iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: XII Month of publication: December 2017

DOI:

Performance Comparison of Machine Learning Models

Sandeep Kumar[1]
[1]Computer Science & Engineering Department, MSIT, GGSIPU.

Abstract: This paper provides an insight about the price prediction scenario of online auctions. Here different ways to examine the price forecasting techniques have been examined. There have been a special emphasize on the machine learning models. Using the classification and regression analysis, the use of existing dataset on different machine learning models is shown. This include the price forecasting graphs under different machine learning models. The dataset used to carry out the experiment uses two types of features one is original one and the other is the set of derived features. At the end the results obtained from different models have been compared to find the scope of improvement. The results can be further utilized to predict a better price forecasting technique for online auctions.

Keywords: Machine learning models, regression analysis, price prediction of online auctions.

I. INTRODUCTION

Different methodologies have been used for predicting the end price in the online auction environment [1]. These approaches try to resolve the forecasting problems faced while using machine learning, functional data analysis and time-series analysis techniques. In a data mining based multi-agent system design [2], a multiple online auction environment for selecting the auction is used, where the traded item is being sold at the lowest price. Here, one of the important questions that seller encounter is how to list their commodities [3]. For online auctions, sellers need to look into many auction settings like starting bid price, reserve price, duration time, and whether to use buy-it-now or advertising option, etc. Many a time seller put a high starting price in order to maximize revenue, however, many times; Sellers keep the price low and emphasize advertising to increase sale probability. In fact, in the pursuit of maximized profit, the seller has to make a trade-off between sale probability and revenue maximization [4].

To find an auction setting that could maximize sellers' profit is a challenging problem. However, it is not easy to ascertain the best auction settings. Given an auction setting, should the current auction setting be used for the given item? Furthermore, if there exists a service that could predict the expected profit, then we might apply such services to ascertain the best auction setting on a commodity, which is advisable for a specific seller [5].

There are several researches on end-price (or closing price) prediction for online auction [6, 7, and 8]. Ghani and Simmons apply three models, including regression, multi-class classification, and multiple binary classification tasks, to predict auction end-price. A dynamic forecasting model based on functional data analysis, which can predict the end price of an "in-progress" auction has been explored by few researchers in this concerned area of online auctions [9, 10]. Such a service is of more significance for bidders to skip auction items with high end-price and focus on others with a potentially lower price. However, for the decision support of commodity listing, dynamic forecasting is not significant since sellers could not change the auction setting when an auction begins. A dynamic is that can forecast the price of an ongoing auction and can update its prediction based on current incoming information. Predicting the price in online auctions is difficult because old forecasting ways cannot sufficiently account for few features of online auction data like the changing dynamics of price and bidding throughout the auction, the unequal intervals of receiving bids. Here one can consider classification and clustering techniques are used to forecast the bid. Predicting end price of an online auction has been strongly considered as a machine learning problem and has been handled using regression trees, multi-class classification and multiple binary classifications [11]. Among these machine learning techniques, showing the price prediction as a series of binary classification can be verified to be one of the best suitable techniques for this task. The past track of an ongoing auction contains significant information and is utilized for the short term forecasting of the next bid by using support vector machines, functional k-nearest neighbour, clustering, regression and classification techniques [12, 13 14].

II. IMPLEMENTATION AND ANALYSIS

The data set used here has been collected through an automated software agent from the e-Bay website [15]. The player reference data have been referred through sportscollector.net. The sentiment data have been collected using twitter API and the text analysis of customer reviews [16]. There are two types of characteristics that have been used in the data set. One is direct, and the second is

the derived feature set of the data. The direct features of data include: Price, starting bid, bid count, Title, Quantity Sold, Seller Rating, StartDate, End Date, Feedback, Has Picture, Has Store, Seller Country etc. The derived dataset features used are Is Hall of Fame, Avg Price, Median Price, Seller, Close Percent. Item Auction Sell Percent, Start of week, End of Week etc.
Here the classification performed, regression and prediction by the following implementations. It also highlights the performance level and the efficiency of corresponding technique.

A. Advanced Price Regression

Models In price determination, it is very important to know about the different variables affecting the final value. Not every chosen value affects the final price equally, but each creates a different degree of impact. After taking up the explanatory variables, those have been taken up against different regression models. The target of regression models is to recognize the price value of the combination of variable names [22].
Now, up to this stage a ready to use data set is available. This dataset is utilized for testing the performance under different regression models. The models that have been taken up for analysis are:

1) Stochastic Gradient Boosting with Xgb Tree
2) Simple Neural Network
3) Boosted Trees
4) CART Method - Decision Trees
5) General Linear Models
6) Ridge Regression
7) Robust Linear Model

Every model here passes through parameter tuning, then model have been trained and in last, the predicted values are shown using graphs plotted. In the below analysis, the same dataset has been used. On this dataset different models have been implemented. These show the graphs plotted towards the price prediction

1) *Stochastic Gradient Boosting with Xgb Tree:* Extreme Gradient Boosting is a very efficient and scalable technique for the implementation of gradient boosting trees. It supports ranking, classification and regression. It is normally 10 times faster than gbm but its customization and result tendency is very good [17, 18].

Fig.1. Predicted Values and Plots for Xg boost

2) *Simple Neural Network:* The simple neural network also provides some good results. Here the results for current data set can be taken up for choosing the best approach [19].

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor :6.887
Volume 5 Issue XII December 2017- Available at www.ijraset.com

Fig. 2. Predicted values in neural networks.

As it can be seen, neural networks can give highly intense results, very stable, but processing time is not effective. Furthermore, there are better methods that deal with the predictive analysis of the price's problem pretty well.

3) Boosted Trees Predicted Values with boosted tree are shown in the figure below:

Fig. 3 Predicted Values in Boosted Trees.

III. CART METHOD - DECISION TREES [20]

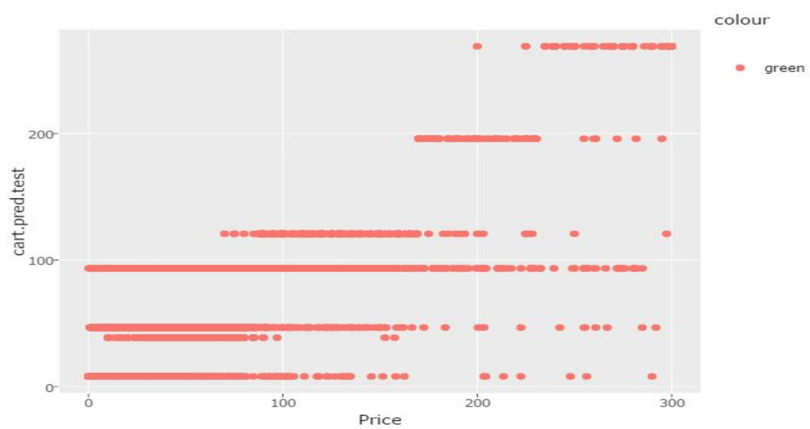

Fig. 4 Predicted values in CART method.

IV. GENERAL LINEAR MODELS

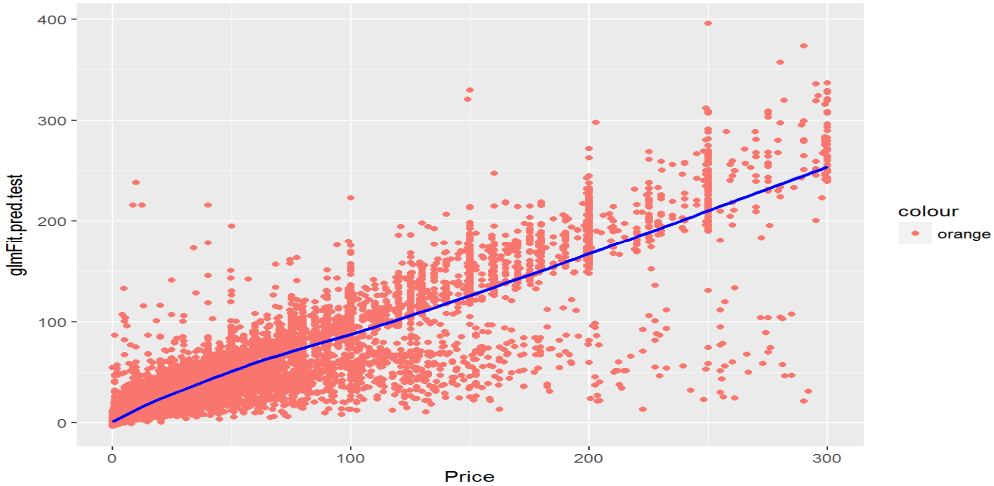

Fig. 5 Predicted values in linear model.

V. RIDGE REGRESSION MODEL [21]

Fig. 6 Predicted values in ridge regression.

VI. ROBUST LINEAR MODEL

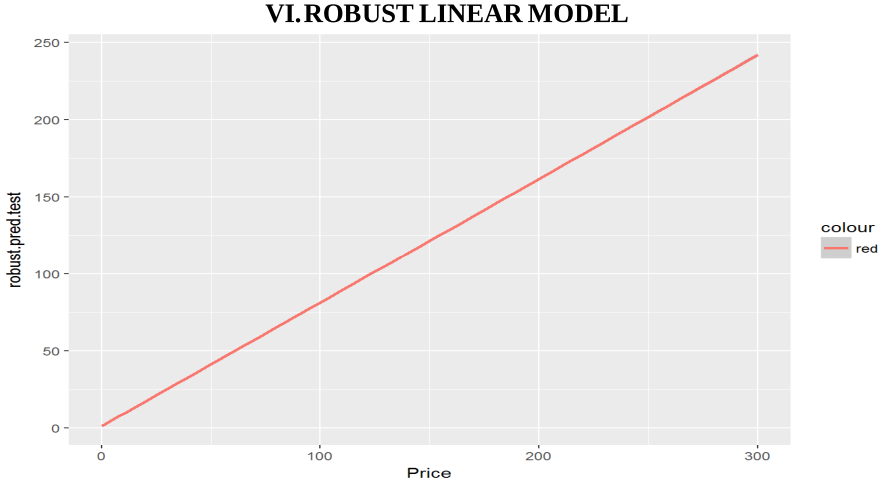

Fig. 7 Predicted values in robust linear model.

VII. CONCLUSIONS

Metrics that is used here includes R-Squared - which tells us how much variabilities of the explanatory variable is explainable by variables and RMSE, i.e. Residual mean square error, which tells us how long distance between real value and value is predicted by the model. The comparison to all models including boosting trees, which are similar to decision trees, but in weak learner class, these models have the best possibilities. The other models tried include linear regression, robust linear regression, and Simple neural network, tend to give the best results, but it will take a long time to tune that model, and it can be over fit on the test data set.

Here, boosted trees have shown very good performance on predicted values. The predicted values do not have to be scaled (as in xgboost, when the result should be multiplied by 95 to indicate a factual value), and this model has better performance, Rsquares comes 0.91, which is very good, and best RMSE from every model that has been tried. The weight of evidence can indicate Good Rate, Bad Rate in each level of the variable. So one can decide to assign "Bad" if predicted value differs from real value more than 19, and "Good" otherwise. Then it has been analysed to see what is a good rate in each group (groups are calculated automatically by combining library), and which group has the lowest good rate. The group with the lowest good rate can tell us that in that group, one can have a higher probability of values that are predicted wrong. In case here, the Group with Higher Probability of the Wrong Prediction is a group where p takes a value higher than 100.

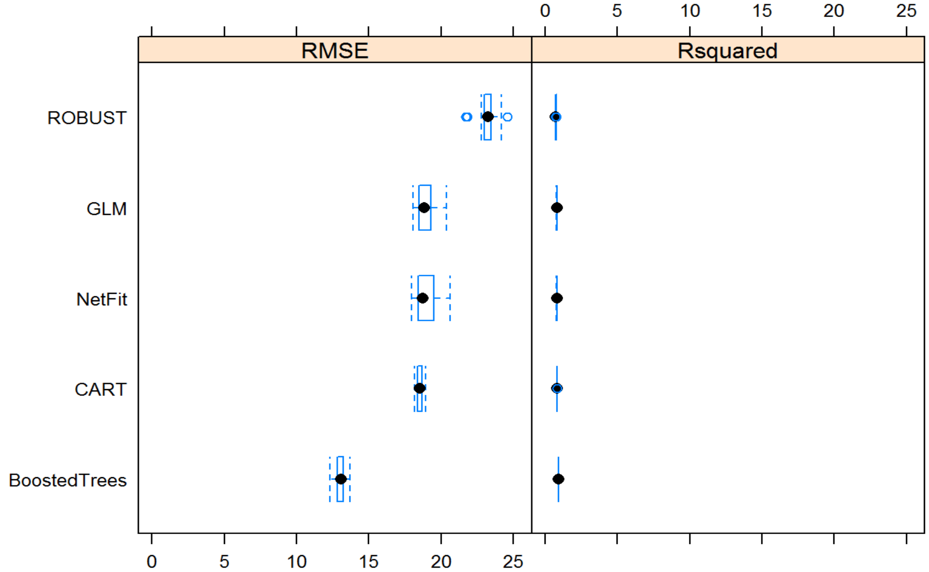

Fig. 8 Efficiency comparison of different models-I.

VIII. ACKNOWLEDGMENT

I would like to thank Dr. Rahul Rishi sincerely. His motivation, positivity and direction always guides me to keep my work going.

REFERENCES

[1] P. Kaur, M. Goyal, and J. Lu, "A Comparison of Bidding Strategies for Online Auctions using Fuzzy Reasoning and Negotiation Decision Functions," IEEE Trans. Fuzzy Syst., 2016.
[2] M. Dewan and F. Lin, "Dynamic pricing mechanism for multi-agent based system of well scheduling," Electron. Vis. (ICIEV), 2016 5th …, 2016.
[3] H. Wang, "Analysis and design for multi-unit online auctions," Eur. J. Oper. Res., 2017.
[4] M. Rudolph, J. Ellis, and D. Blei, "Objective Variables for Probabilistic Revenue Maximization in Second-Price Auctions with Reserve," 25th Int. Conf. , 2016.
[5] Shi, L. Zhang, C. Wu, Z. Li, and F. Lau, "An online auction framework for dynamic resource provisioning in cloud computing," ACM SIGMETRICS Perform., 2014.
[6] N. Chan and W. Liu, "Modelling and Forecasting Online Auction Prices: A Semiparametric Regression Analysis," J. Forecast., 2016.
[7] R. Ghani and H. Simmons, "Predicting the end-price of online auctions," Proc. Int., 2004.
[8] P. Kaur, M. Goyal, and J. Lu, "A proficient and dynamic bidding agent for online auctions," Int. Work. Agents Data, 2012.
[9] N. Chan, Z. Li, and C. Yau, "Forecasting Online Auctions via Self-Exciting Point Processes," J. Forecast., 2014.
[10] S. Wang, W. Jank, and G. Shmueli, "Explaining and forecasting online auction prices and their dynamics using functional data analysis," J. Bus., 2008.
[11] P. Kaur, M. Goyal, and J. Lu, "Pricing analysis in online auctions using clustering and regression tree approach," Int. Work. Agents Data, 2011.
[12] S. Petruseva, P. Sherrod, V. Pancovska, and A. Petrovski, "Predicting Bidding Price in Construction using Support Vector Machine," TEM J., 2016.
[13] T. Pinto, T. Sousa, I. Praça, Z. Vale, and H. Morais, "Support Vector Machines for decision support in electricity markets' strategic bidding," Neurocomputing, 2016.

[14] V. Almendra and D. Enachescu, “Using self-organizing maps for fraud prediction at online auction sites,” Symb. Numer., 2013.
[15] Grossman Jay, “Jay Grossman.” [Online]. Available: http://jaygrossman.com.
[16] S. Kumar and R. Rishi, “Data collection and analytics strategies of social networking websites,” Green Comput. Internet Things , 2015.
[17] C. Tantithamthavorn, S. McIntosh, and A. Hassan, “Automated parameter optimization of classification techniques for defect prediction models,” Proc. 38th, 2016.
[18] T. Chen and T. He, “Xgboost: extreme gradient boosting,” R Packag. version 0.4-2, 2015.
[19] Sutskever, O. Vinyals, and Q. Le, “Sequence to sequence learning with neural networks,” Adv. neural Inf., 2014.
[20] J. Strickland, Predictive Modelling and Analytics. 2014.
[21] S. Chatterjee and A. Hadi, Regression analysis by example. 2015.
[22] S. Kumar and R. Rishi, “Hybrid dynamic price prediction model in online auctions,” Int. J. Appl. Eng. Res., vol. 12, no. 5, pp. 598–604, 2017.