PRUDENTIAL LIFE INSURANCE

Classifying Customer Risk Score

OVERVIEW

INTRODUCTION

Prudential Financial

Business Case

DATA PREPARATION

Data Cleaning

Feature Engineering

MACHINE LEARNING MODEL

Feature Selection

Result evaluation

CONCLUSION

Business Values

Future Improvement

INTRODUCTION

INTRODUCTION

- Global financial services and insurance company founded in 1875
- Provides insurance products to help protect the financial security for millions of customers and businesses.
- Life Insurance, health insurance, asset management,...

PRUDENTIAL FINANCIAL, INC

HOW IT WORKS

Customer submits request

Health check-up

Company evaluates the application

Company informs the result to the customer

Customer decides to sign the contract

BUSINESS CASE

CLASSIFY CUSTOMER RISKS BY LEVELS

REDUCE BUSINESS COSTS

INSURANCE OPERATING GROUPS INDIVIDUALS INSURED

DATA PREPARATION

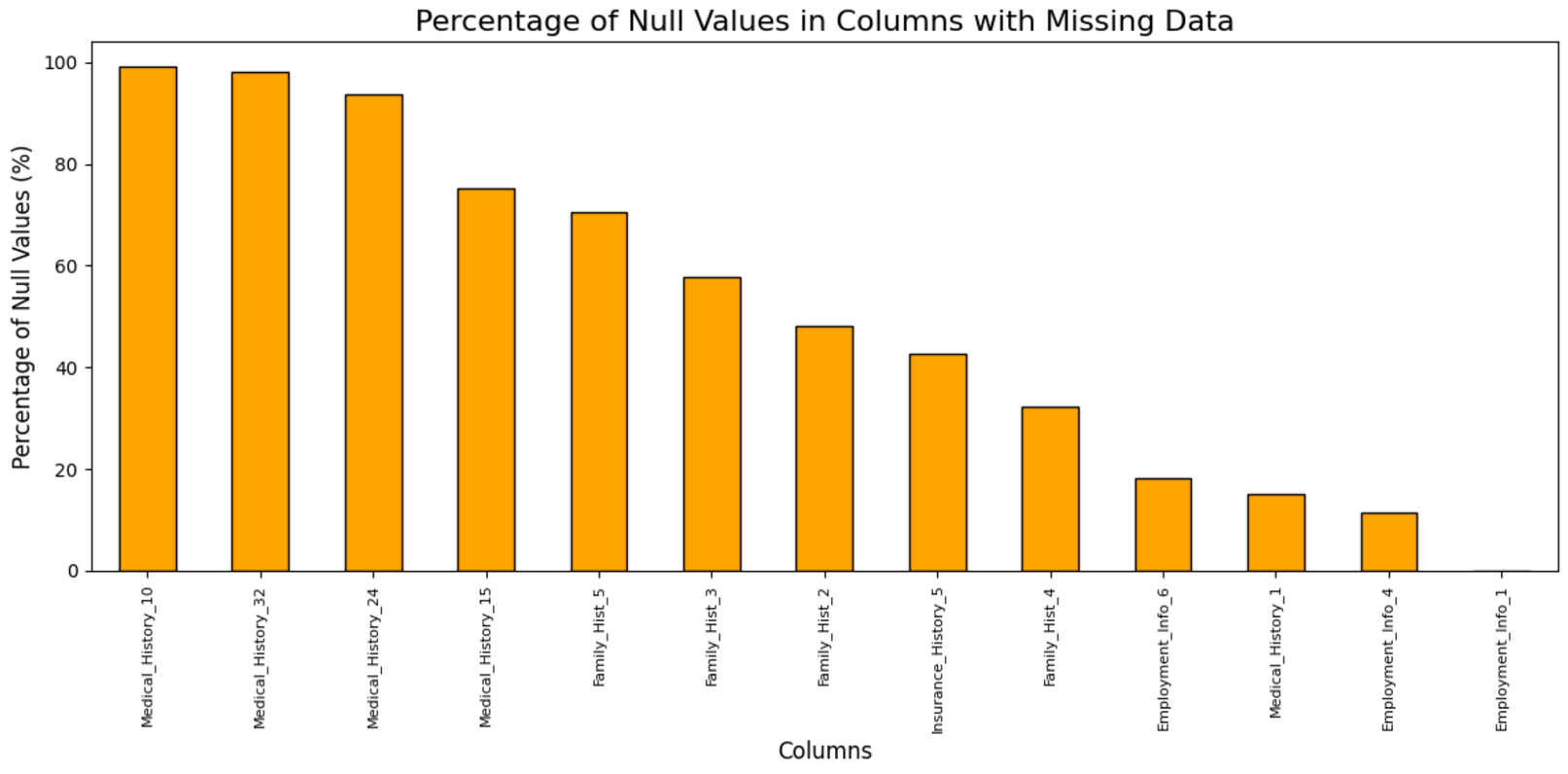
DATASET

128 COL

59382-ROW

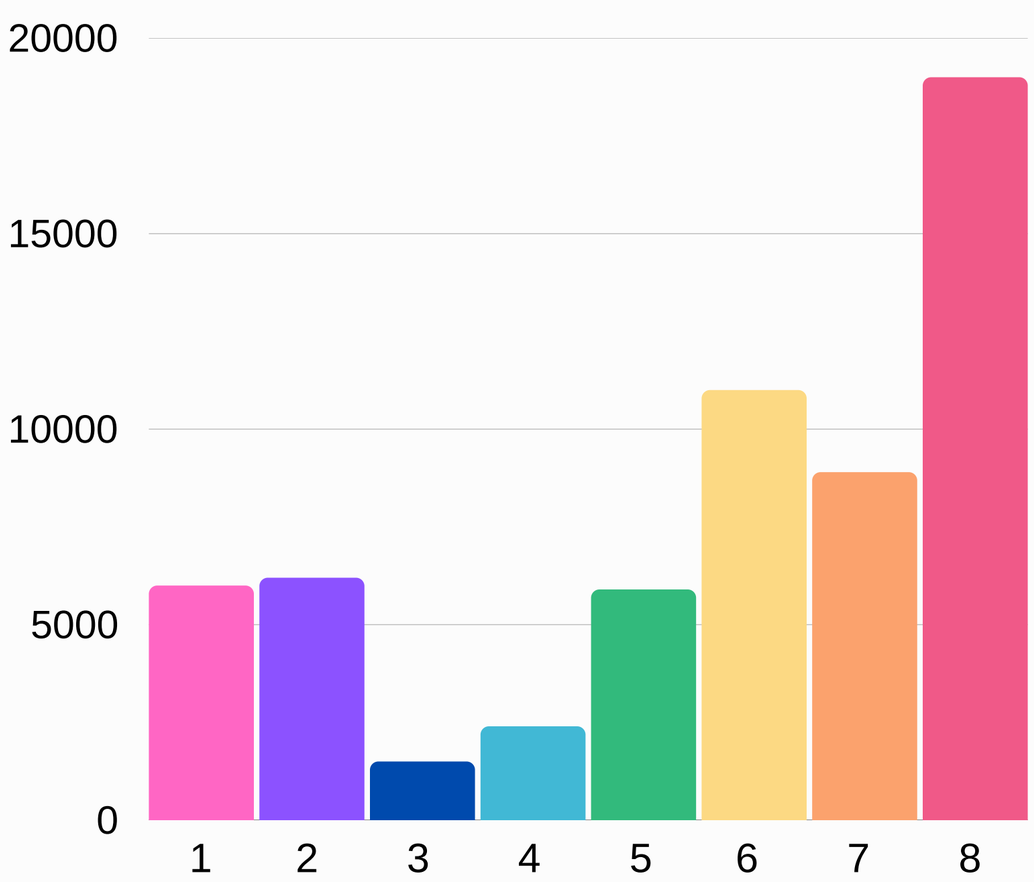
BODY METRICS	Ins_Age, Ht, Wt, BMI
MEDICAL HISTORY	Medical_History, Family_Hist, Medical_Keyword
EMPLOYMENT HISTORY	Employment_Info
INSURANCE HISTORY	InsuredInfo
PRODUCTS	Product_Info
OTHER INFORMATION	Id, Response

DATA CLEANING

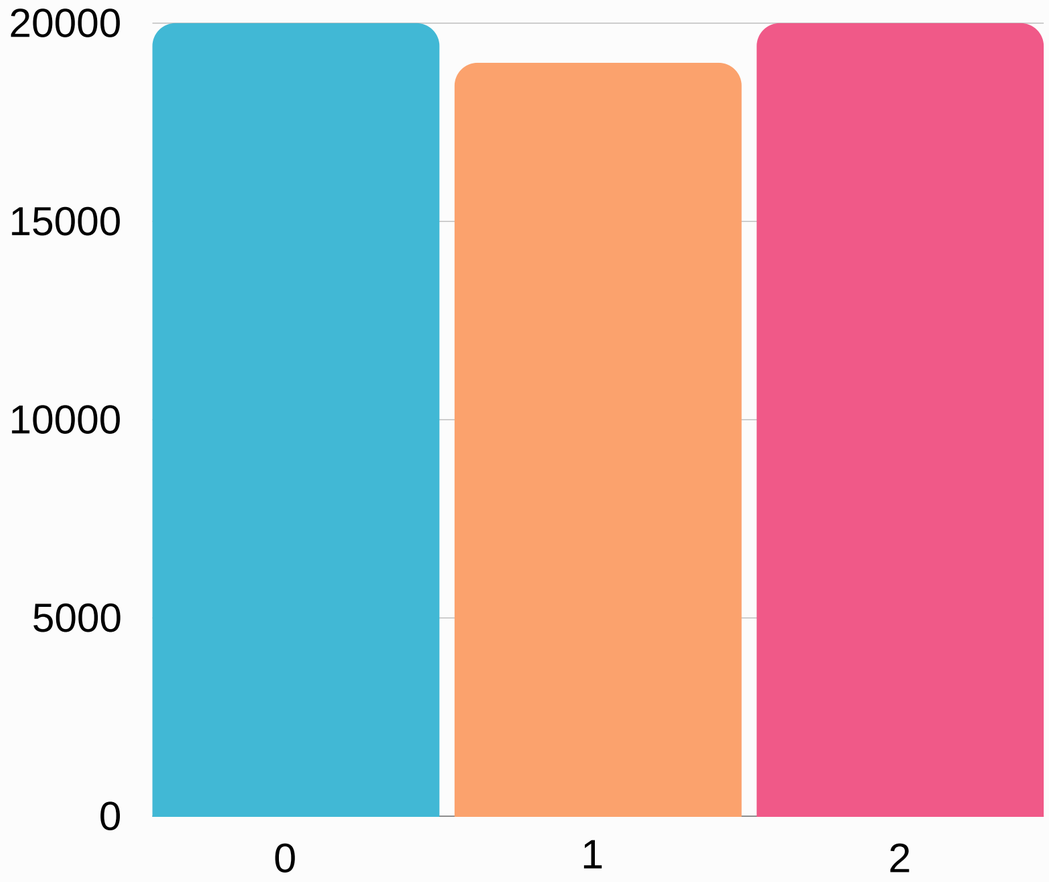


0 Duplicated rows
5 Columns with more than 60% Null

TARGET VARIABLE

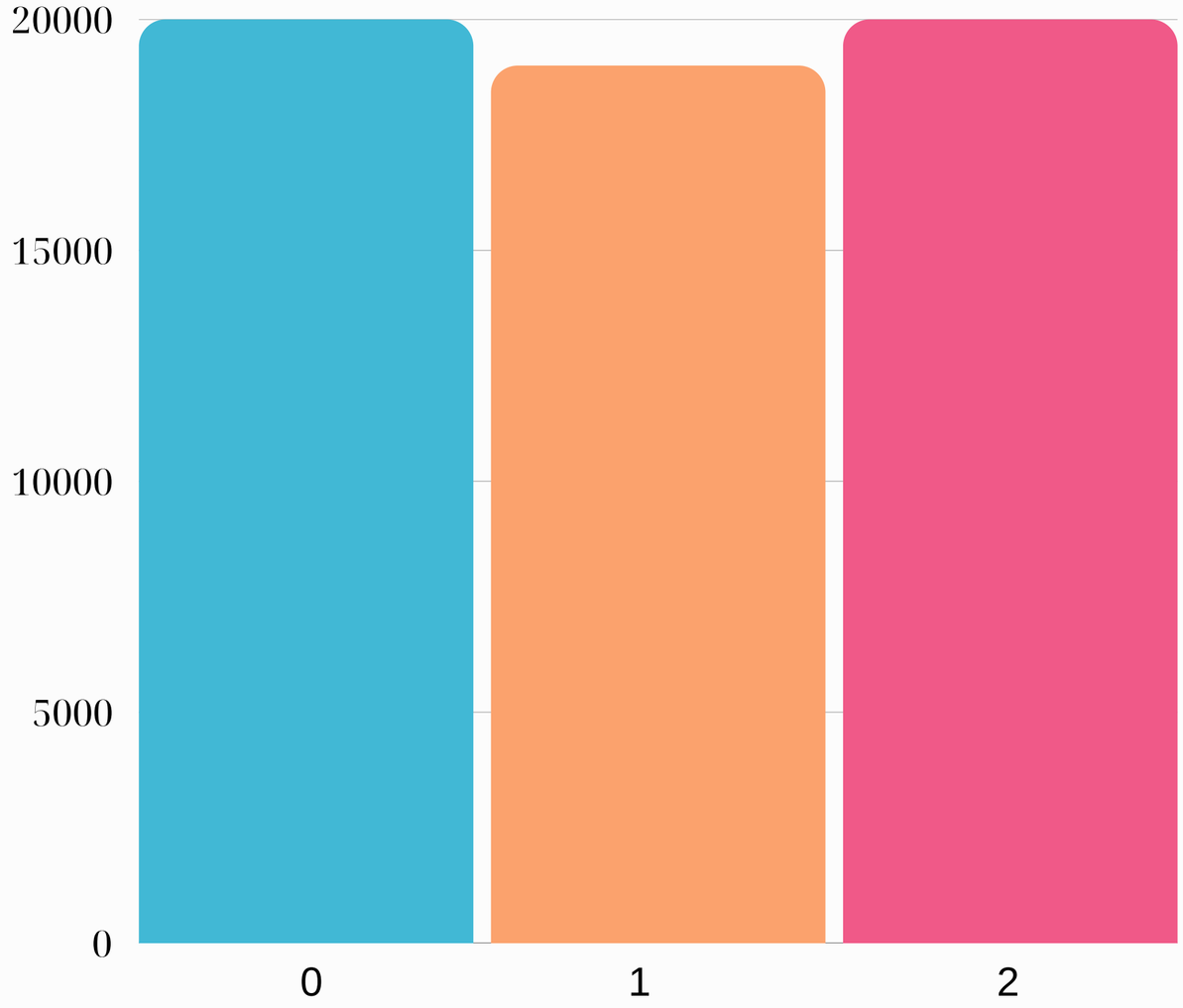


RESPONSE



MODIFIED RESPONSE

TARGET VARIABLE



1-5 Low Risk

6 -7 Medium Risk

8 High Risk

FEATURE ENGINEERING

Med_keyword_count

Total number of medical keywords from each customer

$$\text{Med\_keyword\_count} = \sum_{i=1}^{48} \text{Medical\_Keyword}_i$$

BMI_Age

Describe the interaction between customer's BMI value and Age

$$\text{BMI\_Age} = \text{BMI} \times \text{Age}$$

MACHINE LEARNING MODEL

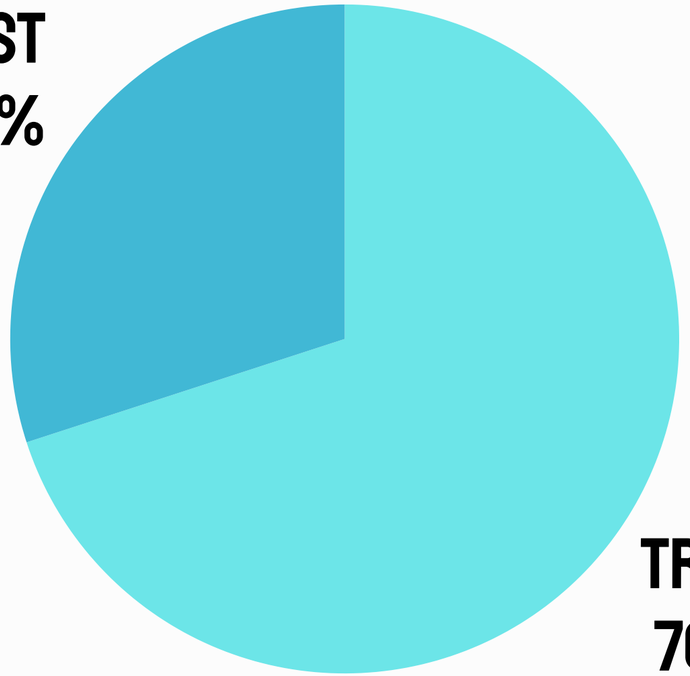
THREE IMPORTANCE COMPONENTS

INPUT DATA

MACHINE LEARNING MODEL

METRICS

MACHINE LEARNING MODEL



RESPONSE

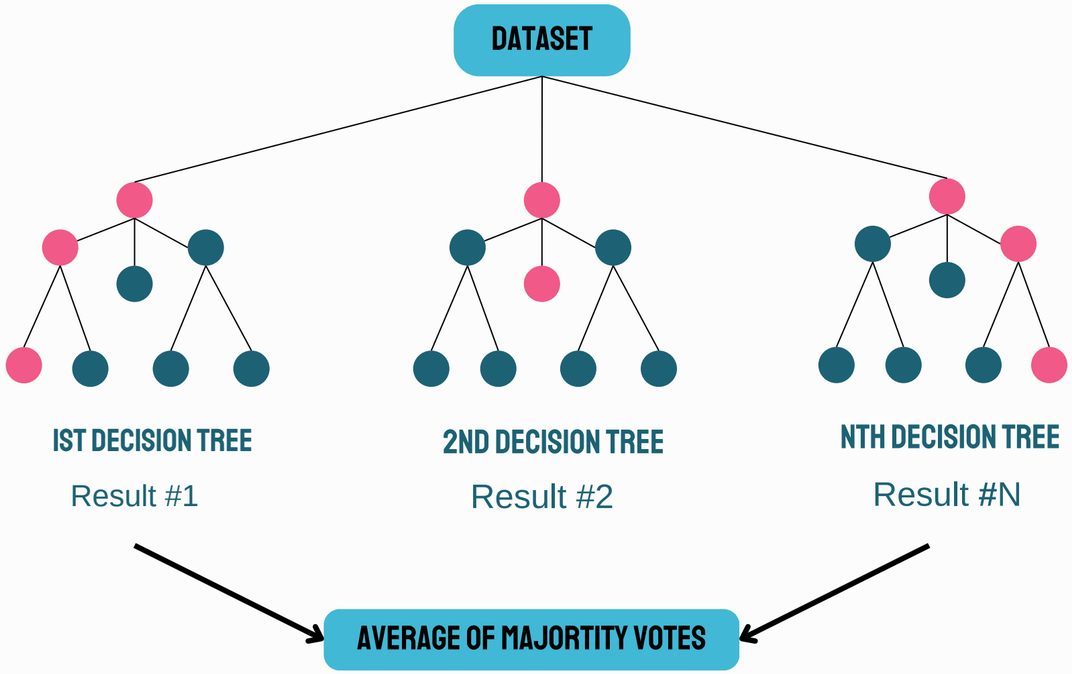
FEATURE SELECTION

- BMI
- WT
- BMI_AGE
- PRODUCT_INFO_4
- INS_AGE
- EMPLOYMENT_INFO_I
- HT', 'FAMILY_HIST_4
- MEDICAL_HISTORY_4
- EMPLOYMENT_INFO_6
- MEDICAL_HISTORY_2
- MEDICAL_HISTORY_I'
- FAMILY_HIST_2
- FAMILY_HIST_3
- INSURANCE_HISTORY_5
- MEDICAL_HISTORY_23
- INSUREDINFO_3
- MED_KEYWORD_COUNT
- MEDICAL_KEYWORD_I5
- EMPLOYMENT_INFO_2
- INSUREDINFO_6
- MEDICAL_KEYWORD_3
- MEDICAL_HISTORY_4

CORRELATION

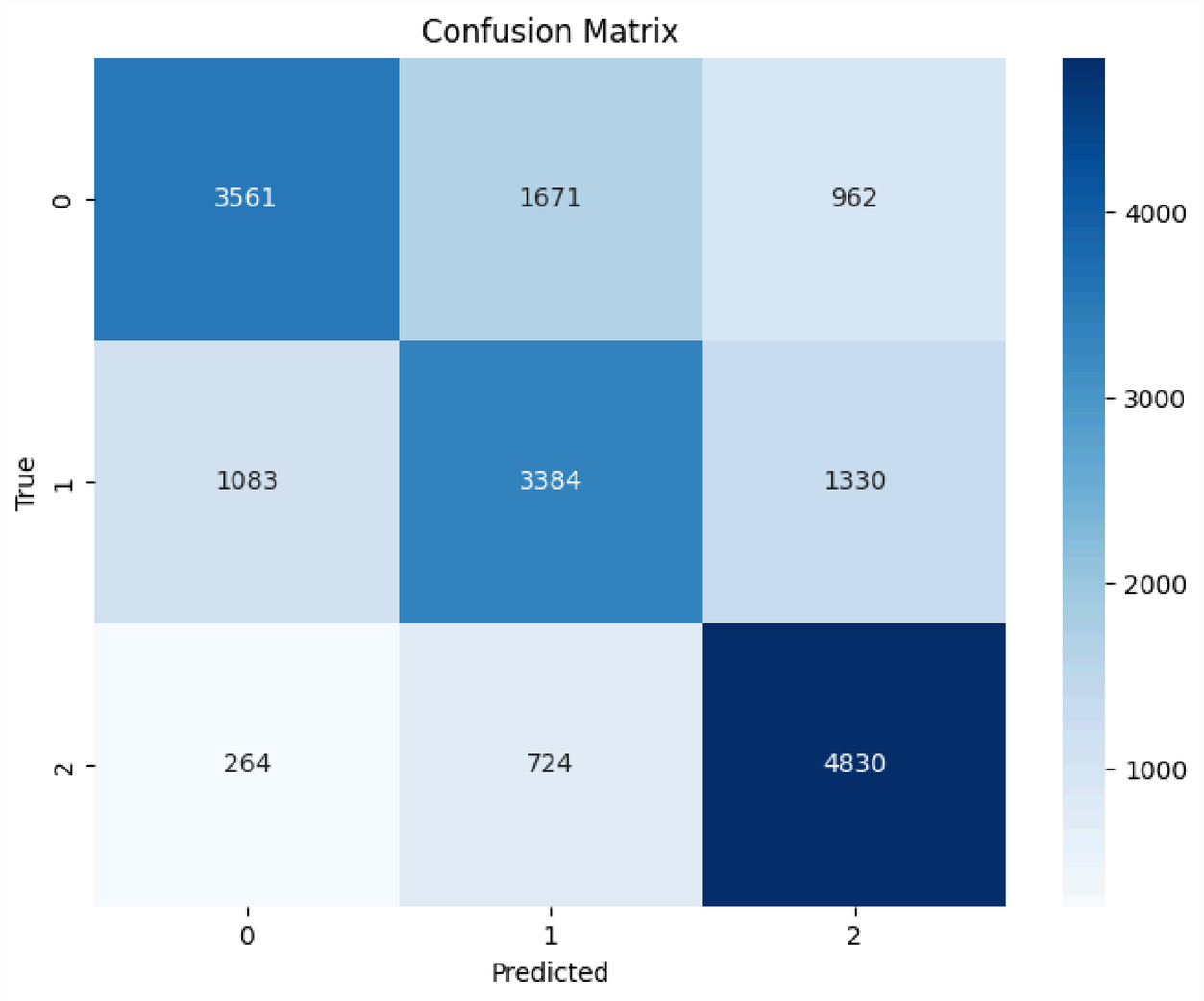
FEATURE IMPORTANCE

RANDOM FOREST



- ABILITY TO HANDLE LARGE AND COMPLEX DATA
- REDUCES OVERFITTING
- IMMUNE TO NOISE

CONFUSION MATRIX



Metrics	
AUC	0.83
Precision	0.66
Recall	0.66
F1-Score	0.657

RANDOM FOREST

	PRECISION	RECALL	FI-SCORE
0	**0.73**	0.58	0.64
1	0.59	0.58	0.58
2	0.68	**0.83**	0.75

VARIOUS MODELS

	PRECISION	RECALL	FI-SCORE
Logistic Regression	0.59	0.60	0.59
Random Forest	0.661	0.663	0.657
XGBoost	0.665	0.663	0.657

NEXT TOP MODEL AWARDS

XGBOOST

	PRECISION	RECALL	FI-SCORE
0	**0.74**	0.56	0.64
1	0.58	0.59	0.58
2	0.67	**0.84**	0.74

CONCLUSION

BUSINESS VALUE

- Reduces processing time to just **one-third** compare to the original method .
- Time = Money
- Improve Customer Experience

FUTURE IMPROVEMENT

COLLECT MORE DATA FROM OTHER RESOURCES

UTILIZE THE DATA THAT WERE GIVEN TO GAIN MORE INSIGHT

DIVE MORE INTO FEATURE ENGINEERING

TRY OUT DIFFERENT FEATURE SELECTION METHODS

APPLYING ADVANCED MACHINE LEARNING TECHNIQUES (STACKING, BLENDING)

TD
ST
HT

THANK You!

HBD T  YOU

COSMOS