Table of Contents

[1. DATASET DESCRIPTION 2](#_Toc437474678)

[1.1 Problem Statement 2](#_Toc437474679)

[1.2 Motivation 2](#_Toc437474680)

[1.3 Feature Description 2](#_Toc437474681)

[1.4 Key Observations in data 2](#_Toc437474682)

[1.5 Data Preprocessing 2](#_Toc437474683)

[2. Results 2](#_Toc437474684)

[2.1 Something 2](#_Toc437474685)

[2.2 Something 2](#_Toc437474686)

[2.3 Something 2](#_Toc437474687)

[3. Analysis of Algorithms 2](#_Toc437474688)

[3.1 Something 2](#_Toc437474689)

[3.2 Something 2](#_Toc437474690)

[3.3 Something 2](#_Toc437474691)

# DATASET DESCRIPTION

## Problem Statement

We are trying to ***predict*** if a car bought at an auction by an auto dealer is a **Good buy** or a **Bad buy.**

## 1.1.1 Background

` When we go to buy a car at auto dealership we expect to get a good selection of car. Also we expect to trust in the condition of the car we are buying. These auto dealerships buy these cars from auctions and they have the same intent as us. However the problem which dealers face is with the cars which have some serious conditions and they turn out to be bad buys. These are called “kicks”, and this can happen due to variety of reasons.

## Motivation

It would greatly benefit both the auto dealers and the end buyers if there is a way to determine a car will be a kicked car. A simple analysis of the same is presented below.

*Note: All values are assumed values as per Auction Direct (a company which deals in second hand cars)*

|  |  |  |
| --- | --- | --- |
| **Legends** | **Amounts** | **Information** |
| Average number of cars bought and sold by a dealer | 15000 | By Auction Direct |
| %age kicked cars | 12.3 % | By Dataset |
| Number of kicked cars | 1845 | By Calculation |
| Average price of car sold | $ 10000 | By Auction Direct |
| Profit on good sale | $ 2000 | Average profit = 20% |
| Profit on bad sale (kicked car) | $ 500 | Due to repairs etc |
| Loss of potential profit | $ 1500$ |  |
| Total loss | **$ 2767500** |  |

This huge amount of potential profit can be converted into actual profit if there exists a model to predict a kicked car. Thus we chose this dataset.

## Feature Description

|  |  |
| --- | --- |
| **Field Name** | **Definition** |
| RefID | Unique (sequential) number assigned to vehicles |
| **IsBadBuy** | **Identifies if the kicked vehicle was an avoidable purchase** |
| PurchDate | The Date the vehicle was Purchased at Auction |
| Auction | Auction provider at which the vehicle was purchased |
| VehYear | The manufacturer's year of the vehicle |
| VehicleAge | The Years elapsed since the manufacturer's year |
| Make | Vehicle Manufacturer |
| Model | Vehicle Model |
| Trim | Vehicle Trim Level |
| SubModel | Vehicle Submodel |
| Color | Vehicle Color |
| Transmission | What type the transmission of the car Auto or Manual |
| WheelTypeID | The type id of the vehicle wheel |
| WheelType | The vehicle wheel type description (Alloy, Covers) |
| VehOdo | The vehicles odometer reading |
| Nationality | The Manufacturer's country |
| Size | The size category of the vehicle (Compact, SUV, etc.) |
| TopThreeAmericanName | Identifies if the manufacturer is one of the top three American manufacturers |
| MMRAcquisitionAuctionAveragePrice | Acquisition price for this vehicle in average condition at time of purchase |
| MMRAcquisitionAuctionCleanPrice | Acquisition price for this vehicle in the above Average condition at time of purchase |
| MMRAcquisitionRetailAveragePrice | Acquisition price for this vehicle in the retail market in average condition at time of purchase |
| MMRAcquisitonRetailCleanPrice | Acquisition price for this vehicle in the retail market in above average condition at time of purchase |
| MMRCurrentAuctionAveragePrice | Acquisition price for this vehicle in average condition as of current day |
| MMRCurrentAuctionCleanPrice | Acquisition price for this vehicle in the above condition as of current day |
| MMRCurrentRetailAveragePrice | Acquisition price for this vehicle in the retail market in average condition as of current day |
| MMRCurrentRetailCleanPrice | Acquisition price for this vehicle in the retail market in above average condition as of current day |
| PRIMEUNIT | Identifies if the vehicle would have a higher demand than a standard purchase |
| AcquisitionType | Identifies how the vehicle was aquired (Auction buy, trade in, etc) |
| AUCGUART | The level guarntee provided by auction for the vehicle (Green light - Guaranteed/arbitratable, Yellow Light |
| KickDate | Date the vehicle was kicked back to the auction |
| BYRNO | Unique number assigned to the buyer that purchased the vehicle |
| VNZIP | Zipcode where the car was purchased |
| VNST | State where the the car was purchased |
| VehBCost | Acquisition cost paid for the vehicle at time of purchase |
| IsOnlineSale | If the vehicle was sold online |
| WarrantyCost | Warranty price (term=36month and millage=36K) |

## Key Observations in data

* Redundant data: VehYear and VehAge mean the same thing
* Poor Quality of variables: PRIMEUNIT only 4.6% records were no
* Class Imbalance: 87.7 % Good Buys, only 13.3 % Bad buys
* There were no Manual transmission vehicles which were bad buys
* Only 0.11% records with RED category in AUCGUART

## Data Preprocessing

* Removed redundant features
* Removed features with more than 95% missing values
* Handles Null/Missing values
  + Continuous data: took average
  + Discrete data: created new catergory NULL
* Normalized all the continuous values in range [0, 1]

# Results

## Something

## Something

## Something

# Analysis of Algorithms

## Something

## Something

## Something

# Conclusion