



eBay Sales Prediction





Outline

- Business Understanding
- Data Understanding
- Data Preprocessing
- Classification Modeling
- Regression Modeling
- Future Considerations
- Conclusion

1

Business Understanding



The Business Environment

Growth in Online Auction Market

The Global Online Auction Market is predicted to grow, in the 2018-2022 period, at a CAGR of 7.2%.

The U.S online auction market alone is expected to see an incremental growth of \$1.04 billion

Advantages of Online Markets

- Offers flexibility to users
- Large user/buyer base
- Product quality and price can easily be compared



The Business Understanding

Goal: Seeking to decrease the uncertainty surrounding online auctions.

Questions

- Is a product likely to sell?
- If a product is predicted to sell, how much will it sell for?

Purpose

- Helps users gain a better understanding of the online market environment.
- Enables third parties, such as advertisers, to be able to determine and target the most profitable auctions.

Narrowing The Scope

- Using a single auction platform and auction category.
- Focusing on historical data from Ebay dealing with signed sports memorabilia.
- Overcome the need for large and complex dataset.

2

Data Understanding



The Dataset

- Contains auction listings and price information, from January to May 2013.
- Exclusive to eBay's sport autograph category
- Player reference data was collected from SportCollectors.Net (every player since 1948).
- Each instance consist of an individual Ebay auction
- Total of 385,172 instances (both complete and incomplete auctions).



Understanding

The Features

- Retrieved from auction's metadata
- 18 original features
 - Price, MemberSince, HasPicture, ReturnsAccepted, etc.
- 15 derived features
 - AvgPrice, isHOF, AuctionCount, etc.

The Target Variables

- Classification: State of auctioned items (QuantitySold)
- Regression: Price of sold item (Price)



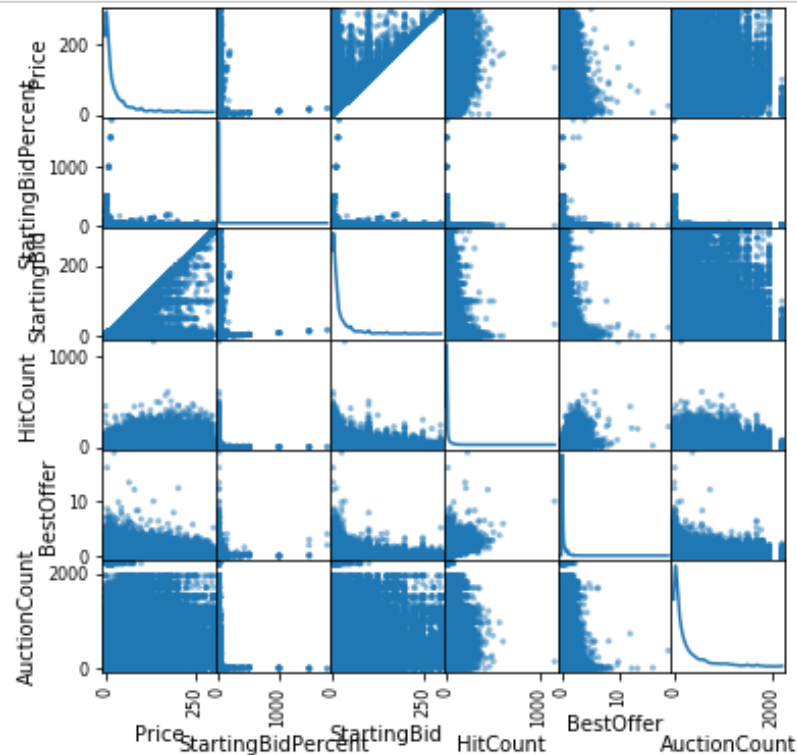
Results

\$29.17
Mean Price

\$0.01-\$300.00
Sale Price

Highly correlated features are:

- Price & Starting Bid
- Starting Bid % & Price Percent
- Hit Count, Quantity Sold & Best Offer



3

Data Preprocessing



Process

Using a preprocessed dataset, which was already divided into a training and a testing set (85:15 ratio), helped limit the amount of data preparation needed.

Our Steps

1. Exploratory data analysis
2. Finding and fixing any missing values
3. Standardizing and scaling the data
4. Simplifying the dataset:
 - Removed highly correlated items
 - Applied the Lasso penalty to retrieve important features
 - Performed PCA



Results (important features)

Classification
HitCount,
AuctionAvgHitCount, BestOffer,

Both
SellerClosePercent, StartingBid,
SellerSaleAveragePriceRatio,
SellerAvg,
SellerAuctionSaleCount

Regression
EbayID,
PricePercent,
StartingBidPercent,
SellerClosePrice,
Category,
ItemAuctionSellPercent, SellerItemAvg,
IsHOF,
PriceBuckets,
AuctionMedianPrice

4

Classification Modelling



Binary Classification Problem

Target Variable --> Quantity Sold: 1 if sale occurs, 0 otherwise

Models Used

- KNN Classifier
- Decision Tree
- Logistic Regression
- Random Forest



Model Results

	Accuracy	Recall	Precision	F-Score
KNN Classifier	0.867	0.744	0.730	0.737
Decision Tree	0.809	0.731	0.598	0.658
Logistic Regression	0.880	0.787	0.713	0.748
Random Forest	0.982	0.954	0.986	0.970

5

Regression Modelling



Model Results

The dataset used for regression only contains the instances where an item was sold.

Target Variable: Price

Results

	Mean Squared Error	Mean Absolute Error	Variance
Multiple Linear Regression	33.76	4	0.98

6

Future Considerations



Future Considerations

Incorporate different types of regression models

Up until now only multiple linear regression models have been used. We will continue to research if any other models should be used

Further explore effects that outliers have on results

Visualize the data such that any potential outliers will stand out

Develop methods to minimize their influence

Run additional trials on parameter inputs

Exploring more values for alpha in PCAs

Create an SVM model for the classification problem

Compare its key metrics to the other models' values

Use cross validation to determine optimal c and γ .

7

Conclusion



Conclusion

Our project proved that it is possible to create models that can help decrease the uncertainty of online auction markets.

Users, third parties and platforms, such as Ebay, can use these models to get a more accurate idea of the auctions with the highest probability of being successful and/or the highest profit potential. This project can also serve as a guide for models aiming to predict other types of online auction markets.

A next step in expanding this project would be exploring the possibility of creating a model or models that can handle various product categories.

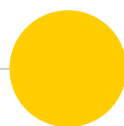
However, this step is heavily reliant on the ability to collect a sufficient amount of data and to have programs that can process large datasets.

8

Appendix

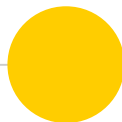
	Price	StartingBidPercent	StartingBid	HitCount	BestOffer	AuctionCount
count	258588.000000	258588.000000	258588.000000	258588.000000	258588.000000	258588.000000
mean	28.961397	1.350498	22.554573	13.124476	0.268529	251.026823
std	44.557177	12.022299	39.921438	24.894116	0.427103	342.166491
min	0.010000	0.000000	0.010000	-1.000000	-0.333333	1.000000
25%	3.290000	0.344500	1.990000	2.000000	0.047619	41.000000
50%	9.990000	0.785200	9.950000	5.000000	0.111111	114.000000
75%	34.010000	1.376900	24.950000	11.000000	0.291667	307.000000
max	299.990000	1799.000000	299.990000	1161.000000	18.750000	2213.000000

Table: Basic statistics of features



	EbayID	QuantitySold	Price	PricePercent	StartingBidPercent	SellerClosePercent	Category	PersonID	StartingBid	AvgPrice
EbayID	1.000000	0.051621	-0.053406	0.000145	-0.001235	0.082350	0.070322	-0.015432	-0.078355	-0.031546
QuantitySold	0.051621	1.000000	0.061080	-0.034879	-0.050568	0.626917	-0.057478	0.010823	-0.172144	0.108310
Price	-0.053406	0.061080	1.000000	0.024128	0.016414	0.088385	-0.305892	0.011476	0.844167	0.397352
PricePercent	0.000145	-0.034879	0.024128	1.000000	0.999446	-0.036416	0.005503	0.001540	0.030438	-0.015804
StartingBidPercent	-0.001235	-0.050568	0.016414	0.999446	1.000000	-0.052484	0.007185	0.001592	0.034863	-0.018725
SellerClosePercent	0.082350	0.626917	0.088385	-0.036416	-0.052484	1.000000	-0.086185	0.012070	-0.163144	0.120629
Category	0.070322	-0.057478	-0.305892	0.005503	0.007185	-0.086185	1.000000	-0.001267	-0.258967	-0.191164
PersonID	-0.015432	0.010823	0.011476	0.001540	0.001592	0.012070	-0.001267	1.000000	0.013099	0.012036
StartingBid	-0.078355	-0.172144	0.844167	0.030438	0.034863	-0.163144	-0.258967	0.013099	1.000000	0.288179
AvgPrice	-0.031546	0.108310	0.397352	-0.015804	-0.018725	0.120629	-0.191164	0.012036	0.288179	1.000000

Table: Correlation coefficients of various features





References

- Global Online Auction Market 2018-2022 | Ease of Bidding to Boost Growth | Technavio. (2018, August 01). Retrieved March 30, 2019, from <https://www.businesswire.com/news/home/20180801005715/en/Global-Online-Auction-Market-2018-2022-Ease-Bidding>
- Global Online Auction Market Size 2018, by Auction Software, Company Profiles, Industry Trends, Types, Strategic Analysis, and Challenges to 2022. (2019, March 31). Retrieved from <https://www.reuters.com/brandfeatures/venture-capital/article?id=50021>