International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 5 (2017) pp. 598-604

Hybrid Dynamic Price Prediction Model In online Auctions

Sandeep Kumar[1] and Rahul Rishi[2]

[1,2]CSED (UIET), MD University,
Rohtak, India.

Abstract

Online auction has turned into an extremely well known e-business exchange sort. The massive business openings pull in a great deal of employment and large number of online stores. In this paper we outlines a hybrid auction model (HDAM) which uses a detailed technique for initial price estimation and predictive analytics. To begin with, HDAM inputs an online auction to participate in and explores its underlying starting price in light of clustering and regression and then prediction is implemented in view of the evaluated starting price, state of mind of the bidders to win the auction and the competitor's appraisal for the bidder's offers. The tests showed value estimation results utilizing the proposed approach.

Keywords: Online Auctions, Social Media, Data mining, Clustering Algorithms, Bidding strategies, Predictive Analysis, Software Agents, Fuzzy reasoning, Price forecasting.

INTRODUCTION

Different methodologies have been presented for end price prediction in the online auctions Environment. These approaches resolved the forecasting problem using machine learning, functional data analysis, and time series analysis techniques. In [1], a data mining based multi-agent system has been designed in favour of a multiple on-line auctions environment for selecting the auction, in which the traded item will be sold at the lowest price. One of the important questions that sellers encounter is how to list their commodities.

For online auctions, sellers need to decide many auction settings like starting bid price, reserve price, duration time, and whether to use buy-it-now or advertising option, etc. Many times sellers put a high starting price in order to maximise revenue, however many times sellers keep the price low and purchase 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 maximisation.

How to find an auction setting that could maximize sellers' profit is a challenging problem. However, it is not easy to ascertain the best auction setting. Thus, we turn to an easier question: 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 a commodity should use for a specific seller.

There are several researches on end-price (or closing price) prediction for online auction [2][3][4]. Ghani and Simmons apply three models, including regression, multi-class classification and multiple binary classification tasks, to predict auction end-price. 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 area [5, 6].

Such a service is more important for bidders to skip auctions items with high end-price and focus on others with potentially low 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.

In dynamic model mean that it can forecast the price of an ongoing auction and can update its prediction based on currently 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 received bids. Our hybrid dynamic forecasting model accounts for these special features by using modern functional data analysis techniques and tools like using inputs from social media as well. Particularly, we calculate an auction's price behaviour by using these dynamics, together with other auction-related information, to develop a hybrid dynamic functional forecasting model.

Classification and clustering is used to forecast the final bid. Predicting end price of an online auction has been considered as a machine learning problem and has been solved using regression trees, multi-class classification and multiple binary classifications [7]. Among these machine learning techniques, showing the price prediction as a series of binary classification has been verified to be the best suitable technique 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 [8, 9 and 10]. However, most of these price prediction methods are static that uses information which is available only in the beginning of the auction. These cannot incorporate price dynamics of auctions for similar items.

HYBRID DYNAMIC MODEL FOR BID FORECASTING

In clustering centred technique is used to project the final price of an online auction for an autonomous agent-based scheme. In the proposed technique, the input auctions are divided into groups of comparable auctions depending on their different features. This partitioning has been done by using k-means clustering algorithm. K-means algorithm is a clustering technique that splits the input feature space into *k* partitions, where each partition signifies a cluster. The key idea of k-means clustering is to describe *k* centroid, one for each cluster. Initially, *k* data are arbitrarily chosen to represent the centroid. Each data point is assigned to the cluster that has the closest centroid. The visualizations of clustering analysis are very useful in showing the estimation of results. After assignment of each data point, positions of the *k* centroid are re-calculated as the average value of every cluster. K-means algorithm returns two vectors as follows:

1. A vector of length *n* for demonstrating the assignment of each data point to one of the clusters.
2. A vector of length *k* for demonstrating the centroid.

The value of *k* in k-means algorithm is determined by using average silhouette approach [11] .Though the number of clusters can be cross verified by using elbow method as well. Based on the converted data after clustering and the characteristics of the current auction, bid selector calculates the starting end-price of the auction by using a regression model. The price forecasting agent is represented in Fig.5.1. This first phase of approach consists of four steps. The data is extracted from the bid server as per the requirements to form the agents' knowledge base for the online auctions.

Let *A* be the set of the attributes collected for each auction then $A= \{a_1, a2 \ldots\ldots a_j\}$ where *j* is the total number of attributes. Then based on the auctions' features, similar auctions are clustered together. Secondly, K-estimator agent determines the best number of partitions for the overall auction data and then the set of similar auctions are clustered together in k groups. Let *C* be the set of clusters then $C = \{c_1, c_2 \ldots . c_k\}$ where *k* is the total number of clusters. Thirdly, based on the transformed data after clustering and the characteristics of the current auction, bid selector nominates the bid evaluator for the end-price forecasting. In the end, the subsequent bid evaluator is invoked for the estimation of the forecasted bid price in first phase.

The bidding agent predicts the final price of the auction using three phases: phase 1 for the initial price estimation phase 2 is for studying the impact of local factors, i.e. sentiment analysis impact and phase 3 for the final price prediction. Phase 1 is responsible for the selection of an ongoing auction for participation and calculating the value of the item in the selected auction. This calculated value appears as the initial price during the second phase of the bidding agent. Phase 2 aids in assessing the sentimental impact on the trending price of the auction. Phase 3 utilizes this estimated price in order to start the predictive path of the final price of the auction based on different bidding strategies of the bidders.

The first phase of the HDAM agent estimates the initial price for an auction and is formally consists of three steps. In first step, comparable auctions are clustered together in *k* groups upon the basis of their price dynamics. Secondly, an auction for participation is selected by nominating a cluster for each ongoing auction using a bid mapping and selection technique. Third, the value of the item in the selected auction is assessed using machine learning techniques. Let ATTRB be the set of the attributes collected for each auction then ATTRB= $\{a_1, a_2 \ldots . a_j\}$ where j is the total number of attributes. Different types of auctions are categorized based on some predefined attributes from the vast feature space of online auctions. To classify different types of auctions, we focus on only a specific set of attributes; opening bid, closing price, number of bids, average bid amount, average bid rate, auction ending behaviour and the seller's reputation. Now ATTRB=$\{OpenB_i, CloseIP_i, NUM_i, AvgB_i, AvgBR_i, AuEND_i, S\_Repu_i\}$, where ATTRB be the set of attributes for an auction OpenBi be the opening bid of the ith auction CloseIPi be the closing price of the i^{th} auction NUMi be the total number of bids placed in the ith auction AvgBi be the average bid amount of the ith auction and can be calculated as $Avg(B_1, B_2, \ldots . B_i)$ where B1 is the 1st bid amount, B2 is the second bid amount, and B_i is the last bid amount for the ith auction. $AvgBR_i$ be the average bid rate of the i^{th} auction and can be calculated as where B_{i+1} is the amount of $(i+1)^{th}$ bid, Bi is the amount of ith bid, t_{i+1} is the time at which (i+1)th bid is placed, and t_i is the time at which ith bid is placed, AuENDi is auction behaviour related to whether auction ends or not and S_Repui tells about the seller trust in the market by using different factors.

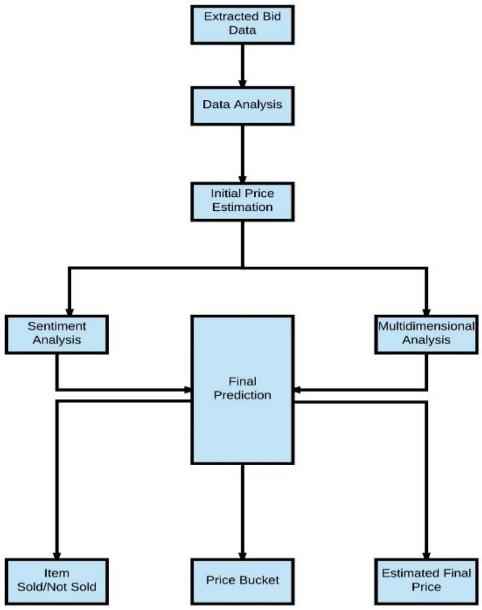

Figure 1. Proposed price forecasting agent for online auctions.

Deciding the value of k in k-means algorithm is a recurrent problem in clustering and is a distinct issue from the process of actually solving the clustering problem. The optimal choice of k is often ambiguous, increasing the value of k always reduces the error and increases the computation speed. The most favourable method to find k adopts a strategy which balances between maximum compression of the data using a single cluster, and maximum accuracy by allocating each data point to its own cluster. K-Means clustering algorithm is used to partition the similar auctions based upon their characteristics. Given a set A of N auctions A= $\{a_1, a2 \ldots a_n\}$ where each auction

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 5 (2017) pp. 598-604

is 7-dimensional real vector ATTRB= [$OpenB_i$, $CloseIP_i$, NUM_i, $AvgB_i$, $AvgBR_i$, $AuEND_i$, S_Repu_i], k means clustering aims to partition N auctions into k clusters (k<N) in such a way, which minimize the within-cluster dispersion. The within-cluster dispersion is the sum of squared Euclidean distances of auctions from their cluster centroid [12]. In order to decide that the current ongoing auctions belong to which cluster, the bid mapping and selection component is activated. Based on the converted data after clustering and the characteristics of the current auctions, it proposes the cluster for each of the ongoing auctions to select the auction for participation. A product for which a customer is going to bid has many aspects. These ranges from the quality-price justification curve, the nature of product, the bid timing, the warranty or service conditions, etc. And more than this is what opinion a customer is going to build about a specific product through a) The sentiments expressed in the social media and Sometime a customer is not given full freedom towards expressing these views so the review about a product has great significance. b) The opinion expressed through the customer reviews in the web world. The Fig. 2 below elaborates this model. The third phase of the HDAM agent predicts the ending price of the input auction by designing bidding strategies for the bidders. These bidding strategies are designed based upon the output of the first phase for the bidders possessing different bidding behaviour. The bidding behaviour which also depends on both the single or multiple bid placements, and the time of bid placements are used to determine the various types of bidders. The bidders who place a single bid are identified as Group1 and Group2 bidders. The bidders who place a single bid are identified as Group1 and Group2 bidders.

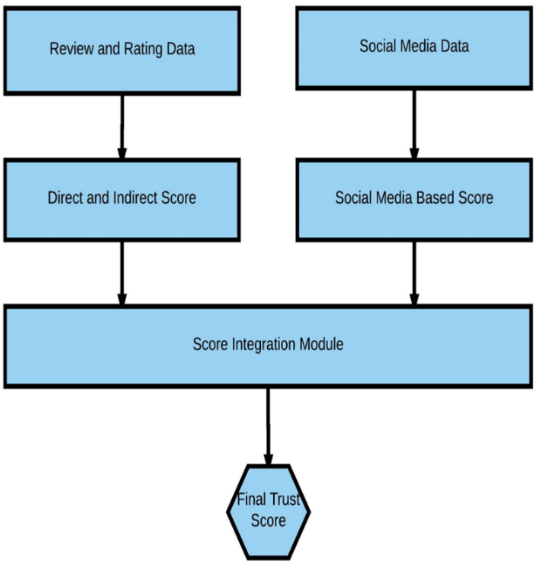

Figure 2. Trust value modelling.

The Group1 bidders place a single bid in the closing moments (last few seconds) of the auction, and the Group2 bidders place a single bid in the last five minutes of the auction. The bidders who place multiple bids in the last hour of the auction are identified as Group3 bidders and these bidders increase their bid amount strategically, based on the bids placed by the other participants. These bidders commonly exhibit two types of behaviour: firstly, they may be desperate to win the item, and secondly; they may be willing to bargain for that item. The bidders exhibiting these behaviours are designated as Determined and Sophisticated bidders respectively.

The bidding strategies are aimed for the fore mentioned bidding behaviour of buyers. The purpose of designing these bidding strategies is to determine how to compute a bid amount at a particular moment of time. This bid is the maximum value that an agent is willing to pay at that particular instant in time. An agent negotiates this value based on bidding characteristics such as auction's attributes, bidder's own attitude and other bidder's attitude.

Auction attributes have been considered while predicting the initial price of the auction during the first phase of the HDAM agent. Now we compute the bid amount at a particular instant in time using bidder’s own attitude and other bidder’s attitude. We study main two types of bidder’s attitude towards achieving the goal; desperate to get the auctioned item (Determined bidders), or he has a desire of a bargain to get the item (Sophisticated bidders). Determined bidder starts with higher bids to maximize its chances to win the auction, while, sophisticated bidder starts with lower bids and gradually increases his bids towards its maximum willingness to pay near to the final moments of the auction. Other bidders’ attitude is used to evaluate competition in the auction by exploiting their previous bids during the auction. In addition, remaining duration left until the auction closure is another important factor, which affects the competition in the auction. The bidders make quick decisions at the end of the auction due to limitations of time. When the Time pressure is more it arises from the bid outside their limits as the end point approaches [13]. Using K-means clustering technique different types of auctions based on some predefined features from the available feature space of online auctions have been considered. The feature space includes the average bid amount, no. of bids, average bid rate, seller reputation under multi dimensions, opening bid, closing bid, available quantity, type of auction, duration of the auction, buyer reputation, item type and many more.

EXPERIMENTAL SETUP:

1. Data Set Used

The data set used here has been collected through an automated software agent from the e-Bay website [14]. The player reference data has been referred through sportscollector.net. The sentiment data has been collected using twitter API and the text analysis of customer reviews [15]. Table showing different item categories having player autographs. There are two types of characteristics that have been used in the data set. One are the direct, and the second is the derived feature set of the data.

2. Results Visualisations:

First step in analysis is to recognize some patterns by simple visualize the data.

i) Price Density Visualization

International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 5 (2017) pp. 598-604

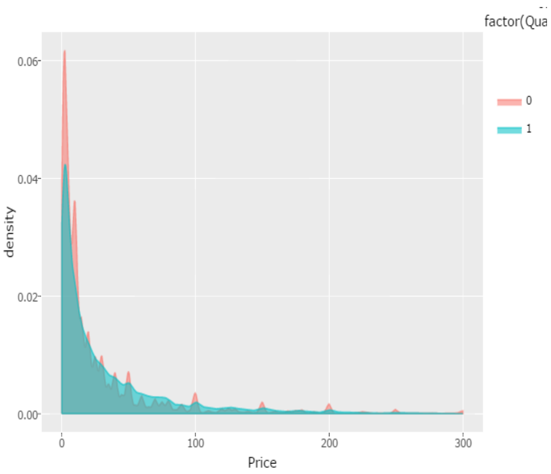

Figure 3. Price density visualization plot

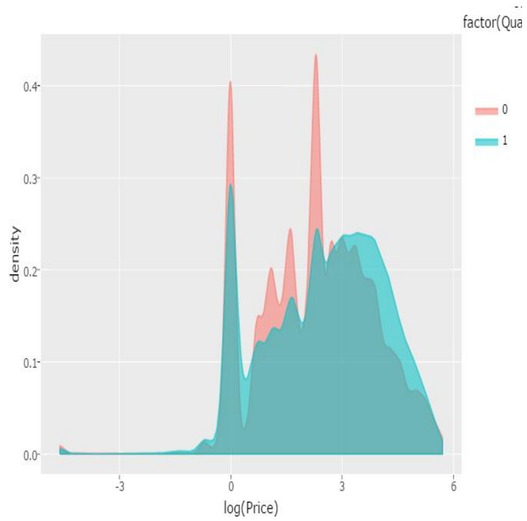

Figure 4. Visualization of log price density.

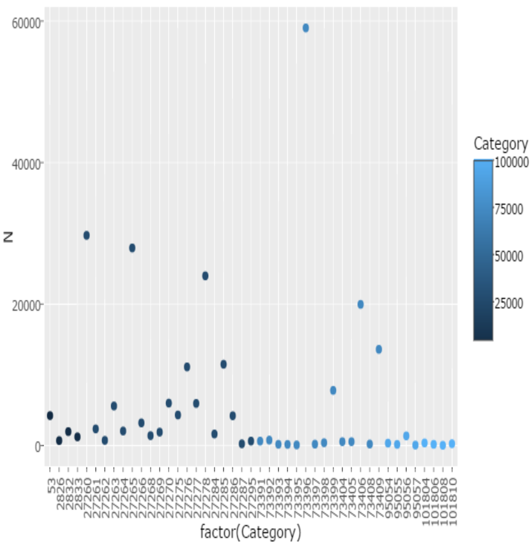

Figure 5. Geom point category frequency.

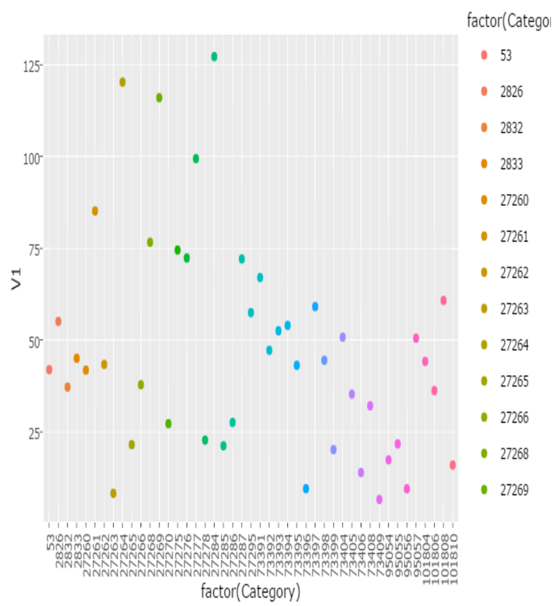

Figure 6. Mean price by category.

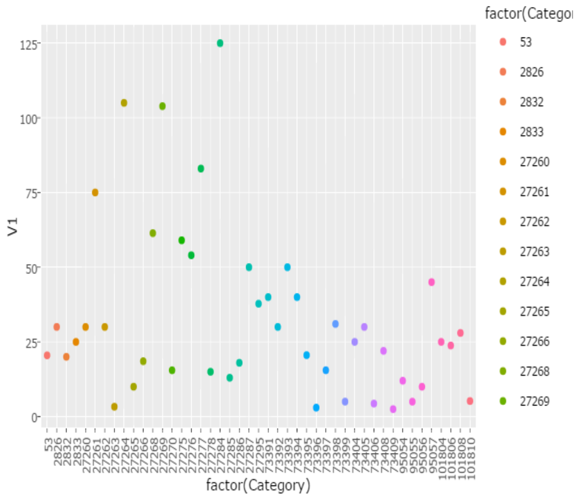

Figure 7. Median price by category.

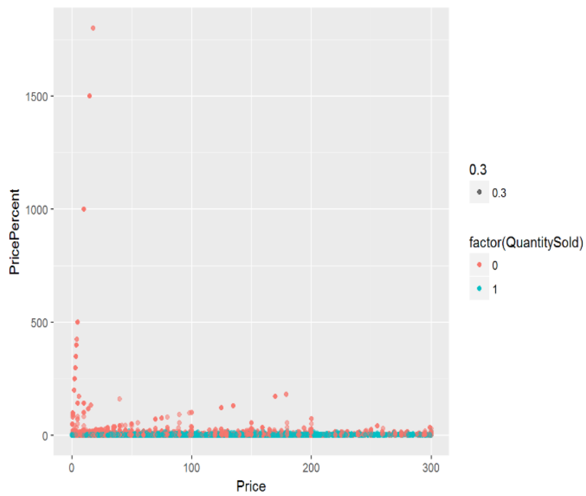

Figure 8. Price vs price percent.

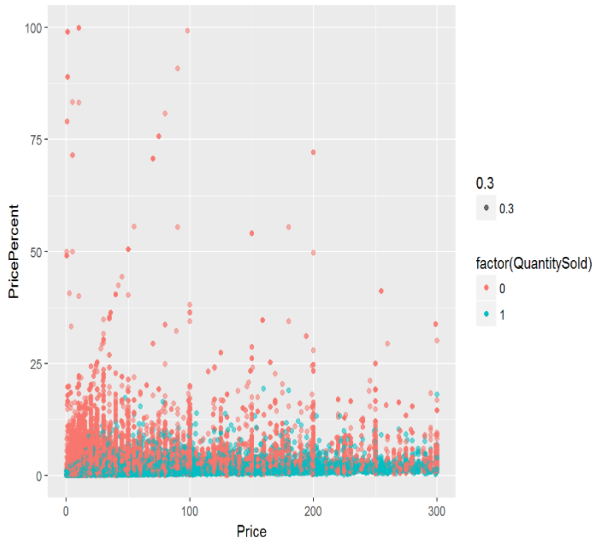

Figure 9. Price vs price percent limit 100.

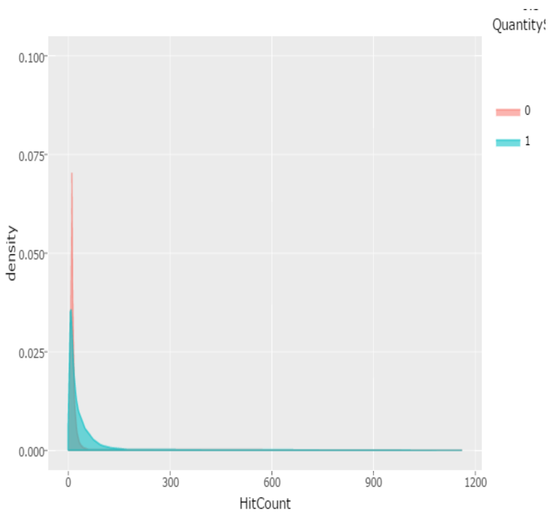

Figure 10. Hit Count Density.

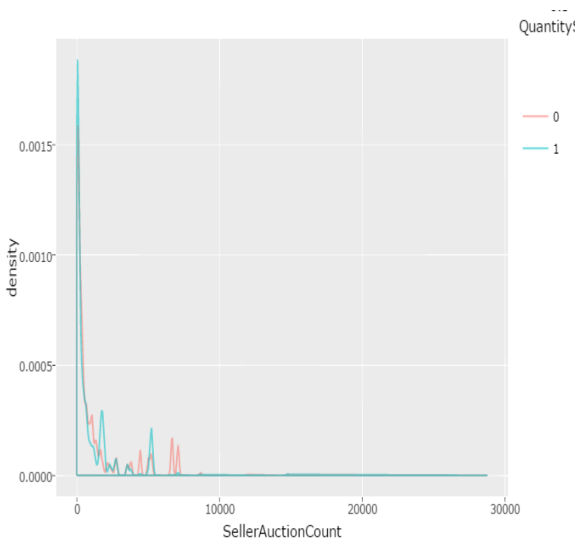

Figure 11. Seller auction density.

Logistic regression with care of multi co linearity, correlations, near zero variance variables: In this section, we created analysis for logistic regression and build the full logistic regression model with stepwise feature selection. For me, the most important part is to detect variables that don't have information value and remove variables, which are correlated at some level (We assume a correlation with a level higher than 0.7 as significant).Next we will Remove Variable which Standard Deviation is Zero [16].

Table 1. Variable with Zero Standard Deviation

Variable	freqRatio	percentUnique	zeroVar	nzv
QuantitySold	2.242279481	0.000773431	FALSE	FALSE
Price	1.452808579	2.820316488	FALSE	FALSE
PricePercent	14.6627907	12.22639875	FALSE	FALSE
StartingBidPercent	11.77209302	11.51406871	FALSE	FALSE
SellerClosePercent	1.253058596	0.929664176	FALSE	FALSE
Category	1.987476013	0.0174022	FALSE	FALSE
StartingBid	2.024820505	0.761056198	FALSE	FALSE
AvgPrice	5.267822736	7.520457252	FALSE	FALSE
EndDay	1.776075821	0.002707009	FALSE	FALSE
HitCount	1.062432196	0.143084753	FALSE	FALSE
...				
...				
SellerAuctionSaleCount	1.00762683762 573	0.15391278790972 5	false	false
AuctionMedianPrice	2.56598143236 074	1.87247668105249	false	false

Result: Remove Returns Accepted variable because it has only one level and has no informational impact on explanatory variable. Search for significant correlation. In my experience, .5 is the safe boundary.

Remove one variable from a pair of correlated Variables Function finds correlation to remove one with highest variance. Next we perform full logistic regression model [17].

Table 2. Hoslem-Lemeshow Goodness of Fit Test Results [18].

```
## Df Deviance AIC
## <none> 174822 174844
## - AuctionHitCountAvgRatio 1 174824 174844
## - AuctionSaleCount 1 174824 174844
## - ItemAuctionSellPercent 1 174890 174910
## - AuctionAvgHitCount 1 174904 174924
## - SellerItemAvg 1 174949 174969
## - AuctionMedianPrice 1 175176 175196
## - BestOffer 1 175859 175879
## - Price 1 179073 179093
## - SellerAuctionSaleCount 1 181025 181045
## - HitCount 1 230748 230768
## Call:
## glm(formula = QuantitySold ~ Price + HitCount + AuctionAvgHitCount +
## ItemAuctionSellPercent + SellerItemAvg + AuctionHitCountAvgRatio +
## BestOffer + AuctionSaleCount + SellerAuctionSaleCount +
AuctionMedianPrice,
## family = binomial(link = "logit"), data = train_caret_min)
## Deviance Residuals:
## Min 1Q Median 3Q Max
## -8.4904 -0.5324 -0.3654 0.0311 4.5398
## Coefficients:
## Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.684e+00 1.435e-02 -187.017 < 2e-16 ***
```

```
## Price                  -2.254e-02  3.838e-04  -58.725  < 2e-16 ***
## HitCount                2.146e-01  1.230e-03  174.394  < 2e-16 ***
## AuctionAvgHitCount      1.665e-03  1.782e-04    9.340  < 2e-16 ***
## ItemAuctionSellPercent -1.788e-01  2.221e-02   -8.049 8.33e-16 ***
## SellerItemAvg          -3.408e-03  2.993e-04  -11.385  < 2e-16 ***
## AuctionHitCountAvgRatio 6.851e-04  3.385e-04    2.024   0.0429 *
## BestOffer               6.660e-01  2.077e-02   32.061  < 2e-16 ***
## AuctionSaleCount       -7.439e-05  6.843e-05   -1.087   0.2770
## SellerAuctionSaleCount  7.863e-04  1.006e-05   78.179  < 2e-16 ***
## AuctionMedianPrice     -8.698e-03  4.598e-04  -18.917  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##     Null deviance: 319529  on 258587  degrees of freedom
## Residual deviance: 174822  on 258577  degrees of freedom
## AIC: 174844
## Number of Fisher Scoring iterations: 8
library(ResourceSelection)
## ResourceSelection 0.3-0   2016-11-04
hoslem.test(step_reglog_min_$y, fitted(step_reglog_min_), g = 6)
##  Hosmer and Lemeshow goodness of fit (GOF) test
## data:  step_reglog_min_$y, fitted(step_reglog_min_)
## X-squared = 4036.9, df = 4, p-value < 2.2e-16
```

Result: At the alpha level 0.05, we clearly reject null hypothesis, which said that model fit to data statistically significantly good.

Table 3. Variance Inflation Factor for Stepwise Regression Model [19].

```
library(car)
vif(step_reglog_min_)
##  Price              HitCount        AuctionAvgHitCount
##           2.369833           1.743201           1.406591
##  ItemAuctionSellPercent     SellerItemAvg AuctionHitCountAvgRatio
##           1.075931           2.522379           2.763537
##           BestOffer     AuctionSaleCount  SellerAuctionSaleCount
##           1.380651           1.267001           1.045371
##     AuctionMedianPrice
##           2.461350
```

Variance Inflation Factor values greater than 2.5 can indicate multi co-linearity. How to deal with multi co-linearity is to remove variables, which are correlated at some level. In this case, we first assumed correlation border at 0.7, but Variance Inflation Factor took values greater than 2.5, so we set correlation level on 0.5, which is safe and acceptable.In the final model, we have to know how well this model separate classes. So we propose density plot based on predicted values from a model. This method can clearly visualize where to set cutoff point and if our model separate classes very well. Now the density plot of two classes for predicted values are taken up for further analysis.

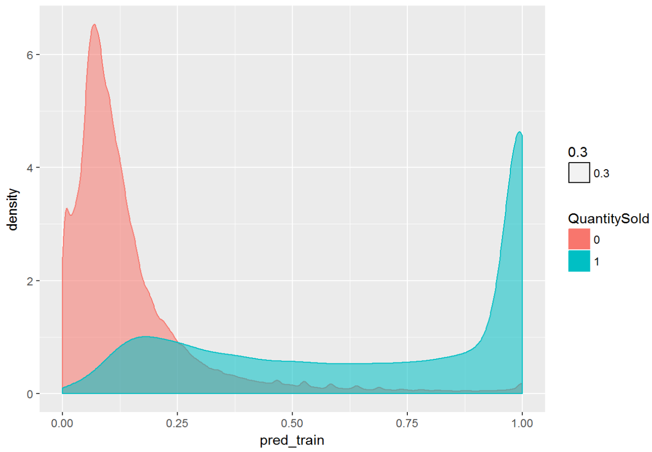

Figure 12. Density plot of training dataset.

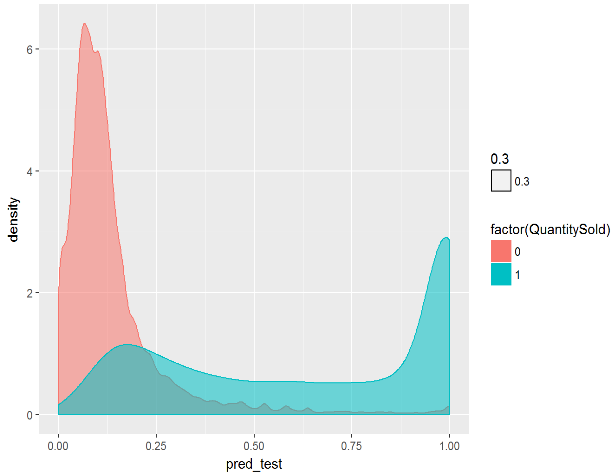

Figure 13. Density plot of test dataset.

Now we shall show the Predicted binary variables and confusion matrix [20].

Table 4. Confusion Matrix for Prediction

```
confusionMatrix(pred_bin,train_caret_min[, QuantitySold])
## Confusion Matrix and Statistics
## Reference
## Prediction      0      1
## 0 171043  27679
## 1   7790  52076
## Accuracy : 0.8628
## 95% CI : (0.8615, 0.8642)
## No Information Rate : 0.6916
## P-Value [Acc > NIR] : < 2.2e-16
## Kappa : 0.6546
## Mcnemar's Test P-Value : < 2.2e-16
## Sensitivity : 0.9564
## Specificity : 0.6529
##Pos Pred Value : 0.8607
##Neg Pred Value : 0.8699
## Prevalence : 0.6916
## Detection Rate : 0.6614
## Detection Prevalence : 0.7685
## Balanced Accuracy : 0.8047
## Positive' Class : 0
```

Accuracy at level 0.86, which is a very good result, but specificity is 0.65, and this is the result that machine learning methods have to improve. We can observe that specificity of test data set is less powerful. It can indicate little over fitting to a model. Logistic regression model can be some kind of "base," which indicate if ML is relatively better than Logistic regression Models.

CONCLUSION

This paper describes a Hybrid Price Prediction Model to know about predictability of online auctions. To establish the R implementations have been used. The historical e-Bay data set to have been utilized to establish the experiments. Under this different machine learning, models have been worked upon. The summaries for results obtained here are shown using visualization plots for different factors. These have been studied for observing their behaviour and prediction patterns.

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] N. Chan and W. Liu, "Modeling and Forecasting Online Auction Prices: A Semiparametric Regression Analysis," J. Forecast., 2016.

[3] R. Ghani and H. Simmons, "Predicting the end-price of online auctions," Proc. Int., 2004.

[4] P. Kaur, M. Goyal, and J. Lu, "A proficient and dynamic bidding agent for online auctions," Int. Work. Agents Data, 2012.

[5] S. Wang, W. Jank, and G. Shmueli, "Explaining and forecasting online auction prices and their dynamics using functional data analysis," J. Bus., 2008.

[6] N. Chan, Z. Li, and C. Yau, "Forecasting Online Auctions via Self-Exciting Point Processes," J. Forecast., 2014.

[7] P. Kaur, M. Goyal, and J. Lu, "Pricing analysis in online auctions using clustering and regression tree approach," Int. Work. Agents Data, 2011.

[8] S. Petruseva, P. Sherrod, V. Pancovska, and A. Petrovski, "Predicting Bidding Price in Construction using Support Vector Machine," TEM J., 2016.

[9] T. Pinto, T. Sousa, I. Praça, Z. Vale, and H. Morais, "Support Vector Machines for decision support in electricity markets′ strategic bidding," Neurocomputing, 2016.

[10] V. Almendra and D. Enachescu, "Using self-organizing maps for fraud prediction at online auction sites," Symb. Numer., 2013.

[11] S. Bollmann, A. Hölzl, M. Heene, and H. Küchenhoff, "Evaluation of a new k-means approach for exploratory clustering of items," 2015.

[12] P. Kaur, M. Goyal, and J. Lu, "A price prediction model for online auctions using fuzzy reasoning techniques," Fuzzy Syst. (FUZZ-IEEE), 2014, 2014.

[13] C. Chang and C. Chen, "Examining hedonic and utilitarian bidding motivations in online auctions: Impacts of time pressure and competition," Int. J. Electron., 2015.

[14] Grossman Jay, "Jay Grossman." [Online]. Available: http://jaygrossman.com.

[15] S. Kumar and R. Rishi, "Data collection and analytics strategies of social networking websites," Green Comput. Internet Things , 2015.

[16] Thiagogm, "Near-zero variance predictors. Should we remove them? | R-bloggers," 2014. [Online]. Available: https://www.r-bloggers.com/near-zero-variance-predictors-should-we-remove-them/.

[17] D. Bates, M. Maechler, B. Bolker, and S. Walker, "lme4: Linear mixed-effects models using Eigen and S4," R Packag. version, 2014.

[18] J. Canary, L. Blizzard, and R. Barry, "A comparison of the Hosmer–Lemeshow, Pigeon–Heyse, and Tsiatis goodness-of-fit tests for binary logistic regression under two grouping methods," Commun., 2016.

[19] D. Montgomery, E. Peck, and G. Vining, Introduction to linear regression analysis. 2015.

[20] S. Kalmegh, "Analysis of WEKA data mining algorithm REPTree, Simple CART and RandomTree for classification of Indian news," J. Innov. Sci. Eng., 2015.