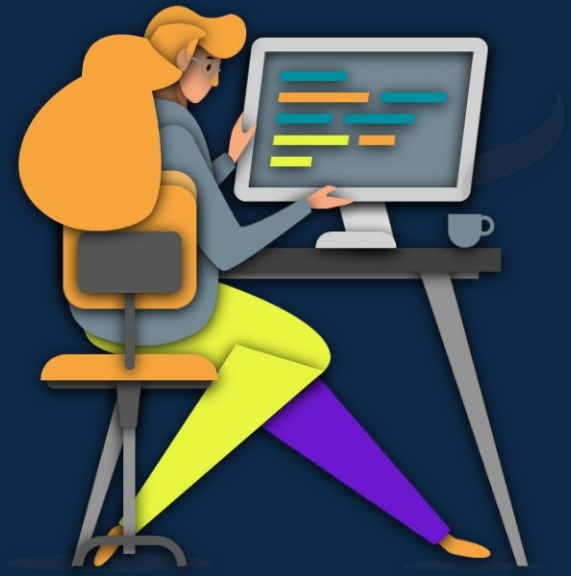
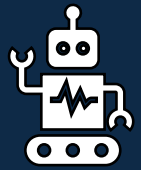
Predicting Motor Insurance Claims Using Machine Learning

Guided By: Shalini Kumari

Submitted By: Subhamita K

CONTENTS

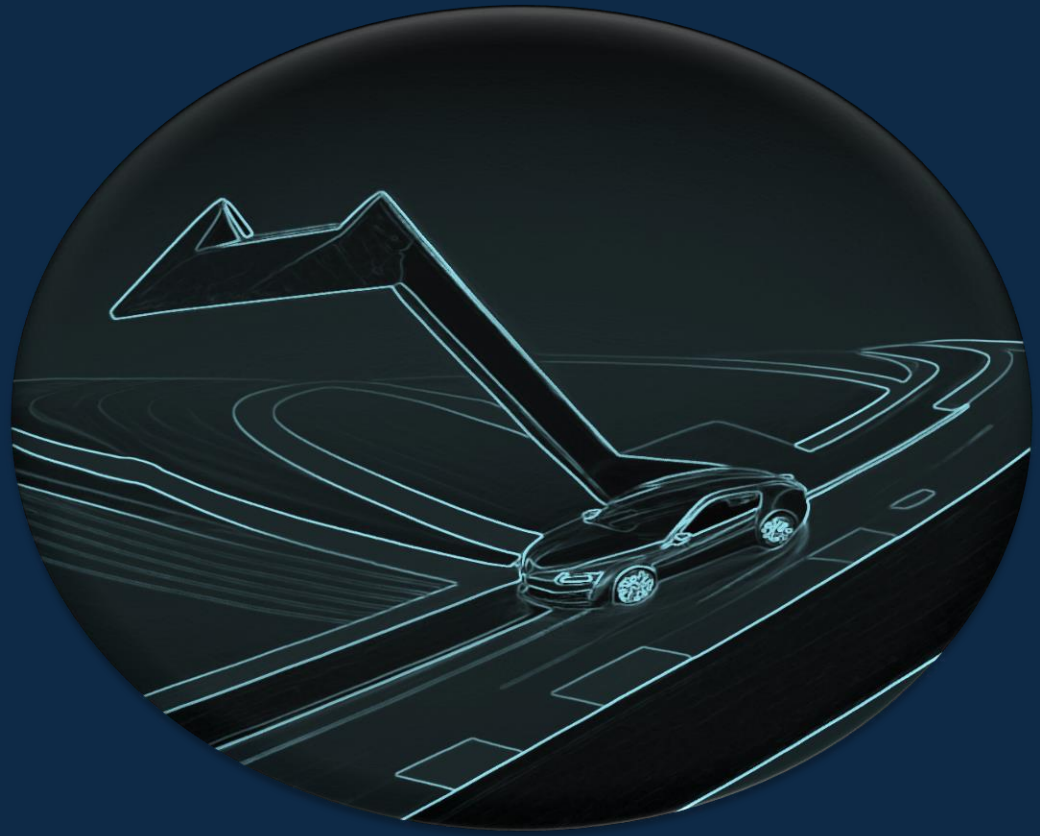




INTRODUCTION

- Motor insurance claims prediction plays a vital role in the insurance industry, enabling accurate risk assessment and improved decision-making.

- The objective of this presentation is to showcase the application of machine learning algorithms for predicting motor insurance claims



Problem Statement

Problem

1

Predicting motor insurance claims accurately.

Challenges

2

Factors such as varying claim amounts, complex relationships between features, and the presence of outliers make accurate prediction challenging.

Need

3

A reliable predictive model that can accurately forecast motor insurance claims, leading to improved risk management and customer satisfaction.

DATA OVERVIEW

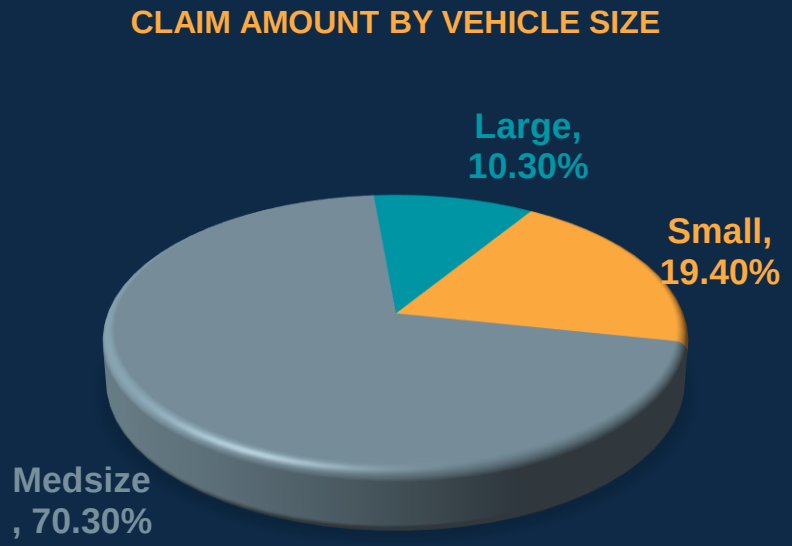
- Dataset: The dataset used for analysis consists of 9134 rows and 26 columns, containing information related to motor insurance claims.

- Customer-related features: Country, State Code, Education, EmploymentStatus, Gender, Income.

- Policy-related features: Coverage, Location Code, Monthly Premium Auto, Months Since Last Claim, Months Since Policy Inception, Number of Open Complaints, Number of Policies, Policy Type, Policy, Sales Channel, Vehicle Class, Vehicle Size.
- The target variable for the prediction task is the 'Claim Amount'. The model will be trained to estimate the claim amount based on the given features.

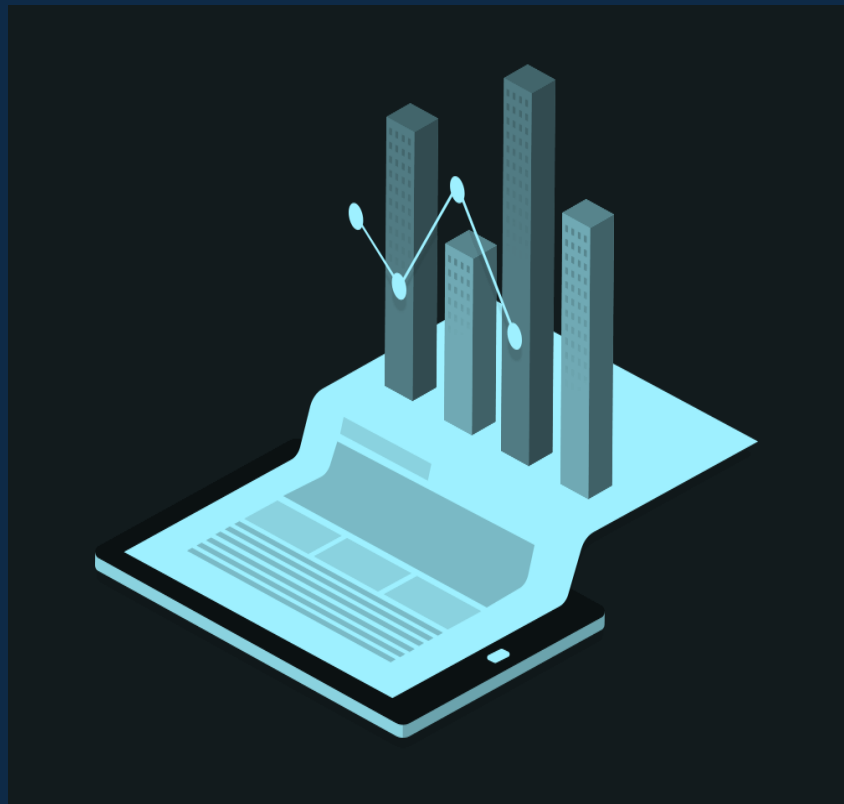
Exploratory Data Analysis (EDA) & Data Pre-processing

- Identified significant correlations between claim amount and variables such as monthly premium auto, income, and number of policies.
- Explored the distribution of categorical variables such as coverage, education, employment status, and vehicle class.
- Utilized various visualization techniques, including histograms, scatter plots, and bar plots, to visualize the data and identify patterns or trends.



Exploratory Data Analysis (EDA) & Data Pre-processing

- Categorical variables were encoded into numerical format using techniques like label encoding or one-hot encoding.
- Detected outliers in the 'Claim Amount' variable using methods like box plots.
- Outliers were treated by either removing the corresponding rows or applying transformation techniques to reduce their impact.



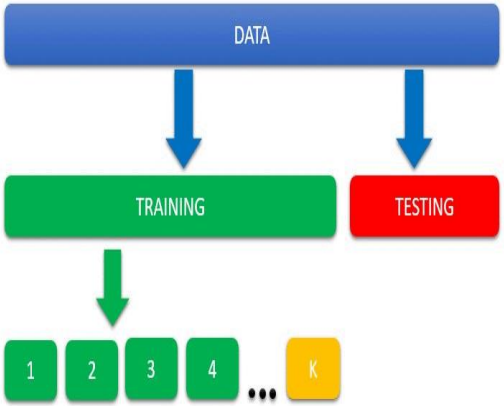
Model Selection

Linear Regression	• A traditional regression model that establishes a linear relationship between the input features and the target variable.
Decision Tree	• A tree-based model that makes decisions based on hierarchical splits in the feature space
Random Forest	• An ensemble of decision trees that combines their predictions to improve accuracy and reduce overfitting
Gradient Boosting	• A boosting algorithm that sequentially builds weak learners and combines their predictions to create a strong learner.

Model Training and Evaluation

- Splitting the Data: We divided our dataset into training and testing sets. The training set was used to train the model, while the testing set was used to evaluate its performance.

- We trained the selected models, including Linear Regression, Decision Tree, Random Forest, and Gradient Boosting, by feeding the training data into each respective algorithm. Each algorithm utilizes different techniques and approaches to learn from the data and make predictions. By evaluating and comparing the performance of these models, we can determine which algorithm is most effective in predicting motor insurance claims



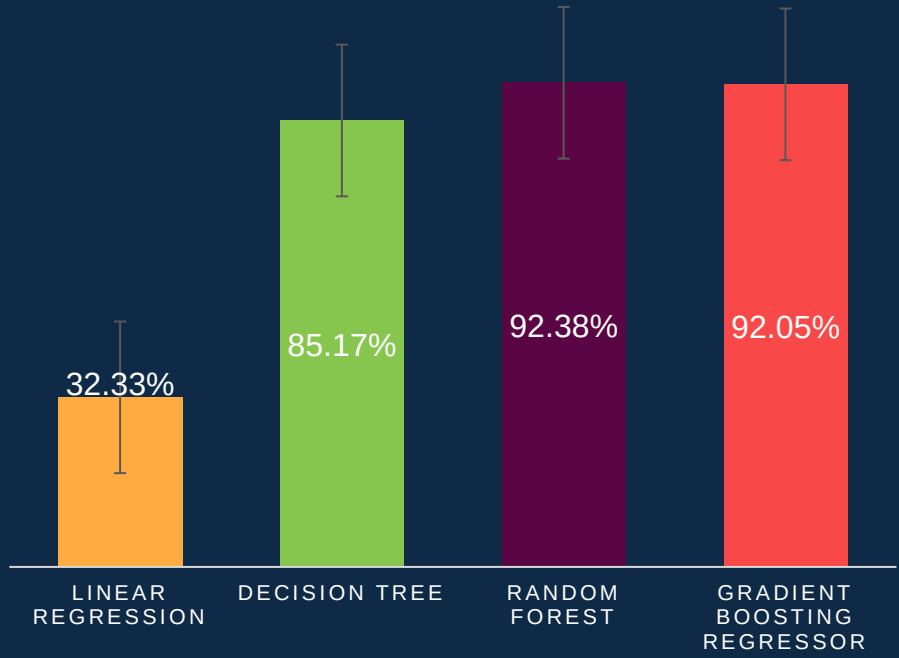
DEMO

Double click on object to view file

C:\Users\lenovo\
nloads\Motor_insu

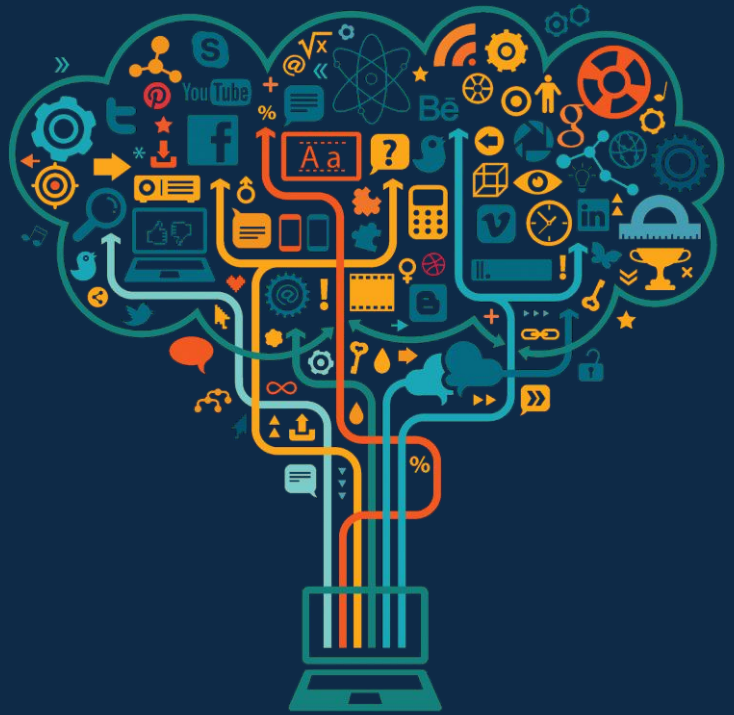
Conclusion

- The chosen model, Random Forest Regression, demonstrated high accuracy and robust predictive capabilities in estimating claim amounts.
- The predictive model can significantly improve the accuracy of risk assessment in motor insurance. It enables insurers to better understand claim likelihood and estimate claim amounts, leading to more informed decision-making processes.
- Incorporating additional data sources, such as telematics data or historical claims data, could enhance the model's predictive power.



Reference

- https://github.com/search?q=motor%20insurance%20claim&type=repositories
- https://medium.com/coders-camp/230-machine-learning-projects-with-python-5d0c7abf8265




Thank you!!!