

RISKY LOAN APPLICANT CLASSIFICATION

Michael Parsaoran

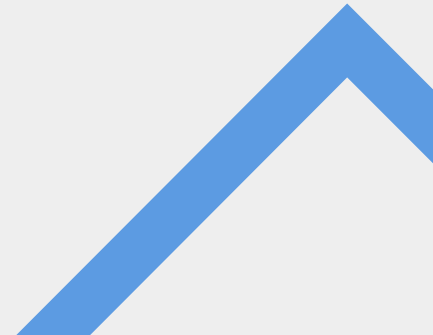
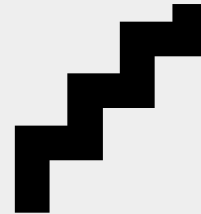


TABLE OF CONTENTS

01

Objective

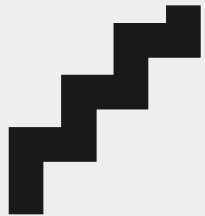
02

**Model Building
Process Overview**

03

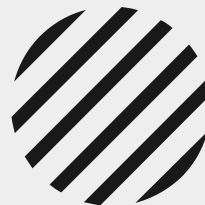
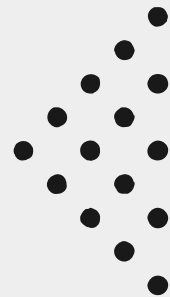
**Conclusion &
Recommendation**





01

OBJECTIVE



OBJECTIVE

Situation

The Bank of India is experiencing a high volume of loan applications, with over 600,000 people applying for loans.

Task

Develop and implement an automated model to predict loan approval for non-risky applicant.

Action

Create a model using machine learning algorithm that processes applicant data real-time and flags whether the applicant a risky or non-risky applicant.

Result

Expected to escalate efficiency in processing loan applicant's data, thus reducing workload of the risk management division.





02

**MODEL BUILDING
PROCESS INTERVIEW**

WORKFLOW

01

Data Collecting

02

EDA

03

Data Preprocessing

04

Model Training

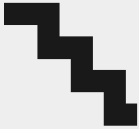
05

Model Evaluation

06

Conclusion

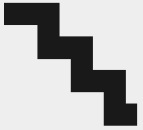




DATA OVERVIEW

| | |
|------------------------|--|
| ID | Unique identifier for each loan applicant. |
| Income | The income level of the applicant. |
| Age | Age of the applicant. |
| Experience | Years of professional experience. |
| Married/Single | Marital status of the applicant. |
| House_Ownership | Indicates whether the applicant owns or rents a house. |

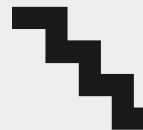




DATA OVERVIEW

| | |
|--------------------------|--|
| Car_Ownership | Indicates whether the applicant owns or rents a house. |
| Profession | Occupation or profession of the applicant. |
| City | City of residence of the applicant. |
| State | State of residence of the applicant. |
| Current_Job_Yrs | Duration of employment in the current job. |
| Current_House_Yrs | Duration of residence in the current house. |





DATA OVERVIEW

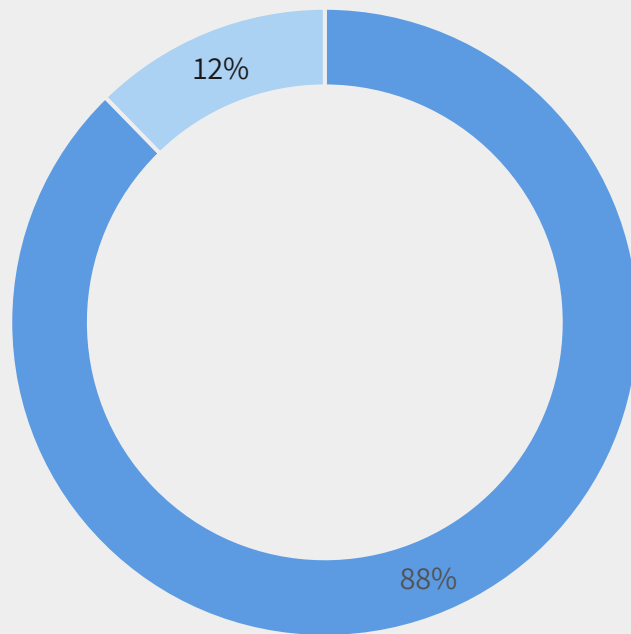
Target Feature

| | |
|------------------|---|
| Risk_Flag | Binary indicator of loan risk, where 1 represent a flagged risky applicant and 10 represents a non-risky applicant. |
|------------------|---|



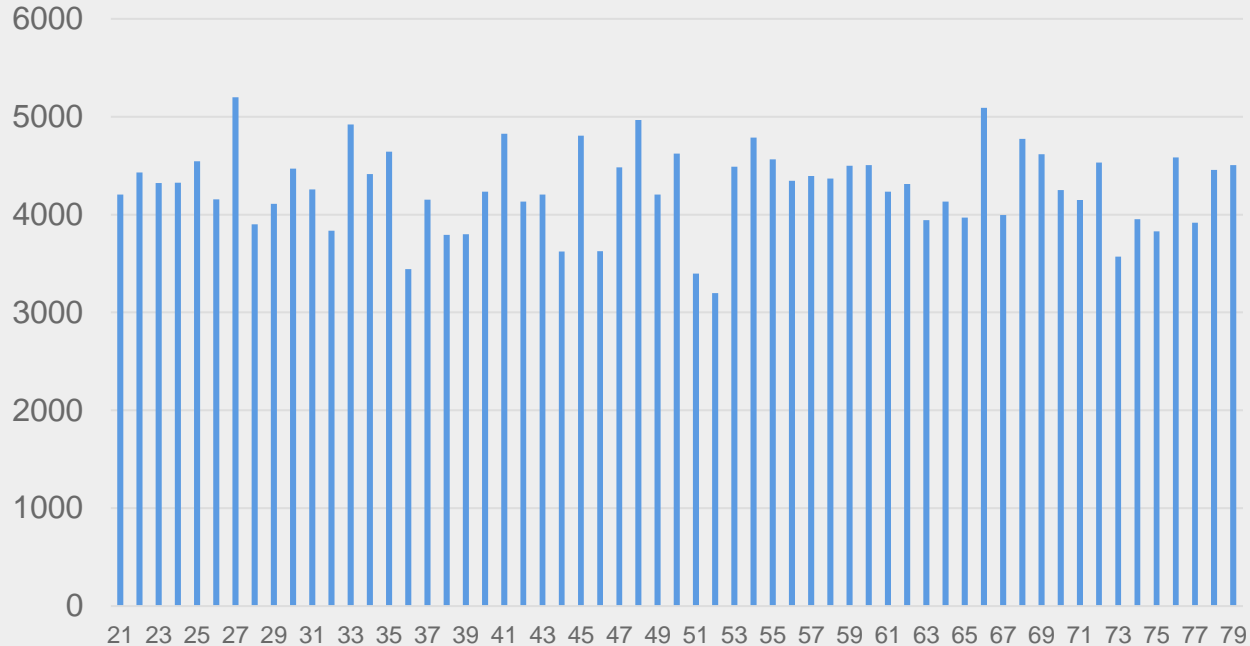


APPLICANT RISK FLAG PERCENTATION



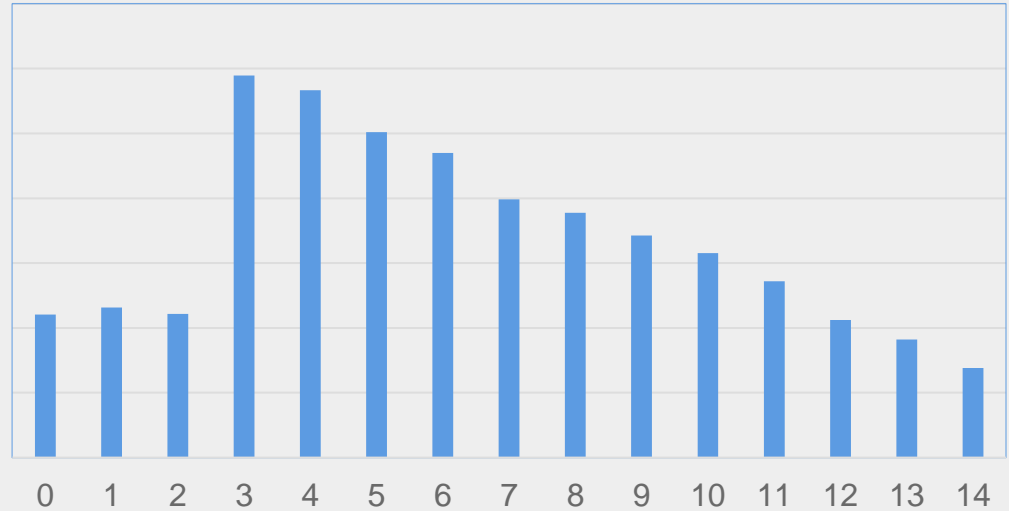
■ Non-risky ■ Risky

+APPLICANTS AGE DISTRIBUTION



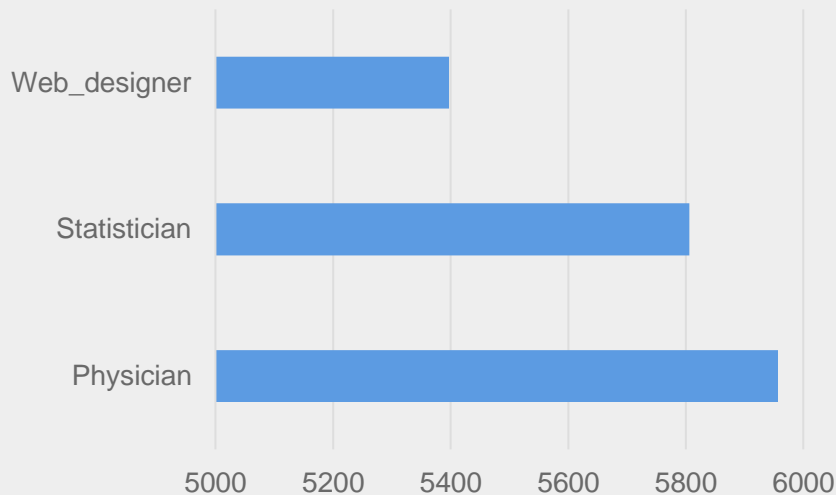


APPLICANT'S CURRENT PROFESSION YEAR OF EXPERIENCE

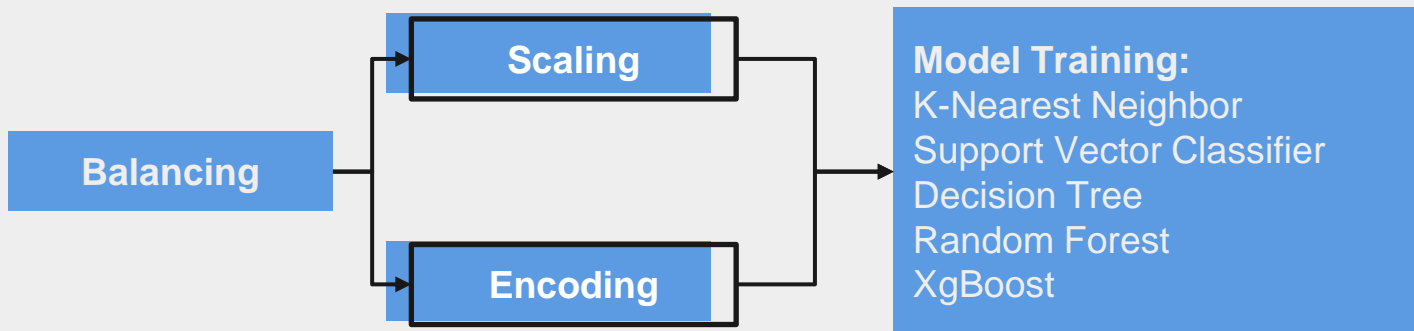




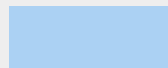
TOP 3 APPLICANT'S PROFESSION



PIPELINE

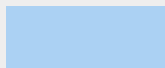


METRICS USED



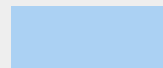
Precision

Precision is a metric that measures the accuracy of the positive predictions made by the model. In the context of predicting loan approvals, precision tells us how many of the loans that the model predicted as approved are actually approved.



Recall

Recall is a metric that measures the ability of a model to identify all relevant instances of a particular class. In the context of predicting loan approvals, recall tells us how well the model is at identifying all the loans that should be approved.



F1 Score

F1 score is a useful metric for measuring the performance for classification models when you have imbalanced data because it takes into account the type of errors.



+MODEL PERFORMANCE

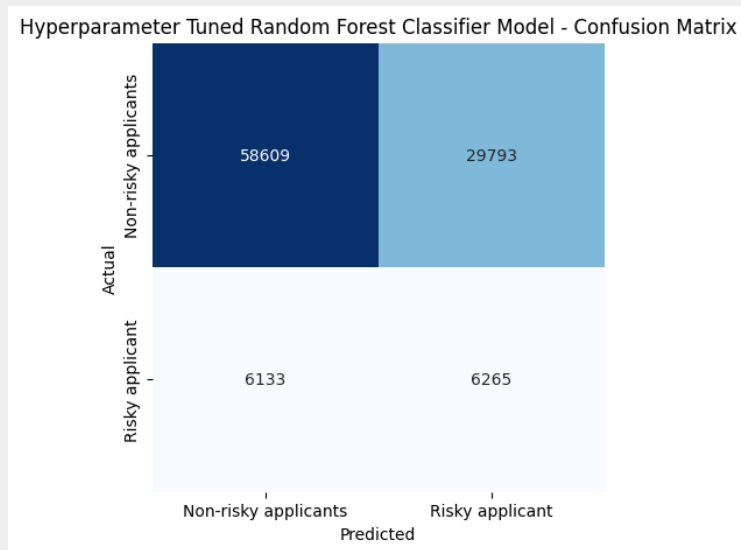
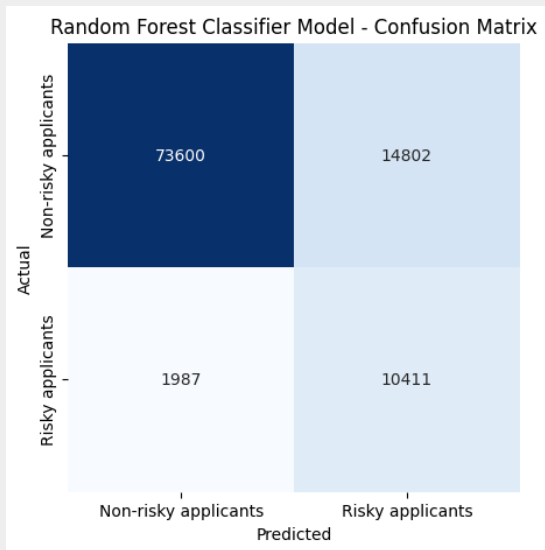
| Base Model Name | KNN | SVC | Decision Tree | Random Forest | XGBoost |
|-----------------|--------|-------|---------------|---------------|---------|
| Train Score | 0.801 | 0.571 | 0.929 | 0.921 | 0.805 |
| Test Score | 0.7632 | 0.502 | 0.82 | 0.835 | 0.748 |
| Precision Train | 0.802 | 0.561 | 0.898 | 0.893 | 0.794 |
| Precision Test | 0.314 | 0.146 | 0.405 | 0.412 | 0.298 |
| Recall Train | 0.799 | 0.648 | 0.967 | 0.963 | 0.825 |
| Recall Test | 0.787 | 0.631 | 0.858 | 0.839 | 0.779 |
| F1 Score Train | 0.800 | 0.601 | 0.931 | 0.927 | 0.809 |
| F1 Score Test | 0.450 | 0.237 | 0.551 | 0.553 | 0.432 |
| CV F1 Mean | 0.763 | 0.594 | 0.835 | 0.830 | 0.763 |

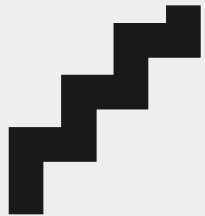
+MODEL PERFORMANCE

| Model Name | Random Forest | Tuned Random Forest |
|-----------------|---------------|---------------------|
| Train Score | 0.921 | 0.571 |
| Test Score | 0.835 | 0.502 |
| Precision Train | 0.893 | 0.561 |
| Precision Test | 0.412 | 0.146 |
| Recall Train | 0.963 | 0.648 |
| Recall Test | 0.839 | 0.631 |
| F1 Score Train | 0.927 | 0.601 |
| F1 Score Test | 0.553 | 0.237 |
| CV F1 Mean | 0.830 | 0.594 |



DATA OVERVIEW

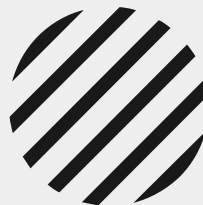




03



CONCLUSIONS



CONCLUSIONS



Base Model

Consistently detects risky applicant with training dataset, and has bad performance when the model uses dataset that the model haven't seen.



Tuned Model

The model has a bad performance when predicting risky loan applicant,

RECOMMENDATION

- Create a more extensive hyperparameter tuning using grid search or random search to find the optimal settings for the model.

Slidesgo

+91 651 736 904

charlie@slidesgo.com



Jane Doe

+91 620 421 838 | jane@freepik.com



Thank You

Any questions?