PROJECT ON CREDIT BANKRUPTCY

Project Report

Submitted in partial fulfilment of the requirements for the award of the degree of

Masters

in

Information Technology and Analytics

by

AJAY VISHNU ADDALA (Net id: **aa2569**)

Master's in information technology and Analytics

Rutgers Business School, Newark

December 2022

INTRODUCTION

- We are given a credit dataset to analyse using different models.
- The model that results in the best benefit is to be suggested.
- Models run:
    - Logistic
    - LDA
    - QDA
    - KNN
    - Tree

DATA CLEANING

- The data can be checked for any errors.
- The purpose column had car0 variables in 7 cases.
- This might be an error, so we changed into the car.

MODELS RUN

- 10 different models were run for each of the 5 models shown above.
- Two sets were run, one considering Recall & Accuracy and the other considering Recall & Precision.
- The 10 models:
    - default ~ checking_balance + months_loan_duration + credit_history + purpose + amount + savings_balance + employment_duration + percent_of_income + years_at_residence + age + other_credit + housing + existing_loans_count + job + dependents + phone
    - default ~ checking_balance + months_loan_duration + credit_history + amount + percent_of_income
    - default ~ checking_balance + months_loan_duration + credit_history + amount + factor(percent_of_income)
    - default ~ checking_balance + months_loan_duration + credit_history + percent_of_income
    - default ~ checking_balance + months_loan_duration + credit_history
    - default ~ months_loan_duration + credit_history
    - default ~ factor(months_loan_duration) + credit_history
    - default ~ checking_balance + factor(months_loan_duration) + credit_history
    - default ~ months_loan_duration
    - default ~ credit_history

INITIAL ASSUMPTIONS AND RESULTS

- Initially, a few assumptions were made to check for the best model.

Assumed Numbers	
Clients	3000
Default	8%
Average Default	250000
Payments	25000

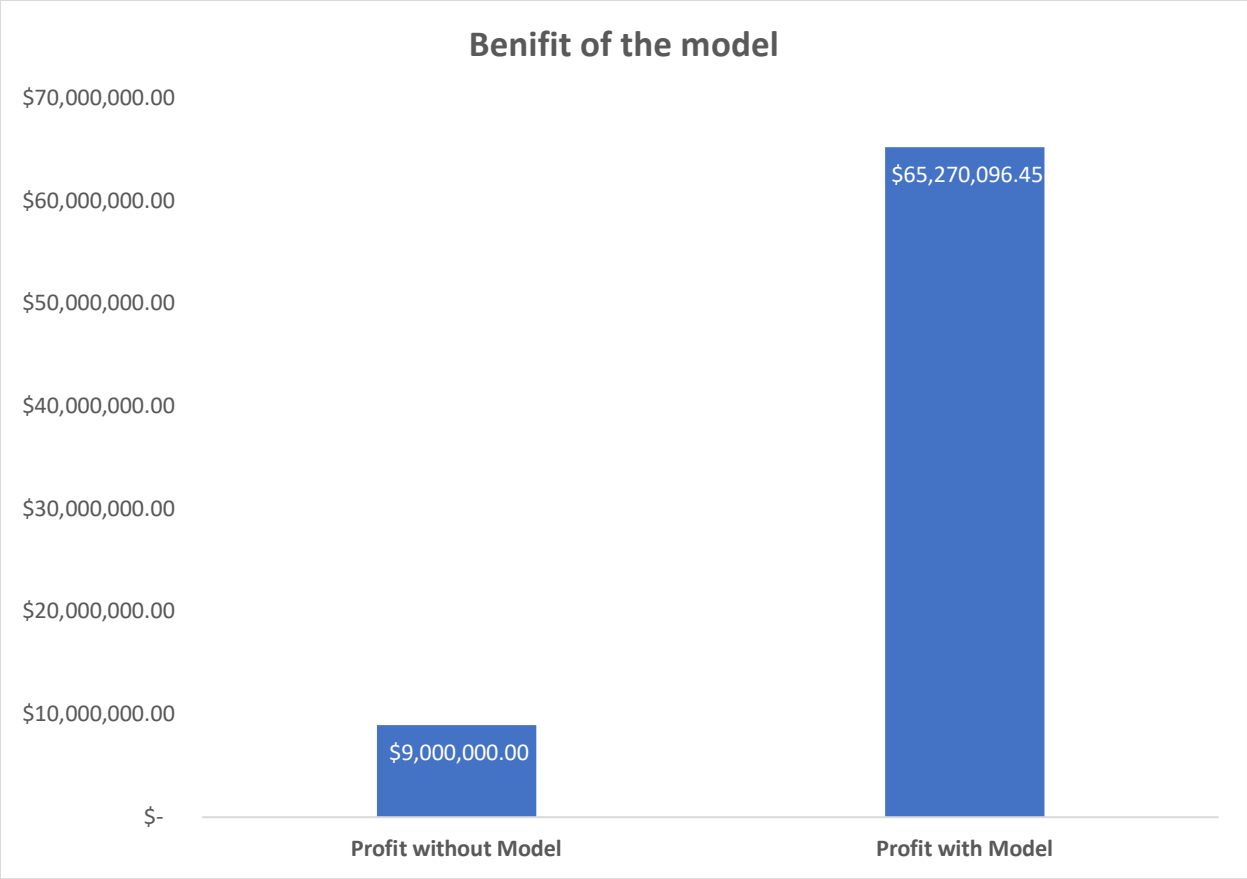
- Results for this model:

Parameters	Precision & Recall	
Model	KNN	
Model No:		4
Accuracy		74.50%
Precision		74.32%
Recall		97.14%
Profit without Model	$	9,000,000.00
Profit with Model	$	65,270,096.45

- Benefit & Increase %

BENEFIT
$ 56,270,096.45

% INCREASE
625%



FUTURE PREDICTION

- Future assumptions were made to check for the best model.

Assumed Numbers	
Clients	5000
Default	10%
Average Default	250000
Payments	25000

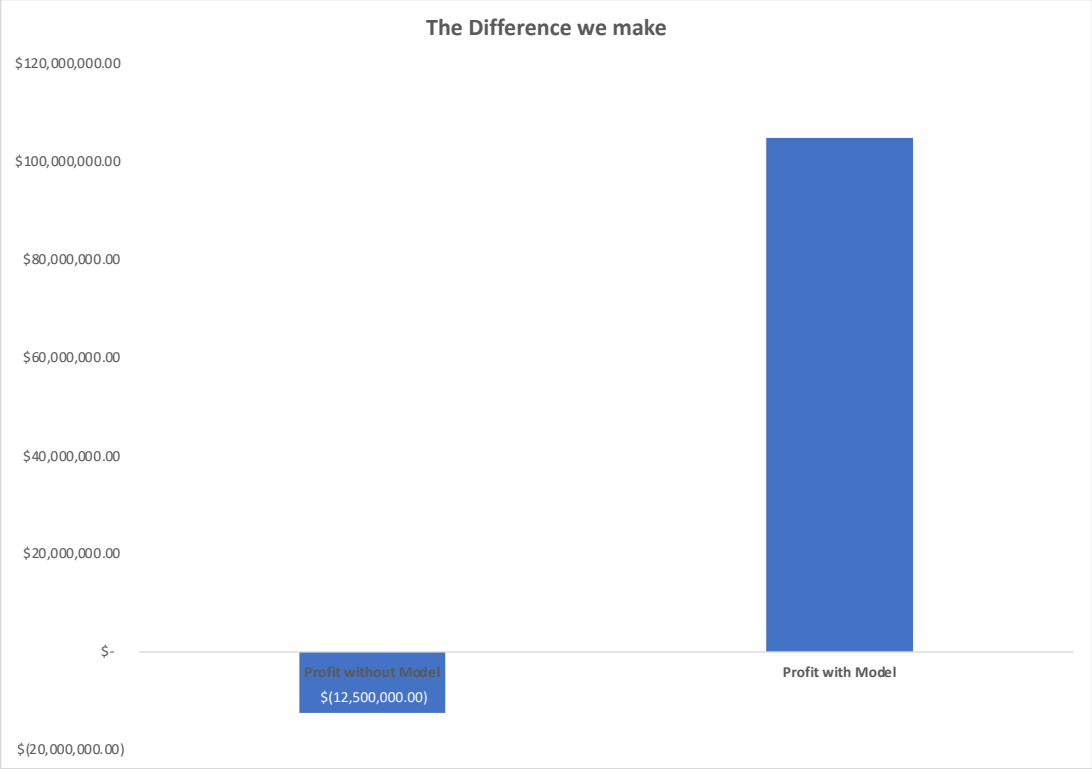
- Results for this model:

Parameters	Precision & Recall
Model	KNN
Model No:	4
Accuracy	74.50%
Precision	74.32%
Recall	97.14%
Profit without Model	$ (12,500,000.00)
Profit with Model	$ 104,729,367.60

- Benefit & Increase %

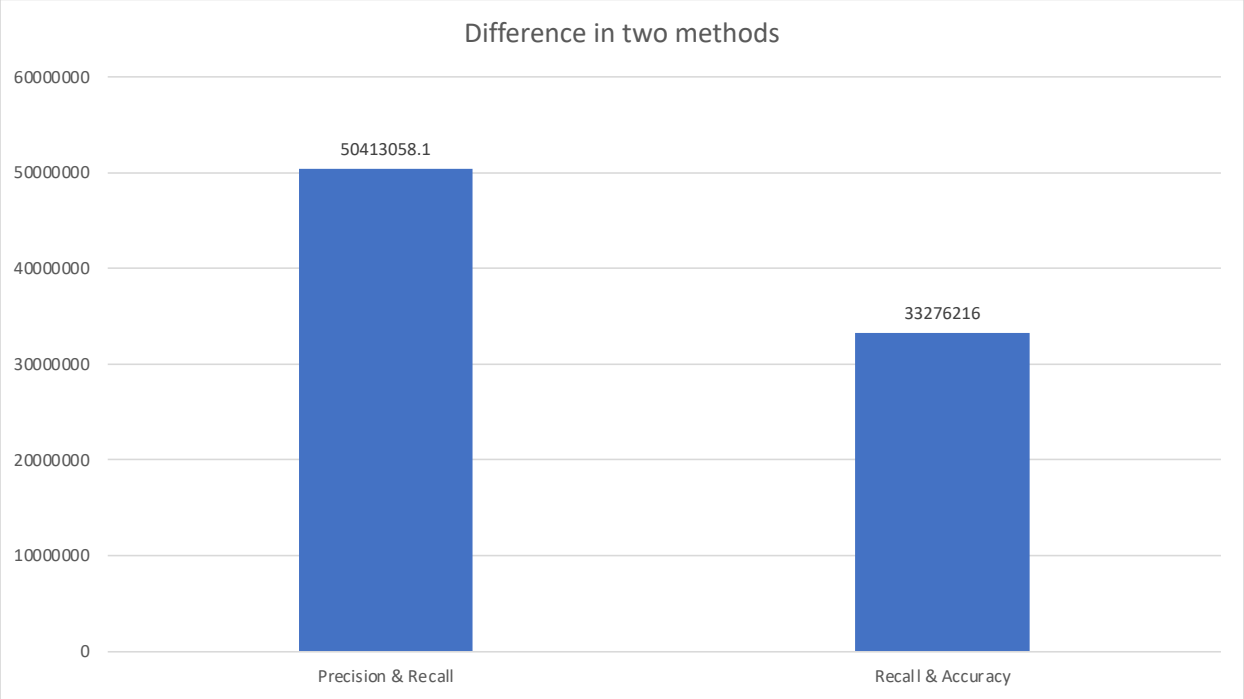
BENEFIT
$ 117,229,367.60

% INCREASE
938%

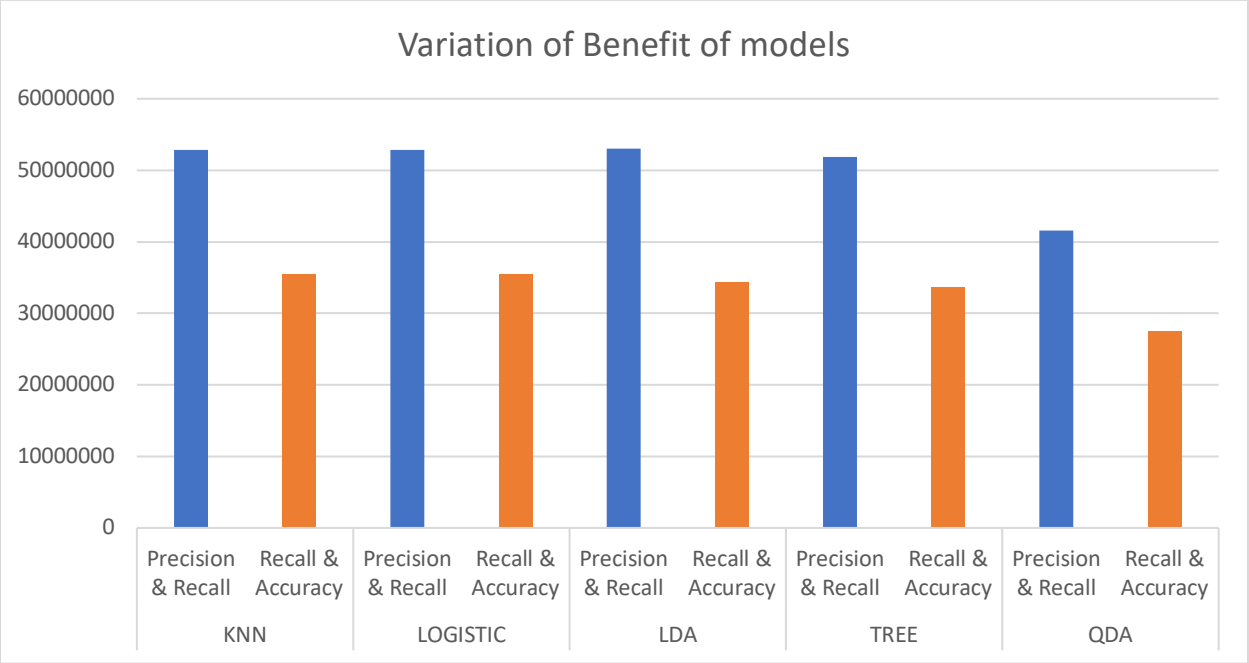


MORE ANALYSIS

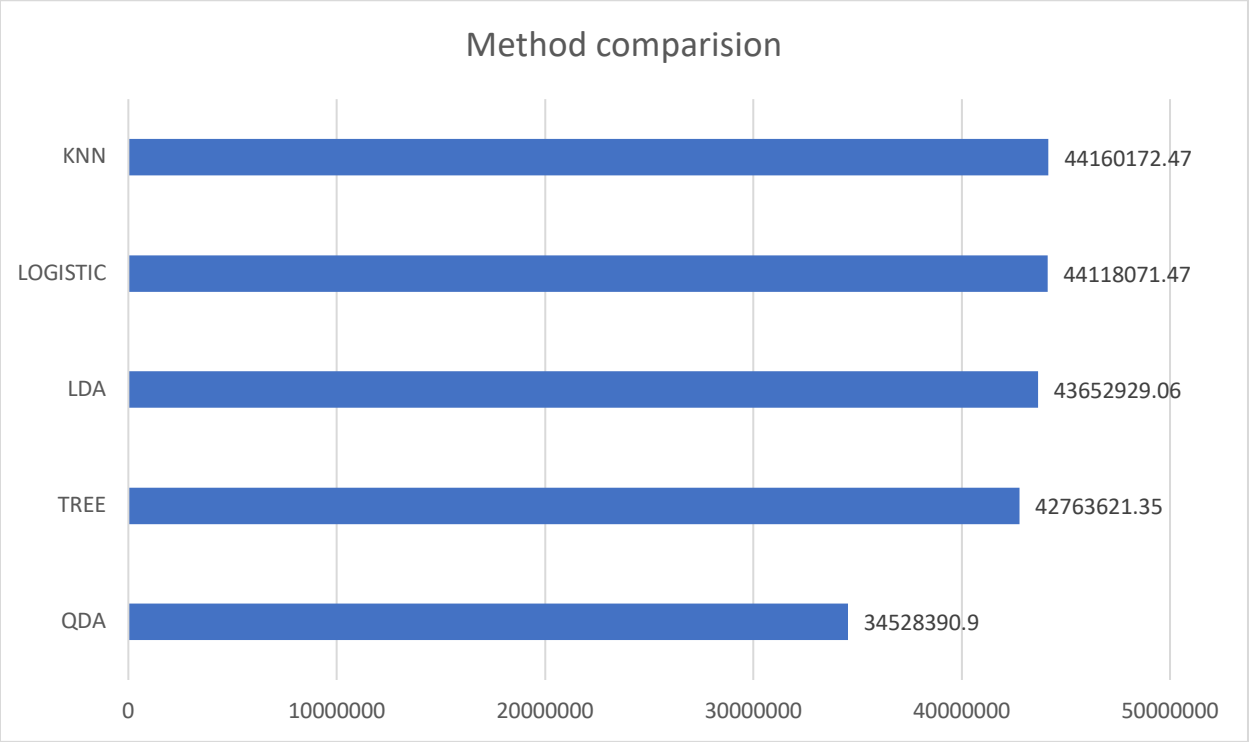
- Checked the average benefit from the two methods



- Difference between methods and models



- Difference between the models



CONCLUSION

- In most cases, KNN gives the best model.
- Recall & Precision have been selected from the models to develop the benefit.
- BEST MODEL: default ~ checking_balance + months_loan_duration + credit_history + percent_of_income
- Considering the following 4 variables gave the best outcome in a KNN model:
    - Checking balance
    - Months loan duration
    - Credit History
    - Present of Income