ONLINE BIDDING FRAUD DETECTION

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

by

SANA ARJUMAN	**(2003A53019)**
JADHAV SATHISH	**(2003A54016)**
ADE RAVIKUMAR	**(2003A54015)**
DUBASI MUKESH	**(2103A51L07)**

Under the guidance of

Mr. S NARESH KUMAR

Assistant Professor, Department of CSE.

Department of Computer Science and Artificial Intelligence

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that this project entitled **“ONLINE BIDDING FRAUD DETECTION**" is the bonafied work carried out by **SANA ARJUMAN (2003A53019), JADHAV SATHISH (2003A54016), ADE RAVIKUMAR (2003A54015), DUBASI MUKESH (2103A51L07)** as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **COMPUTER SCIENCE & ENGINEERING** during the academic year 2021-2022 under our guidance and Supervision.

Mr. S Naresh Kumar
Assistant Professor ,
S R University,
Ananthasagar, Warangal.

Dr. M.Sheshikala
Assoc. Prof. & HOD (CSE),
S R University,
Ananthasagar, Warangal.

ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr. S.Naresh Kumar, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude to Head of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

We have analysed a dataset which is extracted from ebay sales. In day to day life we are evolved with online shopping and online auctions but we are facing many frauds related to prices or could be any other issues. In most of the auctions we have noticed an unusual or can say strange thing which is known as shill bidding fraud which is very hard to recognize. It is difficult to identify due to its similarity tonormal bidding behavior. The complexity of finding and defining SB patterns makesit resistant to discovery. Also, the unavailability of SB datasets that are based onactual e-auctions makes the development of SB detection and classification models challenging. Therefore, the prerequisite task that is necessary to perform, in order to achieve our goals in this work, is to scrape a large number of eBay auctions of apopular product, which we did successfully. We have tried to build a model which would be helpful to detect the fraud. We will be using Decision tree algorithm to analyse the data in the coming thesis of our work. The output variable is CLASS which is used to detect the fraud caused in the auction and reduces the work, it would hold the best accuracy to detect the frauds caused in the bidding by the shill bidder.

Table of Contents

LIST OF TABLES

LIST OF ACRONYMS

ML	Machine Learning
SB	Shill Bidding
BT	Bidder Tendency
EB	Early Bidding
BR	Bidding Ratio
LB	Last Bidding
ASP	Auction Starting Price
SOB	Successive Out Bidding
WR	Winning Ratio
AB	Auction Bids
B	Participated Bidder
S	Auction Seller
A	Targeted Auction

1. INTRODUCTION

1.1 OVERVIEW

Over the past three decades, electronic commerce has witnessed a significant increase in the trade of goods and services. This kind of business has created an attractive environment for fraudsters who can perform various types of fraudulent activities to illegally raise money. One such activity is SB, which is prevalent across many auction sites and, unfortunately, this type of fraud is challenging to detect due to its similarity to normal bidding behaviour as well as the difficulty of systematically defining it. Furthermore, genuine bidders are not aware of this fraud until it is too late. Thus, this type of fraud has become one of the most critical fraud perpetrated on auctions. Implementing a SB detection model is very difficult and requires a deep understanding of the bidding behaviour. Moreover, the lack of SB datasets makes the implementation of SB detection and classification systems troublesome. In order to build an efficient SB detection model, we, first, created an SB dataset from data collected from real auctions of commercial sites that were most likely infected by malicious moneymakers. Thus, the initial phase in this study was scraping a large number of eBay auctions of a popular product. The raw eBay auction data collected.

1.2 PROBLEM STATEMENT

During the last three decades, the exchange of goods and services over the Web has increased significantly. According to the World Trade Organization, the merchandise sold online globally between 1995 and 2014 was worth over $18 billion 1 . With technology in hand, people’s lifestyles are moving faster in various regards, such as communications, advertisements, education, etc. One of the most critical aspects is commercial agreements over the Internet, such as online auctions. Buying and selling products and services through online auctions is convenient for many people, since it saves time and effort. However, this creates an attractive environment for fraudstero carry out their suspicious activities, due to the vulnerability of online auctions to cyber crimes. As stated in [CC11], auction fraud has persisted since 2004 as one of the top two cyber crimes.

The harm caused by malicious moneymakers targeting online auctions not only affects the victims (genuine bidders) but the online auction businesses and the authorities as well. Regarding the victims who are regular users, thousands of complaints have been filed with the authorities and millions of dollars have been lost. As the federal Internet Crime Complaint Center (IC3) stated, 9,847 victims lost $ 11 million due to online auction fraud in 2014 2 . Also, the IC3 reported that 5% of the Internet complaints were associated with automobile auctions scams, where $51 million was lost in 2013. This undoubtedly frustrates the authorities, as a vast amount of work is required to track down the fraudsters. On the business side, if the number of victims keeps increasing, online auctions will be abandoned and will eventually go out of business. Thus, to protect society and provide safe online auctions, it is important for e-auction administrators to employ and support researchers who investigate, understand, and handle the methods, tricks, and security bugs that fraudsters use to engage in their criminal activities.

Typically, fraud operations related to online auctions fall under three categories:
1) pre-auction fraud, where the fraud is conducted before the auction starts;
2) inauction fraud performed during the auction bidding;
3) post-auction fraud done after the auction is completed. The second category of fraud is considered the most challenging to handle, since bidders are unaware that they have been victimized. For example, Shill Bidding (SB) is a form of in-auction fraud where a bidder keeps outbidding others to inflate an item’s price, in order to maximize the seller’s profits. In addition, as the number of bids increases, regular bidders are deceived that the item is desirable. As a result, many buyers end up paying more than what a product is actually worth. Unfortunately, SB is difficult to discover at the right time, since it looks similar to normal bidding. After the payment is processed and the item has been received, then the buyer may find that she or he has been duped, but it is too late.

Online auction systems are complex and involve highly sophisticated security techniques that control the systems’ users such as sellers, bidders, banks inquiries, and processing online payments transactions. Nevertheless, various of fraud activities could be committed such as inflating the products prices, selling misrepresented items, non-delivered of purchased items, and non-payment received for delivered item. Systems like feedback reputation, fraud awareness (fraud prevention system), and user authentication are not enough to deter scam attempts, due to the computational complexity of these types of fraud (such as SB).

The scope of this work is about fraud is online auctions. More precisely, we concentrate on the SB fraud that is widely associated with online auctions, and which is exceedingly burdensome to systematically detect, due to its similarity to genuine bidding behaviour. To implement an effective and robust e-auction system that protects actual bidders from being deceived by SB activities, the following points must be adequately addressed:

• Analyze sellers' and bidders' behaviours to identify the SB patterns.
• Recognize fraudsters and their various ways of committing fraud, in order to know how to deal with it.
• Define the most effective SB metrics for measuring the SB behaviour.
• Handle a huge volume of bidding transactions, which include bidder name, auction ID, bid amount, a bid time, number of bids, etc.
• Determine the most feasible ML methods to help developers build fraud detection systems, such as the best classification algorithm for detecting SB.
• Develop an ability to systematically learn from existing fraudulent characteristics data.
• Extract and preprocess actual data with a relatively large number of bids to effectively handle and manage the above-mentioned points.

1.3 OBJECTIVES DESCRIPTION

The objective of this research is to use ML in online auctions to determine user behaviour. More specifically, we would like to employ instance-incremental learning to recognize the auctions, sellers, and bidders involved in SB. These classification methods will allow us to systematically base fraud detections models on precise, wellproven SB patterns. In 2017, we collected data from the most popular e-auction site that is eBay to create a high-quality SB dataset; to do so, the extracted raw data were hard-loaded during preprocessing, in order to obtain clean, useful auction data. One of the most critical phases when building classification models is labelling multidimensional data. Indeed, classification quality relies on how accurately labelled the instances are in the dataset. Moreover, there is no available training SB datasets; thus, one of the research goals is to create a high-quality labelled SB dataset and make it public for researchers and developers from the same domain. Since the lazy learning classifiers are claimed to be useful for data streaming mining, we investigate the feasibility of some of these classifiers by applying them to our labelled SB dataset. In addition, one of our objectives is to find out which technique for handling imbalanced data is the most suitable for the imbalanced SB dataset. Therefore, we examine the most practical handling imbalanced techniques (on both data and algorithm levels) for our case study. These techniques are integrated with the instance-based classifiers to build the detection models based on the produced SB dataset. Finally, we perform a comprehensive comparison between the experimental results.

1.4 OVERALL ARCHITECTURE

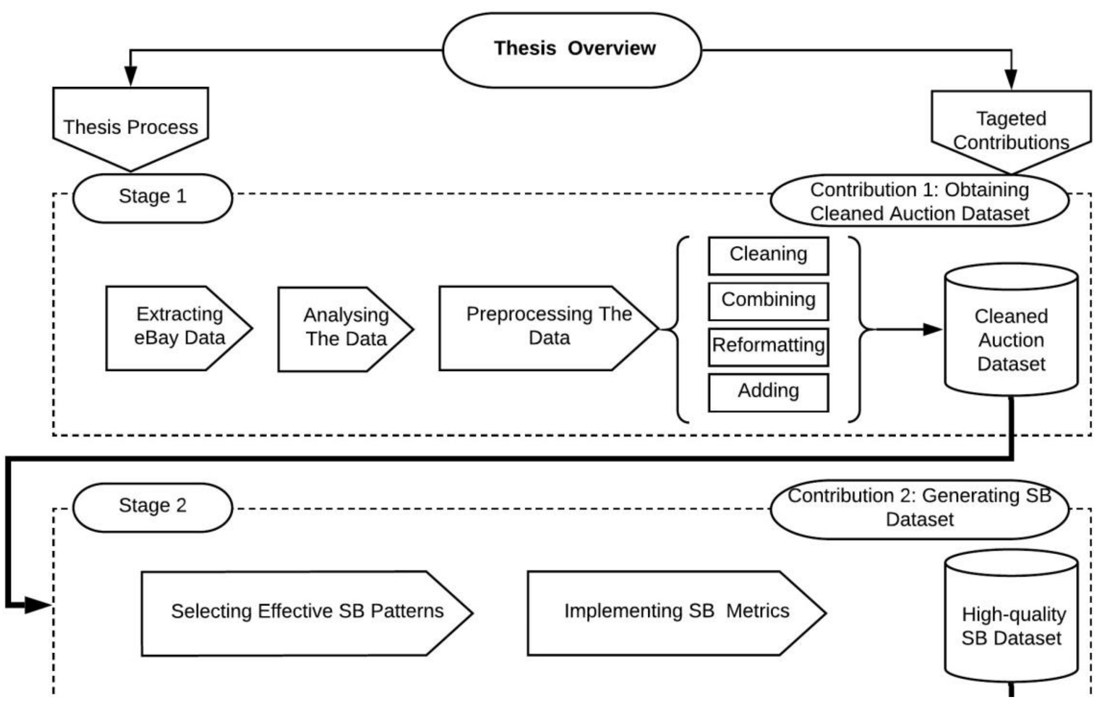


Fig 1.4.1 ARCHITECTURE

As shown in the fig 1.4.1 we have successfully extracted actual and recent online auction data for about three months (end of March to beginning of June, 2017). In accordance with the nature of raw data, the extracted data was a mess. There were over a half-million bids, yet, unfortunately, there were many duplicate bids, missing values, and inconsistent bidding transactions records. Thus, our first contribution in this work is to obtain a cleaned, organized, processable online auction dataset. We have preprocessed the extracted data using cleaning, combining, reforming and adding. In binary classification, each record presented in a dataset is classified into a specific class. This is not the case with the bidders in online auction datasets, where a bidder may participate more than once in an auction; thus, we cannot classify each individual bid transaction. To overcome this problem, instead of looking at each bid in each auction in the cleaned auction dataset, we study the behaviour of each bidder in each auction in that dataset. This is done by computing eight different SB patterns that efficiently describe bidder and seller behaviours. Thus, we generate a new highquality SB dataset that accurately defines the behaviour of each participating bidder in each auction.

2. LITERATURE SURVEY

Shill Bidding (SB) is the most common auction fraud but the most difficult to detect due to its similarity to normal bidding behaviour . A shill bidder is a malicious user (the fraudulent seller and/or his accomplices) who bids aggressively in order to drive up the price of the product only to benefit the owner of the auction. SB may cause a massive money loss for genuine sellers and bidders in the context of high priced products and also products with unknown value in the market, such as antiques . Excessive SB could lead to a market failure. Online auctions may affect the users' confidence, which may negatively impacts the auctioning business. In fact, several sellers and their accomplices have been prosecuted due to SB activities, including:

In 2001, three sellers were charged of SB fraud worth a pay-off of $300,000 through 1100 auctions of art paintings. The fraud was conducted on eBay with more than 40 fake accounts .

In 2007, a jewellery seller was accused of conducting SB fraud on eBay, and had to pay $400,000 for a settlement. Also, he and his employees were prevented from engaging in any online auctioning activities for four years.

In 2010, a seller faced a £50,000 fine after being found outbidding himself on eBay. He claimed that : "eBay let me open up the second account and I gave all my personal details and home address to do so."

In 2012, the online auction ”TradeMe” had to pay $70,000 for each victim after the investigation discovered SB fraud conducted by a motor vehicle trader in Auckland. The fraud was carried out for one year, and caused a significant loss for the victims. Trade Me blocked this trader from using their site, and referred the case to the Commerce Commission for a further investigation.

In 2014, a lawsuit was filled against ”Auction.com” by VRG in California claiming that the website allowed SB. The bid of $5.4 million should have secured the property as the plaintiff declared, and yet the winning price was 2 million more. Auction.com was accused of helping the property's loan holder, which is not fair for genuine bidders. The California state passed a law on July 1, 2015, which requests the property auctioneers to reveal bids they submit on a seller's behalf. The spokeswoman for the California Association of Realtors said: "To the best of our knowledge, we're the only state to pass this sort of legislation, even though we believe SB to be prevalent all over the country."

eBay policy states that "Shill bidding can happen regardless of whether the bidder knows the seller. However, when someone bidding on an item knows the seller, they might have information about the seller's item that other shoppers aren't aware of. This could create an unfair advantage, or cause another bidder to pay more than they should. We want to maintain a fair marketplace for all our users, and as such, shill bidding is prohibited on eBay." This statement clearly demonstrates that SB is troublesome and tough to be addressed.

3. DATA PROCESSING

3.1 DATASET DESCRIPTION

Table 3.1.1 – Data Set Information

Number of auctions	807
Number of instances	6321
Number of bidder IDs	1054
Avg. bidders in low starting price auctions	10
Avg. bidders in high starting price auctions	6
Winners and not aggressively participated	536 (8.5%)
Not winners and aggressively participated	1659 (26.25%)
Not winners and not aggressively participated	3817 (60.4%)

Bidder Tendency: a shill bidder participates exclusively in auctions of few sellers rather than a diversified lot. This is a collusive act involving the fraudulent seller and an accomplice. The latter acts as a normal bidder to raise the price.

Early Bidding: a shill bidder tends to bid pretty early in the auction (less than 25% of the auction duration) to get the attention of auction users.

Bidding Ratio: a shill bidder participates more frequently to raise the auction price and attract higher bids from legitimate participants.

Last Bidding: a shill bidder becomes inactive at the last stage of the auction (more than 90% of the auction duration) to avoid winning the auction.

Auction Starting Price: a shill bidder usually offers a small starting price to attract legitimate bidders into the auction.

Successive Outbidding: a shill bidder successively outbids himself even though he is the current winner to increase the price gradually with small consecutive increments.

Winning Ratio: a shill bidder competes in many auctions but hardly wins any auctions.
Auction Bids: auctions with SB activities tend to have a much higher number of bids than the average of bids in concurrent auctions (i.e. selling the same product). Therefore, sellers of these auctions have a high probability of colluding with shill bidders to increase their profits.

3.2 DATA CLEANING

Firstly, we need to remove noisy data possessing the following characteristics:

- Irrelevant and duplicated attributes: several attributes in the raw dataset are not needed to compute the SB metrics, such as the product location and ID, feedback ratings of sellers and bidders, and bidders' account links. Also, some data are displayed twice on the main auction page and inside the auction link, such as the seller name, auction starting time, number of bids, and seller rating. There attributes are removed.

- Duplicated records: during the scrapping process, some data have been collected more than once, e.g., when a bidder participates more than once in an auction, the crawler collects his history each time. For example, let us suppose a bidder has 10 records in his bidding history, and he participated two times in an auction, then the crawler will grab his history each time. As a result, we would receive 20 records of that bidder but 10 of them have been already captured.
- Records with missing values: there are several rows without the bidders' IDs; and we are not certain whether it was caused by eBay or the scrapper itself. These IDs cannot be generated by using the inputting techniques. So, we need to delete them.

- Auctions with less than 5 bids: these auctions did not engage any SB fraud because the very few placed bids are genuine. In some of these auctions, the items were sold using the “Buy-It-Now” feature on eBay (bidders can pay off the price directly by clicking Buy-It-Now button). In some other auctions, sellers canceled their sales due to reasons like the items were not available anymore for sale, or error discovered in the listing.

- Auctions with inconsistent data: several attributes contain incompatible values. For instance, the last submitted bid is greater than the winning price, or the starting price is greater than the winning price. So, we decided to remove these auctions to not mislead the fraud classifiers.

3.3 DATA AUGMENTATION

In our original data, a specific URL represents an auction, which is not of a proper format for measuring SB patterns. To overcome this problem, a new attribute called AuctionID is given to provide a unique integer identifier for each auction. However, the AuctionID is repetitive in the dataset w.r.t. number of bids in that auction. This will make the computation of the SB patterns time consuming. So, the AuctionID cannot be the primary key for the records in the auction dataset. Thus, there is a need to represent several attributes: AuctionID, BidderID and Bid Submit Time Sec, with a single identifier for each record. Thus, a new attribute is added to the dataset called recordID to uniquely identify each individual record. The relevant set of auction attributes are presented in Table 3.1.1

Table 3.2.1: Auction Attributes for Shill Bidding

AUCTION LEVEL		
1	Auction ID	Unique identifier of an auction
2	Seller ID	Unique identifier of a seller
3	Number of Bidders	Number of bidders in an auction
4	Starting Price	Starting price set up by a seller
5	Auction Duration Sec	How long an auction lasted
6	Start Time Sec	Time on which an auction started
7	End Time Sec	Time on which an auction ended
BID LEVEL		
1	Bidder ID	Unique identifier of a bidder
2	Bid Amount	Bid price placed by a bidder
3	Bid Submit Time Sec	Time where a bid was submitted by a bidder
4	Number of Bids	Number of submitted bids in an auction
5	Winning Bid	Final price of an auction
6	Record ID	Unique identifier of a record in the dataset

3.4 DATA VISUALIZATION

1 26.25% of bidders aggressively outbidded themselves and others but did not win any auctions. The behavior of those bidders indicates that they committed SB.

2 8.5% of bidders did not highly participated in the auctions but won. Those bidders were fairly active at the last auction stage. All these refer to genuine behaviour.

3 60.4% of bidders looked normal and did not win. Indeed, those bidders did not aggressively outbid others and did not submit successive bids.

4 4.9% of bidders extremely outbidded others and won the auctions. This indicates their desire to win the auction.

5 The average number of bidders in auctions with a low starting price is 10, whereas the average is 6 in regular auctions with a regular starting price.

6 11.5% of bidders submitted bids at an early stage and aggressively outbidded others but did not win any auctions. This indicates their intention to increase the item price.

Table 3.2.2 - SB Patterns in the Auction Dataset

SB Pattern	**High Value (> 0.7)**
Bidder Tendency	209 (3.3%)
Early Bidding	2112 (33.4%)
Bidding Ratio	43 (0.68%)
Last Bidding	1976 (31.26%)
Auction Starting Price	2944 (46.6%)
Successive Outbidding	1968 (31.13%)
Winning Ratio	1659 (26.24%)
Auction Bids	221 (3.5%)

As presented in Table 3.2.2, SOB pattern is relatively high since 31.13% of samples/bidders aggressively outbidding others. LB pattern also shows that 31.26% of bidders remained inactive at the last auction stage whereas 33.4% of bidders participated at the early stage. A good number of bidders (46.6%) tends to bid in auctions with a low starting price.

In [17]:
```
x1=data['Record_ID']
y=data['Class']
print(x1,y)
from matplotlib import pyplot as pt
pt.scatter(x1,y)
```

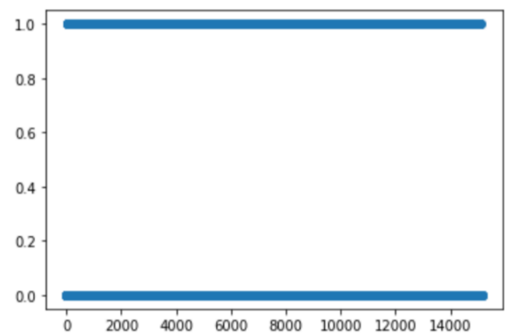


Fig 3.3.1

Fig 3.3.1 shows the scatter plot between Record Id and Class

In [18]:
```
x2=data['Auction_ID']
y=data['Class']
print(x2,y)
from matplotlib import pyplot as pt
pt.scatter(x2,y)
```

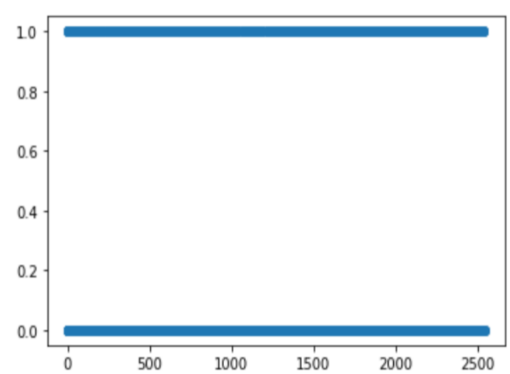


Fig 3.3.2

Fig 3.3.2 shows the scatter plot between Auction Id and Class

In [19]:
```
x3=data['Bidder_Tendency']
y=data['Class']
print(x3,y)
from matplotlib import pyplot as pt
pt.scatter(x3,y)
```

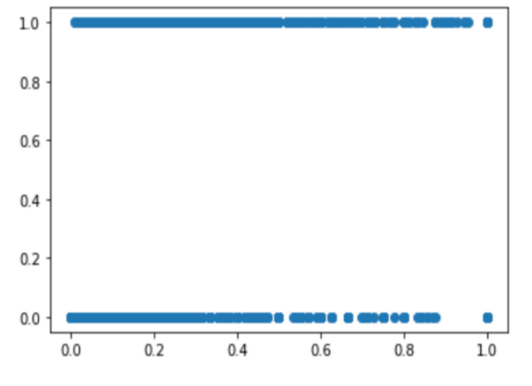


Fig 3.3.3

Fig 3.3.3 shows the scatter plot between Bidder_Tendency and Class

In [20]:
```
x4=data['Bidding_Ratio']
y=data['Class']
print(x4,y)
from matplotlib import pyplot as pt
pt.scatter(x4,y)
```

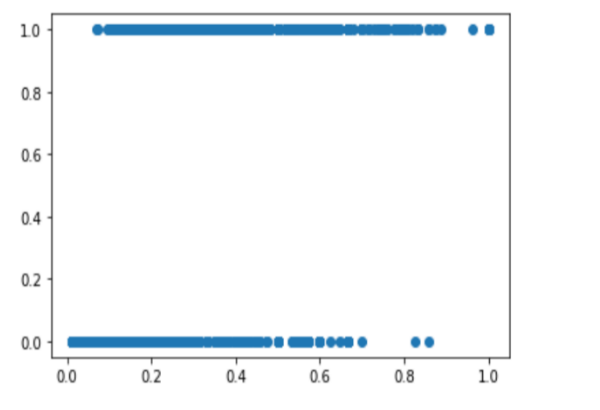


Fig 3.3.4

Fig 3.3.4 shows the scatter plot between Bidding_Ratio and Class

In [22]:
```
x5=data['Successive_Outbidding']
y=data['Class']
print(x5,y)
from matplotlib import pyplot as pt
pt.scatter(x5,y)
```

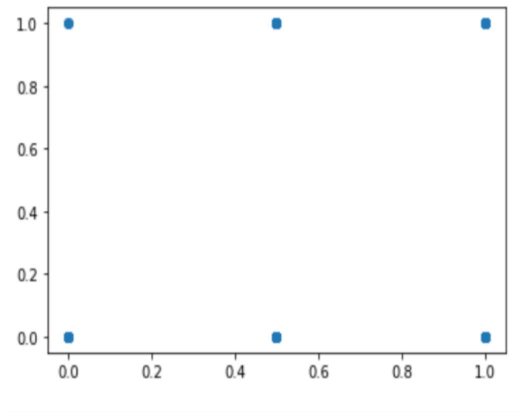


Fig 3.3.5

Fig 3.3.5 shows the scatter plot between Successive_Outbidding and Class

In [24]:
```
x6=data['Last_Bidding']
y=data['Class']
print(x6,y)
from matplotlib import pyplot as pt
pt.scatter(x6,y)
```

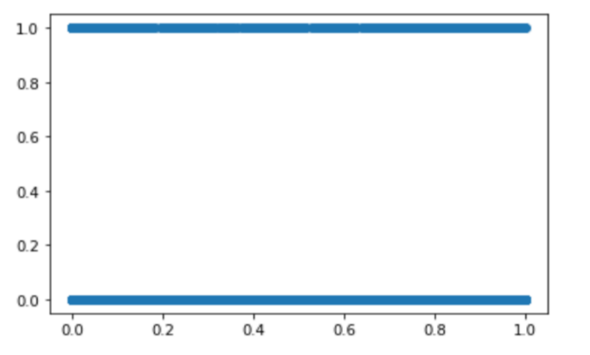


Fig 3.3.6

Fig 3.3.6 shows the scatter plot between Last_Bidding and Class

4. METHODOLOGY

4.1 MODEL DESCRIPTION

To overcome this problem, instead of looking at each bid in each auction in the cleaned auction dataset, we study the behaviour of each bidder in each auction in that dataset. This is done by computing eight different SB patterns that efficiently describe bidder and seller behaviours. Thus, we generate a new highquality SB dataset that accurately defines the behaviour of each participating bidder in each auction.Shill Bidding training datasets of online auctions are highly imbalanced. Imbalanced data have proved to deteriorate the accuracy of the classification models, especially for the minority class. This is a serious concern in fraud detection problems where the minority class, representing the fraud class, has the highest misclassification cost. To the best of our knowledge, this is the first time, the effect of imbalanced data is investigated with incremental classification algorithms. In our empirical study, we have examined the performance of several instance-based classifiers combined with oversampling and under-sampling techniques.

In the study the attribute CLASS helps us to indentify whether the bid is normal or fraud. If the outcome is '0' then it is a normal behaviour of bidding and if the outcome is '1' then there is a fraud undertook in the bidding.

These classification methods will allow us to systematically base fraud detections models on precise, wellproven SB patterns. In 2017, we collected data from the most popular e-auction site that is eBay to create a high-quality SB dataset; to do so, the extracted raw data were hard-loaded during preprocessing, in order to obtain clean, useful auction data. One of the most critical phases when building classification models is labelling multidimensional data. Indeed, classification quality relies on how accurately labelled the instances are in the dataset.

4.2 MODEL ARCHICTURE

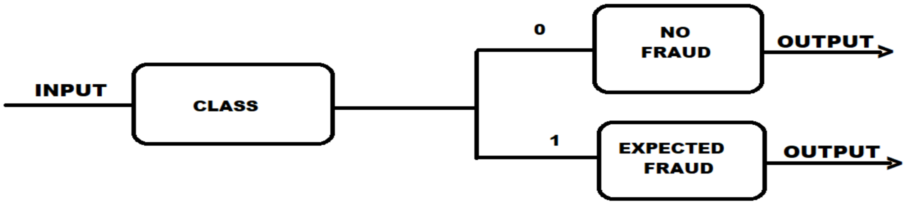
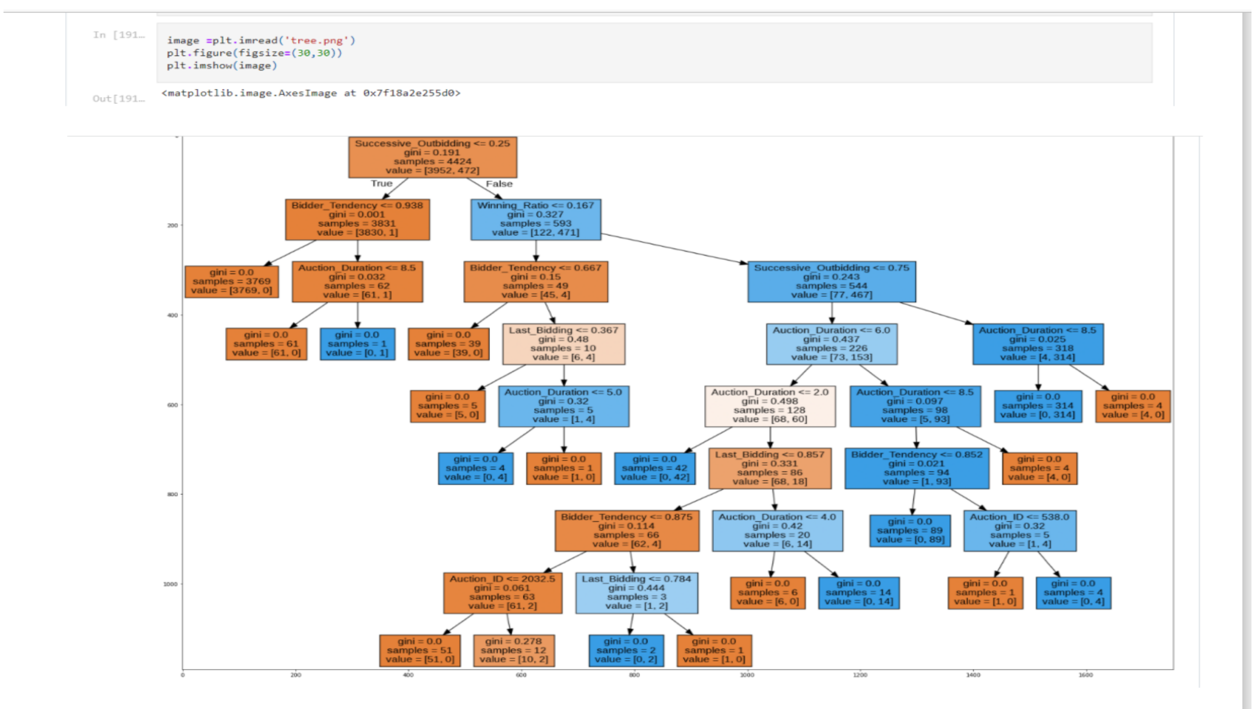


Fig 4.2.1

- From fig 4.2.1 in the study the attribute CLASS helps us to indentify whether the bid is normal or fraud. If the outcome is ‘0’ then it is a normal behaviour of bidding and if the outcome is ‘1’ then there is a fraud undertook in the bidding



FINAL DECISION TREE

4.3 SOFTWARE DESCRIPTION

a. PYTHON - Python is an interpreted, high-level, general-purpose programming language. Python's design philosophy emphasizes code readability with its notable use of significant whitespace. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small and large-scale projects. Python is dynamically typed AND supports multiple programming paradigms, including procedural, object-oriented, and functional programming.

b. GOOGLE COLAB – Colaboratory or “Colab” for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education.

5. RESULTS AND DISCUSSIONS

As in the case of any fraud detection applications, the SB dataset obtained from online auctions is imbalanced. Thus, we have to tackle this issue before implementing fraud classification models. Classifying each instance in the training dataset is another critical and sensitive step in ML, as the quality of a model's predictions depends on this step. This is another challenging task faced in this research because the obtained high-quality SB training dataset is not labelled. Therefore, one of the objectives of this work is to address the above-mentioned problem by providing a labelling approach for the SB pattern dataset to be used in SB detecting models based on ML.

Shill Bidding training datasets of online auctions are highly imbalanced. Imbalanced data have proved to deteriorate the accuracy of the classification models, especially for the minority class. This is a serious concern in fraud detection problems where the minority class, representing the fraud class, has the highest misclassification cost. To the best of our knowledge, this is the first time, the effect of imbalanced data is investigated with incremental classification algorithms. In our empirical study, we have examined the performance of several instance-based classifiers combined with oversampling and under-sampling techniques.

In the study the attribute CLASS helps us to indentify whether the bid is normal or fraud. If the outcome is '0' then it is a normal behaviour of bidding and if the outcome is '1' then there is a fraud undertook in the bidding.

6. CONCLUSION AND FUTURE SCOPE

CONCLUSION

In day to day life we are evolved with online shopping and online auctions but we are facing many frauds related to prices or could be any other issues. In most of the auctions we have noticed an unusual or can say strange thing which is known as shill bidding fraud which is very hard to recognize. It is very difficult to analyse the each and every record. We have tried to build a model which would be helpful to detect the fraud. We will be using Decision tree algorithm to analyse the data in the coming thesis of our work. The output variable is CLASS which is used to detect the fraud caused in the auction and reduces the work, it would hold the best accuracy to detect the frauds caused in the bidding by the shill bidder.

In the study the attribute class helps us to indentify whether the bid is normal or fraud. If the outcome is ‘0’ then it is a normal behaviour of bidding and if the outcome is ‘1’ then there is a fraud undertook in the bidding. We got 99% of accurac

Therefore, this project helps us to identify and reduce the fraud in online shopping and online auctions.

FUTURE SCOPE

We would like to provide an overview of the future contributions in ML generally, and SB fraud detection in particular.

Preprocessing raw datasets is a challenging and important step that consumes a lot of time and effort, but it must be addressed. Therefore, it is important to develop software that helps data scientists to track what has been done to the data, which is one of our planned future projects. Preprocessing raw data can contain different types of tasks and queries. So, by tracking the modifications that have been done in the dataset, we can save time and efforts for preprocessing future raw data. Also, we will automate the data preprocessing to address the issue of hard-loaded work.

Identifying and defining more bidder and seller behavioural patterns is another work we plan to do in the future.

Our future work will concentrate on deriving a complete online auction fraud detection system that instantly discovers shill bidders at the last stage of an auction i.e. just before processing the payment of the products

7. REFERENCES

1. Albert Bifet. Weka: Advanced data mining with weka. https://www. cs.waikato.ac.nz/ml/weka/mooc/advanceddataminingwithweka/ slides/Class2-AdvancedDataMiningWithWeka-2016.pdf, 2016. Accessed: 2018-06-10.
2. Phiradet Bangcharoensap, Hayato Kobayashi, Nobuyuki Shimizu, Satoshi Yamauchi, and Tsuyoshi Murata. Two step graph-based semisupervised learning for online auction fraud detection. In Proceedings of the Machine Learning and Knowledge Discovery in Databases - European Conference, ECML PKDD 2015, Porto, Portugal, September 7-11, 2015, Proceedings, Part III, pages 165–179, 2015.
3. Mile Brandly. What is shill bidding? https: //mikebrandlyauctioneer.wordpress.com/2010/11/25/ what-is-shill-bidding/, 2010. Accessed: 2017-07-30.
4. Rodel Balingit, Jarrod Trevathan, Yong Jin Lee, and Wayne Read. A software tool for collecting data from online auctions. In Proceedings of the 6th International Conference on Information Technology: New Generations, pages 922–927. IEEE, 2009.
5. Jonathan Burez and Dirk Van den Poel. Handling class imbalance in customer churn prediction. Expert Systems with Applications, 36(3):4626– 4636, 2009.
6. Nitesh V. Chawla, Kevin W. Bowyer, Lawrence O. Hall, and W. Philip Kegelmeyer. SMOTE: synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, (16):321–357, 2002