
| --- |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | | Model |  | | --- | --- | | best\_score | best\_params |
| 1 | svm | 907602 | {'C': 1, 'kernel': 'rbf'} |
| 2 | random\_forest | 0.905621 | {'n\_estimators': 10} |
| 3 | logistic\_regression | 0.910348 | {'C': 10} |

According to the result we can predict the best model for the above project will be Logistic regression as best score for the logistic regression is 0.91% among all best.

# Create a report stating the performance of multiple models on this data and suggest the best model for production.

logistic Regression

# Report on Challenges faced

(1) Data Quality and Quantity:

Insufficient or poor-quality data can significantly impact the performance of machine learning models. Imbalanced datasets may lead to biased or inaccurate predictions. Here in this data unknown values doing the same effect on data.

#Missing value 'job' variable: 330

#Missing value 'marital' variable: 80

#Missing value 'education' variable: 1730

#Missing value 'default' variable: 8596

#Missing value 'housing' variable: 990

#Missing value 'loan' variable: 990

#Missing value 'contact' variable: 0

#Missing value 'month' variable: 0

#Missing value 'day\_of\_week' variable: 0

#Missing value 'poutcome' variable: 0

(2) Feature Engineering:

Identifying and selecting relevant features is a crucial but challenging task. Dealing with high-dimensional data and extracting meaningful features can be complex. Here in this dataset out of 21 columns only some are relevant data other than that not useful according to requirement which effect the result of machine learning result.

(3) Model Selection and Hyperparameter Tuning:

Choosing the right algorithm for your specific problem and tuning its hyperparameters can be time-consuming. Overfitting or underfitting may occur if the model is not appropriately tuned. here in this project time consuming for model selection and for hyperparameter is very high might.

(4) Interpretability and Explainability:

Understanding and interpreting the decisions made by complex machine learning models is often challenging. Explainability is crucial, especially in applications where model decisions impact human lives or have legal implications. here in this project related to financial occupancy so due to that very sensitive depending project.

(5) Computational Resources:

Training deep learning models can be computationally intensive and may require powerful hardware. Optimization of algorithms to run efficiently on available resources is a common concern.

(6) Deployment and Integration:

Translating a trained model into a production-ready system can pose challenges. Integrating machine learning solutions with existing workflows or applications may require additional effort.

# Conclusion

As we can see from the analysis there is enough information to predict customer behavior with high accuracy. Due to the unbalanced nature of the dataset, we need to pay attention to the minority class, i.e. call duration. here in this dataset age reliving depends some of the results which we can able to understand by using default classifiers, there was 70% bal. accuracy was achieved, after introducing the class weight, model improved by .10, with total accuracy 87%. Final model is LogisticRegression(C=1.0,classweight = balanced).