The Effects of Product Characteristics and Auction Format on Auction Revenue *

Xiaotong Fu, Peter Yang

Abstract

In this paper, we investigate how to explain and predict the winning bid as well as the probability of whether the car is sold or not by using key auction and product characteristics. The paper uses the dataset (Lewis, 2011) that contains information on eBay used car auctions and machine learning methods such as basic linear model, Support Vector Machines, and Random Forest. The results show that the auction settings, the information disclosed by sellers, and the product characteristics all have significant effects on the price of the winning bid, and the probability of the vehicle being sold. The recommendations for sellers about auction settings and information disclosed for the most popular car model will also be addressed in the paper.

Keywords— Auction, Winning Bid, Revenue, Support Vector Machines, Random Forest

1 Introduction

Over the last several decades, people are increasingly inclined to put their used cars up for auction on eBay. According to Lin of Oberlo, in the year 2020, the number of things listed on eBay topped one billion with total worth of goods bought and sold on eBay's marketplace platform reaching 22 billion. These online auctions on eBay are structured as English auctions. Sellers are able to modify the auction format such as creating a "Buy it Now" option, adjust the start bid, and amend the length of the auction. Auctioneers can also give additional information of the cars such as text illustration and photos. Finally, sellers can set a reserve price. If this reserve price isn't met, the vehicle won't be sold. How will these format logistics affect the auction? There are previous literature that have used econometric methods to investigate which key variables affect auction prices (Lewis, 2011) and which variables affect the probability of a car be sold in the online auction (Andrews and Benzing, 2007). We were curious if we could use machine learning methods to contribute to this growing body of research.

Our project will use advanced machine learning methods such as support vector machines and random forest, in comparison with linear regression models, to look at eBay motor sales and predict the winning bid as well as provide listing recommendations for future car dealers. The data set we use is carefully processed to eliminate any large outliers and null values. It contains key variables that would affect bid price such as length of the auction, car model type, number of photos, etc. The results indicate that the auction settings such as start bid, number of bidders and whether there is reserve price will significantly affect the winning bid and the probability of the car be sold. Also, product characteristics such as the number of photos offered by the sellers

*A Machine Leaning paper for the course Econ 725 at University of Wisconsin-Madison. We thank Prof. Lorenzo Magnolfi and Prof. Christopher Sullivan for their helpful suggestions through this project.

also influence the winning bid. We also give recommendations for the sellers who sell the most popular car model, Ford Mustang, which can help them obtain higher bids and sell them out easier. The suggestions will contains how to set the auction format.

This paper is organized as follows. After a basic introduction to auction and analysis method background, section 2 describes the basic model designs for different analytical purposes. Section 2 will also contain description of the data and corresponding variables. Section 3 analyzes the impact of auction format and product characteristics on the price of the winning bid and probability of sell. Section 4 investigates several key factors that are vital for the auction and proposed effective strategies to gain maximum return for prospective sellers of the Ford Mustang. We will conclude our paper in section 5 with an overall summarization of the paper.

2 Model Design

2.1 Data And Summary Statistics

Our data comes from Professor Gregory Lewis of Harvard University, who obtained the data by downloading the auction web-pages for selected car models over an 8-month period (March-October 2006), and then implementing a pattern matching algorithm to pull variables of interest from the web-page html code (Lewis, 2011). This initial data set contains over 140000 auctions each with 500 different characteristics. We then modified this data set in several ways. The original data set had variables describing whether key phrases such as ding, dent, crack, and scratch were present in the owner description of the car and modifiers for how they were used. From these variables, we created new dummy variables that would be key part of our covariates in later models. Furthermore, we had number of car characteristics that were deemed mandatory by eBay Motors for every car auction. This included things like age, model, type of transmissions, and number of doors. These key characteristics would also be vital covariates for the models. Finally, we filtered through our observations to eliminate some auctions. This includes observations where the car on auction has negative number of miles or has one/five doors. We also eliminated cases where car model descriptors didn't match with the car itself. For instance, Ford F-100's only have four doors, thus we eliminated all auctions that advertised a non-four door Ford F-100. Our final data set contained just over 60,000 observations with 33 key characteristics. Essential information to note is that the average highest bid on a car is about ten thousands dollars. On average, the cars are not old (about 7.5 years on average) and well traveled (about 80,000 miles on average). Also, the general auctioneer photos provides about 16 photos, and supply lengthy text descriptions (about 5000 characters).Table 1 in the appendix will summarize the other key variables of this data set.

The full sample consists of 60,763 observations of 18 models of cars. We divided the data into two groups, the "Sell=1" group with 16,580 observations, and the "Sell=0" group with 44,183 observation. According to Gregory Lewis, all of the information is standardized and mandatory which means the seller must provide the vehicle when the seller lists the car on eBay (Lewis, 2011).

2.2 Variable Selection

Dependent Variable - Winning bid: In this project, our models contained two main dependent variables. Our main dependent variable was named "Winning bid". Winning bid is a continuous variable that represents the highest bid received during the auction. As shown in table A1 in the Appendix, the average car had their highest bid be around 10,000 dollars. We used this as our dependent variables in our basic linear, random forest, and SVM models. The purpose of these models will give us answers to questions such as does winning bid depend on time of year or does winning bid depend on auction characteristics such as buy it now, text, photos. We will also use a random forest model to provide a recommendation for auction logistics for any seller of a used mustang.

Dependent Variable - Sell: The second main dependent variable was named "sell". Sell is a binary variable and if equals 1 it means the car was sold to the highest bidder. As shown in Table A1, 27.3 percent of cars auctioned ended up being sold. We can find that the average of the highest bid in the "sell=1" group is lower than the average of the highest bid in "sell=0" group. We used the sell variable as our dependent variable in random forest, and SVM models. Using sell in these models will determine what pivotal product and auction attributes help the car get sold. As Andrews and Benzing (2007) did in their paper, they also investigate what kind of characteristics can affect the binary sell variable.

Independent Variables - Key Car Characteristics: This project's independent variables are also split into two main groups. One part is the key car characteristics which includes variables such as model, doors, trans, age, warranty, inspection, miles. These covariates are mandatory and required by ebayMotors which means the seller must provide the information when listing a car. These variables were chosen to be part of the models for two main reasons. First, these variables were strongly suggested to be important by the project description. Logically, a car's age and miles will affect the final car price. Old, more used cars tend to be cheaper. Second, previous literature found that seller reputation, miles, age, presence of a warranty were key factors in the price premium of the vehicle (Andrews and Benzing, 2007)

Independent Variable - Auction Logistics: In addition, we collected information about the auction, such as the start bid, the highest bid, the end date, whether the car was sold, how long the auction last, whether the auction ended on Sunday, and the number of photos taken. We also parsed the item description for key phrases and then modifiers for how they are used. For example "no scratches" or "scratch-free" is seen as the negation of the phrase "scratch". We will use these to create dummy variables of terms ding, scratch, crack, broken, dent, problem, and rust.

2.3 Auction And Data Description

The auctions we observed in the data are second price English auctions. For each auction, the price starts from an auctioneer determined starting bid and increased as each additional bidder provides a higher bid. At the end of the auction, the person with the highest bid would pay the highest bid price if everything checks out. The bidder with the highest valuation wins the auction and pays an amount slightly higher than the value of the second highest bidder, so this is a type of second-price auction. Some of the auctions have a secret reserve price. These reserve price auctions which are the specific kind of the English auction only result in vehicle sales when the winning bid exceeds the reserve price.

As we mentioned in the last section, in Table A1 in the Appendix, the number of bidders in each auction is around 5 on average, and the "Sell=1" group has more number of bidders in each auction on average (about 7 on average) than number of bidders in the "Sell=0" group (about 4 on average). In addition, the average of the highest bid is about $10,005, and the average of the highest bid is about $7,977 in the "Sell=1" group, which is lower than the average of the highest bid in the "Sell=0" group (about $10,766 on average). On average, the age is about 7.5 years and the miles is about 80,438 which means cars are not old and well traveled. With no surprise, both the average value of age (nearly 10 years) and the average value of miles (nearly 110,000 miles) in the "Sell=1" group is larger than the average value of age (nearly 7 years) and sample mean of miles (nearly 70,000 miles) in the "Sell=0" group, and the higher values of both variables may lead to the lower average value of the highest bid level in "Sell=1" group. And for photos and text, the average number of photos is 16 and the average number of text characters is 4932. Also, the length of the auction is about 7 days on average. Moreover, nearly half of sellers are dealers (nearly 0.5 on average), and the "Sell=0" group has more sellers who are dealers. Finally, the summary table suggests most of the auctioned cars do not have the warranty and inspection (only about 0.27 and 0.28 on average respectively).

We also observed two interesting facts in Table A1. The first thing is that compared with the value of buy it now in the "Sell=0" group (0.465 in Table 1), we can find that the "Sell=1" group has very small buy it now which is nearly 0 (0.0001 in Table 1), indicating that almost none of the auction has the buy it now option in the "Sell=1 group", whereas nearly half of the auctions in the "Sell=0" group have the buy it now option. The other interesting thing is that the value of the reserve in the "Sell=1" group is zero, but the value of the reserve in the "Sell=0" group is about 0.89, and the value is nearly equals one, which indicates that almost all auctions in the "Sell=0" group has the reserve price. Hence, buy it now and reserve price may be two most important variables that affect whether the car is sold.

2.4 Linear Regression Model

Basic linear models are one of the simplest yet universal models in economics. It captures the basic relations between dependent and independent variables. In this project, we use the OLS linear model as an exploratory tool that will help us identify correlations between our independent and dependent variables. Raviv (2006) uses the linear model to investigate how the auction factors affect the highest bid in his research. Wan and Teo (2001) evaluate the effect of duration of auction and seller's feedback rating on the auction price. The basic form of such models used in this research can be expressed below:

$$y_i = x_i'\beta + \epsilon_i \tag{1}$$

where
y_i is dependent variables;
x_i is a $k \times 1$ vector of independent variables,
β is a $k \times 1$ vector of coefficients,
ϵ_i is the error term.

An ordinary least squares (OLS) method can be applied to obtain the estimators for $\boldsymbol{\beta}$. The OLS result would be unbiased and consistent under the assumptions of (1) $E(x_i\epsilon_i) = 0$. (2) (y_i, x_i) i.i.d. (3) $E(x_i x_i')$ is finite and non-singular. (4) $E(\epsilon_i^2|x_i) = \sigma^2$

2.5 Support Vector Machines

Support Vector Machines (SVM) training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. We decided to use SVM's because of ease of interpretation compared to a Neural Network. Furthermore, SVM is the only linear model that can classify binary data without the data being linearly separable thus making it a perfect solution to the problem at hand.

In order to understand SVM's, we must first define a separating hyperplane as:

$$\begin{aligned} &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p < 0, &&\text{if } y_i = -1 \\ &\beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p < 0, &&\text{if } y_i = 1 \end{aligned} \tag{2}$$

With this setup, the marginal maximal hyperplane solves the following constrained optimization problem:

$$\max_{\beta, M} M \tag{3}$$

$$\text{s.t.} \quad \sum_{j=1}^{p} \beta_j^2 = 1 \tag{4}$$

$$y_i(\beta_0 + \beta_1 x_{i1} + ... + \beta_p x_{ip} \geqslant M, \qquad \forall i = 1, ..., n \tag{5}$$

2.6 Random Forest

The random forest algorithm provides a way of creating less correlated trees, and then average over them. Compared with neural network, random forest possesses the advantages of little concern for over-fitting as well as relatively quick computation. Furthermore, Random Forest are much easier to interpret and understand exactly what the model has 'learned'. The basic procedure can be expressed in the form below:

1. For $b = 1$ to B:
   (A). Draw a bootstrap sample $\mathbf{Z}$ of size N from the training data.
   (B). Grow a random-forest tree T_b to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size $n_m in$ is reached.
      a. Select m variables at random from the p variables.
      b. Pick the best variables/split-point among the m.
      c. Split the node into two daughter nodes.
2. Output the ensemble of trees $\{T_b\}_1^B$.

To make a prediction at a new point x:

$Regression : \hat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^{B} T_b(x)$

$Classification$: Let $\hat{C}_b(x)$ be the class prediction of the bth random-forest tree. Then $\hat{C}_{rf}^B(x) = majority\ vote\{\hat{C}_b(x)\}_1^B$.

3 Revenue and Product/Auction Characteristics

In this section, we will investigate the relationships between Winning Bid and Product/Auction Characteristics. We will also explore how characteristics affect the sell probability. We trained our model on a sample of 50763 observations and tested it on a sample of 10000 observations. The linear model results show the relationships between Winning bid and Characteristics. The Random Forest shows the results of two dependent variables: Winning bid and binary sell. The SVM will only be used with the binary sell variable.

3.1 Linear Regression Model Results

The linear model will be used here to investigate how these auction and product characteristics affect the winning bid. In Table 1, we can see the number of bidders has a significant positive effect on the winning bid. That is, the more people get involved in the bidding, the higher the winning bid tends to be. The start month variable has a significant negative effect on profit with respect to the time of auction start. This result suggests that the later months of the year are more likely to generate higher winning bid. Afterwards, the coefficient of the variable "Buy it Now" is significantly negative. Therefore, the auctions with the option "But it Now" tend to have a lower winning bid. The auction duration has a significant positive effect on the winning bid. So, this suggests that the longer the auction lasts, the more likely it is to produce a higher bid.

Table 1: Linear Regression Model Results

DEPENDENT VARIABLE = Winning Bid					
	Estimate	Std. Error	t value	Pr(>|t|)	
Miles	-0.017	0.001	-31.756	0.000	***
Buy it Now	-1649.349	67.303	-24.506	0.000	***
# of Photos	42.951	3.066	14.008	0.000	***
Age	325.350	10.687	30.445	0.000	***
# of Bidders	386.428	8.073	47.865	0.000	***
Professional Platform	-715.446	109.989	-6.505	0.000	***
Featured	1199.822	113.377	10.583	0.000	***
Book Value	0.946	0.006	162.665	0.000	***
Start Bid	-0.105	0.004	-24.614	0.000	***
Warranty	783.298	81.407	9.622	0.000	***
Inspection	-171.504	63.854	-2.686	0.007	**
Reserve	1814.995	68.630	26.446	0.000	***
Added Information	460.975	71.057	6.487	0.000	***
% Negative Feedback	-10.356	4.712	-2.198	0.028	*
Auction Duration	0.025	0.009	2.657	0.008	**
Dealer	-338.487	71.248	-4.751	0.000	***
Ding Group	5.766	30.898	0.187	0.852	
Scratch Group	-102.844	26.614	-3.864	0.000	***
Crack Group	-126.520	41.583	-3.043	0.002	**
Broken Group	-131.191	62.959	-2.084	0.037	*
Dent Group	-115.937	32.564	-3.560	0.000	***
Problem Group	-59.382	35.830	-1.657	0.097	.
Rust Group	-168.742	39.072	-4.319	0.000	***
Auction Duration	141.464	17.964	7.875	0.000	***
End Sunday	68.385	100.867	0.678	0.498	***
End Day	2.664	18.088	0.147	0.883	
Start Month	-145.581	13.061	-11.146	0.000	***
Car Model ID	34.096	2.131	16.002	0.000	***
(Intercept)	-5634.055	231.979	-24.287	0.000	
Adjusted R-squared:	0.3801				
MSE:	42791027				
Significance:	0 ***	0.001 **	0.01 *	0.05 .	

The coefficient of the photo variable is 42.9 and it is statistically significantly positive. If the seller uploads more photos on the website, then the auction will tend to have a higher winning bid. Furthermore, according to this model, the text length has a significant positive effect on the winning bid, and that means describing this car in more words will significantly increase the highest bid. Moreover, the results indicate that if the seller put an added information in the website, the auction will have higher winning bid. Other vehicle properties such as warranty and book-value also had a significant positive effect on the highest bid. If a car has a higher book value, bidders will tend to bid a higher price. A car with warranty is also a pull factor of the highest bid. Finally, the cars with higher miles are more likely to have a lower winning bid in the auction.

We also investigated the dummy variables like ding, problem, crack, dent, scratch, and rust as group variables which are contained in the text of the auction. It's understandable that more defects in cars like crack, dent, scratch, and rust have a significant negative effect on the winning bid of the car. On the other hand, ding and problem being reflected in the text description did not have significant negative coefficients in the result.

3.2 Random Forest Results

We use the Random Forest model to investigate the importance of each characteristics on the winning bid and dummy variable Sell/not. The MSE (7969841) of the Random Forest model is lower than the MSE (42791027) of the OLS model when analyzing the winning bid. Note that the large size of MSE is due to the large sample size and the magnitudes of winning bid. The smaller MSE suggests that the Random Forest model predicts the winning bid with better accuracy. As the text book Introduction of Statistical Learning, we can obtain an overall summary of the importance of each predictor (James et al., 2020). A large value indicates an important predictor.

In the Table 2 column 1, the results show the importance of each variable on the winning bid. The book value variable has the highest importance coefficient. This is reasonable because the value of the car determined the bid price of it. And the number of the bidders has the second importance coefficient. The third one is miles and the forth one is age, which represent the product characteristics. The variable with the fifth highest importance coefficient is the start bid, this suggests that the seller should choose this value seriously and it does influence the highest bid. In the results, the crack and broken have the relative low importance coefficients which indicates that variables play a minor role in determining the winning bid.

Table 2 column 2 illustrates the results of importance coefficients on the binary variable Sell. The variable with highest importance coefficient is reserve price. This indicates that the reserve price of seller has a vital impact on whether the car will sell or not. If the highest bid does not reach the reserve price of the seller, then this car may not be sold. The variable with second highest importance is the number of bidders. Auction settings by the seller such as "Buy it Now" and Start Bid are the third and fourth highest importance coefficients, respectively. The car characteristics like dent, crack, and broken has a lower importance coefficients in the Random Forest with binary sell as the dependent variable.

3.3 Support Vector Machine Results

In the Table 2 column 3, the results show the weights of each variable in the SVM model for the Sell or not variable. We can see the MSE of the SVM model on the binary sell variable is 0.0065 and it has a good prediction accuracy. The table shows the coefficients of the normal vector of a linear SVM as attribute weights, whose linear combination predicts the value of dependent variable.

The variable with highest weight is the reserve price. An auction with a reserve price is highly unlikely to sell. The second highest weight factor is the number of bidders, this variable also can influence whether the auction have a deal. Other higher weights are whether to set "Buy it Now" and how to set the Start Bid. The order of the first four weights is the same as the order of the first four importance coefficients we showed in the Random Forest Results. This means that in both Random Forest and SVM, these four variables (reserve price, number of bidder, Buy it Now, and Start Bid) all play relatively large roles in whether sell or not in an auction. Moreover, text variables such as rust, broken, and problem of cars are also have relative high weights. Text length and photos do not have relatively high weights.

Table 2: Random Forest And SVM Results

Winning Bid—Random Forest			Sell—Random Forest			Sell—SVM	
Variable	Scaled Importance	Percentage	Variable	Scaled Importance	Percentage	Variable	Weights
Book Value	1.000	0.360	Reserve	1.000	0.623	Reserve	444.824
# of Bidders	0.582	0.209	# of Bidders	0.245	0.153	# of Bidders	196.357
Miles	0.377	0.136	Buy it Now	0.172	0.107	Buy it Now	124.077
Age	0.209	0.075	Start Bid	0.055	0.034	Start Bid	109.571
Start Bid	0.170	0.061	Book Value	0.053	0.033	Rust Group	81.543
Car Model ID	0.129	0.046	Age	0.031	0.019	Broken Group	77.718
Warranty	0.099	0.035	Miles	0.023	0.014	Problem Group	63.024
Photos	0.029	0.010	Warranty	0.007	0.004	% Negative Feedback	52.230
Reserve	0.028	0.010	Car Model ID	0.004	0.003	Age	51.134
Text	0.025	0.009	Text	0.003	0.002	Dent Group	42.637
Start Month	0.024	0.008	photos	0.003	0.002	Text Length	41.499
End Day	0.015	0.005	Professional Platform	0.002	0.001	Crack Group	38.461
Buy it Now	0.015	0.005	Scratch Group	0.001	0.001	Miles	36.332
% Negative Feedback	0.012	0.004	Start Month	0.001	0.001	Featured	31.476
Auction Duration	0.010	0.004	% Negative Feedback	0.001	0.000	Ding Group	31.402
Scratch Group	0.006	0.002	Dealer	0.001	0.000	Added Information	27.160
Dealer	0.006	0.002	Added Information	0.001	0.000	Dealer	24.205
Rust Group	0.005	0.002	End Day	0.001	0.000	Scratch Group	20.837
Ding Group	0.005	0.002	Ding Group	0.000	0.000	# of Photos	17.604
Added Information	0.005	0.002	Auction Duration	0.000	0.000	Book Value	14.755
Problem Group	0.005	0.002	Problem Group	0.000	0.000	End Sunday	13.641
Featured	0.005	0.002	Rust Group	0.000	0.000	Inspection	12.920
Inspection	0.005	0.002	Dent group	0.000	0.000	Car Model ID	12.736
Professional Platform	0.004	0.002	Crack group	0.000	0.000	Professional Platform	7.435
Dent Group	0.004	0.001	Featured	0.000	0.000	Warranty	7.300
End Sunday	0.003	0.001	Inspection	0.000	0.000	End Day	6.623
Crack Group	0.002	0.001	Broken Group	0.000	0.000	Start Month	3.623
Broken Group	0.001	0.000	End Sunday	0.000	0.000	Auction Duration	2.399
MSE	7969841		MSE	8.277267e-05		MSE	0.0065

4 Recommendations for Auctioneers

The most popular car during the period observed was the 2 doors, automatic, Ford Mustang. In total, of the nearly 100,000 vehicles posted 11915 was this type of model. In this section we will provide some recommendations for the sellers of this car model on how to structure the vehicle auction in order to get the highest bid.

4.1 New Variable Description

In this section, we are mostly concerned with auction structure, variables that deal with product attributes are not included in this model. This includes covariates like warranty and inspection. Furthermore, in order to facilitate the computational ability of our random forest we restructured some variables into categorical covariates. We changed start month into Start Quarter splitting the 12 months into 4 seasons. Start Quarter =1 means that the auction started between January and March, Start Quarter = 2 means April - June, Start Quarter = 3 means July-September, while Start Quarter = 4 means October-December. We also categorized text creating new variable Text-level. Text-level splits text in to 5 different categories each sized 1500 characters. For example, if Text-level = 1 then the amount of characters in the automobile text description is between 0 and 1500. We also changed start bid into a category variable Start Bid-level. Start Bid-level is ranged 1 to 10 where Start Bid-level = 1 if the starting bid is between 0 and 100 dollars. Start Bid-level = 2 if bid between 100 and 500, Start Bid-level = 3 if starting bid between 500 and 1000. Start Bid-level = 4 if starting bid between 1000 and 2000, Start Bid-level

= 5 if between 2000 and 3000, Start Bid-level = 6 if bid between 3000 and 5000. Finally, Start Bid-level = 7 if the starting bid was between 5000 and 10000 dollars while anything greater than 10000 dollars would mean that Start Bid-level = 8. Finally, we re-classified photos in to 10 levels creating the new variable Photos-level. EBay Motors allows at most 50 photos thus we evenly divided chunks of 5 photos into a class. Photos-level = 1 when the number of photos is between 0 and 5. Photos-level = 10 when the number of photos taken by the auctioneer is between 45 and 50.

4.2 Predicting Results

We will first use the random forest model in order to investigate which auction characteristics would lead to the biggest bid for Ford Mustangs. We trained our model on a sample of 9914 observations and tested it on a sample of 2000 observations. Table 4 shows the results of our random forest model and its top 3 recommendations for auction structure in order to gather the highest bid and maximize revenue. We define revenue as the highest bid conditioned on the car actually getting sold in the auction. In order to get a higher bid, this model recommends the auctioneer to have a buy it now price option, set a secret reserve price, have a start bid price of between 0 and 100 dollars, have the auction last 10 days, and have between 15 and 25 photos. The model does not recommend the auctioneer to use a professional listing platform or to upgrade your eBay package in order to get the auction featured. The model predicts that following these recommendations could net the auctioneer at most an 84,436 dollar bid. In order to get maximum revenue, the recommendation is to respond to prospective customers, start your auction between the months of January and March, describe your product with between 3000 - 4500 characters, have the auction last 10 days, and have a start bid between 5000 and 10000 dollars. It also does not recommend for the seller to hire a professional platform or to have the car featured. However, unlike the recommendation for the highest bid, the advice is to not set a reserve price and not have a buy it now option. Following these suggestions would result in car being sold for model forecast 69,328 dollars.

Compared to our Random Forest model, Table 3 suggests that the Linear model has a lower MSE. However, we decided against using a linear model mainly because a linear model would often predict negative winning bids. Otherwise, the MSE's for all of the prediction were very good. Of note is the very small MSE for Random Forest and SVM when trying to predict the Sell variable. This corroborates our analysis on the effectiveness of the chosen machine learning models in section 2.

Table 3: MSEs of Strategy Effectiveness

Model	MSE
Linear for winning bid	47659707
Random Forest for winning bid	55285459
Random Forest for Sell	0.036
SVM for Sell	0.042

Sell_RF and Sell_SVM are the predictions of whether the car can be sold in the mentioned setup based on random forest model and SVM model, respectively. We could see that in Table 4, these two models yield to the same prediction for all 6 scenarios.

Table 4: Recommendation for Maximizing Winning Bid

	Top 3 bids			Top 3 winning bids when car is sold		
Buy it Now	1	1	0	0	0	0
Professional Platform	0	0	0	0	0	0
Featured	0	0	0	0	0	0
Reserve	1	1	1	0	0	0
Added Info	0	0	0	1	1	1
End Sunday	0	0	0	0	0	0
Auction Duration	10	10	7	10	10	10
Problems	0	0	1	1	1	1
Start Quarter	3	4	2	1	1	1
Text-level	4	4	2	3	3	3
Start Bid-level	1	1	8	8	8	8
Photos-level	5	5	4	5	6	7
Sell_RF	0	0	0	1	1	1
Sell_SVM	0	0	0	1	1	1
Biddy_RF	93145	88668	85767	65321	61141	60512
Revenue_RF	0	0	0	65321	61141	60512
Revenue_SVM	0	0	0	65321	61141	60512

5 Conclusion and Further Research

To sum it all up, the auctions we observed on eBay are English auctions, and some of them have reserve prices. According to the OLS model, we can find that as the number of bidders increases, the winning bid will rise. In addition, the best time to start the auction is around the first three months, and the winning bid increases as auction duration increases. The seller's settings for the auction, such as the higher Start Bid and Buy it Now option, have negative impacts on winning bid. The more information, such as photo and text, there are, the higher winning bid the car will tend to have. In terms of the condition of the car, mileage, crack, dent, scratch, and rust have significant negative effects on the highest bid of the auction.

From the Random Forest and SVM model, in order to have a higher winning bid price, we recommend that the seller may set a secrete reserve price and had better not set a very high start bid price. In addition, the number of uploaded photos should be proper and relative description should be attached. Also, it might be better if the start time of an auction was between January to March, and setting a Buy it Now option is very likely to have a negative effect on the final deal price. Moreover, we can find that the four most important factors that lead to a successful auction are the reserve price, the number of bidders, Buy it Now option, and Start Bid price. Hence, it is probably better for sellers to have no reserve price and no Buy it Now option and set the Start Bid prices carefully, which could increase the sellers' profits.

Finally, we would like to provide a recommendation for future research. The winning bid may not be equal to the revenue the seller gets, since there are many cars that failed to be sold. In this project, we assumed that these cases were in part due to the presence of a higher secret reserve price. However, in reality, cars are often not sold due to a multitude of different reasons. If we had more time, we would dive deeper and provide more specific guidance on how to generate the most profit.

References

[1] Andrews, T., & Benzing, C. 2007. The Determinants of Price in Internet Auctions of Used Cars. *Atlantic Economic Journal*, 35(1): 43-57.

[2] Chan, C. C. H. 2005.Online Auction Customer Segmentation Using a Neural Network Model. *International Journal of Applied Science and Engineering*, 3(2): 101-109.

[3] Chow, V. 2020. Predicting Auction Price of Vehicle License Plate with Deep Recurrent Neural Network. *Expert Systems with Applications*, 142: 113008.

[4] Hasker, K., & Sickles, R. 2010. eBay in the Economic Literature: Analysis of an Auction marketplace. *Review of Industrial Organization*, 37(1): 3-42.

[5] James, G., Witten, D., Hastie, T., & Tibshirani, R. 2010. An Introduction to Statistical Learning. *New York: springer.*, Vol. 112: p. 18.

[6] Khadge, M. R., & Kulkarni, M. V. 2016. Machine Learning Approach for Predicting End Price of Online Auction. In 2016 International Conference on Inventive Computation Technologies *International Conference on Inventive Computation Technologies*, Vol. 3: 1-5.

[7] Lewis. G. 2011. Asymmetric Information, Adverse Selection and Online Disclosure: The Case of eBay Motors. *American Economic Review*, 101: 1535-1546.

[8] Raviv, Y. 2006. New evidence on Price Anomalies in Sequential Auctions: Used cars in New Jersey. *Journal of Business & Economic Statistics*, 24(3): 301-312.

[9] Wan, W., & Teo, H. H. 2001. An Examination of Auction Price Determinants on eBay.

Appendix

Table A1: eBay SUMMARY STATISTICS

	All Sample		Sell=1		Sell=0	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Sell	0.273	0.445	1.000	0.000	0.000	0.000
Winning Bid	10,004.590	10,092.730	7,976.801	8,446.491	10,765.530	10,544.940
Miles	80,437.680	67,671.040	110,003.800	69,478.130	69,342.790	63,524.280
Buy It Now	0.338	0.473	0.0001	0.008	0.465	0.499
# of Photos	16.279	12.340	14.739	11.426	16.856	12.618
Age of Cars	7.509	4.512	9.898	4.269	6.613	4.268
Number of Bidders	5.175	4.147	7.050	4.614	4.471	3.722
Pro	0.346	0.476	0.223	0.416	0.392	0.488
Featured	0.080	0.271	0.076	0.265	0.081	0.273
Book Value	11,522.320	8,510.067	7,355.421	6,555.444	13,085.970	8,632.033
Start Bid	5,705.624	8,994.499	2,606.615	5,638.395	6,868.550	9,714.650
Warranty	0.265	0.441	0.109	0.312	0.323	0.468
Inspection	0.278	0.448	0.244	0.429	0.290	0.454
Reserve	0.648	0.478	0.000	0.000	0.891	0.312
Added Information	0.215	0.411	0.292	0.455	0.186	0.389
% Negative Feedback	1.687	5.992	1.720	5.883	1.674	6.033
Text	4,931.791	5,384.259	3,942.345	4,920.780	5,303.088	5,502.534
Dealer	0.542	0.498	0.475	0.499	0.568	0.495
Ding Group	0.395	0.986	0.490	1.097	0.359	0.938
Scratch Group	0.522	1.160	0.651	1.270	0.473	1.112
Crack Group	0.182	0.692	0.297	0.873	0.138	0.605
Broken Group	0.077	0.452	0.120	0.575	0.061	0.395
Dent Group	0.316	0.902	0.460	1.072	0.261	0.822
Problem Group	0.329	0.814	0.424	0.925	0.293	0.765
Rust Group	0.227	0.763	0.394	1.005	0.164	0.639
Length	6.844	1.596	6.679	1.578	6.906	1.598
End Sunday	0.169	0.375	0.198	0.399	0.159	0.365
End Day	3.923	2.089	4.025	2.126	3.885	2.073
Start Month	6.011	2.337	5.883	2.354	6.058	2.329
Car Model ID	29.373	13.494	26.779	14.139	30.346	13.113
Model	33.793	4.618	32.996	4.826	34.092	4.501
Doors	2.799	0.974	2.804	0.971	2.797	0.975
Transmission	2.291	0.454	2.301	0.459	2.287	0.452
N	60,763		16,580		44,183	