

INSURANCE PREMIUM DEFAULT PROPENSITY

Capstone Project

Kudakwashe Nyikadzino

CONTENTS

- ❑ Executive Summary
- ❑ Typical Customer
- ❑ Defaulting Customers
- ❑ Default, Income and Cash Premiums
- ❑ Analytical Approach
- ❑ Model Evaluation and Results
- ❑ Final Model Selection
- ❑ Insights
- ❑ Recommendations
- ❑ Final Word

EXECUTIVE SUMMARY

The Problem

- ❑ The impact of insurance premium defaults on revenues.

The Approach

- ❑ Build a machine learning model to identify defaulting customers.

The Outcome

- ❑ Provide insights and recommendations on reducing defaults.

WHO IS THE TYPICAL CUSTOMER

Age - 51

Residence - Urban

Risk Score – 99,07

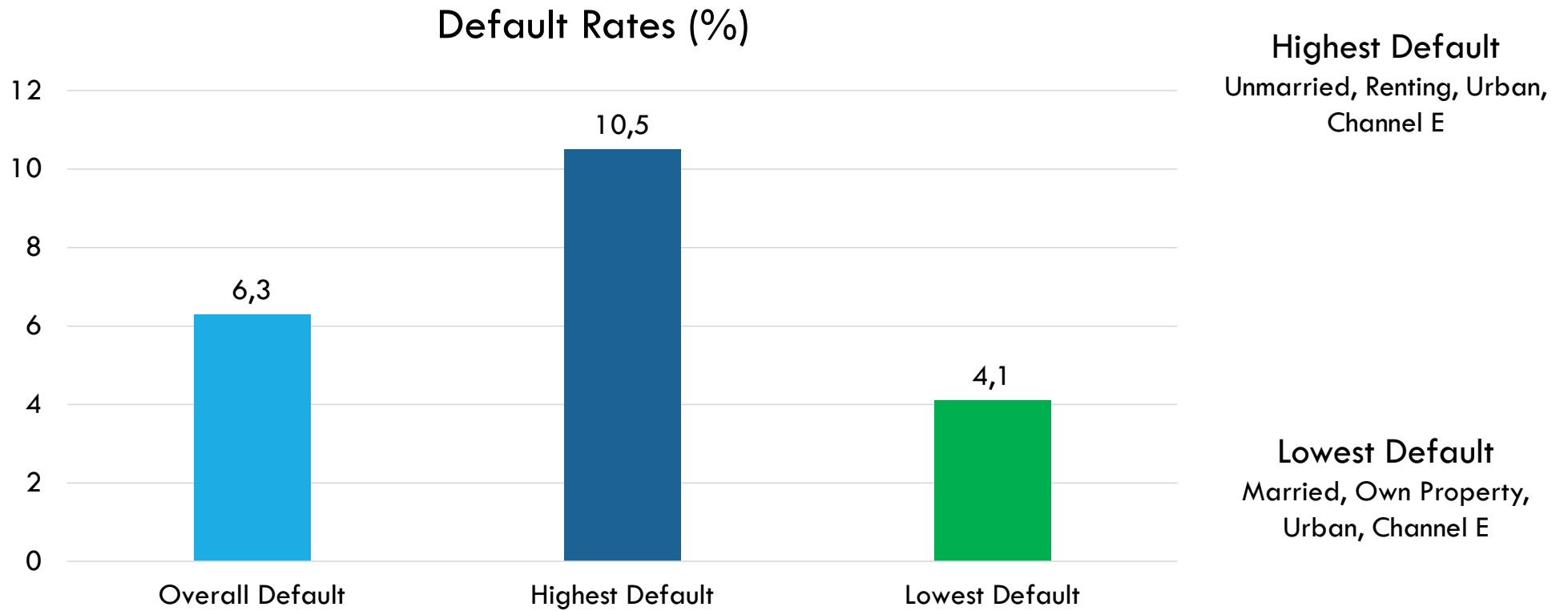


Sourcing Channel - A

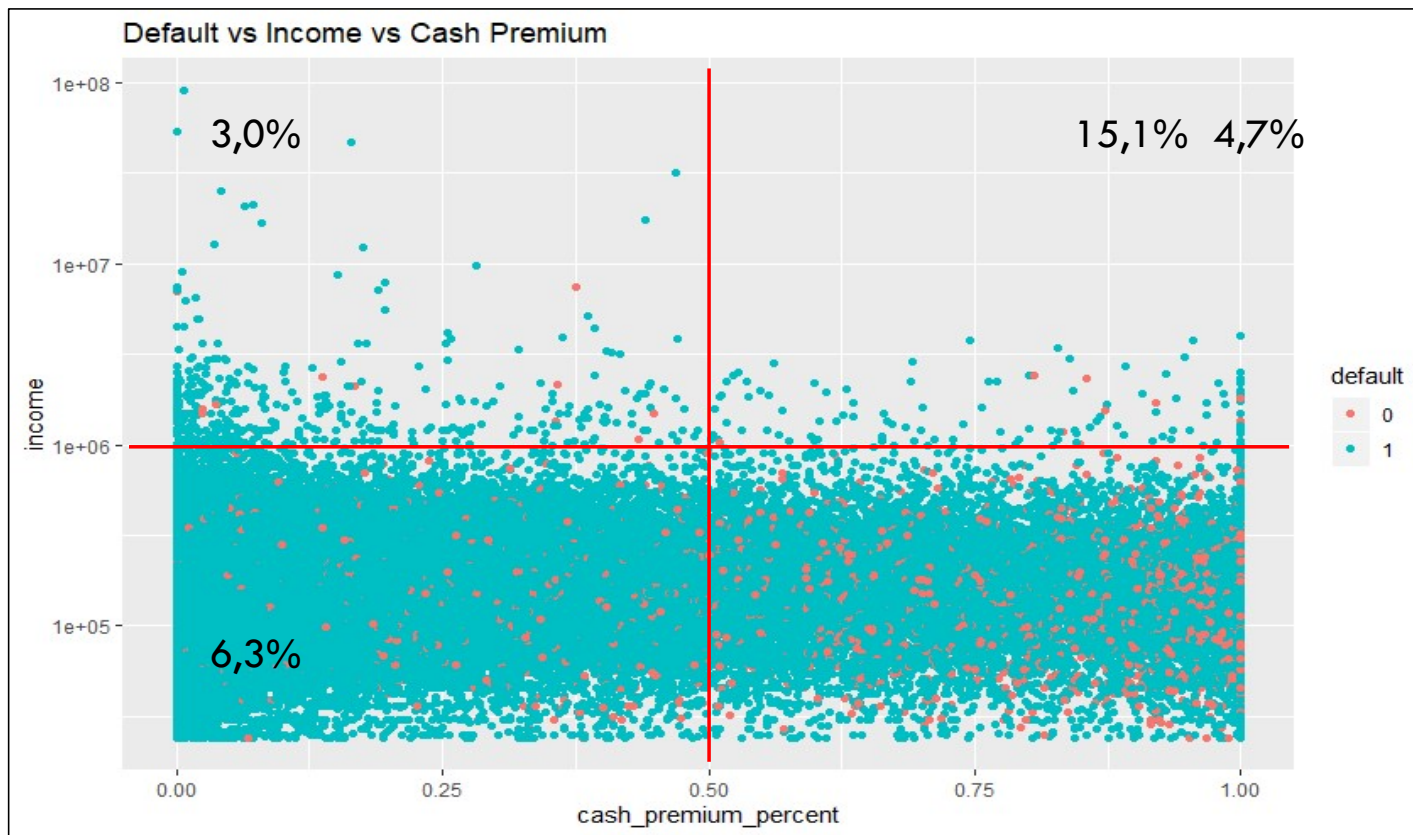
Cash Premiums % - 32

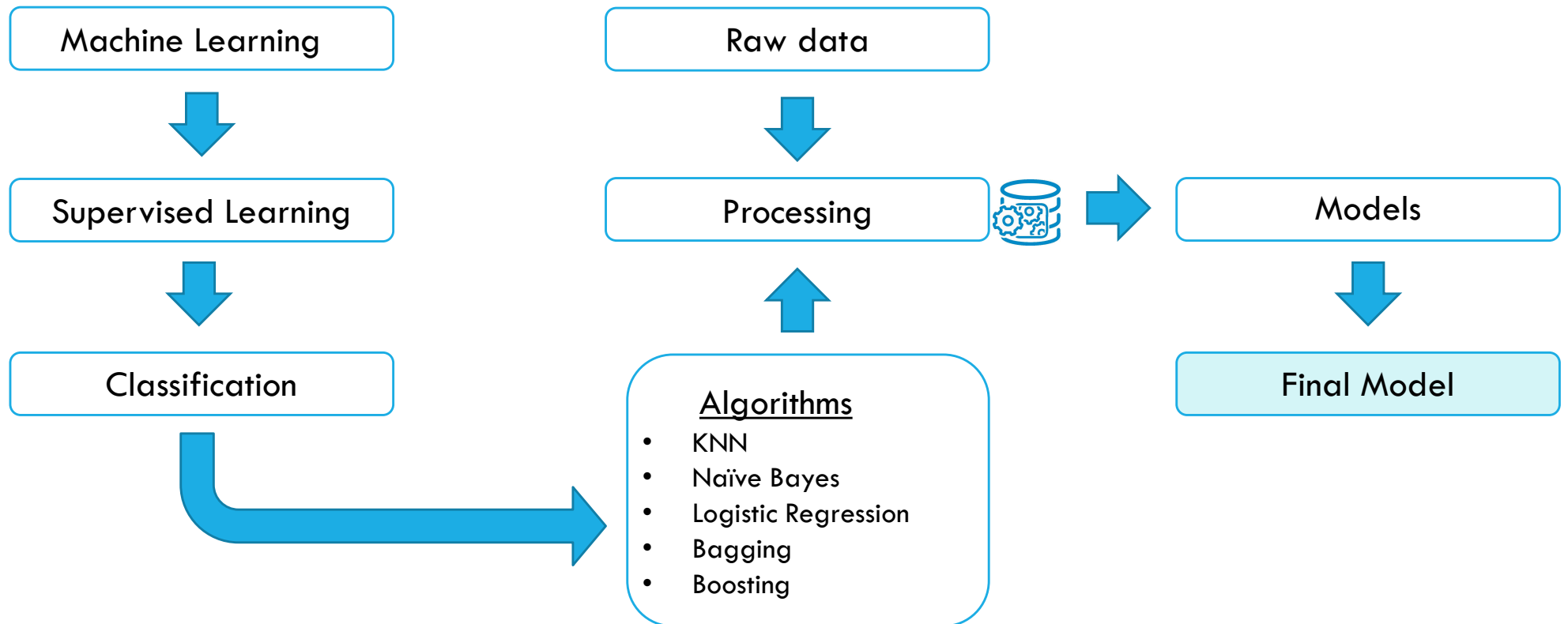
Premiums Paid - 11

DEFAULTING CUSTOMERS



DEFAULT, INCOME AND CASH PREMIUMS





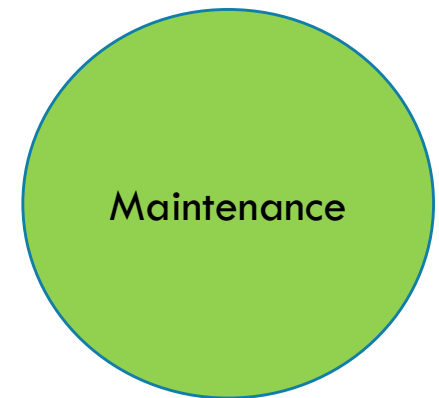
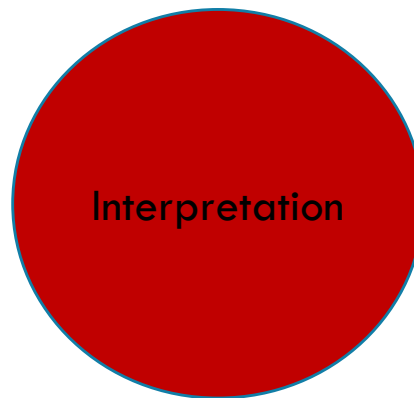
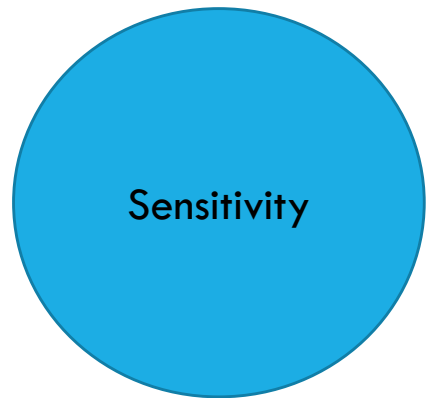
MODEL EVALUATION AND RESULTS

- Accuracy, Sensitivity and AUC as performance measures

Model	Accuracy	Sensitivity	Specificity	AUC
KNN	0.938	0.063	0.996	0.760
Naïve Bayes	0.937	0.000	1.000	0.832
Logistic Regression	0.854	0.597	0.872	0.823
Bagging	0.894	0.428	0.926	0.801
Boosting	0.924	0.314	0.965	0.814

FINAL MODEL SELECTION

Why Logistic Regression



INSIGHTS

- ❑ Percentage of premium paid as cash the most important variable
- ❑ Not all variable are important – e.g. residential area type
- ❑ Accuracy alone is not the only basis to select the model
- ❑ Positive relationship between default and the number of late payments

RECOMMENDATIONS

- ☐ Promote use of non-cash payment methods for premiums
- ☐ Use percentage of premiums as a qualifier for providing insurance
- ☐ Develop a communication system that notifies customers when premiums are due
- ☐ Focus marketing efforts on sourcing channels with lower default rates
- ☐ Remove unnecessary information in the application process

FINAL WORD

- ❑ There will be missed opportunities based on the model in use
- ❑ Consider other variables like employment type e.g. part-time or full-time
- ❑ Ensure accurate data is captured and continuous model improvement