Capstone Two - Project Proposal: Car Insurance Claim Prediction

Problem Statement Formation: Car insurance companies encounter difficulties in accurately predicting and assessing insurance claims, which can result in potential financial losses and inefficiencies in claims processing. The problem is to accurately predict and assess car insurance claims using machine learning algorithms, thereby enabling insurance companies to improve their claims processing efficiency and mitigate financial losses.

Context: By leveraging historical claims data, customer demographics, vehicle information, and driving behavior, insurance companies can improve their ability to predict the likelihood and severity of future insurance claims. This project seeks to address this need by utilizing data science techniques to analyze the provided dataset and develop a predictive model.

Criteria for Success: The success of this project will be evaluated based on the model's accuracy in predicting car insurance claims. Specifically, the project aims to achieve a high level of classification metrics in identifying potential insurance claims. The deliverables will include the developed machine learning model, code implementation, documentation outlining the methodology and findings, and potentially a presentation summarizing the project outcomes. A GitHub repo will contain the work for each step of the project, including a slide deck and a project report.

Scope of Solution Space: The solution space encompasses the development and implementation of a machine learning model capable of predicting car insurance claims based on various factors such as customer demographics, vehicle details, and historical claim data. The approach involves preprocessing the dataset, performing exploratory data analysis, feature engineering, selecting appropriate machine learning algorithms, training and validating the model, and evaluating its performance. This iterative process will involve fine-tuning the model to optimize its predictive accuracy.

Constraints:
- Availability and quality of the dataset
- Computational resources for model training and evaluation

Stakeholders:
- Car insurance companies
- Policyholders
- Regulatory authorities overseeing insurance practices

Datasources: The project will utilize a provided dataset containing car and policyholder details, as well as an insurance claim label. Link to the dataset: https://www.kaggle.com/datasets/ifteshanajnin/carinsuranceclaimprediction-classification?select=train.csv