Predictive Analytics

Predicting The Best Starting Price for Ebay Auctions

Course Project

Jiacheng Liao[1], Yi Wan[1], Shuang Zhou[1], Zhaoyin Zhu[2]

1 Department of Computer Science, New York University, New York, NY 10012, USA

2 Division of Biostatistics, NYU School of Medicine, Street, New York, NY 10016, USA

Abstract: Online auctions are one of the most popular methods to buy and sell items on the internet. With more than 100 million active users globally, eBay is the worlds largest online marketplace, where anyone can buy and sell anything. In order to successfully selling products on ebay, a reasonable starting price does not only determine whether the product will be sold or not but also affects the profit you can make from the transaction. In this project, we use the historical auction data collected from eBay from April 2013 to the first week of May 2013 which contains information about 296,048 successful and unsuccessful auctions. Different statistical models and machine learning algorithms will be utilized to study online auction patterns and predict the starting price that maximizes profits. Furthermore, we will compare the performance of different methods and summarize the pros and cons in different situations.

Keywords: Ebay auction • Predictive Analytics • Data Mining

1.   Introduction

Online auctions are one of the most popular methods to buy and sell items on the internet. With more than 100 million active users globally, eBay is the worlds largest online marketplace, where anyone can buy and sell anything. In order to successfully selling products on ebay, a reasonable starting price does not only determine whether the product will be sold or not but also affects the profit you can make from the transaction. In this project, we use the historical auction data collected from eBay from April 2013 to the first week of May 2013 which contains information about 296,048 successful and unsuccessful auctions. Different statistical models and machine learning algorithms will be utilized to study online auction patterns and predict the starting price that maximizes profits. Furthermore, we will compare the performance of different methods and summarize the pros and cons in different situations.

2.   Data and Business Understanding

Data Understanding involves collecting initial data, describing the data in terms of amount, type and quality of data, exploring data using available tools and verifying data quality.

Business Understanding involves determining and defining business objectives in business terms, translating these to data mining goals and making a project assessment and plan.

3.   Data Preparation

Data Preparation is an important and time-consuming part of data mining which can take up 5070% of the project's time and effort. It involves selecting data to include, cleaning data to improve data quality, constructing new data that may be required, integrating multiple data sets, and formatting data.

4.   Modeling

Modeling involves selecting suitable modeling techniques, generating test designs to validate the model, building predictive models and assessing these models.

A predictive model is a mathematical function that predicts the value of some output variables based on the mapping between input variables. Historical data is used to train the model to arrive at the most suitable modeling technique. For example, a predictive model might predict the risk of developing a certain disease based on patient details. Some commonly used modeling techniques are as follows: Regression analysis that analyzes the relationship between the response or dependent variable and a set of independent or predictor variables. Decision trees that help explore possible outcomes for various options. Cluster analysis that groups objects into clusters to look for patterns. Association techniques that discover relationships between variables in large databases.

5.   Evaluation

Evaluation involves evaluating the results against the business success criteria defined at the beginning of the project.

6.   Deployment

Deployment involves consolidating the findings, determining what might be deployed and planning the monitoring and maintenance required to keep the model relevant.

7.   Conclusion

TODO

Acknowledgements

References

[1] Han, Jiawei, Micheline Kamber, and Jian Pei. *Data mining: concepts and techniques: concepts and techniques.* Elsevier, 2011.

[2] Rajaraman, Anand, and Jeffrey D. Ullman. *Mining of massive datasets.* Vol. 77. Cambridge: Cambridge University Press, 2012.

[3] Bari, Anasse, Mohamed Chaouchi, and Tommy Jung. *Predictive analytics for dummies.* John Wiley & Sons, 2014.