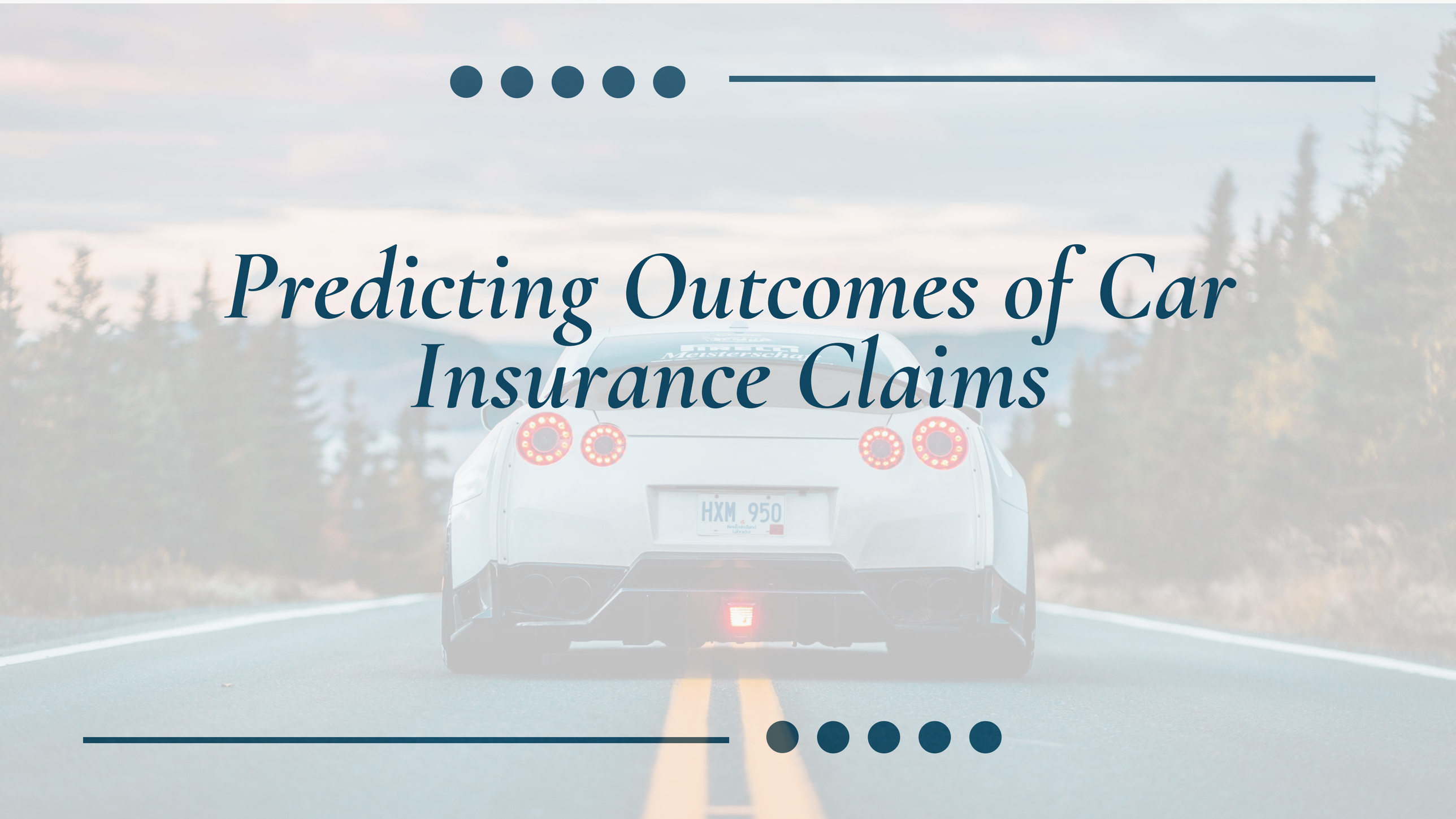

Predicting Outcomes of Car Insurance Claims

Overview

Insurance companies invest significant time and resources into optimizing their pricing strategies and accurately predicting the likelihood of customers filing claims. In many countries, car insurance is legally required to drive on public roads, making the market quite large.

With this in mind, On the Road car insurance has sought in building a model to predict whether a customer will file a claim during their policy period. Since they have limited expertise and infrastructure for deploying and managing machine learning models, they've asked to identify the single most important feature that leads to the best model performance, based on accuracy, so they can start with a simple model in production.

Problem Statement

On the Road car insurance aims to develop a simple and effective model to predict whether a customer will file a claim during their policy period. They lack extensive expertise and resources for complex machine learning models.

Goals

01.

Identify the single most important feature for predicting claim filing using a logistic regression model.

02.

Build a simple and accurate model for claim prediction suitable for deployment with limited resources.

Dataset

The dataset used for this project is named "car_insurance.csv". It likely contains information about car insurance policyholders and their history of filing claims.

Outputs

- Determine the one feature in the data that best predicts whether a customer will file a claim (indicated by the **outcome** column), excluding the **id** column.

- Name the result as a DataFrame named **best_feature_df**, which should include two columns: **best_feature** for the name of the top-performing feature and **best_accuracy** for its corresponding accuracy score.

Data Findings

The **"driving_experience"** feature is identified as the single most important feature for predicting claim filing using a logistic regression model.

This means that among all the features considered in the model, **"driving_experience"** has the strongest correlation with the outcome variable (whether a claim will be filed or not).

By incorporating this feature into a logistic regression model, On the Road car insurance can potentially build a simple yet effective model to predict claim filing.

Data Findings

The table presents the results of a machine learning experiment aimed at identifying the most important feature for predicting insurance claims. The table has two columns:

- **best_feature:** The name of the feature that yielded the highest accuracy in predicting claim outcomes.
- **best_accuracy:** The accuracy score associated with the best feature.

	best_feature	best_accuracy
0	driving_experience	0.7771

Based on the table, the feature **"driving_experience"** has been identified as the most important predictor of claim filing, achieving an accuracy of 0.7771. This means that a model built solely on this feature can correctly predict whether a customer will file a claim approximately 77.71% of the time.

This finding suggests that **"driving experience"** is a strong indicator of claim likelihood.

Insights

01.

The importance of features can vary depending on the modeling technique used. Here, logistic regression identified a single most important feature, but other models might prioritize different features.

02.

Simple models can be a good starting point, especially when resources are limited. They can provide valuable insights and be relatively easy to deploy.