

Chenlong Jiang(2204207)

by Chenlong Jiang

Submission date: 10-Mar-2023 12:08PM (UTC+0000)

Submission ID: 198994790

File name: 42003_Chenlong_Jiang_Chenlong_Jiang_2204207_482892_730024523.rtf (58.43K)

Word count: 594

Character count: 3199

Source:Car Evaluation Data Set

Abstract: Derived from simple hierarchical decision model, this database may be useful for testing constructive induction and structure discovery methods.

Data Set Information:

Car Evaluation Database was derived from a simple hierarchical decision model originally developed for the demonstration of DEX, M. Bohanec, V. Rajkovic: Expert system for decision making. Sistemica 1(1), pp. 145-157, 1990.). The model was designed to evaluate cars according to the following concept structure.

The Dataset:

The dataset consists of 1728 records, each containing 6 attributes, and the number of Web Hits is 1717593.

This program evaluates the current value of vehicles based on

1. Overall price 1.1 Purchase price (vhigh, high, med, low)
 1.2 Maintenance price of the vehicle in the database(vhigh, high, med, low)
2. Technical characteristics. 2.1 Comfort. 2.1.1 Number of doors. (2, 3, 4, 5more)
 2.1.2 Personnel carrying capacity.(2, 4, more)
 2.1.3 Luggage compartment size (small, med, big)
 2.2 Estimated safety of vehicles(low, med, high)
3. Class Values: unacc, acc, good, vgood

Why I choose this data set:

In China, the public acceptance of second-hand cars has gradually improved, and there is huge potential for development now and in the future. However, the problems of the second-hand car market have gradually emerged with its prosperity and development. The characteristics of "one car, one situation" of second-hand cars determine the complexity of its transactions. There are many second-hand car brands, and different brands of cars are complex, and the degree of damage is also different. Their prices are affected by many factors, It is difficult to accurately estimate and set the value. At the same time, in the second-hand car trading market, the information asymmetry between the buyer and the seller undoubtedly increases the difficulty of the second-hand car price evaluation. In addition, the country has not yet issued a unified standard for the second-hand car evaluation. Therefore, it is urgent to establish a reasonable evaluation system, adopt a more scientific and accurate evaluation model to help regulate the pricing business of the second-hand car market, and improve the credit system of the second-hand car market, To ensure the smooth transaction of second-hand cars

What I am hoping to find within it:

1. Based on the analysis of some features and combined with the actual sales scenario, suggestions are put forward to speed up the sales of vehicles in the store.
2. Prepare for learning how to establish a quality evaluation system.
3. Whether there are other important factors affecting the price of second-hand cars, and if so, how to combine them with the analysis methods of this data set to produce a more advanced and accurate evaluation method

Why I feel I have the necessary background knowledge:

The author believes that he majored in vehicle engineering, so he has a certain understanding of the production and sales of vehicles, and can better understand the significance of various data in the process of analyzing this data set.

Plan steps to analyze this dataset:

1. Data collection and data correlation analysis
2. Data processing based on correlation
3. Model training
4. Conclusion

Relevant Papers:

M. Bohanec and V. Rajkovic: Knowledge acquisition and explanation for multi-attribute decision making. In 8th Intl Workshop on Expert Systems and their Applications, Avignon, France. pages 59-78, 1988.

B. Zupan, M. Bohanec, I. Bratko, J. Demsar: Machine learning by function decomposition. ICML-97, Nashville, TN. 1997 (to appear)

Video: The Chinese Internet has a firewall, and students in China cannot access youtube unless we cross the wall, which is illegal, so I will send the video to your mailbox later.

FINAL GRADE

GENERAL COMMENTS

Instructor

57 /100

PAGE 1

PAGE 2

GRADING FORM: DATA MINING/ VISUALISATION

CHENLONG JIANG

57

KNOWLEDGE AND UNDERSTANDING

Dataset Suitability (What features of the dataset lend themselves to data mining? Has it been mined previously? Why does the student feel that they have the contextual background to work with this data?)



Reason for choice of dataset is valid.Prior dataset analysis not made explicit.

20

COGNITIVE AND INTELLECTUAL

Objectives (What does the student plan to investigate/expect to find? Are these SMART?)



Proposed questions are measurable and would provide a great starting point for analysing data. But more information needed to be provided for said data for readers of document to have a clearer picture of proposed analysis.

8

PRACTICAL APPLICATION SKILLS

Dataset Description (Where did the data come from? What does the data describe? Is the sample large enough? What data formats and types are involved?)



No link to dataset is provided. Very limited description of dataset in summary document.

25

TRANSFERABLE SKILLS FOR LIFE

Presentation/Slides (Clarity, readability, flow, accuracy, completeness, performance)



No space between headings and content. Headings needed to be embolden.

4