International Journal of Advances in Computer Science & Its Applications
Volume 6 : Issue 1 [ISSN 2250-3765]
Publication Date : 18 April, 2016

A Web-based Decision Support System for Used-Car Pricing

Ayhan Demiriz

***Abstract*— Used-car trade has a significant portion in overall automobile market and determining the values of the cars is an important problem. This study proposes a new methodology for determining the market value of the used-cars by observing the classifieds in an e-commerce site. This type of data acquisition plays an important role to build pricing models and to conduct further analysis in our approach. In data acquisition stage, a set of new listings are chosen randomly each day from an e-commerce site (a web site like Craigslist), then these listings are observed until a pre-determined period (e.g. thirty days) or delisting time, whichever comes first. The crucial part of our approach is the assumption of a sale event when the listing is no longer available i.e. delisted from the e-commerce site. The proposed methodology may potentially be used for pricing any used item based on the web listings. A web site was developed to help clients/users for determining the market values of their cars as a decision support tool that can assess the likelihood of selling a particular car at a certain price. We also presented the applicability of predictive models to determine the likelihood of selling a car within thirty day period based on the price set by the owner.**

***Keywords*— Used-Car Pricing, Web-based Decision Support, Logistic Regression**

I. Introduction

According to recent statistics, used-car sales generate approximately half of the total vehicle trade annually in terms of volume in the U.S. [1]. Therefore it is important to study the used-car market where price is one of the most important determinants of sales [2]. Setting the asking price high may discourage any potential buyer, which practically hinders the buyer to visit the dealership and/or contacting the private seller. On the other hand, lowering the asking price too much will definitely expedite the sales by lowering the profit of the seller/dealership [3].

We consider the used-car listings from an e-commerce site as our subjects and we follow them until they are delisted or a 30-day period expires. We assume that the delisted cars are sold at the time of delisting. Basically, delisting i.e. sale of a car is considered as death (or failure) in survival analysis analogy. Since lifespan of web listings may resemble lifespan of any subject that is observed during a survival analysis study, survival analysis is also an important component of our methodology. In survival analysis [4], subjects (laboratory animals, people, light bulbs, machines, and so on) are followed during the experiment until a specific event occurs (in most cases failure or death) or until the experiment ends. The data in consideration are primarily event times as well as some characteristics of the subjects that may or may not change over time. The data for subjects that have not experienced the event (such as those who survive) are censored. The exact event time is not known for the censored data, but it is known to have occurred after the censored time. Survival analysis is highly popular in medical sciences, especially in new treatment and related drug studies.

In [5] and [6], survival analysis based pricing models were introduced. In this study, we introduce a web site as a decision support tool to aid buyers/sellers for determining the used-car prices.

II. Data Acquisition

As the Internet has changed so many things around us, it practically affected used-car markets as well. First of all, the Internet reduced the cost of getting information about the used-cars on the market dramatically [7], [8]. On the other hand, the Internet offers successful referral services helping buyers to redeem the additional discounts on used-cars [8]. There is strong empirical evidence that the Internet enables the car buyers to be more informed [8], [7] regarding the prospective used-cars on the market which practically alleviates the negative effects of information asymmetry about the used-cars. Thus, the Internet has effectively replaced the traditional media such as newspaper and magazine listings around the globe for the used-car classifieds. In contrast to some early work [9], [10] that used newspaper listing as their data source; it is now common to collect used-car related data from the Internet-based sources [11]. In order to drive price response function correctly, we need to follow the prices of used-cars while they are on the market [3].

Considering that used-car web sites such as mobile.de, autotrader.com, motors.co.uk are popular destinations around the globe, we have collected data from the most popular e-commerce site for the second-hand items (e.g. cars, computers, real estate and so on) in Turkey, Sahibinden.com (translates as "from the owner"). In this study, we observed randomly selected used-car listings up to thirty days to collect car related data including sellers' location and price changes. Between September 30, 2013 and February 10, 2014, approximately a hundred new listings were selected each day to gather data from the e-commerce site, sahibinden.com. In addition to new randomly selected car listings each day, we fetched web data to observe the car listings that were previously added to our database. Basically each used-car listing was observed up to thirty days or until they were delisted. Approximately 6,800 car listings were observed in our data acquisition stage. In addition to some crucial data about the listings, such as make,

Ayhan Demiriz
Department of Industrial Engineering, Sakarya University
TURKEY

model, trim line, year, odometer reading, listing date, location and price, buyers can see other information about the cars including engine size and power, transmission type, color, body type, owner type (private, dealer), trade in possibility and warranty at this particular e-commerce web site. There is also a text box for the owners to give additional information about the car on the web page. We did not fetch data from the text box area for this particular study.

Throughout the data acquisition stage, we randomly picked approximately 100 new listings every day, excluding luxury cars that were priced above 150,000 Turkish Lira (TRL). Occasionally, some listings may have slipped this filter due to the price changes during the data acquisition stage. In preliminary data acquisition stage, we noticed that dealers have a tendency to pull back their listings arbitrarily and relist them after a while. Therefore we preferred to use the listings of the private sellers. After the completion of the first stage, the collected data were cleaned by removing some unreasonable cases such as the listings with very low mileage but very old cars that were potentially misleading.

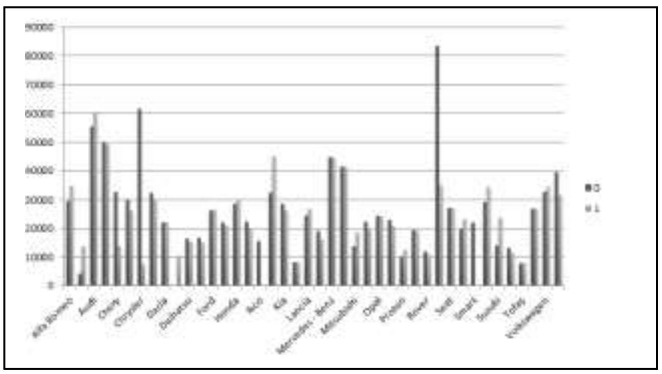

Figure 1 Average Asking Price By Make (0-Unsold Cars, 1-Sold Cars)

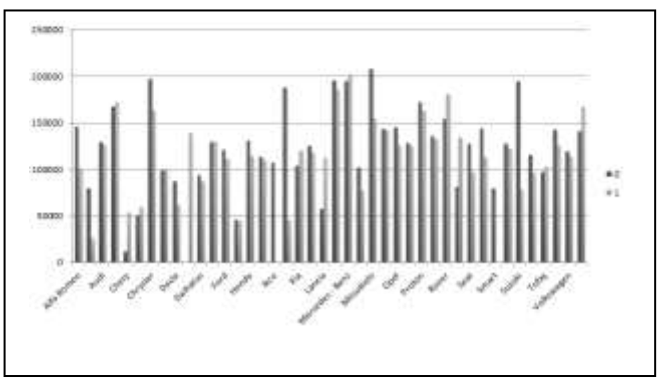

Figure 2 Average Odometer Reading By Make (0-Unsold Cars, 1-Sold Cars)

Figure 1-Figure 4 summarize some statistics from the data. Average asking prices by different makes are depicted in **Figure 1** for both sold (“1”) and unsold (“0”) cars. Most car brands have average asking prices around 20,000 TRL. Saab cars have the highest average asking price. Notice that there were very few Saab cars on the market and most of them were not sold. It was definitely an outlier. On the average, many listed car brands have odometer readings above 100,000 km mark as seen in **Figure 2**. Some brands have higher averages of odometer reading for the unsold cars compared to the sold cars. In **Figure 3**, we summarized the average age by car brands that are on the market. Many brands are between 5 and 10 years old on average. **Figure 4** shows the numbers of sold and unsold cars by most frequent brands in our dataset. Tofaş is a defunct car brand that licensed Fiat cars to manufacture in Turkey between 1970s and 1990s. Fiat no longer manufactures cars under license agreement with a local manufacturer; but it manufactures cars with its own brand since late 1990s. Renault, Fiat, Ford, Hyundai, Toyota, and Honda have their own assembly plants in Turkey. Opel used to have an assembly plant; but it seized the operations in the early 2000s. However, it is still a popular brand in Turkey as seen from **Figure 4**.

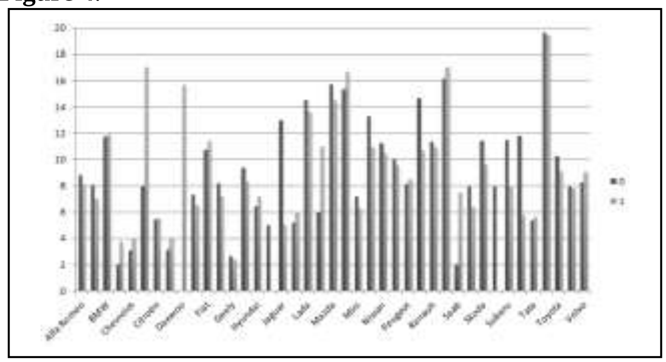

Figure 3 Average Age By Make (0-Unsold Cars, 1-Sold Cars)

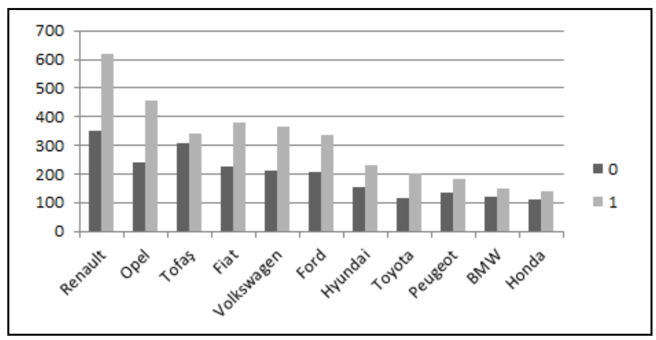

Figure 4 Sales Figures of Popular Makes

III. A Decision Support Tool for Pricing Used-Cars

A simple decision support tool may help buyer/seller better in some cases to determine the fair market value of a car than a complex pricing model. In this section, we introduce a web site, www.arabamkaca.net, which contains several modules to aid a pricing decision for the used-car buying and selling. This particular web site is designed from an academic perspective and is limited only for the collected data at the moment.

The web site contains following modules:

1. Market Value
2. Sales Ratio
3. Visualization
4. Predictive Model.

International Journal of Advances in Computer Science & Its Applications
Volume 6 : Issue 1 [ISSN 2250-3765]
Publication Date : 18 April, 2016

Figure 5 shows the result of a query of a particular make (brand) and series from year 2004. Market values are returned based on average asking prices in corresponding car type. Average kilometers (mileage) are also returned. The average results from only sold cars and also from the full data are both reported. In this particular query, the market value of year 2004 Opel Astra cars is returned. As seen in **Figure 5**, sold cars have lower mileage and lower prices. In other words, some cars with higher mileage and higher prices were probably not sold within the 30-days period.

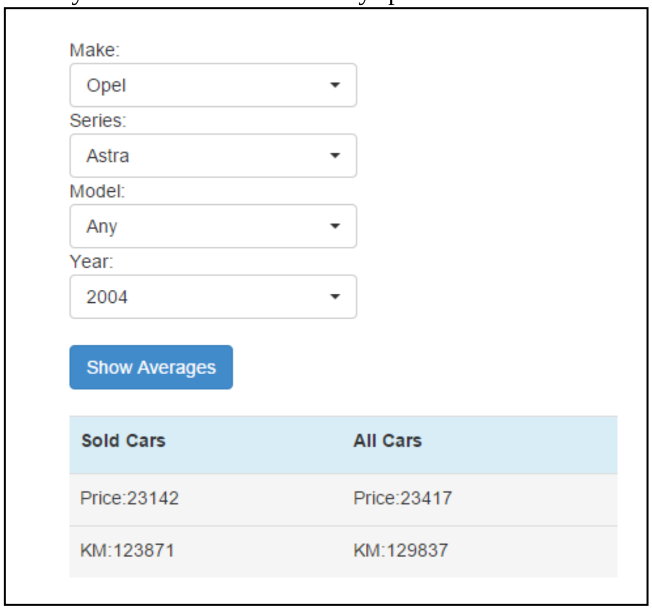

Figure 5 Presenting Average Values

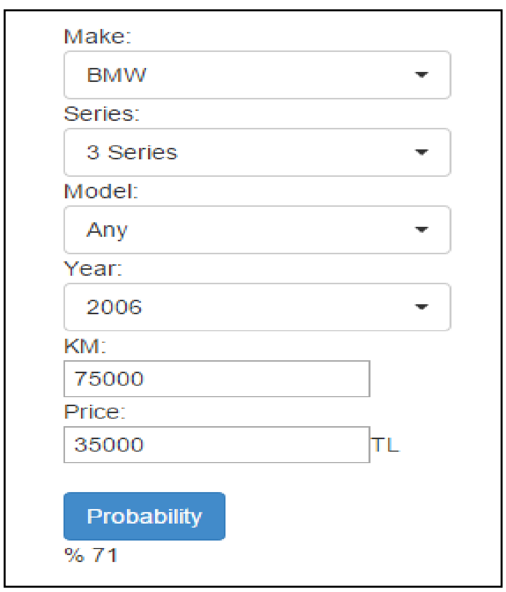

Figure 6 Sales Probabilities

Figure 6 depicts an outcome of this study that can be considered innovative. None of the current web sites, that cater the needs of car buyers'/sellers' by providing price information, provides an assessment of likelihood of selling a particular car with a particular price. However, our decision support tool can return a probability of selling a particular car with a set price as seen from **Figure 6**. The figure shows that a year 2004 BMW 3 Series car with 75K kilometers at 35K TRL has a likelihood of 71% being sold within 30 days. One can easily find that probability by taking ratio of number of sold cars that are similar model, series and age with higher mileage and price to number of all similar cars.

Figure 7 visualizes average prices and odometer readings of cars on the market in a two-axis plot. One can pick a particular make and model to visualize its average market values from various years. Such a two-axis plot is not currently provided by any used-car related web sites like www.kbb.com . This is also a useful tool to determine the right market value of a car.

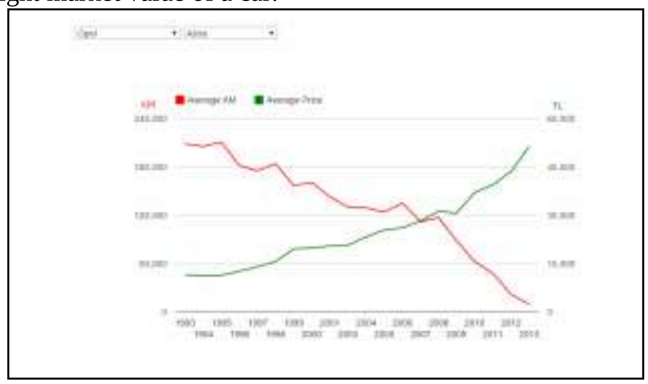

Figure 7 Visualizing Average Prices and Odometer Readings per Year

We introduced three different modules from a web based decision support tool in this section. Sometimes user may prefer predictive models that can assess the probability of selling a particular car at a certain price. In the next section, a logistic regression based model is implemented as a part of decision support tool introduced in this section.

IV. A Logistic Regression Based Predictive Model

In certain cases, one may desire to model a binary outcome such as our case i.e. sold cars and unsold cars. In this section, a predictive model based on logistic regression is introduced. Due to the nature of logistic regression, our model can yield a probability of event i.e. a sale of the car.

Logistic regression is used when the computation of a class probability is needed where this probability is a linear function of predictor variables [12]. A binary classification case can be specified by a **logit** (log-odds) transformation as follows:

$$\text{logit}\, p = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k$$

in which p is the probability of the event, k is the number of predictors (independent variables) and **logit** $\boldsymbol{p}$ is equal to ***log(p/(1-p))*** (i.e. *log odds*). Notice that the probabilities of event (p) and non-event ($1-p$) sum to one. Parameters of logistic regression can be computed by maximum likelihood estimation. The choice of logistic regression in this study as a predictive model is not arbitrary. First of all, it uses a linear function in predictor variables. In addition, it can give us probability of the event which is an important assessment for a potential car buyer/seller. Once the parameters are estimated, then the probability of the event, p, can be computed as in following:

$$p = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)} \, .$$

Figure 8 shows a prediction result of a particular car where the specifics of the car can be seen. In this example, the probability of the sale is almost 0.90. Notice that the specifics of the car can be chosen in this module including asking price and odometer reading. Logistic regression models were run by using **LOGISTIC** procedure of SAS for each make and model pair separately. Estimated parameters were store in a SQL table to enable predicting for any particular car later on. We believe that this module is also an important component of our decision support tool as seen from **Figure 8**. A predictive model always needs to be tested for unseen points, but our primary aim in this paper is to show the usability of collecting data from classified web listings. For useful and sound predictive models, one certainly needs sufficient data. On the other hand, we believe that this paper has sufficient merit to show the applicability of our methodology.

Figure 8 Implementing Logistic Regression

V. Conclusion

In addition to statistical pricing models implemented in [5] and [6], providing the users with decision support tools may improve better determining the used-cars' market values. We introduced some new tools that may be very innovative and useful in terms of the value to the decision makers (buyers/sellers).

Collecting further data may certainly help in building better predictive models that are proven to be good for the unseen data. Our tools show the importance of a predictive model. Data acquisition process was not fully automated in this study. We plan to fully automate this process to collect further data to build better predictive models to determine whether a car with certain characteristics and asking price will be sold within a period or not. This is indeed a very crucial step that our methodology may improve best, which may enable us to build a commercial quality web site in future for the decision support.

Acknowledgment

I want to thank to Sumeyye Kaynak for her help in coding the web site.

References

[1] Bureau of Trasportation Statistics, "New and Used Passenger Car Sales and Leases." 2010.

[2] S. Singh, B. T. Ratchford, and A. Prasad, "Offline and Online Search in Used Durables Markets," *J. Retail.*, vol. 90, no. 3, pp. 301–320, 2014.

[3] A. Jerenz, "Survival Analysis: Estimation of the Price-Response Function," in *Revenue Management and Survival Analysis in the Automobile Industry*, Gabler, 2008, pp. 63–96.

[4] D. G. Kleinbaum and M. Klein, *Survival Analysis A Self-Learning Text*, Second. New York: Springer-Verlag, 2005.

[5] A. Demiriz, S. Sen, and F. Gokcegoz, "Analyzing Classified Listings at an e-Commerce Site by Using Survival Analysis," in *European Conference of Technology and Society (EuroTecs2013)*, 2013.

[6] A. Demiriz and S. Kaynak, "Using Survival Analysis For Determining Used-Car Market Values," in *International Symposium on Intelligent Manufacturing and Service Systems (IMSS'14)*, 2014.

[7] F. S. Morton, F. Zettelmeyer, and J. Silva-Risso, "Internet Car Retailing," *J. Ind. Econ.*, vol. 49, no. 4, pp. pp. 501–519, 2001.

[8] F. Zettelmeyer, F. S. Morton, and J. Silva-Risso, "How the Internet Lowers Prices: Evidence from Matched Survey and Automobile Transaction Data," *J. Mark. Res.*, vol. 43, no. 2, pp. 168–181, May 2006.

[9] J. G. Odink and E. van Imhoff, "Prices of used cars in West-Germany before and after the second energy crisis," *Economist (Leiden).*, vol. 130, no. 3, pp. 381–396, 1982.

[10] P. Kooreman and M. Haan, "Price Anomalies in the Used Car Market," *Economist (Leiden).*, vol. 154, no. 1, pp. 41–62, 2006.

[11] J.-D. Wu, C.-C. Hsu, and H.-C. Chen, "An expert system of price forecasting for used cars using adaptive neuro-fuzzy inference," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7809–7817, 2009.

[12] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2001.

About Author:

Ayhan Demiriz received his Ph.D. degree from the Department of Decision Sciences and Engineering Systems at Rensselaer Polytechnic Institute, Troy, NY in 2000. He is currently teaching in the Department of Industrial Engineering at Sakarya University. His current research interests are in Merchandising, Markdown Optimization and Call Center Product Recommender Solutions.