



---

# BANK CUSTOMER CHURN CLASSIFICATION

# Summary

- The project sought to assess whether we can accurately predict which customers likely to churn.
- Information is vital to Bank Executives, Marketing and Retention Team and Customer Service Representatives.
- Took an iterative approach to modeling by building multiple models:
- A parameter tuned Random Forest algorithm chosen as the final model
- Performed the best in terms of accuracy, precision & RMSE..
- The model could accurately predict which customers are likely to churn with an accuracy of 85%, precision of 87.5% and RMSE of 0.39.

# Outline

1

**Business Problem**

2

**Data**

3

**Modelling**

4

**Results**

5

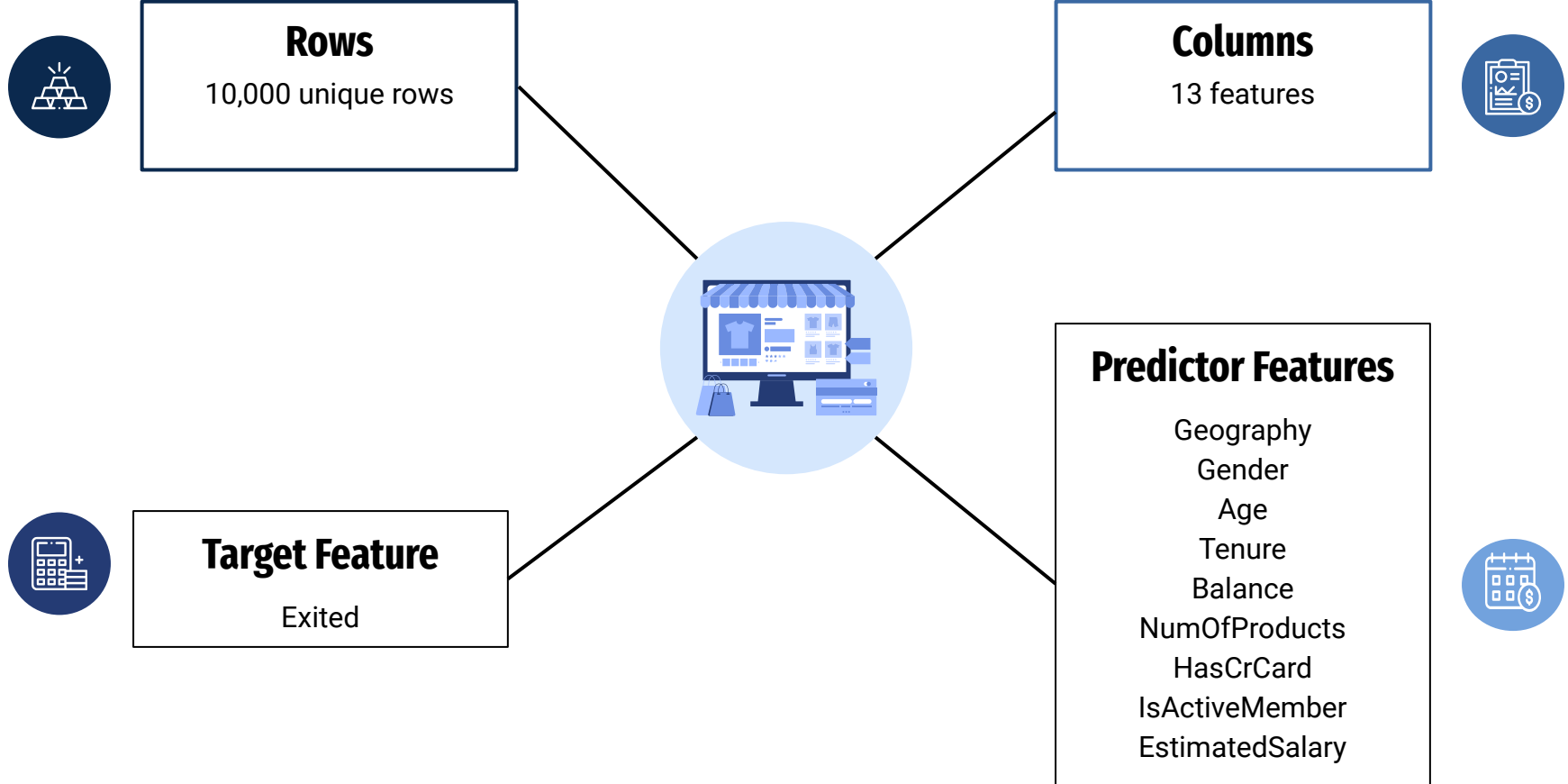
**Conclusion**



# Business Problem

- Model aims to identify patterns and indicators that suggest a higher likelihood of churn.
- Goal is to enable bank prioritize retention efforts and develop tailored strategies to retain customers at risk of churn.
- Key stakeholders are Bank Executives, Marketing and Retention Team and Customer Service Representatives
- Objective: "To accurately predict which customers are likely to churn based on their age, location and account information."

# Data Understanding



# Data Understanding

- No missing values
- 80% of entries are for customers who did not churn, 20% for customers who churned
- 50% of the records are for French customers, while the rest are for German and Spanish customers.

# Methods

## Method Used:

- Iterative Approach to Classification Modelling

## Technologies

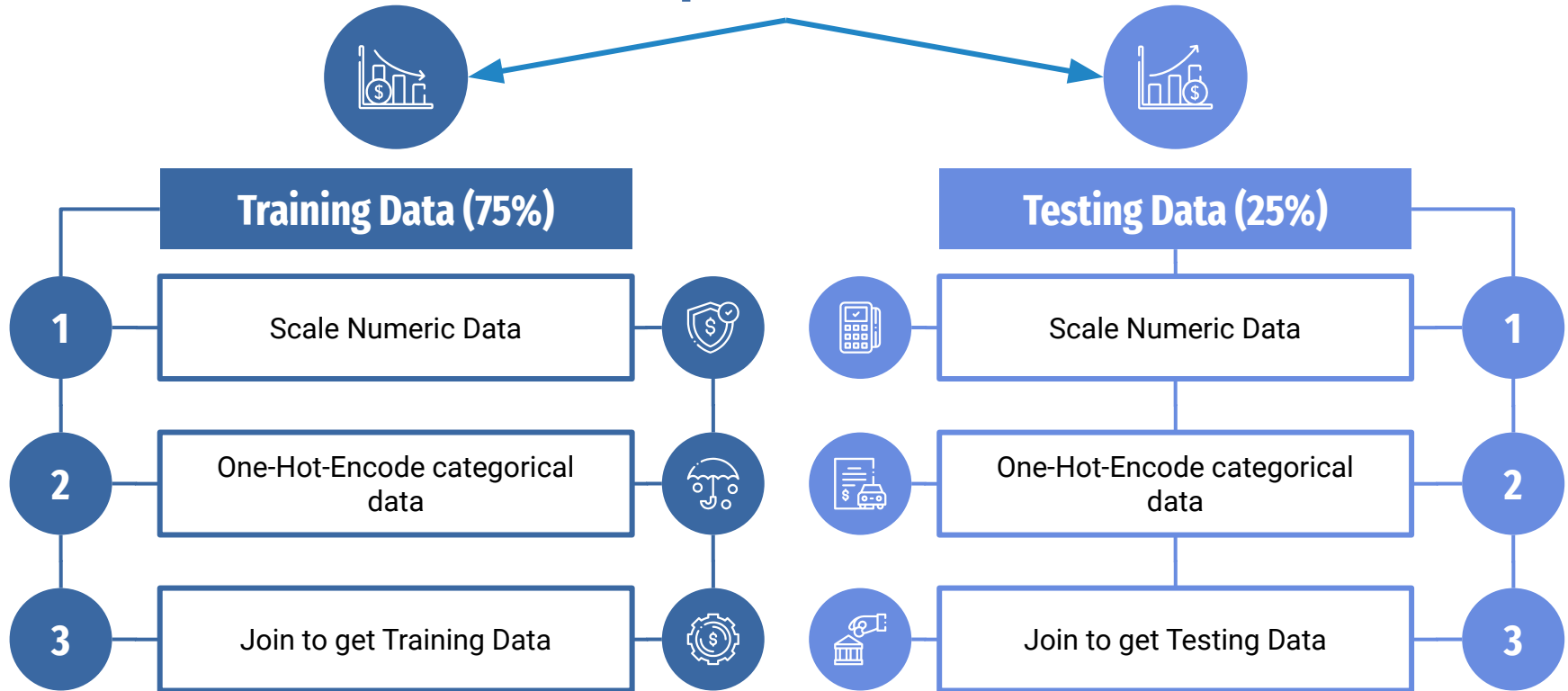
- Python Programming Language
- Jupyter Notebook
- Libraries: Pandas, Matplotlib, Seaborn, Scikit-learn

## Evaluation Metrics

- Precision, Accuracy, Root Mean Squared Error

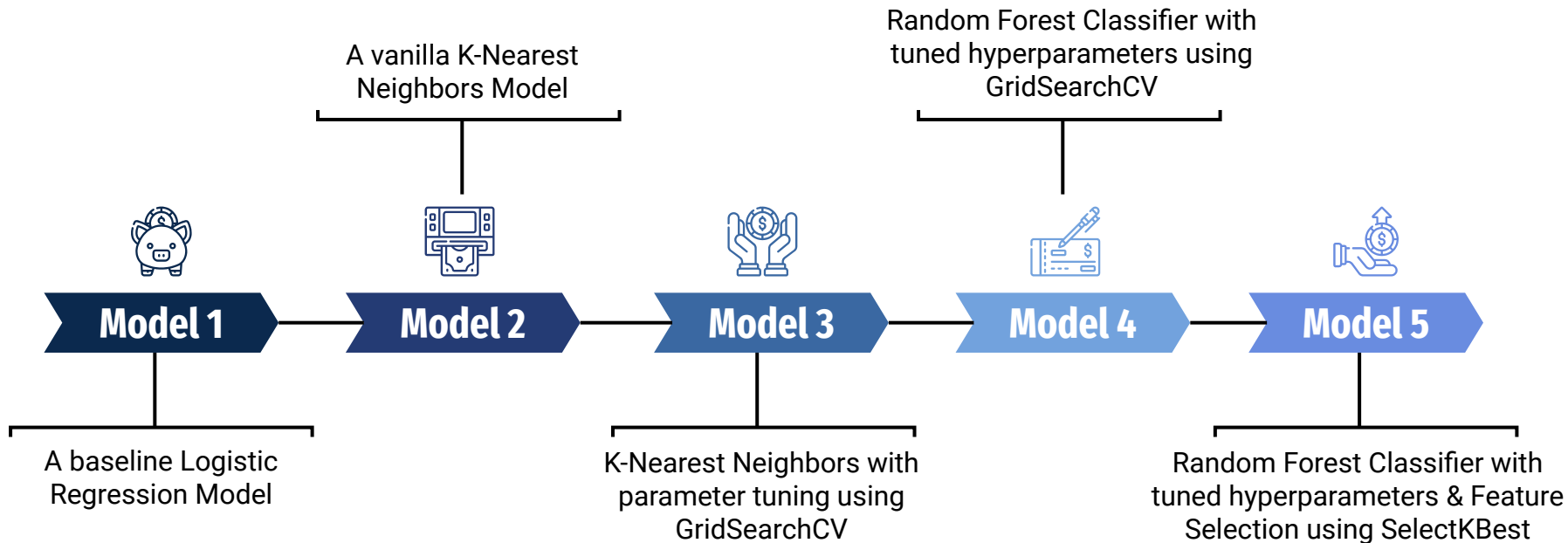
# Data Preparation

## Split Data

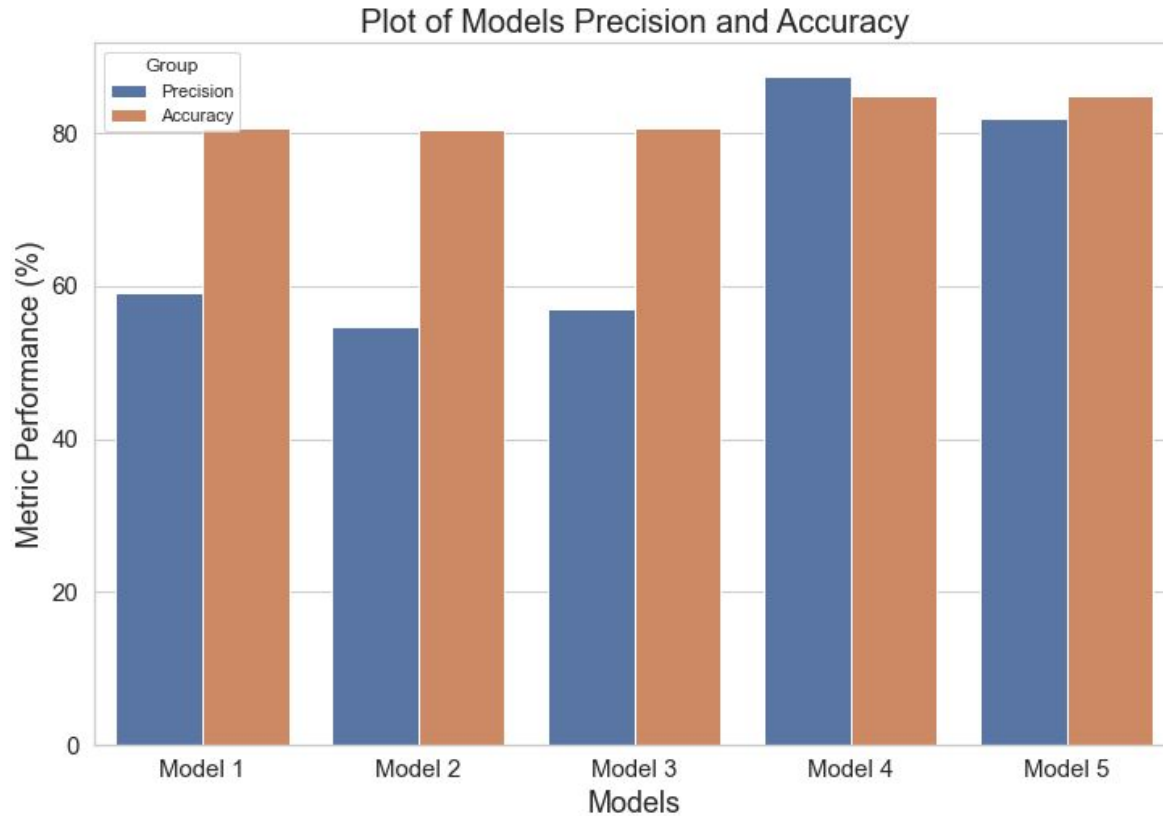




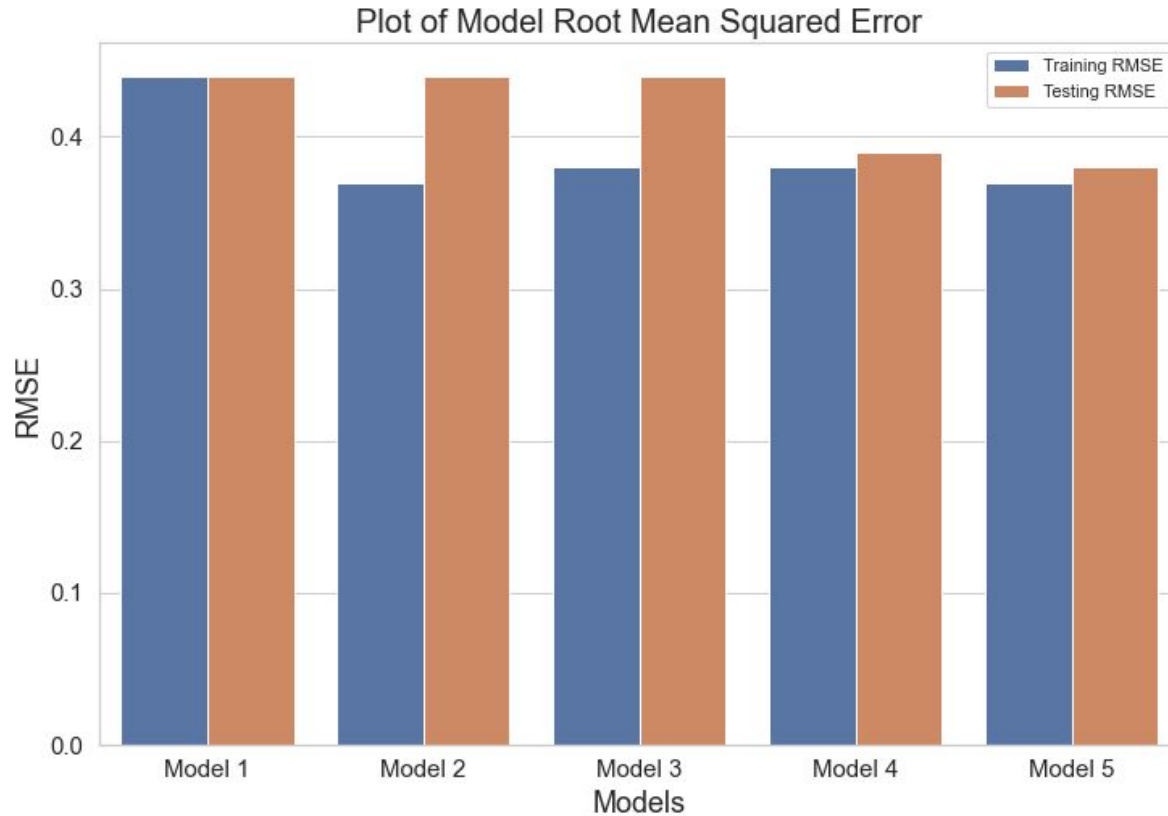
# Modelling



# Results

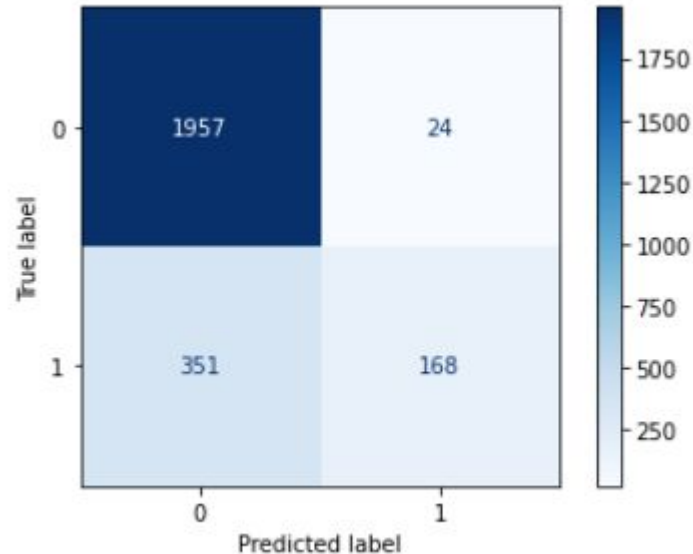


# Results



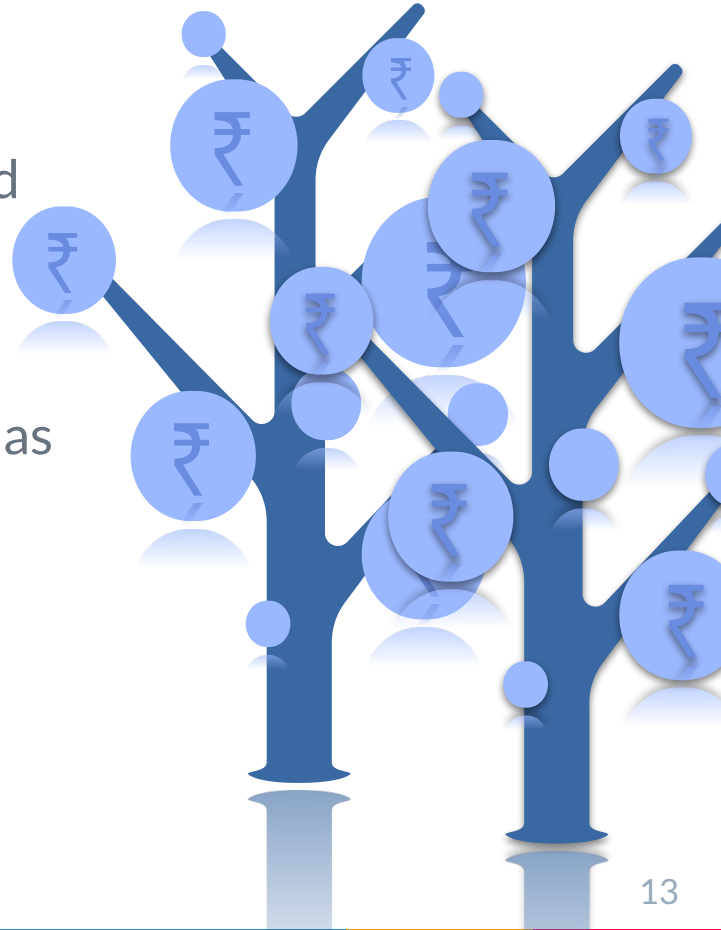
# Results

- **Model 4** chosen as the FINAL MODEL
- It is a **Random Forest Classifier** that is **hyperparameter tuned**
- Accuracy = 85%, Precision = 87.5%, Training RMSE = 0.38, Testing RMSE = 0.39
- Confusion Matrix:

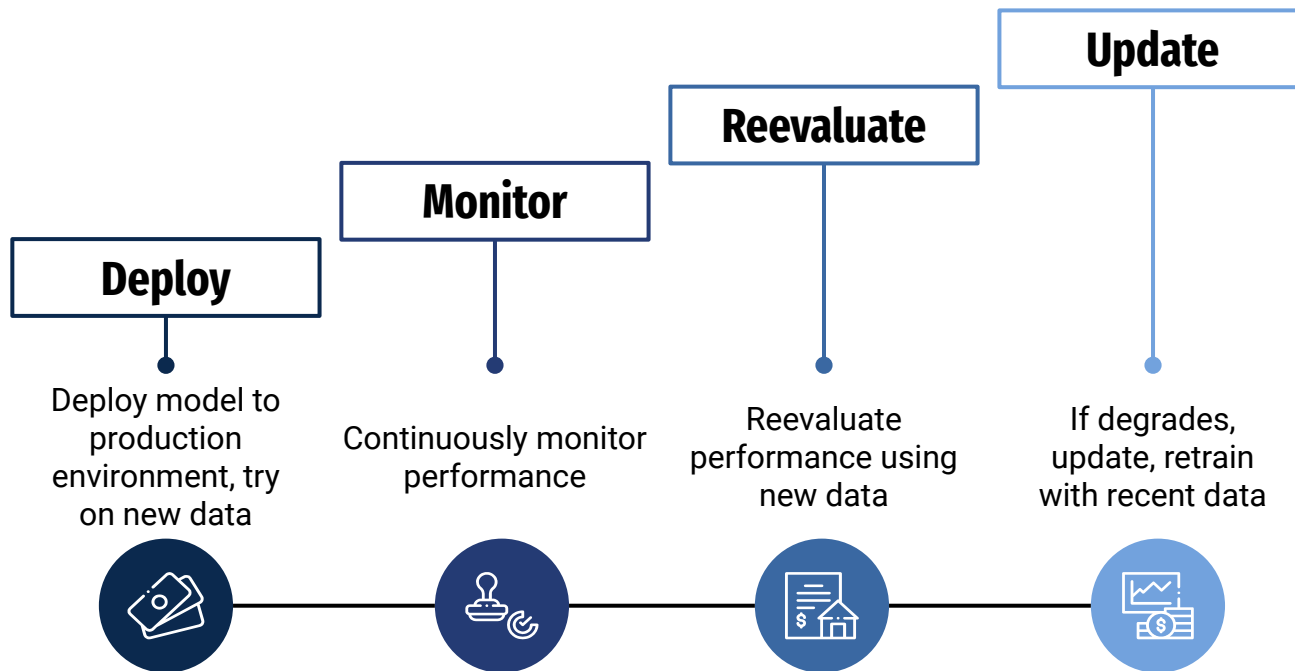


# Conclusion

- Hyperparameter tuned **Random Forest** model demonstrated power and effectiveness of analyzing the dataset and predicting customer churn.
- Model's performance was validated on unseen data,
- Deployment of this classification model has potential to offer tangible benefits to the bank executives.
- Study challenges: Limited Training Data



# Next Steps



# Thanks!

## Any questions?

You can find me at:

Email: [githinjisteve96@gmail.com](mailto:githinjisteve96@gmail.com)

GitHub: <https://github.com/stevegithinji>

LinkedIn: [linkedin.com/in/steve-githinji-10ba0114a](https://www.linkedin.com/in/steve-githinji-10ba0114a)