# Efficient Claims Management at Safiri Insurance

Muthuri Abigail Mwendwa

# TABLE OF CONTENTS


01 OVERVIEW

02 BUSINESS UNDERSTANDING

03 DATA UNDERSTANDING

04 MODELLING

05 EVALUATION

06 RECOMMENDATIONS

# 01 OVERVIEW

# OVERVIEW

Safiri Insurance, a leading provider of car insurance, has been experiencing a steady increase in the number of insurance claims filed by its customers. While the company has been successful in growing its customer base, the rise in claims has significantly impacted its profitability.

To sustain its competitive edge and improve its bottom line, Safiri Insurance needs to better understand the factors driving insurance claims and develop a robust predictive model to identify high-risk customers.

# 02 BUSINESS UNDERSTANDING

# STAKEHOLDERS

## SAFIRI INSURANCE COMPANY

As the primary stakeholder, Safiri Insurance is interested in improving its operational efficiency, reducing financial losses due to insurance claims, and enhancing customer satisfaction.

## POLICYHOLDERS

Policyholders are directly impacted by the predictions of car insurance claims. Accurate prediction allows Safiri Insurance to offer fair premiums based on the risk profile of each policyholder.

## UNDERWRITERS AND ACTUARIES

Underwriters and actuaries are responsible for assessing risks associated with insurance policies and determining appropriate premiums.

# BUSINESS OBJECTIVE

The primary objective of this project is to develop a predictive model that can accurately forecast the likelihood of a policyholder filing a car insurance claim. By analyzing historical data and identifying patterns, the model will be trained to predict the probability of future claims for each policyholder.

This predictive capability will enable Safiri Insurance to:

- Identify High-Risk Customers
- Allocate Resources Efficiently
- Enhance Customer Experience

# BUSINESS PROBLEM

The primary business problem for Safiri Insurance is to reduce the number of fraudulent or unnecessary claims while maintaining customer satisfaction.

This involves accurately predicting which customers are likely to file a claim. By identifying high-risk customers, Safiri Insurance can take proactive measures, such as personalized risk assessments, tailored premium pricing, or even targeted interventions to mitigate risk.

# 03 DATA UNDERSTANDING

# DATASET

The dataset used in this analysis comprises 10,302 records and 27 columns, which include customer demographics, car details, claim history, and other relevant information.

# DATA ANALYSIS

To ensure the accuracy and reliability of the data, we performed a thorough cleaning process. This involved converting dates and monetary values into a consistent format. We also addressed missing information by removing records that lacked crucial data. Additionally, duplicate records were eliminated to maintain the integrity of the dataset.

During the exploratory data analysis phase, we examined the data to understand the characteristics and relationships within it. We looked at the distribution of claims to see how many customers filed claims and identified any unusual values. We also created visual representations of the data to help us gain insights into the factors that might influence whether a customer files a claim.

To prepare the data for modeling, we transformed categorical information into a format suitable for analysis. These steps allowed us to include all relevant information in our predictive model, ensuring it could accurately identify patterns and make predictions.

# 04 MODELLING

# Baseline Model - Logistic Regression

To start our modeling process, we used a Logistic Regression model as our baseline. This model is straightforward and interpretable, making it a good starting point for our analysis. Logistic Regression works well for binary classification problems like predicting whether a customer will file an insurance claim.

By using this model, we established a benchmark performance level to compare more complex models against. Our initial results showed that Logistic Regression could provide a reasonable level of accuracy, helping us understand which features were most influential in predicting claims.

# Complex Model: Random Forest

After establishing our baseline, we implemented a more complex model: Random Forest. This model is an ensemble learning method that builds multiple decision trees and combines their outputs to improve predictive accuracy. Random Forest is particularly effective at handling a large number of features and capturing complex interactions between them.

By using this model, we aimed to improve the accuracy of our predictions and gain deeper insights into the factors that drive claim filing. The Random Forest model demonstrated superior performance compared to the baseline Logistic Regression, indicating its effectiveness in capturing the nuances of our data.

# Hyperparameter Tuning on Logistic Regression

To further enhance the performance of our Logistic Regression model, we conducted hyperparameter tuning. This process involves adjusting the model's settings to find the optimal combination that maximizes its predictive accuracy.

By experimenting with different parameters we were able to fine-tune the model and achieve better results. This step ensured that our baseline model was operating at its best, providing a solid foundation for comparison with more advanced models like Random Forest.

# 05 EVALUATION

# MODEL EVALUATION

## Logistic Regression Model

Our Logistic Regression model achieved a decent accuracy of 0.80, indicating that it can correctly predict a significant portion of instances.
The model showed a good balance between precision and recall, which is crucial for identifying true positives and minimizing false positives.

## Random Forest Model

The Random Forest model outperformed the Logistic Regression with an accuracy of 1.0, indicating its superior predictive power.
With higher precision and recall, the Random Forest model was better at correctly identifying true positives and minimizing false negatives.

# HYPERPARAMETER TUNING

Hyperparameter tuning significantly improved the performance of the baseline Logistic Regression model.
Tuning resulted in enhanced accuracy, precision, recall, F1 score, and ROC-AUC score, making the model more robust and reliable.
This demonstrates the importance of optimizing model parameters to maximize performance and ensure optimal predictive capabilities.

# 06 RECOMMENDATIONS

**Offer Tailored Coverage for Sports Car Owners:**

- Given that owning a sports car has a significant positive coefficient, Safiri can design specialized insurance packages tailored to sports car owners. These packages could offer enhanced coverage options and services specifically suited to the needs and risks associated with sports car ownership.

**Implement Policies to Address Revoked Licenses:**

- Since having a revoked license is a top influential feature, Safiri should implement policies to address this risk factor. This could include targeted interventions such as driver education programs, stricter underwriting criteria, or higher premiums for individuals with a history of license revocation.

**Focus on Parents with Teenage Drivers:**

- Given the positive coefficient for having a parent as the primary driver, Safiri should target marketing efforts towards parents with teenage drivers. Offering comprehensive coverage options, safe driving incentives, and parental guidance resources can help mitigate risks associated with young drivers.

**Customized Plans for High Claim Frequency Individuals:**

- Since claim frequency has a positive coefficient, Safiri can develop customized insurance plans for individuals with a history of frequent claims. These plans could incorporate personalized risk management strategies, proactive claims support, and incentives for safe driving behavior to reduce claim frequency over time.

**Tailor Policies for Families with Multiple Children Driving:**

- Safiri should offer tailored insurance policies for families with multiple children driving to address the unique risk profile associated with this demographic. Family-friendly discounts, flexible coverage options, and safety-focused initiatives can appeal to this customer segment.

**Consider Home Environment in Risk Assessment:**

- Safiri should consider the home environment when assessing insurance risks. Factors such as household safety measures, parental supervision, and neighborhood characteristics can influence risk levels and premium calculations.

# THANK YOU!!

ANY QUESTIONS ??