Project Proposal

Yang Xiao

November 5, 2023

1 Introduction

1.1 Data Profiling

The dataset contains columns about cars: name, year, selling_price, km_driven, fuel, seller_type, transmission, and owner.

1.2 Description of the Problem

The used car market offers insights into consumer preferences and vehicle value depreciation. This project will use Bayesian statistical methods to analyze the Car Dekho dataset to identify factors affecting car valuations and survival times.

1.3 Motivation

The second-hand car market significantly impacts consumer choices and the economy, making its analysis vital for informed decision-making and understanding sustainability in terms of car lifespan.

2 Problem Statement

The goal is to model the selling price of used cars based on attributes and determine cars' longevity in the market, providing value and lifespan insights.

3 Method and Analysis Plan

3.1 Bayesian Regression Model

To predict selling prices based on car features, employing normal priors for coefficients and inverse-gamma for variance, with MCMC sampling for posterior estimation.

3.2 Hierarchical Bayesian Model

To consider the nested data structure, using hierarchical priors for make and model level variations, with MCMC sampling for estimation and WAIC or LOO-CV for model selection.

3.3 Bayesian Survival Analysis Model

To examine the time until a car is sold or deemed unsellable, using priors for Weibull or exponential distribution parameters, with MCMC methods for censored survival data and posterior predictive checks for model validation.

4 Model Implementation

Bayesian models will be implemented using Stan or PyMC3, ensuring computational feasibility and robust sampling.

5 Desired Result

This Bayesian analysis aims to uncover factors that influence second-hand car prices and market lifespan, offering valuable insights for consumers and dealers.