

**AUTOMOBILES DATA ANALYSIS**

to

**Prof. Viswanathan V**

by

|  |  |  |
| --- | --- | --- |
|  | GROUP MEMBERS  **V Jai Prashanth 14BCE1019**  **K Akhil Reddy 14BCE1222** |  |

**Table of Content**

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | Data Content | 3 |  |
| 2 | Objectives | 4 |  |
| 3  4 | Methods  Attributes | 6  8 |  |
| 5 | Output | 15 |  |
|  |  |  |  |

**Data Content**

* This data set consists of three types of entities:
* (a) the specification of an auto in terms of various characteristics,
* (b) its assigned insurance risk rating,

* (c) its normalized losses in use as compared to other cars. The third factor is the relative average loss payment per insured vehicle year.

This value is normalized for all autos within a particular size classification (two-door small, station wagons, sports/speciality, etc...), and represents the average loss per car per year.

**Objectives**

* To classify the cars based on the Risk Factor involved and the Normalized Losses of a each automobile.
* To predict the Engine Size using linear multiple Regression with attributes.

**Methods:**

**K-Means Clustering**

K-Means clustering intends to partition *n* objects into *k* clusters in which each object belongs to the cluster with the nearest mean.

This method produces exactly *k* different clusters of greatest possible distinction.

The best number of clusters *k* leading to the greatest separation (distance) is not known as a priori and must be computed from the data.

The objective of K-Means clustering is to minimize total intra-cluster variance, or, the squared error function.

**Multiple Linear Regression**

Regression is a data mining technique used to predict a range of numeric values (also called *continuous values*), given a particular dataset.

For example, regression might be used to predict the cost of a product or service, given other variables.

Regression is used across multiple industries for business and marketing planning, financial forecasting, environmental modeling and analysis of trends.

**Attributes**

* **Symboling** - This rating corresponds to the degree to which the auto is more risky than its price indicates. Cars are initially assigned a risk factor symbol associated with its price.

Then, if it is more risky (or less), this symbol is adjusted by moving it up (or down) the scale. Actuarians call this process "symboling".

* A value of +3 indicates that the auto is risky, -3 that it is probably pretty safe.
* **Bore** is the diameter of each cylinder.
* **stroke** is the length that it travels when moving from bottom position to the top position.
* The static **compression ratio** of an internal combustion engine or external combustion engine is a value that represents the ratio of the volume of its combustion chamber from its largest capacity to its smallest capacity. It is a fundamental specification for many common combustion engines.

**Output:**

**Clustering**

![A screenshot of a cell phone

Description generated with very high confidence]()

![A screenshot of a social media post

Description generated with very high confidence]()

![A screenshot of a social media post

Description generated with very high confidence]()

![A screenshot of a cell phone

Description generated with very high confidence]()

![A screenshot of a cell phone

Description generated with very high confidence]()

![A picture containing screenshot

Description generated with high confidence]()

![A screenshot of a cell phone

Description generated with very high confidence]()

![A screenshot of a social media post

Description generated with very high confidence]()

![A screenshot of a cell phone

Description generated with very high confidence]()

![A screenshot of a cell phone

Description generated with very high confidence]()

![A screenshot of a social media post

Description generated with very high confidence]()

![A screenshot of a cell phone

Description generated with very high confidence]()