CHAPTER ONE
INTRODUCTION

1.1 Background to the Study

A few decades ago, auctions were done in auction houses and the bids were made with the auctioneer delegating the bids and this method required the physical presence of the bidders, thus it resulted in a number of limitations (Jurjees *et al*., 2018). This led to the use of online auctioning which allowed for the auctions to be carried out over the internet from anywhere in the world. However, with the development of the Internet, it is now simple to purchase goods through online auction systems that were previously only available in brick-and-mortar markets. Online auctions are among the most influential and effective electronic business applications. A lot of people now choose to buy and sell goods at their convenience through various online auction websites. Even most companies have started procurement projects of their own, thereby improving their purchasing and sales channels (Dong *et al*., 2019).

An E-auction or Online auction system is described as a system that performs online auction for various products and services on a website. It is regarded as a point where sellers and buyers gather to trade almost anything conveniently on the internet within a time frame and a commencement price inclusive. However, it comes with its own risk. Shill bidding stands as the most common manifestation of auction fraud. It occurs when a seller places a bid on their own auction, aiming to artificially boost the final winning bid amount. This action can be executed either directly by the seller or indirectly through collaboration with a third party, who submits fake bids on behalf of the seller. Due to the sheer volume of auctions and bids taking place, manually detecting instances of shill bidding becomes a formidable task. The optimal approach to mitigating the negative impact of auction fraud lies in promptly detecting shill bidding as it occurs, in real-time. In fact, vulnerable to dishonest money makers who commit different types of malicious activities such post auction and in-action fraud. This vulnerability exits due to high anonymity of users, flexibility in bidding and cheap auction services. Commercial auction sites (eBay, Amazon, Trust Me etc.) provided a tremendous volume of public data, and researchers can employ machine learning methods to effectively examine data of auctions and users in order to determine the trends and online patterns of users.

Considering the challenges inherent in orchestrating an online auction and the diverse array of individuals engaged in the process – encompassing buyers, sellers, auctioneers, third-party vendors, acquiring banks, and online payment systems, it has become of immense importance

to develop different fraud monitoring methodologies to analyze and track existing e-auction frauds. However, fraud prevention mechanisms like fraud awareness, user authentication and feedback-based reputation systems can sometimes be reliable for small scale frauds but are not full proof enough to protect users from becoming victims of the frauds which are sometimes complex to detect. Various fraud detection approaches have been researched on for effective identification and the elimination of fraud in online auctions (Thiagu *et al.,* 2014).

In the proposed system neutral network is use for detecting fraud and genuine bide in store. In the last decade, neural network, which is an area in artificial intelligence, has drawn the attention of researchers and analysts. Neural network has been used successfully in developed economies in predicting stock price movement and investment banking (Leung *et al*., 2001). Neural network comprises of a set of input, output and hidden layers connected with the help of nodes whose weights are adjusted and tuned during learning process. The system is trained using different parameter (dataset). During training process input vectors are supplied to the input layer which further propagate through hidden layers and reaches the output layer. At the output layer actual output is compared with desired output and an error is generated which is further propagated backwards thereby adjusting connection weights of the nodes.

1.2 Statement of the Problem

E-auction fraud is a type of online fraud that occurs when sellers manipulate the bidding process or the outcome of an online auction, such as eBay, to gain an unfair advantage or to deceive others. Shill bidding is a common e-auction fraud where a seller or someone associated with them bids on their own item to inflate the price or to create a false sense of demand. E-auction fraud can cause significant losses and damages for both sellers and buyers, as well as undermine the trust and reputation of the online platforms. Therefore, it is important to have an automated system that can detect and prevent fraudulent activities in real time, or flag them for further investigation. The proposed architecture is versatile enough to suit both forward and reverse auctions, incorporating a distinct security module designed to actively oversee and identify shill activities during the course of the auction. This module will accomplish its task by employing diverse techniques to recognize suspicious bidding patterns and by tracking IP addresses.

1.3 Aim and Objectives

The aim of this study is to develop a machine learning model for fraud detection in an e-auction system. The study objectives include to:

1. Develop a model for detecting shill bidding fraud in an e-auction system.
2. Design the proposed model using machine learning algorithm.
3. Train and test the data set obtained from ebay website using logistic regression technique.
4. Implement the proposed shill bidding detection model using python programming language.
5. Evaluate the model performance using metrics such as accuracy, precision, recall and f1-score.

1.4 Significance of the Study

Fraud constitutes a major challenge in e-commerce platform. Shill bidding is a prevalent method for auction fraud. In shill bidding, item prices are increased by fake bidders. Online shill bidding detection is a prevalent crime and very difficult to detect because of its very similar behavior during real-time bidding. This study offers the following significance;

1. Pinpointing areas of research deficiency within current models for recognizing and averting shill bidding tendencies that are yet to be addressed both in literature and commercial solutions.
2. Strengthening the security of auction transactions by developing a shill bidding authentication model that can efficiently distinguish between normal and dishonest bidders.
3. This study will be helpful to both bidders and the e-auction companies that faces yearly millions of dollars loss and fraud reports.
4. This study will develop a machine learning model to find the shill bidding in the real-time auction in which logistic regression will be used for prediction and to decide whether the shill bidding is committed or the bid is normal.
5. Outlines potential study directions and difficulties in detecting shill bidding in live e-commerce transactions.

1.5 Scope of the Study

This study covers the detection of an in-auction fraud such as the shill bidding fraud in order to guarantee user authentication and verification. This research work is inspired by the work done by Rohan Pillai, (2021). It is an online auction system based on the Block chain and smart contract technology. In the existing system, Seller can create a single page where their product can be uploaded for buyer to see. While the buyer can directly access the sellers page or the web site page. It was considered that the sites that allowed sellers to sell without forcing the seller to register on the site. The smart contract contained important information about the transaction details such as buyer and seller bank details. The seller initiates the transaction by sending a request to the block application. While client invokes Hash to submit the transaction proposals to the endorsing peer. These endorsement peers receive the transaction proposals, simulate the transaction by executing chain code specified by the transaction, and verify it with buyer details. After verification, the endorsing peer executes the transaction in the ledger and returns the response result to the Fabric platform. In the existing system, the seller information is not open for the buyer to see along with other relevant information about a product easily. The system is faced with a problem of the admin not been able to view or detect user details on the platform so to avoid counterfeit, infringing, stolen or other questionable products on its platform. However, in this proposed system, we would use deep learning neural network algorithm to tackle user authentication challenges.

1.6 Definition of Terms

Auction: In an open or closed style, an auction is a sales event where prospective buyers submit competing bids on goods or services.

E-auction: This is also called online auction or e-auction, which is held over the internet.

Internet: Computers all over the world are connected by a huge global network called the Internet. People can share information and converse via the Internet from any location with an Internet connection.

Online auction system: A web-based application called the online auction system allows users to publish ads and purchase or sell any item they wish. Users of this application will be able to list their goods for auction; interested parties can register and place bids on any item that is up for grabs.

Electronic Commerce: The term electronic commerce (e-commerce) refers to a business model that allows companies and individuals to buy and sell goods and services over the Internet.

Security: This involves limiting others' freedom to act in order to protect oneself against or to be resilient against potential harm (or other undesirable coercive change) perpetrated by others.

Fraud: Fraud encompasses a purposeful action (or omission of action) intended to attain an unauthorized advantage, whether for personal gain or on behalf of an entity, through the utilization of deceit, misleading ideas, concealment of truth, or other unethical methods that others accept and depend upon.

Fraud Detection: Fraud detection is a set of activities undertaken to prevent money or property from being obtained through false pretenses. Fraud detection is applied to many industries such as banking or insurance.

Credit Card Fraud: Credit card fraud is the unauthorized use of a credit or debit card, or similar payment tool to fraudulently obtain money or property. Credit and debit card numbers can be stolen from unsecured websites or can be obtained in an identity theft scheme.

Bidder A person or organization making a formal offer for something, especially at an auction.

Model: It is used as an example to follow or imitate.

Graphical User Interface: It is a human–machine interface through which a user interacts with electronic devices, such as computers and smart devices.

Algorithm: It refers to a method or formula designed to resolve a problem by following a series of defined steps. A computer program can be thought of as a sophisticated algorithm in action.

Binary Classification: *Binary classification* is a form of classification which involves the process of predicting categorical variables where the output is restricted to two classes.

Supervised Learning: It is also known as **Supervised Machine Learning** is an algorithm that learns from labeled training data to help you predict outcomes for unforeseen data. You train the computer in supervised learning with data that has been accurately "labeled." It indicates that some material has previously been appropriately tagged.

Neural Network: Deep learning is a machine learning technology that uses interconnected neurons or nodes in a layered framework to mimic the human brain.

Shill Bidding : It takes place when a seller in a forward auction wants to sell his item for more, or a buyer in a reverse auction wants to buy the item for less than the usual price.

CHAPTER TWO

LITERATURE REVIEW

2.1 Review of Auction System

Online auction sites are target for fraud due to their anonymity, number of potential targets and low likelihood of identification. Researchers have developed methods for identifying fraud. However, these methods must be individually tailored for each type of fraud, since each differs in the characteristics important for their identification.

Auction is gotten from the Latin word “augeo” which means increase. It was first recorded in the year 193AD, when praetorian guards had an entire empire of “Rome” to sell. After year 500 BC, an auction was held to sell women (Rashesh et al., 2015). According to Bunnell (2001), the first web-based online commercial activity for online auctions began in 1995, when two independent auction sites were established using different business concepts. Jerry Kaplan launched Onsale.com, the first online auction site, in May 1995. According to New York Time (2019), French-Iranian computer scientist Pierre Omidyar launched eBay in September of the same year. These two businesses both used ascending bid.Utilizing automated bids through electronic forms, employing a search engine for streamlined product identification, and allowing users to conveniently browse items by categories represent only a selection of the supplementary advantages facilitated by the internet. In distinction from on sale, where the company acted as the vendor, eBay pioneered person-to-person transactions in the realm of online auctions.

The earliest form of online auctions diverged from today's website-hosted ones. In the past, online auctions were conducted through email and bulletin boards. The auctioneer or seller would post a notification featuring the item available for bidding, inclusive of the minimum bid and closing time, on a system bulletin board. Presently, there are websites available for online auctions. Items like antiquated stuff, computers, automobiles, clothes, luxury items and machinery can be found on auction sites.

2.1.1 Types of Auction

There are different types of auctions that exist but the most significant ones are;

1. **English auction:** English auctions involve the announcement of bids by either the auctioneer or the participants themselves. The successful bidders then pay the winning bid amount to acquire the item. Among third-party online auction formats, English auctions are frequently favored due to their apparent simplicity. This format usually follows an open-auction model with ascending bids that are visible to the public. The

victor is determined as the highest bidder who also offers the highest price. While there exists the possibility of various forms of deceit, this format extends beyond the limitations of a traditional in-person English auction where bidders must be physically present.

2. **Dutch Auctions:** The Dutch auction is similar to reverse auctions but with a slight twist. Unlike reverse auctions, the Dutch type has only one seller selling items to the bidder. During the auction, bidders will place lower price bids, and a single seller will offer items that match the bid. This auction type has only one seller making the auctioning process a little easier. In particular, startups opt for Dutch auctions to sell their products and services.
3. **First-Price Sealed-Bid:** First-price sealed-bid auctions occur when all bidders submit a single bid, and the highest bid secures the auction, obliging the winner to pay the amount bid. The pivotal difference from English auctions is the absence of public revelation or showcasing of bids, which typically fosters a competitive atmosphere. From a game-theoretic viewpoint, the first-price sealed-bid auction holds a strategic equivalence to the Dutch auction, implying that participants will employ identical bidding strategies in both auction types.
4. **Vickrey Auction:** The foundational concept underlying a Vickrey auction, often called a second-price sealed-bid auction, closely resembles that of a first-price sealed bid auction. In this type of auction, the highest bidder, who ultimately emerges as the victor, is obligated to pay only the amount proposed by the second-highest bidder. Because the Vickrey auction compels participants to honestly reveal their true assessment of the item's value, it's recommended as a method to counteract the tendency for buyers to strategically manipulate their bids. If a bidder places an exceptionally high bid, they might be bound to honor it, while a bidder submitting a lower bid risks forfeiting the item. As a result, the individual who places the highest value on the item emerges as the winner. Both the seller and the winner reap benefits from this arrangement (Amit, 2017).
5. **Reverse auction:** In a reverse auction, there is a reversal of roles between the buyer and seller. Prices typically decrease as fresh offers are presented, with multiple sellers vying for the buyer's patronage. These auctions contrast with conventional ones as the buyer can assess all the offers and choose their preferred one. Businesses mostly use reverse auctions for procurement purposes. Unique bid auctions, often misconstrued with reverse auctions, resemble traditional auctions more closely, featuring a solitary seller and numerous buyers. While the pricing strategy is akin, each bid remains confidential, and the lowest individual bid consistently emerges as the victor (Jap, 2003).

6. **Bidding fee auction:** In a bidding fee auction, often known as a penny auction, customers are required to make payments for their bids. Each bid they place increments the auction price by one monetary unit. As customers can expend significant amounts of money without securing any items, certain individuals have drawn parallels between this auction format and gambling(Simone, 2012).

2.1.2 Dynamics of Online Auction

Auction is held when there are number of buyers available for a single item or number of sellers available for a single buyer- this is called a "Reverse auction". Online auctions shatter and eradicate the geographical, temporal, spatial, and limited audience constraints inherent in traditional auctions (Bapna, 2001). This heightened accessibility has concurrently facilitated the occurrence of illicit activities within auction contexts (Albert, 2002).

2.1.3 Features of Online Auctions

The electronic auctions are an important channel for selling and buying for many companies and individuals. The use of the internet as a basic infrastructure helps auctions to be realized with reduced transaction costs. The same contributes for increasing the profitability to those who buy and those who sell products and services. On an auction platform, there are three different types of users, similar to a straightforward marketplace model: service providers or vendors, buyers, and website administrators.

Figure 2.1 shows the screenshot of a typical model of an online auction site. As earlier stated one of the major infrastructure is the internet. Also the platform must have an appealing graphical user interface that can be used by anyone even a novice to internet technology.

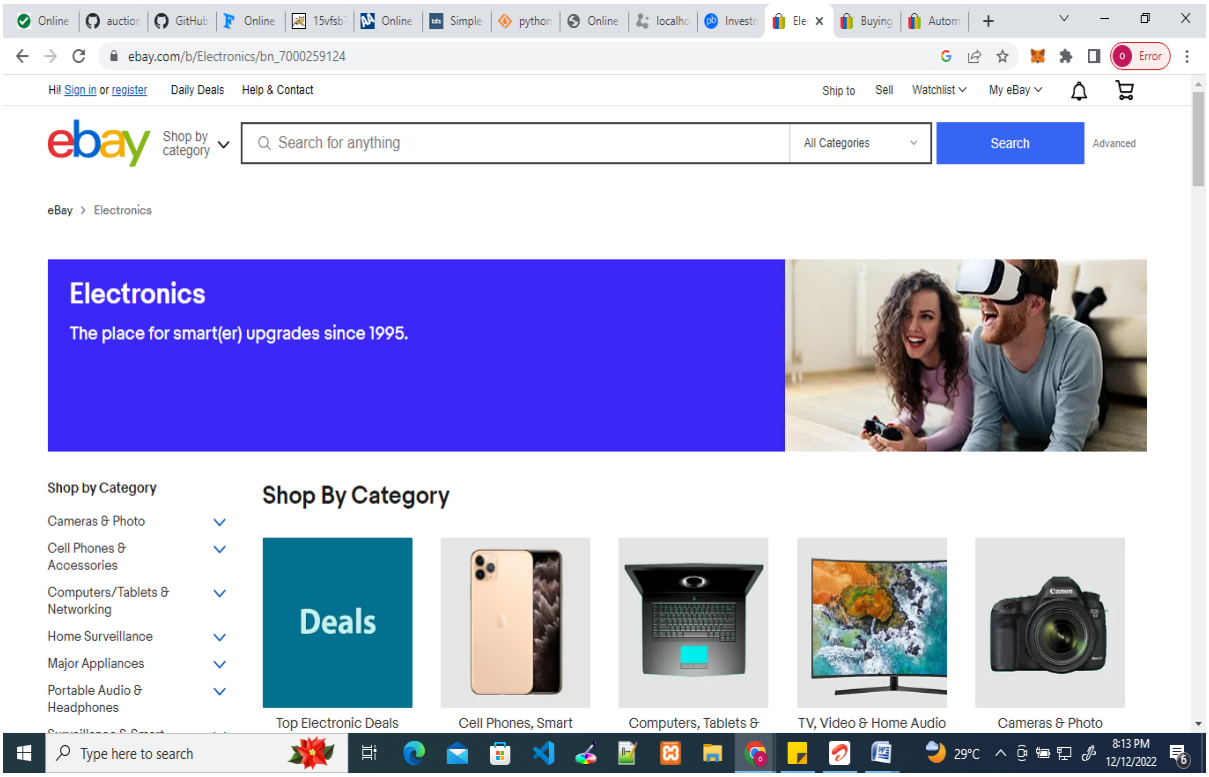


Figure 2.1: Model of s (Source: ebay).

2.1.4 Benefits of E-auction

There are many benefits of e-auction . These include:

1. Convenience: One can buy an item even at the comfort of your house or in the office, there is no need to go to a shop or store to participate in an auction.
2. Access of full information: The buyer can gain access to check all the details about an item on sale. One can even get to review the terms and conditions of payment and feedback rating of seller.
3. Time saving: It saves time in many ways like you don't need to go to any place for active participation in an auction.
4. Multiple choices: The bidder has multiple choices for items because from a single location one can participate in multiple item auctions and also have a choice to view same item or related items.
5. Wide geographical area: It covers large geographical area. It means you can take part in an auction event that the items seller is out of country.
6. Less paper work: Auction held online reduces paper work.
7. Less complexity: There is little or no complexity in business transaction.

2.2 Fraud

Auction fraud is a significant problem for this modern commercial medium, despite the fact that it is drawing more sellers and purchasers than ever before (Greg and Scott, 2008). As a result, both bidders and sellers may engage in auction fraud for personal gain. Wells, 2011 defined fraud as 'any crime which uses deception as its main modus operandi'. Fraud is also described as improper or criminal deception meant to produce financial or personal benefit in the Oxford Dictionary. It is useful to first categorize the many types of online auction fraud according to the three time periods in which the fraudulent behavior can occur: prior to, during, and after the auction.

As shown in Figure 2.2, we identify pre-auction fraud—misrepresentation of products, selling of black-market commodities, and triangulation—as such; post-auction fraud—non-delivery of goods and fee stacking—occurs after the auctions have ended. Investigation of such frauds relies more on real-world evidence than on online prevention and detection techniques because both pre-auction and post-auction frauds entail offline actions that are frequently simple for buyers and sellers to identify. However, in-auction fraud occurs while transactions are still being completed; as a result, it may take place without directly leaving behind physical proof and, what's worse, it can go unnoticed by the victims. Due to the difficulty of detecting in-auction fraud, it has received significantly less attention (Gavish and Tucci, 2008). In-auction fraud merits additional consideration and effort from mechanism designers and information technology experts in order to lessen the loss to victims and to protect online business participants. Due to the nature of internet applications such as a high degree of anonymity, incomplete legal constraints and lower barriers to entry and exit, it is difficult to combat in-auction fraud where one party of the transaction has more or better information than the other party. In an auction, sellers usually have better knowledge about the item than buyers; therefore shills who are associated with the sellers also have better knowledge than other buyers. The fraudulent participants intentionally make use of this information asymmetry to obtain benefit and deliberately hide their fraudulent or opportunistic behaviors so as to avoid being detected and caught. Since most of the fraudsters sign up on auction sites with fake identity information, when fraudulent behaviors are exposed, it is hard for investigators to find out the real identity under the meaningless net IDs. Afterwards the fraudsters could start over again using new net IDs so as to conceal their identities and reduce the chance of being punished.

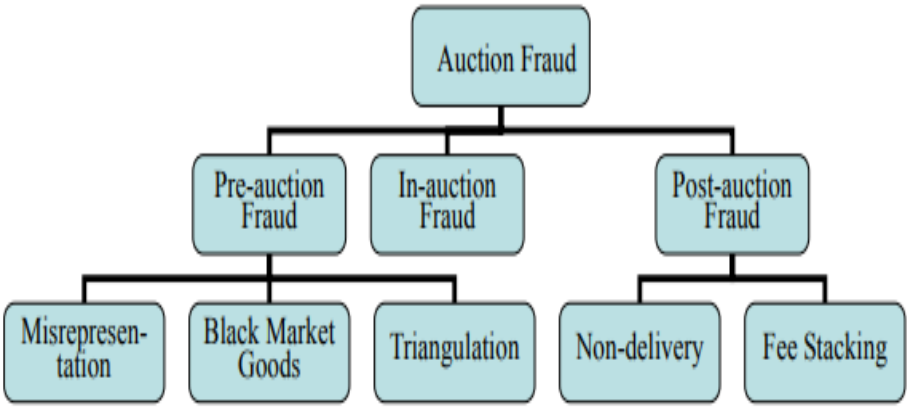


Figure 2.2: Categories on Online Auction Fraud *(Dong et al., 2019*)

This study is centered more on In-auction fraud which posed a devastating effect on auctioning system functionality.

Online auction platforms sometimes lack a robust dedication to proactively tackling the problem of auction fraud. (*Chua et al., 2018*) observed that these platforms grant fraudulent individuals the capability to manipulate the system, aiming to deceive buyers. This is attributed to the prevailing simplistic evaluation approach that centers on transaction feedbacks. (Chua and Wareham, 2004). Online auction regulations often establish predefined time frames for auctions, leading to extended durations for online auctions compared to their traditional counterparts. This stands as another distinguishing factor between online and traditional auction formats. As an illustration, eBay users can opt for auction periods of 1, 3, 5, 7, and 10 days, offering sellers the flexibility to attract a larger pool of bidders. However, these adjustable auction durations also grant deceitful buyers and sellers increased chances to engage in fraudulent activities. The longer the auction lasts, the more chances for fraudsters to make their fraudulent behaviors look normal. With enough time, cheaters could potentially behave like normal bidders with few clues left for fraud investigation.

2.2.1 In-Auction Fraud

When an auction is taking place, in-auction fraud happens. This kind of fraud could take place without any clear indication of direct physical involvement, and the victims might not even be aware of it(Majadi et al., 2017). Some examples of in-auction fraud are bid shading, false bidding, shill bidding, multiple bidding, and bidding rings. Shill bidding is one of the trickiest types of auction fraud to spot. This study mainly concentrated on shill bidding detection and prevention methods. Shill bidding occurs at public auctions where bidders are

permitted to place numerous bids in competition with one another. It could not happen in a sealed-bid auction due to the nature of concealed bids. Bid shading and false bidding are specific terms for cheating behaviors for first-price auctions and second-price auctions respectively. In-auction fraud can be categorized into sellers' fraud and buyers fraud (as seen in figure 3). False bidding, competitive shilling, reserve price shilling, buy-back shilling, and competitive shilling are all examples of seller fraud, while bid shading, multiple bids, and bidding rings are examples of buyer fraud (Dong et al., 2019).

The various classes of fraud are explained as follows:

1. **Competitive Shilling:** This is a bidding behavior that artificially drives up the bidding price of the auctioned item with no intention of actually buying. The aim is to make a legitimate winner pay more than this person would otherwise pay, so that the seller can gain more profit than expected.
2. **Reserve Price Shilling:** In (Kauffman and Wood, 2005) reserved price shilling is described as a bidding behavior motivated by the desire to avoid payment of the reserve price fee. By accepting the reserve price service, a seller is agreeing to pay the auction house a fee for the service. To minimize the payment of auction house fees but still reserve the item below a certain price, some high volume sellers will not set an "official" reserve price but instead engage shills to place bids on their auctions.
3. **Buy-back shilling:** This is defined as a bidding behavior employed by sellers or other shills who act as agents of the seller, when the legitimate bidders couldn't bid an acceptably high price. The seller or their shills would rather buy back the item and sell it again than sell the item immediately at a low price that does not reach their expectation. However, shills behave as a normal bidder in order to buy the item at a bargain price and complete the transaction. Such activity cheats the other bidders since they could not purchase an item at a bargain price.
4. **False bidding:** In an auction featuring sealed bids and a second-price outcome, every bidder submits a solitary bid, and the victor is required to pay the second-highest bid, rather than the first. An auctioneer can help a seller profitably cheat by examining the bids under the table after all buyers have submitted their bids. Knowing all bids, the auctioneer can submit an extra bid to make the second highest price very close to the current highest price such that the seller can gain more profit.

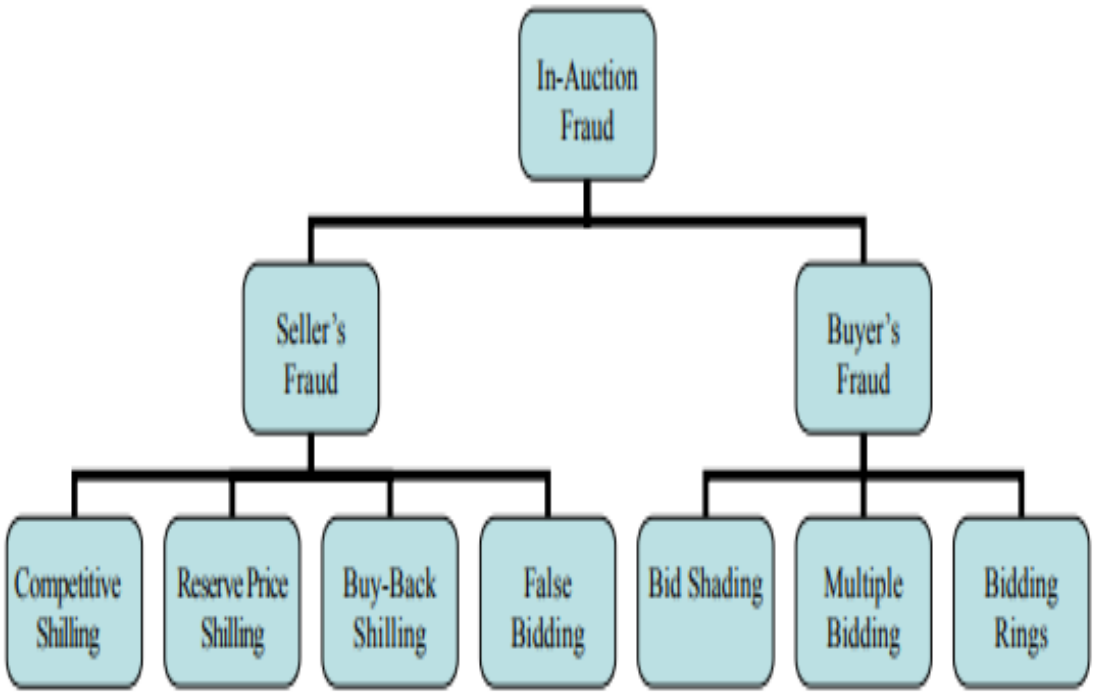


Figure 2.3 : Types of In-auction Fraud (Dong et al., 2019).

5. **Bid shading:** In first-price sealed-bid auctions, the victorious bid is the one used for payment. If a bidder had knowledge of the leading bid before it was officially revealed through unethical means, they could strategically place a bid slightly above that highest offer. As a result, the deceitful bidder would enhance their likelihood of winning while simultaneously reducing their payment obligation to the seller(Porter and Shoham, 2003).
6. **Multiple bidding:** This is also called bid shielding.In contrast to shill bidding, it involves dishonest action on the part of buyers rather than sellers. In order to put numerous bids on the same item, the buyers create a number of aliases and use them. The bidders deter other possible competitors by raising the price using several auction identities. They then withdraw all high bids, leaving the auction's lowest winning bid. The winning bidder ultimately receives the auctioned item at a substantially lesser price.
7. **Bidding Rings:** Bidding rings are collectively fraudulent practices that numerous bidders engage in during auctions. A group of dishonest bidders known as a "ring" forms, and the members of the ring agree not to bid against one another, either by

refraining from participating in the auction or by putting fictitious (phantom) bids. As a result, the winning bidder might purchase the auctioned item for a very low sum.

2.3 Shill Bidding Fraud

Shill bidding fraud stands out as the most widespread and intricate form of deceit due to its close resemblance to genuine bidding actions. Shill bidding entails the act of introducing counterfeit bids into an auction with the intention of elevating the price beyond the prevailing rate. As outlined in (Majade *et al*., 2017), shill bidding typically involves collaboration between the auction's seller and one or more bidders. Alternatively, a seller might establish multiple fictitious bidding accounts to place fraudulent bids. When the item holds a substantial value or possesses an uncertain market valuation, the buyer can incur significant financial losses (Anowar et al., 2018).Shill bidding exerts a detrimental and lasting impact on the online auction community, compelling sincere bidders to incur inflated costs for items. In accordance with Majade *et al.* (2017), shill bidding can be categorized into two forms: Reserve Price Shilling and Competitive Shilling. Reserve Price Shilling occurs when a seller sets a reserve price lower than the anticipated market value to minimize the listing fee. Subsequently, it employs shill bids to drive up the ultimate price of the auctioned item.

As a result, the seller can avoid paying higher reserve price fees to the online auction site. Competitive shilling is a bidding behavior that artificially boosts the bidding price of the auctioned item. This usually happens to make legitimate bidders pay extra money to win their auctions rather than paying a fair price. Comparatively, competitive shill bidding causes greater harm to the online auction sites than the reserve price shilling as the latter typically only affects the online auction sites; while the competitive shilling may affect all the normal bidders in an auction market.

2.4 Shill Bidding Fraud Detection and Prevention Mechanisms

Due to the ease and rapidity with which individuals can establish multiple accounts on online auction platforms, detecting and preventing shill bidding becomes a complex endeavor. Moreover, these platforms offer a high degree of anonymity, often requiring only an email address for account creation. Online auctions are especially susceptible to shill bidding due to the uncomplicated process of setting up accounts and the level of anonymity achievable. Determining whether an individual has participated in shill bidding can pose challenges. Online auction platforms claim vigilance in monitoring shill bids, yet the methods employed remain undisclosed. Consequently, the quandary of preventing shill bidders from engaging in such activities becomes even more perplexing. Approaches to detect and prevent shill bidding

can be classified into two groups: offline or static methods (where detection/prevention mechanisms come into play post-auction) and real-time or dynamic methods (where detection/prevention mechanisms are active during the auction).The fundamental distinction between these two varieties of methods for detecting and preventing shill bidding lies in their operation. An offline mechanism permits the occurrence of fraud and subsequently provides evidence of shill bidding after the fact. In contrast, real-time systems possess the capability to enforce penalties during an ongoing auction if shill bidding is identified.

2.4.1 Agent-based Trust Management Approach

This can provide a solution for detecting and preventing shill bidding auction fraud, applicable both in offline scenarios and in real-time. Traditionally, the foundation of the online auction model rests on trust among its participants. Specifically, participants are presumed to adhere to the rules of the auction with honesty. However, the possibility for certain participants to engage in fraudulent activities underscores the inadequacy and inefficacy of this trust-based approach. Ford et al. in 2012 introduced an innovative model-based methodology aimed at delineating intricate and adaptable bidding strategies for bidding agents. The authors introduced a formal specification for these versatile and complex strategies, labeled as a Layered Bidding Strategy Model (LBSM). This LBSM takes the form of a visual strategy model, internally stored as an XML file. Once a strategy is identified, the LBSM is transformed into a Rule-based Bidding Strategy Model (RBSM), constituted by a collection of production rules. Additionally, the authors constructed a prototype agent-based online auction system using the Java Agent Development (JADE) framework and elucidated how the execution of specifications for flexible and intricate bidding strategies can be effectively achieved.

Similarly, Patel *et al*., 2018 introduced shill monitoring for agent-based online auctions using rule-based access control mechanisms. The study proposed a trust management module where a user could configure an agent for initiating an auction as a seller or for submitting bids on an auctioned item as a bidder. User identifications are sent by the user-configured agent to the trust management module for security reasons. The agent can make requests for actions connected to auctions once the authentication and authorization procedures are complete. Throughout the auction process, a configured agent had the ability to check the status or ratings of other configured agents. It made it possible to choose the best auction through wise decision-making. At the same time, a security agent monitors each auction transaction to look for signs of shill bidding and responds appropriately. Additionally, a

history module was employed to keep track of users' past auction activity over a predetermined time period, including access details, bidding patterns, and feedback ratings.After a transaction is successfully completed, the user-configured agent updates all the data in the history module. The history module was used by the security agent and the trust management module to make decisions. However, the effectiveness of their suggested security agent depends on user feedback, which may be easily falsified through the use of false identities and group shill bids. Additionally, their suggested trust management architecture relies solely on security agents to identify shill bidders, which becomes inefficient when the number of participants grows significantly.

2.4.2 Data Mining Mechanism

Data mining or knowledge discovery is a powerful computer-assisted process designed to analyze and remove relevant information from historical data (Han and Kamber, 2001). It allows users to analyze data from different dimensions or perspectives in order to uncover consistent patterns, anomalies and systematic correlations between data elements. The ultimate goal of data mining is to predict future behaviors and trends based on the discovered patterns and association rules. Several researchers have adopted data mining methods to detect shill associations and suspicious patterns. Consequently, Shah et al., 2003 applied data mining techniques in detecting shill behaviors in eBay video game console auctions. They analyzed the relationships between buyers and sellers over time and discovered that some users would only participate in auctions that were hosted by a specific seller and would seldom win, such as once or twice in all of the transactions in which they were involved. They speculated that these might be instances of shill bidding. As with reputation procedures, data mining techniques take a very long time to complete and require the analysis of enormous amounts of historical data. The ultimate goal of data mining is to forecast future behaviors and trends using patterns and association rules that have been identified. The majority of researchers have started using data mining techniques to look for suspicious trends and shady links (Pandit *et al*., 2007).

2.4.3 Formal Methods

Model checking is one of the many automated formal methods that are used to check the correctness of a system. Specifications about a system are expressed in the form of logic formulas, and efficient algorithms are used to verify certain properties of a system by means of an exhaustive search of all possible states that it could enter during its execution. For the specification, creation, and verification of software and hardware designs, formal approaches

employ mathematical tools like formal notations and logic. Several researchers utilized this technique to detect in-auction fraud. Xu, et al., 2007 designed a formal model checking approach which was used to detect shilling behaviors, especially the competitive shilling behaviors. The authors first derive a formal auction model from real world auction data according to a predefined auction model template.

2.4.4 Detection Techniques Based on Reputation Systems

Feedback based reputation system is a one of the simplest fraud detection systems. Reputation systems help honest bidders to decide whether to purchase a product from a particular seller based on the feedback score or not. After the completion of auction, both the seller and the buyer can put down ratings and feedback comments on the other party. These comments and feedback build up in the trader's transaction history, of which the feedback score is one part. So this type of reputation system is simple and easy to understand. However, the existing reputation system for auction houses has already shown its weakness in providing trustworthy information. This detection technique suffered some major drawbacks; firstly, the positive feedback is overwhelming but the negative feedback is deflated. Regardless of the actual behaviors of dishonest auction users, they usually take advantage from the weakness of current rating mechanisms in reputation systems by helping each other to build up a good reputation history artificially.

Secondly, existing reputation systems can be easily manipulated whereby fraudsters could gather a high feedback score by selling low value goods, and then deal high value goods by using that good reputation score. Finally, there is a little information available about a seller's degree of honesty in the existing reputation system. However, it is not really practicable to browse through the feedback page by page when trading with a seller that has a long history. Honest auction consumers may find information about auction fraud in the feedback.

Therefore, the current reputation systems are not able to satisfy peoples' need when evaluating trustworthiness in online auctions. A reputation method was given by Trevathan and Read (2018) to assess the level of shill bidding in online auctions. For each bidder, the authors created a score dubbed the "Shill Score" to indicate the likelihood of them being a possible shill bidder after considering a number of auctions for identifying bidding trends for a specific user. The six ideal bidding actions of a shill bidder are used to determine the shill score, which includes

(i) Display a substantial number of bids;

(ii) Exhibit a limited count of auction wins;

(iii) Place bids shortly following a legitimate bidder to extend their response window;

(iv) Employ marginal bid increments; and

(v) Initiate bids early in an auction to stimulate activity and cease bidding toward the auction's conclusion to decrease their odds of winning (Trevathan and Read, 2018). However, if users were equipped with more proactive measures, like an automated fraud detector, the online auction process could become more secure. Auction houses employ the strategy of rating after transactions conclude, allowing traders to receive ratings on a scale encompassing positive, neutral, and negative feedback.

Several methodologies have been proposed by researchers for identifying shill bidders in online auctions. Mamun and Sadaoui (2019) introduced a three-tier software architecture named the Shill Free1 auction system, designed to safeguard auction systems from shill bidding in both forward and reverse auctions. Similarly, Sadaoui et al. (2015) proposed a comprehensive framework capable of real-time monitoring for multiple ongoing auctions. This framework vigilantly observes the progression of auctions to swiftly take corrective measures and thwart the success of shill bidders. Monitoring activities are carried out at various intervals depending on the duration of the auction, whether measured in hours or days.

2.4.5 Social Network Analysis (SNA).

This describes a set of research procedures for identifying structures in systems based on the relations among actors. Grounded in graph and system theories, this approach has proven to be a powerful tool for studying networks in physical and social worlds, including on the web. In the recent years social network approach has been increasingly applied in fraud detection in internet. With the development of web technologies, there is an increasingly greater amount of interaction by people while on the internet. Social Networks bring in a multi-disciplinary approach to solve problems in this domain. Social Network Analysis focuses on using quantitative measurement to study the contact among the members to profile the structure of the neighborhood and its members. It has been developed and practiced in the domain of social studies for decades. Recently, SNA are employed by social scientists to investigate the dynamics of a social group.

2.4.6 Triangulation Detection System

In this system, verification is submitted for incoming transaction. The system receives the details of the card and the purchase value to verify whether the transaction is genuine or

otherwise. The types of goods that are bought into that transaction are not known to the system. It tries to find any irregularity based on the transactions pending profile of the cardholder. If the system confirms the transaction to be malevolent, it raises an alarm and the issuing bank to cancel the transaction. Here mainly used four steps; information building, comparison, prediction and detection. Next step is compared present user behavior and already proven fraudulent behavior. And then forecast the result using both behaviors. Finally, using the above steps the system can tell whether the user is legitimate or fraudulent (Thiagu *et al*., 2014).

2.5 Binary Classification

The most frequently used models for binary classification include;

2.5.1 Support Vector Machine (SVM)

Support Vector Machine (SVM) was invented in 1963 as a supervised ML technique for analyzing data related to classification and regression analysis. It works as a non-probabilistic binary classifier which helps in classifying instances into one of the two classes. SVM was primarily designed to address linearly separable datasets. But later with the introduction of kernel functions, non-linear classification has been made possible. The method has been improved further in several ways to satisfy different other problems (as seen in fig 2.5.1) in the domain of data classification and regression. These problems are clustering (unsupervised learning), multi-class classification, transductive or semi-supervised learning and incremental learning.

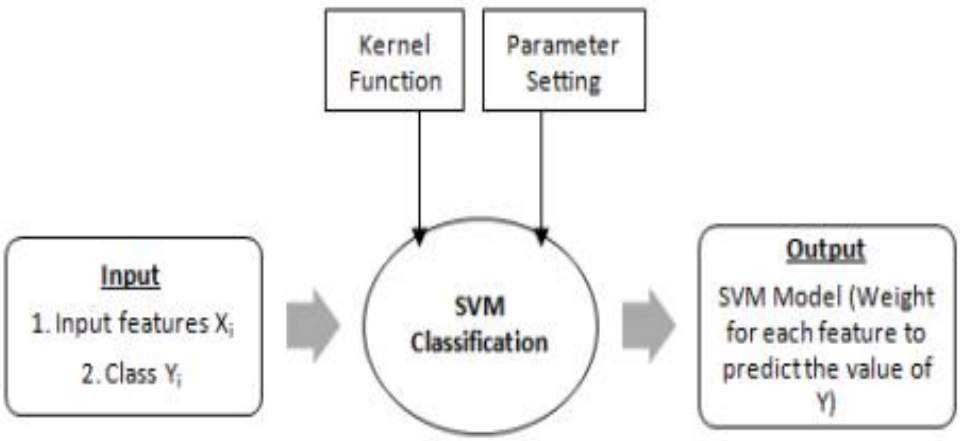


Figure 2.4: SVM Classification Model (Ganguly, 2016)

2.5.2 Decision Trees

A recursive partition of the instance space is how a decision tree expresses itself as a classifier. The nodes that make up the decision tree are known as a rooted tree, which is a directed tree containing a node named "root" that has no outgoing edges. Each of the other nodes has just one incoming edge. An internal or test node is a node having outgoing edges. Leaves (also known as terminal or decision nodes) are all other nodes. Each internal node in a decision tree divides the instance space into two or more sub-spaces in accordance with a particular discrete function of the values of the input attributes. Each test considers a single attribute in the simplest and most typical scenario, dividing the instance space according to the attribute's value. The condition is a range in the case of numeric attributes. Each leaf is given a class that corresponds to the best target value. As an alternative, the leaf might store a probability vector that shows the likelihood that the target characteristic will have a specific value. According to the results of the tests along the path, instances are categorized by moving them from the tree's root all the way down to a leaf. A decision tree in Figure 2.5.2 explains how to determine whether or not a prospective consumer will reply to a direct mailing. Circles are used to symbolize internal nodes, while triangles are used to indicate leaves. The decision tree in this example includes both nominal and numerical attributes. Given this classifier, the analyst may comprehend the behavioral traits of the entire population of possible customers for direct mailing, as well as forecast the response of a potential client (by sorting it down the tree). Each node is given a name for the attribute it is testing, and each branch is given a label for the attribute's corresponding value.

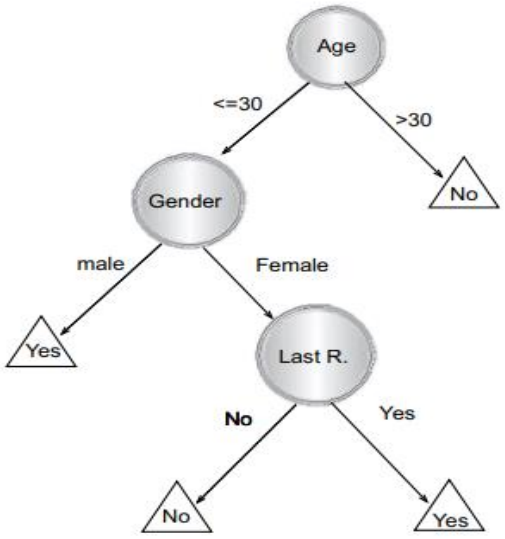


Figure2.5: Decision Tree

Decision trees can be geometrically seen as a collection of hyper planes, each orthogonal to one of the axes, in the case of numerical characteristics. Naturally, decision-makers like simpler decision trees since they may be regarded as being easier to understand. Additionally, according to Breiman et al. (1984), the correctness of the tree is significantly impacted by its complexity. The stopping criterion and pruning technique are directly in charge of the tree complexity. Typically, one of the following metrics—total nodes, total leaves, tree depth, or number of characteristics used—is used to assess the complexity of a tree.

2.5.3 The K-Nearest-Neighbours

The K-Nearest-Neighbours (KNN) is a non-parametric classification algorithm, i.e, it does not make any presumptions on the elementary dataset. It is known for it's simplicity and effectiveness. It is a supervised learning algorithm. A labeled training dataset is provided where the data points are categorized into various classes , so that class of the unlabeled data can be predicted. In Classification, different characteristics determine the class to which the unlabeled data belongs. KNN is mostly used as a classifier. It is used to classify data based on closest or neighboring training examples in a given region. This method is used for its simplicity of execution and low computation time. For continuous data, it uses the Euclidean distance to calculate its nearest neighbors. For a new input the K nearest neighbors are calculated and the majority among the neighboring data decides the classification for the new input. Despite the fact that this classifier is straightforward, the value of 'K' is crucial for identifying the unlabeled data. There are numerous ways to choose the values for "K," but we can just run the classifier with various values several times to determine which value produces the best results. Due to the fact that all calculations are done when classifying the training data rather than when it appears in the dataset, the computation cost is a little higher. Because it only stores the training data and memorizes the dataset when the dataset is being trained, it is a lazy learning method.On the training dataset, generalization is not done. Therefore, when in the testing stage, the whole training fundamental dataset is needed. KNN forecasts continuous values in regression. The average of this value's K closest neighbors' values makes up this value.

2.5.4 A Naive Bayes classifier

A Naive Bayes classifier is a straightforward probabilistic classifier built using the Bayes theorem and naive independence assumptions from Bayesian statistics. The phrase "independent feature model" might be a better way to describe the underlying probability model. A naïve Bayes classifier, to put it simply, thinks that the presence (or lack) of one

feature in a class has nothing to do with the presence (or absence) of any other feature. A fruit might be categorized as an apple, for instance, if it is red, rounded, and around 4" in diameter. A naive Bayes classifier regards each of these characteristics as individually adding to the likelihood that this fruit is an apple, even though they all depend on one another or the existence of additional variables. In a supervised learning environment, naive Bayes classifiers can be trained quite effectively depending on the specifics of the probability model. It is possible to operate with the naive Bayes model without accepting Bayesian probability or applying any Bayesian techniques because parameter estimation for naive Bayes models frequently employs the maximum likelihood method. naïve Bayes classifiers have performed admirably in a wide variety of challenging real-world circumstances, despite their naïve design and ostensibly oversimplified assumptions.

2.5.5 Neural Network

A neural network, an artificial intelligence method, guides computers to analyze data in a way resembling the functioning of the human brain. Deep Learning, a subset of machine learning, harnesses interconnected neurons or nodes arranged in layered structures to emulate the human brain's operation. This methodology cultivates an adaptable system that computers employ to iteratively learn from mistakes and enhance their performance. As a result, artificial neural networks strive to tackle intricate challenges, like document summarization or facial recognition, with increased accuracy. Computers can make intelligent decisions with minimal human intervention thanks to neural networks. This is why they can learn and model complex, nonlinear relationships between input and output data. Numerous sectors and use cases for neural networks include the following:

- Performing medical diagnosis through the classification of medical images
- Implementing targeted marketing using social network filtering and analysis of behavioral data
- Conducting financial predictions via the analysis of historical data from financial instruments
- Forecasting electrical load and energy demand
- Overseeing process and quality control
- Identifying chemical compounds

A fundamental neural network comprises interconnected synthetic neurons distributed across three layers:

Input Layer: The input layer serves as the entry point for external data into the neural network. Input nodes process, evaluate, or categorize the information and subsequently transmit it to the subsequent layer.

Hidden Layer: Hidden layers receive input from the input layer or other hidden layers. Neural networks can encompass multiple hidden layers. Each hidden layer assesses the output from the previous layer, processes it further, and then transfers it to the next layer.

Output Layer: The output layer furnishes the ultimate outcome of all data processing accomplished by the neural network. It can consist of either a single node or multiple nodes. For instance, in a binary (yes/no) classification scenario, the output layer would comprise a single output node, yielding a result of 1 or 0. Conversely, in a multi-class classification scenario, the output layer might incorporate more than one output node.

2.5.6 Types of Neural Networks

The way that data moves from the input node to the output node distinguishes different artificial neural networks. Here are some illustrations:

i. **Feedforward neural networks:** Data is processed in one direction, from the input node to the output node, by feedforward neural networks. Each node in one layer is linked to each node in the layer above it. To make better predictions over time, a feedforward network employs a feedback process.\

ii. Backpropagation algorithm: Artificial neural networks use corrective feedback loops to continuously learn and enhance their predictive modeling. You can think of the data flowing through the neural network's numerous channels as it moves from the input node to the output node. In order to translate the input node to the right output node, only one path can be considered proper. The neural network employs a feedback loop that functions as follows to identify this path:

1. Every node in the path makes a prediction about the subsequent node.
2. It verifies that the guess was accurate. Node pathways that result in more accurate guesses are given greater weight values, whereas those that result in inaccurate guesses are given lower weight values.
3. The nodes repeat Step 1 after making a new forecast for the following data point using the higher weight pathways.

iii Convolutional neural networks: In convolutional neural networks, the hidden layers carry out convolutions, which are specialized mathematical operations like filtering or summarizing. They can extract pertinent elements from images that are helpful for image recognition and classification, making them particularly valuable for image classification.

Without sacrificing elements that are essential for producing an accurate prediction, the new form is simpler to process. Edges, color, and depth are just a few of the various image properties that each buried layer extracts and analyses.

Learning Neural Networks

The process of instructing a neural network to carry out a task is known as neural network training. In the beginning, neural networks process numerous sizable sets of tagged or unlabeled data. They can then more precisely process unknown inputs by leveraging these examples.

2.6 Supervised Learning

In supervised learning, data scientists submit labeled datasets to artificial neural networks that already know the solution. These datasets are used by the neural network to steadily accumulate knowledge. After the network has been taught, it begins speculating about the racial or emotional makeup of fresh results that it has never previously analyzed.

2.7 Related Works

We highlighted further studies in this area that used machine learning techniques to detect fraud in e-auction systems. Based on the problems covered by each survey in Table 2.1, we compiled these linked surveys into a handy reference.

In their work, Wurmann et al. (2017) presented a software architecture for online English auctions, catering to both software and human agents. Their suggested auction server, known as the Michigan Internet Auction Bot, enabled the adaptable configuration of auctions, accommodating diverse parameters. This allowed researchers working with agents to investigate various auction mechanisms. However, the authors did not elaborate on the development process of their auction system.

Jurjees *et al.*, (2018) developed a Web-based android application for auction based e-commerce system. The authors provided a simple and elegant design for an online auction system based on UML. They presented an analysis and design for the auction system illustrating key system components using UML diagrams. Their developed online service platform, called Deal Daddy was used for selling and buying their products (brand new products). The platform provided an account for every user where they could customize their products. It worked on the basis of product auction (Bidding). Every product was assigned a base price for bidding. Then the individual users of the Deal Daddy had to bid for the product at their own price. The user with the highest bid price bought the product. The above prescribed bidding mechanism applied to both old and new products (also for antiques). Deal

Daddy has a facility for categorizing for user and to categorize the product needs. Every user had to register or have to create their own account on Deal Daddy to begin their transaction. Deal Daddy assigned a unique user code to every user. Every product transaction took place based on the user id. The authors did a good job but failed to categorize the various auction frauds using their auction platform.

Madhuri *et al*., (2020) proposed an analysis, design and implementation of a secure online auction system using detaining techniques. This system made consumers more effective and efficient in their behavior and had driven businesses to a new level, forcing many to make the necessary adjustments and changes to reach the new market of knowledgeable consumers. Rapid growth of has resulted in an e-transformation in the global retail infrastructure. The authors did a good job but failed to implement the discussed issue to an architectural model for further clarification.

Didwania and Chattaraj (2015) devised an enhanced system, driven by considerations surrounding the design of websites. This led to the creation and implementation of architectural software. The architecture enabled bidding on available commodities at any time through the site. The proposed system demonstrated the ability to manage a wide array of products effectively. Additionally, time management emerged as a crucial concern within such an online auction system, which the proposed design meticulously addressed. Moreover, the system was constructed with the goal of facilitating necessary adaptations in existing setups and ensuring future maintenance without substantial rework. Despite introducing a versatile and resilient system, equipped to concurrently manage diverse items and navigate the complications of time for both bidders and auctioneers, the enhancement of security by incorporating the option of Blacklisting defaulting bidders was not achieved.

Deorukhakar *et al* (2014) described the indicators of in-auction fraud, and pointed out that because no single indicator was accurate enough to assure the presence of in-auction fraud, a combinatorial way using multiple indicators would be more effective and precise. Increasing use of internet, online shopping and online auction have gained more importance and at the same time the criminals are also taking benefits to perform undesired activities to gain illegal profit. Therefore, to uncover and thwart such illicit and fraudulent undertakings, fraud-detection moderation systems are routinely employed. Additionally, the authors put forward an online probity model framework that amalgamates online feature selection, coefficient constraints derived from human insights, and the incorporation of multiple instances learning.

Upon scrutiny of actual online auction fraud detection data, the proposed model demonstrated an enhanced capacity to identify possible fraud cases and notably curbed customer grievances, surpassing the performance of various baseline models and the human-crafted rule-based system. Nevertheless, the authors did not execute the outlined matter into an architectural model.

Lin and Khomnotai (2017) presented online auction fraud detection in privacy-aware reputation systems. The study explained two crucial factors essential to the success of an online auction website. Firstly, was how easily buyers could find sellers and secondly, was the trust in the website facilitates through its reputation system. The reputation systems in online auctions were essentially recommendation systems. The study proposed two privacy-related attributes to quantify the proportion of anonymous ratings that a user received. Additionally, the research indicated that both characteristics enhanced the effectiveness of the fraudster detection approach. The reputation system utilized a more intricate technique for computing the reputation score, such as allocating reduced and elevated weights to anonymous and non-anonymous evaluations, correspondingly. However, while the reputation score was accessible to all users to assess a buyer's reliability in real time, its influence was significant. Notably, the study omitted the explanation of how the reputation score was computed, which was a deliberate measure to prevent an inflated reputation.

In their work, Tyagi and Sharma (2021) introduced an application of machine learning techniques for identifying fraudulent sellers in online marketplaces. The authors highlighted the consistent growth of e-commerce's share in global retail expenditure, reflecting a clear transition from physical stores to online shopping. Their intention was to address the specific challenge of creating an interactive model to recognize auction fraud in the realm of e-commerce. This need is particularly relevant as the most prominent businesses in this domain involve retailers and auction participants. As customers increasingly engage in online transactions, concerns about fraudulent activities arise, jeopardizing the integrity of honest parties. Hence, implementing proactive fraud detection systems becomes essential to counter such illicit behavior. While the authors made commendable efforts, they did not execute the discussed concept into an architectural model that could further elucidate the operational process of detecting fraudulent sellers within the online auction system.

Sekar *et al*., (2019) presented a survey on online auction using data mining techniques. The authors described fraud as a critical factor in auction systems. Various data mining techniques such as ID3, C4.5,C5.0, CART, Neural networks etc. were studied and analyzed. Buyers and

sellers alike benefitted from the great opportunities that online auctions provided, but these auctions also provided criminals the opportunity to perpetrate fraud. Scam and fraudulent bidders abound in the online auction market, and the losses brought on by these unethical practices are significant. By utilizing data mining and machine learning techniques, the issue of shill bidding and fraudulence was resolved. Additionally, this study supported the various auction kinds and data mining and machine learning techniques that may be utilized to stop the counterfeiting that is common in online auctions.

Benjamin et al. (2013) introduced an incremental neural network technique within the realm of data mining to detect questionable bidders in online auctions. In this methodology, the neural network utilizes a training dataset comprising historical auction data, continuously updated with incoming data. This approach discerns bidder behavior through the dataset, aiding in the identification of suspicious bidders and enhancing result precision. Nevertheless, the authors adopted a practice of reinitializing the network with each influx of new data, rendering the process resource-intensive and time-consuming.

Dong *et al*., (2009) proposed two main techniques; Model checking technique and Dempster-Shafer theory. The model checking technique was used to detect suspicious shill bidders in auction based on bidding pattern-based temporal formulas. The Dempster-Shafer theory justified the shill suspects by collecting the large amount of evidences about bidders and it was combined with new evidences. Based on the evidences it will give shill score to each bidder which will classifies the trusted and suspect bidder. Using Dempster-Shafer theory gives more accurate result but using Bayesian inference method it improves the system performance.

Chang *et al*., (2010) proposed Clustering technique for discovering the fake transaction records in auction. The X-means (extended version of K means) clustering technique analyzes the behavioral changes of the fraudsters and identifies them in the earlier phase. The authors also used C4.5 algorithm which was applied for explaining the rules of the labeled fraudulent behaviors. These techniques identify fraudulent behavior in earlier phase and along with that apply different instance-based learning algorithms to refine the behavioral changes of fraudsters.

Ponhu and Kummari (2018) used various machine algorithms such as Artificial Neural Networks (ANN) and Support Vector Machines (SVM). They conducted experiments with the datasets and noticed the accuracies between them. Then used pandas and various packages to load datasets and performed the numerical calculations. Both the ANN and SVM

gave better results when compared to the other machine learning algorithms. He proposed the Parameter tuning (Grid Search technique) which is used to find the best quality parameters. The Grid Search test several combinations of quality parameters and returned the best selection that gives best accuracy. The proposed system was applied to both the ANN and SVM algorithms but the ANN has more accuracy than the SVM.

Qadeer *et al*., (2009) proposed a system which used dynamic packet filtering and buffering to enable the effective bulk recording of large traffic streams. Then the congestion is encountered by the system. Moreover, the authors discussed the concept of Very Fast Decision Tree Learner(VFDT) used for Data mining based on Hoeffding Algorithm which could handle data to tune millions of packets implemented the Bitmap indexing in data mining with Fastbit, which is the first robust end-to-end system. But it takes lesser time to mine data in lesser time than it takes to input them from disk. And also it does not store data in the main memory and hence storage space is required only for that tree and associated statistics. Kumar (2018) used several machine learning algorithms for Stock market predictions. The different data mining models were developed and their performances were compared. He proposed the simple architecture, which involved feature extraction from the given dataset, supervised classification of training dataset and test dataset, and result evaluation. For the comparative study of Supervised learning algorithms, he calculated the accuracy and F-measure. By the comparison of accuracies of different algorithms he concluded that Naive Bayes performs best for small datasets and Random Forest performs best for large datasets. He concluded that reduction in the technical indicators reduces the accuracy of each algorithm.

Aishwarya (2020) presented Binary classification model for fraudulent credit card transactions. Unusually, large transactions or the ones that happened in a typical locations evidently deserve additional verification which could be detected by looking at on-surface and evident signals. The authors employed a fraud detection system which not only detects fraud but further makes it essential to model past credit card transactions with the ones that are typical or turned out to be fraud. The study aimed to identify the typical patterns found in transactions by employing a supervised learning approach and train a binary classification model that can identify the features of these transactions and sort the data into one of two transaction classes: fraudulent or valid, based on provided historical data.

Gadade *et al*., (2014) built online models for the auction fraud moderation and detection system. The online framework was extended to other applications like web spam detection,

content optimization and so forth. The study provided the responsibility of selling the trustful products by the website itself managed by the admin. When a customer wishes to buy a product, he will get an idea about the product to how much extent he can believe in that product. If he has faced any problem he can make others aware of that product by complaining about the product. The proposed system could detect sellers selling fraudulent products to an extent but failed to include the adjustment of the selection bias in the online model training process and to deploy the online models described in the study to the real production system, and other applications.

Thiagu *et al*., (2014) presented an effective method for identifying online auction fraud using triangulation detection techniques. With the reliable increase in Internet auction fraud tiding early and the increasing number of internet users and Internet auction participants, it is impatiently apparent that fraudulent activities are not going to disappear anytime soon and demands the attention of all law enforcement and internet auction house officials. The study described three methods to detect online auction frauds it has own advantages and disadvantages, so that the public and law enforcement can be made more aware and informed about it. Continual research needs to be conducted on the crime, to raise awareness of the current trends, latest statistics, and the tools that are being developed to prevent the crime.

Deep learning (DL) models can be helpful in accurately identifying fraudulent transactions, according to Mubalaike and Adali's (2018) research. A mobile money service provider in Africa's real financial records from one month were used to create the dataset, which had information on more than six million transactions. On the preprocessed data, the most effective deep learning and ensemble of decision tree (EDT) classifiers, stacked auto-encoders (SAE), and restricted Boltzmann machines (RBM) classifiers were used. Based on metrics including accuracy, sensitivity, specificity, precision, confusion matrix, and ROC values, the created classifier models' performance was assessed. 90.49%, 80.52%, and 91.53% were the optimal accuracy results, respectively. The results of the comparisons clearly showed that the constrained Boltzmann machine outperforms the other methods.

Carta *et al*., (2019) proposed new data intelligence technology-based in Multi Consensus Prudential model which in accordance with a double criterion, probabilistic and majority-based effectiveness of various state-of-art classification algorithms. The aim was to optimize the model's efficacy in the identification of fraudulent transactions irrespective of some data mismatch. The study model was validated with observations in a broad real-world dataset,

which indicates that the model proposed outperforms other state-of-art solutions, both in terms of ensemble models or classification approaches.

Chang and Chang (2012) introduced an innovative approach to early detection that incorporates both accuracy and timeliness considerations. They developed a modified wrapper technique aimed at selecting a subset of attributes from an extensive array of potential attributes. This selection process was designed to effectively distinguish between legitimate traders and fraudulent entities. The chosen attributes would serve as a supplementary modeling tool, facilitating the retrieval of the latest trader history segment. This in turn streamlines the process of data collection and model development, minimizing the demands on time and resources. The creation of decision gems or instance-based learning represents an initial model for fraud detection. The study's outcomes demonstrated that the selected attributes outperformed other attribute sets, with the integration of hybrid complement model phases significantly elevating the accuracy of fraud detection.

Kim *et al*., (2019) examined deep learning ability to help financial risk management decisions. In order to predict how separate distribution dealers protect future profits, they developed a deep learning model. This involved traditional problems of risk and behavior. Conventional machine learning includes data that are a representative of functionality-objective relationship and that rely on often costly development, maintenance & analysis of handicraft functionalities. Hence it is a difficult task to model highly variable, large and diverse models such as trader conduct. Profound thinking was adopted as a solution and the knowledge of hierarchically distributed data representations automatically (e.g. behavior risks). It revealed generative features which defined the target (e.g. productivity of traders), Avoided but was more robust to changes (e.g. changing market conditions) mechanical functional engineering. The outcomes of a comprehensive risk operating network demonstrated the capabilities of deep learning for apps, provide guidance in the design of the correct network architecture and demonstrates the dominance of deep learners over machine learning or expectations based on laws

A study to identify financial fraud using deep learning techniques was presented by Singh and Jat (2020). The number of fraudulent transactions is steadily increasing as society transitions further into a digital economy. The study placed more attention on systems that use in-depth learning and prompt analysis of already-in-use techniques for payment fraud detection. The survey was conducted with the intention of routinely comparing techniques for identifying fraud in online transaction volumes for industry. The outcomes of the study's tests

and results showed that various techniques for spotting fraud perform realistically in the market. It is possible to identify the fundamental challenges in implementing a thorough understanding of fraud.

Vegunta and Minni (2013) introduced a Bayesian Probit online model framework designed for the detection of auction fraud. In the context of Internet auction platforms, auction fraud predominantly encompasses instances where fraud stems from either the non-delivery of purchased items or the distortion of product descriptions in sales advertisements. Dishonest sellers might present non-existent items for bidding with misleading descriptions to deceive buyers regarding the actual value. They then request direct payments to themselves. Likewise, malicious buyers could engage in fraudulent transactions using compromised credit card details that don't match the shipping address. The ramifications of online auction fraud extend to both consumers and merchants, in addition to the commercial auction websites themselves.

The study delved into the intricacies of constructing models for online auction fraud detection systems, which inherently undergo dynamic evolution over time. The research highlighted that the model framework was grounded in real-world online auction fraud detection data, integrating bounded coefficients derived from expert knowledge, online feature selection, and numerous instances of learning. This comprehensive approach exhibits substantial improvements over baseline models and human-tuned approaches. However, the study fell short of translating the discussed concept into an architectural model.Read (2009) designed an algorithm to detect collusive shill bidding where multiple shill bidders shill in a group. But the two problems of duplicate identity shill bidding and consecutive multiple bidding are not addressed. Moreover, the algorithm does not work in live auctions.

Chau *et al*., (2009) applied data mining and trust propagation techniques to detect fraudulent users in online auction systems. Generally, these techniques suffer from two drawbacks. Data mining related approaches deal with a large amount of historical data; thus they may have limited value in detecting shill bidding in a time-efficient manner. Pattern matching based and model-checking based approaches do not regularly update prior knowledge with the presence of new evidence. Therefore, they frequently generate false positive results.

Goel (2010) used a multi-state Bayesian network to verify detected shill suspects. Similar to the D–S theory based approach, bayesian networks are capable of reasoning under uncertainty and can be used to calculate the probability of a bidder being a shill. The author

did a good job but the technique suffers from being time consuming in their investigation of bidders.

Kavu et al. (2015) introduced a tool for the detection of fraud. The surge in online auction usage has provided fraudsters an opportunity to exploit these platforms for their advantage, engaging in tactics like shilling, where they inflate prices through their own auctions. Consequently, legitimate bidders find themselves obligated to pay more than their original intended offer. This situation necessitated the development of a shill detection algorithm. Addressing this shilling concern, the research devised a shill detection algorithm seamlessly integrated with an online auction system. The algorithm's effectiveness was demonstrated, and it exhibited real-time capability suitable for platforms with large user bases. While the study succeeded in rectifying the shilling issue, it omitted the incorporation of certain classification mechanisms such as user-friendliness, real-time processing, and adaptability into the proposed algorithm. These enhancements could have bolstered the accuracy of classifying fraudulent transactions.

In their study on fraud in planning and bidding for the purchase of goods and services, Rohmaniyah *et al*. (2020) looked at these processes. An avenue for bridging Indonesia's development is through the procurement of products and services. In the planning and bidding phases of the purchase of products and services, fraud practices were examined. The writers performed a decent job, but they didn't provide a thorough analysis of scams or incorporate it into an architectural model to provide further insight into the topic at hand.

Vitor *et al*. (2021) introduced a novel approach for detecting auction fraud, utilizing a Fuzzy Neural Network anchored in a uni-null neuron framework. The model integrated a fuzzification technique based on Typicality and Eccentricity Data Analysis operators and a parallel processor for handling stream samples. Neuro-fuzzy models were employed to assess and contrast against a publicly accessible dataset in the subject domain. The outcomes derived from the model proposed in the study exhibited remarkable superiority over the other examined models, reaching a test accuracy of nearly 96%. Additionally, the fuzzy rules underscored the model's capacity to extract meaningful insights. The authors also offered suggestions for potential future endeavors, which encompassed the exploration of fresh training techniques to refine the model's evolution and the exploration of alternative neural structures to enhance the model's accuracy. Bhamare *et al.*, (2019) researched on Online Auction. The authors described Online Auction as a composite of customer interface and admin interface. Customer panel permits a customer to upload a product for sale and bid on a

particular product to buy. The survey also provided an online display of the things they wanted to buy or sell, organized by category. An administrator can manage the entire bidding system using the admin interface. The administrator can control the registered consumers as well as authorize products based on categories. The authors provided a thorough explanation of online bidding along with its uses and advantages, but they neglected to mention any potential drawbacks.

Elshaar and Sadaoui (2020) delved into the realm of Semi-Supervised Classification (SSC), a method well-suited for addressing our fraud detection challenge due to its efficiency in learning from limited labeled data. The research also conducted a series of experiments to statistically determine the minimum requisite amount of labeled data for achieving optimal accuracy. Moreover, the investigation examined instances of misclassified bidders to pinpoint where these misclassifications occurred. Nonetheless, the empirical analysis revealed that SSC mitigates the labor-intensive process of labeling SB data. The study trained the SSC algorithms with balanced data, incorporating synthetic data into the fraud class while reducing data in the normal class. However, a notable limitation was the absence of a developed cost-sensitive semi-supervised classification model capable of systematically addressing the imbalanced nature of SB data.

Dornadula and Geetha (2019) carried out a research on credit card fraud detection using machine learning algorithms. In order to create a profile for each cardholder, the study created a revolutionary fraud detection system that groups clients according to their transactions and extracts behavioral patterns. Then different classifiers are applied on three different groups later rating scores are generated for every type of classifier. The system was able to quickly adapt to new cardholder transaction behaviors as a result of dynamic parameter changes, and concept drift was addressed using a feedback mechanism. An architectural model for further clarity of the discussed topic was not included in the study.

Khan *et al.* (2012) outlined a theoretical framework for the detection of fraudulent activities. The authors introduced models tailored to the well-known types of auctions, forward and reverse auctions. The model also addressed issues of account dependency arising during transactions. Furthermore, the model implemented high levels of trust and security across various communication levels to thwart potential harm from malicious sellers and bidders.

Chotni *et al*. (2015) conducted an evaluation of online auctions, scrutinizing their advantages and disadvantages. The study explored various auction types, analyzing the benefits and drawbacks associated with them. Additionally, the research delved into the challenges

experienced during transactions on these platforms. While limited to a review, the study did not extend to implementing a comprehensive system.

Anowar and Sadaoul (2019) investigated auction fraud detection in commercial sites. Their study introduced a classification model for identifying legitimate bidders versus shill bidders with efficacy. A robust hierarchical clustering technique and a semi-automated labeling approach were integrated to label shill bidding instances. The research also evaluated advanced oversampling, undersampling, and hybrid-sampling methods, comparing their performance using diverse classification algorithms.

Yoshida and Ohwada (2010) put forth a semi-automated approach that combined one-class SVM and Decision Tree for shill bidder detection. Utilizing one-class SVM learning, the approach initially identified outliers, primarily representing shill bidders. These outliers were then assessed by a Decision Tree classifier to avoid misclassifying innocent bidders. Authors manually examined classified data from each Decision Tree node, adjusting labels where misclassifications occurred. This refined Decision Tree was deemed the optimal model for detecting shill bidders. Experimenting with data from a Chinese website named Wowma, the study detected around 77.5% of shill bidders using this approach. However, it should be noted that the evaluation was limited to two attributes, feedback rating, and bidding history, which may not provide a comprehensive view for shill bidder detection. The study lacked details on preprocessing, training steps, and implementation of the model.

Ford *et al*. (2012) introduced a shill bidding classifier based on an Incremental Feedforward Backpropagation-Artificial Neural Network (IFB-ANN). The research involved retrieving eBay product auctions and deploying a hierarchical clustering technique to categorize bidding data. Clusters were manually labeled as suspicious or normal based on specific attributes. This labeled dataset initialized the fraud classifier. While the model showed promise, it could suffer from local minima and incurred computational expense due to its numerous parameters. The classification approach might also misclassify innocent bidders as it evaluated participation across all auctions. This design could potentially hinder the prevention of financial loss.

Gupta and Mudra (2015) devised a generalized approach using Hidden Markov Model (HMM) for in-auction fraud detection. The authors employed only two parameters, bid number, and bid values, for simplicity. The registration phase proposed restricting users from creating multiple accounts using legal parameters such as credit card information. The model's architecture encompassed a training layer and a detection layer. K-means clustering

generated initial probabilities for understanding bidding behavior. The detection layer categorized users into fraud categories based on behavioral analysis. HMM was then applied to detect fraud. However, the use of only two parameters proved insufficient for in-auction fraud detection. The type of bidding fraud wasn't specified, and the proposed approach lacked experimental validation.

Tsang *et al.* (2014) devised a synthetic data generator to construct a classification model for detecting competitive shilling fraud. The generator employed agent-based simulation to fabricate data reflecting fraudulent behavior. The training data was synthesized through agent-defined behaviors, yielding under-sampled datasets due to the high imbalance of normal and fraudulent bidders. Decision Trees were used for classification, although the study didn't elaborate on shill bidders' features. It's important to note that the use of synthetic data may not fully represent actual auction scenarios.

Table 2.1: Summary of Related Works

S/N	Author(s)/Year of Publication	Research Carried out	Summary of Findings
1.	Wurmann *et al*., (2017).	proposed a software architecture that supports both software and human agents for English auctions conducted online.	The Michigan Internet Auction Bot, their suggested auction server, offered flexible definition of auctions taking many aspects into account.
2.	Jurjees *et al*., (2018).	Developed Web-based android application (DealDaddy) for auction based e-commerce system.	They presented an analysis and design for the auction system illustrating key system components using UML diagrams.
3.	Madhuri *et al*., (2020).	proposed employing detaining approaches to analyze, build, and execute a safe online auction system.	This system has made consumers more effective and efficient in their behavior.
4.	Didwania and Chattaraj (2015).	Designed an improved system.	The proposed system was capable of handling a wide range of goods, which is crucial in real life. The suggested design gave serious consideration to time management issues.
5.	Deorukhakar *et al* (2014).	Described the indicators of in-auction fraud.	The authors put forth a paradigm for an online probity model that simultaneously considers multiple instances of learning, human knowledge-based coefficient constraints, and online feature selection.
6.	Lin and Khomnotai (2017).	presented privacy-aware reputation algorithms for online auction fraud detection.	They suggested privacy-related variables to measure a user's percentage of anonymous ratings.

S/N	Author(s)/Year of Publication	Research Carried out	Summary of Findings
7.	Tyagi and Sharma (2021).	Proposed an implementation of fraudulent sellers detection system of online market places using machine learning techniques.	The authors explained a steady increase in e-commerce proportion in global retail expenditure, showing an obvious shift from brick and mortar to retail clicks.
8.	Sekar *et al*., (2019).	Presented a survey on online auction using data mining techniques.	Various data mining techniques such as ID3, C4.5, C5.0, CART, Neural networks etc. were studied and analyzed.
9.	Benjamin *et al* (2013).	Proposed the data mining technique incremental neural network approach to identify the suspicious bidder in online auction.	The neural network used training dataset which contains the historical auction reversible in new data every time.
10.	Dong *et al*., (2009).	Proposed two main techniques are model checking technique and Dempster-Shafer theory.	Using temporal calculations based on bidding patterns, the model checking technique was utilized to identify suspect shill bidders in an auction.
11.	Chang *et al*., (2010).	proposed Clustering technique for discovering the fake transaction records in auction.	The X-means (extended version of Kmeans) clustering technique analyzes the behavioral changes of the fraudsters and identifies them in the earlier phase.
12.	Ponhu and Kummari (2018).	Analysis of the performance of gender classification algorithms using machine learning.	The authors conducted experiments with the datasets and noticed the accuracies between them. Then used pandas and various packages to load datasets and performed the numerical calculations. Both the ANN and SVM gave better results when compared to the other machine learning algorithms.
13.	Qadeer *et al*., (2009).	Tools for Data Mining and Retrieval in High Volume Data Stream Comparative Analysis.	Proposed a system which used dynamic packet filtering and buffering to enable the effective bulk recording of large traffic streams. Then the congestion is encountered by the system.
14.	Kumar (2018).	Used several machine learning algorithms for Stock market predictions.	He proposed the simple architecture, which involved feature extraction from the given dataset, supervised classification of training dataset and test dataset, and result evaluation.

S/N	Author(s)/Year of Publication	Research Carried out	Summary of Findings
15.	Aishwarya (2020).	A binary classification model for phony credit card transactions is presented.	The goal of the study was to use supervised learning to identify the typical patterns present in transactions. It also aimed to train a binary classification model that could recognize the characteristics of these transactions and categorize the data into one of two transaction classes—fraudulent or valid—based on the historical data that was provided.
16.	Gadade *et al*., (2014).	Built online models for the auction fraud moderation and detection system.	The online framework was extended to other applications like webspam detection, content optimization and so forth. The study provided the responsibility of selling the trustful products by the website itself managed by the admin.
17.	Thiagu *et al*., (2014).	presented an effective method for identifying online auction fraud using triangulation detection techniques.	The study described three methods to detect online auction frauds it has own advantages and disadvantages, so that the public and law enforcement can be made more aware and informed about it.
18.	Mubalaike and Adali (2018).	presented how deep learning (DL) models can be benevolent in detecting fraudulent transactions with high accuracy.	A mobile money service provider in Africa's real financial records from one month were used to create the dataset, which had information on more than six million transactions. On the preprocessed data, the most effective deep learning and machine learning methods, including stacked auto-encoders (SAE) and Restricted Boltzmann Machines (RBM) classifiers, were applied.
19.	Carta *et al*., (2019).	Proposed new data intelligence technology-based in Multi Consensus Prudential model which, in accordance with a double criterion, probabilistic and majority-based effectiveness of various state-of-art classification algos.	The aim was to optimize the model's efficacy in the identification of fraudulent transactions irrespective of some data mismatch. The study model was validated with observations in a broad real-world dataset, which indicates that the model proposed outperforms other state-of-art solutions, both in terms of ensemble models or classification approaches.

S/N	Author(s)/Year of Publication	Research Carried out	Summary of Findings
20.	Chang and Chang (2012).	Offered a new early detection method that simultaneously takes into account accuracy or timeliness.	A modified wrapper technique was built to pick a subset of attributes from a large pool of candidate attributions, in order to identify the attributes best suited to differentiate between everyday traders and fraudsters.
21	Kim *et al*., (2019).	Examined deep learning ability to help financial risk management decisions.	In order to predict how separate distribution dealers protect future profits, they developed a deep learning model.
22.	Singh and Jat (2020).	presented a poll to test deep learning methods for detecting financial fraud.	The test findings and study results showed that several fraud detection techniques operate realistically in the field. It is possible to identify the fundamental challenges in implementing a thorough understanding of fraud.
23.	Vegunta and Minni (2013).	Proposed a Bayesian Probit online model framework for auction fraud detection.	The study analyzed the problem of building models for online auction fraud detection system.
24.	Read (2009).	Detected Shill Bidding in Online English Auctions.	created a detection method for collusive shill bidding, which occurs when several shill bidders work together to shill.
25.	Chau et al., (2009)	Detected Fraudulent Personalities in Networks of Online Auctioneers.	The study applied data mining and trust propagation techniques to detect fraudulent users in online auction systems.
26.	Goel (2010).	Shill Verification in Online Auctions Using a Multi-State Bayesian Network.	Used a multi-state Bayesian network detected shill suspects.
27.	Kavu *et al*., (2015).	Presented a fraud detection tool in .	The study designed a shilling detection algorithm integrated with an online auction. The algorithm proved to be effective and worked on real time s with large user base.
28.	Rohmaniyah *et al*., (2020).	Examined fraud in planning and bidding for procurement of goods and services.	The study analyzed fraud practices at the planning and bidding stages of procurement of goods and services.
29.	Vitor *et al*., (2021).	proposed a model based on a uni-null neuron-based fuzzy neural network to detect auction fraud.	The model utilized a parallel processor for stream samples and a fuzzification technique based on Typicality and Eccentricity Data Analysis operators. A public dataset on the subject was tested and compared using neuro-fuzzy models.

S/N	Author(s)/Year of Publication	Research Carried out	Summary of Findings
30.	Bhamare *et al*., (2019).	Researched on Online Auction.	According to the creators, Online Auction is a hybrid of the admin interface and consumer interface. A customer can upload a product for sale on the customer panel and place a bid to purchase it.
31.	Elshaar and Sadaoui (2020).	Explored Semi-Supervised Classification (SSC).	The study carried out several experiments to determine statistically the minimal sufficient amount of labeled data required to achieve the highest accuracy. In addition, investigation of misclassified bidders was done to detect misclassification.
32.	Dornadula and Geetha (2019).	Carried out a research on credit card fraud detection using machine learning algorithms.	The study developed a novel fraud detection method where customers were grouped based on their transactions and extract behavioural patterns to develop a profile for every cardholder.
33.	Khan *et al*., (2012).	Theoretical model for .	The authors proposed models for most famous ; forward and reverse auctions. The model handled the account dependency issue during transaction.
34.	Chothani *et al*., (2015).	A review of online Auction and its Pros and Cons.	The study reviewed types of Auction benefits and disadvantages of .
35.	Anowar and Sadaoul (2019).	Detection of Auction fraud in commercial sites.	The study combined a robust clustering technique and semi-automated labeling approach to label the shill bidding samples.
36.	Yoshida and Ohwada (2010).	Proposed a semi-automated approach that combine one-class SVM' and Decision Tree to detect shill bidders.	Based on the one-class SVM learning, first this approach finds all the outliers whose are basically shill bidders. After that, the outliers were sent to the Decision Tree classifier for further analysis to prevent misclassification of some innocent bidders as outliers.

S/N	Author(s)/Year of Publication	Research Carried out	Summary of Findings
37.	Ford *et al*.,(2012).	Introduced a shill bidding classifier that can adapt to new auction data based on an Incremental Feed forward Back propagation- Artificial Neural Network (IFB-ANN).	They retrieved eBay product auctions, then, defined a hierarchical clustering technique to partition the bidding data. After that, they manually labeled each cluster (either suspicious or normal) based on the values of ten SB attributes (such as Average Outbid Time, Elapsed Time before First Bid and Seller Feedback Rating).Then, the labeled dataset is used to initialize the fraud classifier.
38.	Gupta and Mudra (2015).	Developed a general approach to detect in-auction fraud by using the Hidden Markov Model (HMM).	In their approach, only two parameters (number of bids and bid values) were used for the sake of simplicity.
39.	Tsang *et al*., (2014).	Implemented a synthetic data generator to build a classification model to detect competitive shilling fraud.	The synthetic data generator used an agent-based simulation to generate data based on several fraudulent behaviors. Training data was done in three ways.

CHAPTER THREE
METHODOLOGY

3.1 Methodology Adopted

The approved methodology for the analysis and system design in this research is the Object-Oriented Analysis and Design Methodology (OOADM). The choice of the methodology is motivated by the emphasis OOADM places on system improvement approach, empowering and encouraging re-utilization of programming segments. With this approach, a system is utilized and created on a premise which empowers the re-utilization of existing segments and encourages the sharing of its segments by different systems. Certain OOADM tools used in the design of the proposed system integrate data and processes modeled after the objects that the system interacts with. The tools used in the study include: use case diagram, sequence diagram, class diagram, flowchart diagram and the system architectural design.

3.2 Analysis of the Existing system

Following the review from existing systems in online auction system, we picked interest in the work developed by Rohan Pillai, (2021). It is an online auction system based on the Block chain and smart contract technology. In the existing system, a seller can create a single page where their product can be uploaded for buyer to see. While the buyer can directly access the seller page or the web site page. All that the user has do is to choose the product, add it to the basket and follow the steps to the end of the purchase. The smart contract contains important information about the transaction details such as buyer and seller bank details. The seller initiates the transaction by sending a request to the block application. While client invokes Hash to submit the transaction proposals to the endorsing peer. These endorsement peers receive the transaction proposals, simulate the transaction by executing chain code specified by the transaction, and verify it with buyer details. After verification, the endorsing peer executes the transaction in the ledger and returns the response result to the platform. In the existing system, the seller information is not open for the buyer to see along with other relevant information about product easily. The system is faced with a problem of administration not being able to view or detect user details on the platform so to avoid counterfeit, infringing, stolen or other questionable products on its platform.

3.2.1 Disadvantages of the Existing system

i. Poor management of record
ii. No validation and verification of user sellers and buyers
iii. Slow in retrieving of record (poor time management)

3.2.2 Architecture of the Existing System

Figure 3.2.2's architecture of the existing system, which depicts the evolution of block chain models with previously unheard-of adaptability and freedom, enables learning through the data set's information. This strategy aids in the clear illustration of how to create a model that can spot some patterns in a studied problem. The lodgers are in charge of database extraction. Its training is based on the idea of weight definition using a stepwise, recursive least squares method.

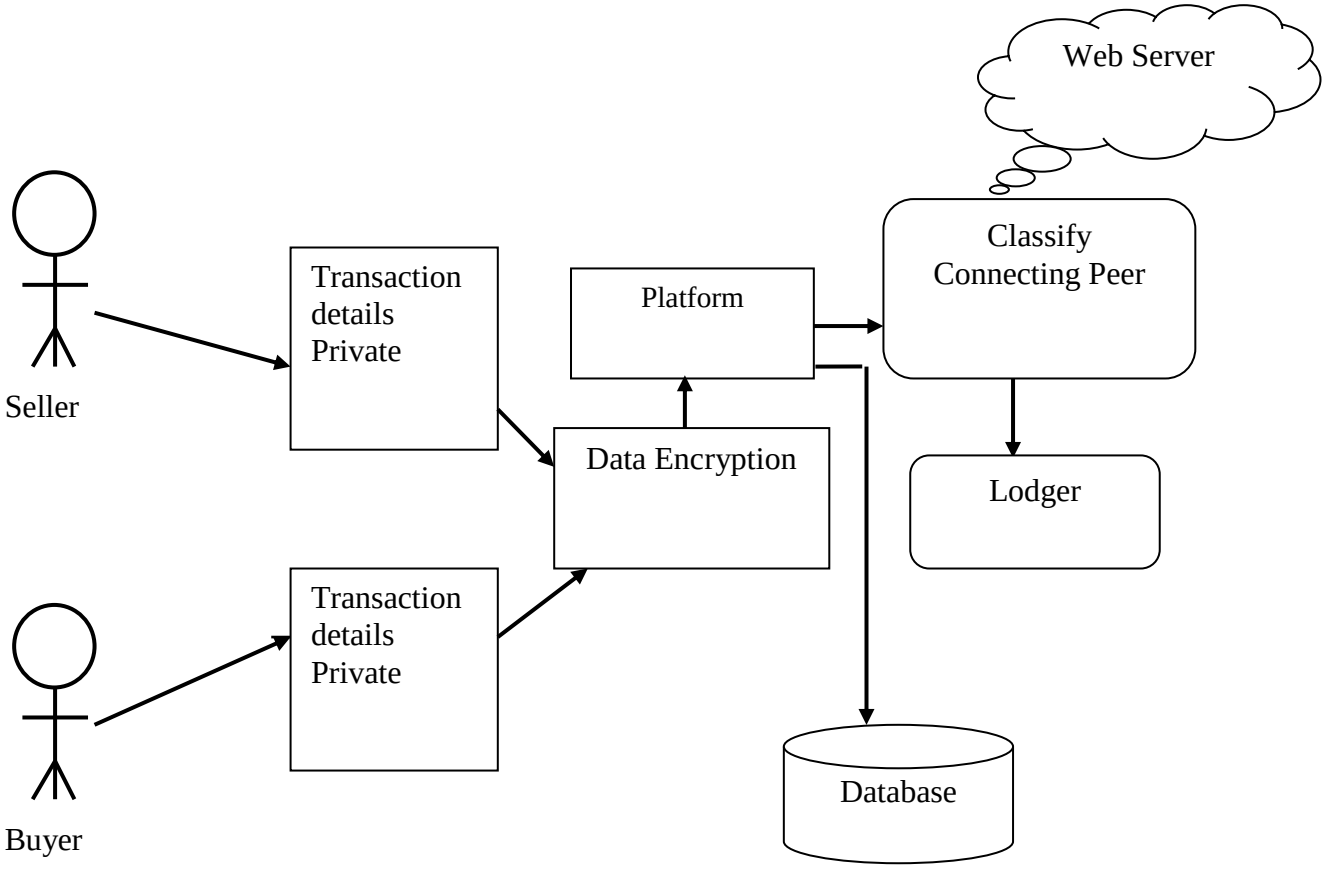


Figure 3.1: Architecture of the Existing System

3.3 Analysis of the Proposed System

The proposed system is a secured auction system using neural network binary classification technique to detect fraud. It is an enhancement of the system developed by Rohan Pillai, (2021). Our prospective system adopts a similar method to the existing system but it has additional features and a web based which makes it secured for sellers and buyers. Sellers have to register with the system; email verification will be sent to the seller for authentication. The seller logs in to his/her email to continue the process. When a seller post a product, the administrators would authorize the product for auction, set the auction dates & minimum auction amount for that product and verify seller and all the bid details. The binary

classifier with extract the seller Internet protocol address (IPS), submet mask and the browser save it on the database. The system will check the seller's past and current transaction. Comments from other buyers would be noted, if it was negative or positive including the seller profits to determine if the product is fake or genuine. Based on the comments, we construct the features of items. The system uses comments in order to validation using the IP address to check if buyer is fake or genuine.

The proposed system is creating online auction system. The users take part in the bidding process by registering and confirming their secure code number. The secure code number verification will be done through a confirmation link that they will receive in their email inbox after registration. Without this verification the users cannot access to the website. The users must verify their identities to reduce the fraud made by users and producing higher level of confidently. Thus, a person is known who she/he is. By this way the users can trust each other, and providing confidence and willingness to deal with each other. The verification of user's identities is done by asking user to input his/her national card number at registration time.

3.3.1 Advantages of the Proposed System

1. Detection of anomalies faster
2. Better prediction
3. Saves time and money

3.3.2 Architecture of the Proposed System

The system's architecture is depicted in figure 3.1, comprising three key phases: Data Preparation, Binary Classifier Learning, and Fraud Detection. The Data Preparation phase involves handling seller and buyer information. Preprocessing data remains an essential stage in any machine learning process. The dataset is divided into distinct training and test sets. Binary classification is employed to categorize set elements into two groups—fake and genuine—based on classification rules. In this proposed system, sellers and buyers can register domestically before system utilization. Following registration, users log in using a password. The validation of buyer and seller details during transactions is executed using the Neural Network algorithm to determine their authenticity. This validation is conducted within a secure area implemented through session variables, which function similarly to cookies, storing information for a specific duration. Session variables retain values until the session concludes, enabling repeated access during the session.

Binary classification categorizes the proposed system into two groups—input variable (genuine) and output variable (fake). A supervised learning algorithm constructs a model using the response variable across known data to assess model accuracy. The Classification and Regression Tree (CART) algorithm, a form of supervised learning, is employed for identifying patterns in data collections without predefined categories, distinguishing similar and dissimilar behaviors.

Patterns that exhibit consistent behavior, such as specific IP addresses, usernames, passwords, and software tokens, are established during the first stage of training for legitimate users. When users attempt to access their accounts using login details, input data is compared to trained data. A match indicates normal, non-malicious entry/activity, granting access. In contrast, if a dissimilarity is detected between input and trained data, the entry/activity is classified as malicious, resulting in access denial. In the subsequent discussion, mathematical notations are introduced to facilitate understanding, followed by the presentation of word-level, semantic, and structural features. Transactions that violate this validation are marked as suspicious with a user-defined weight, Wa.

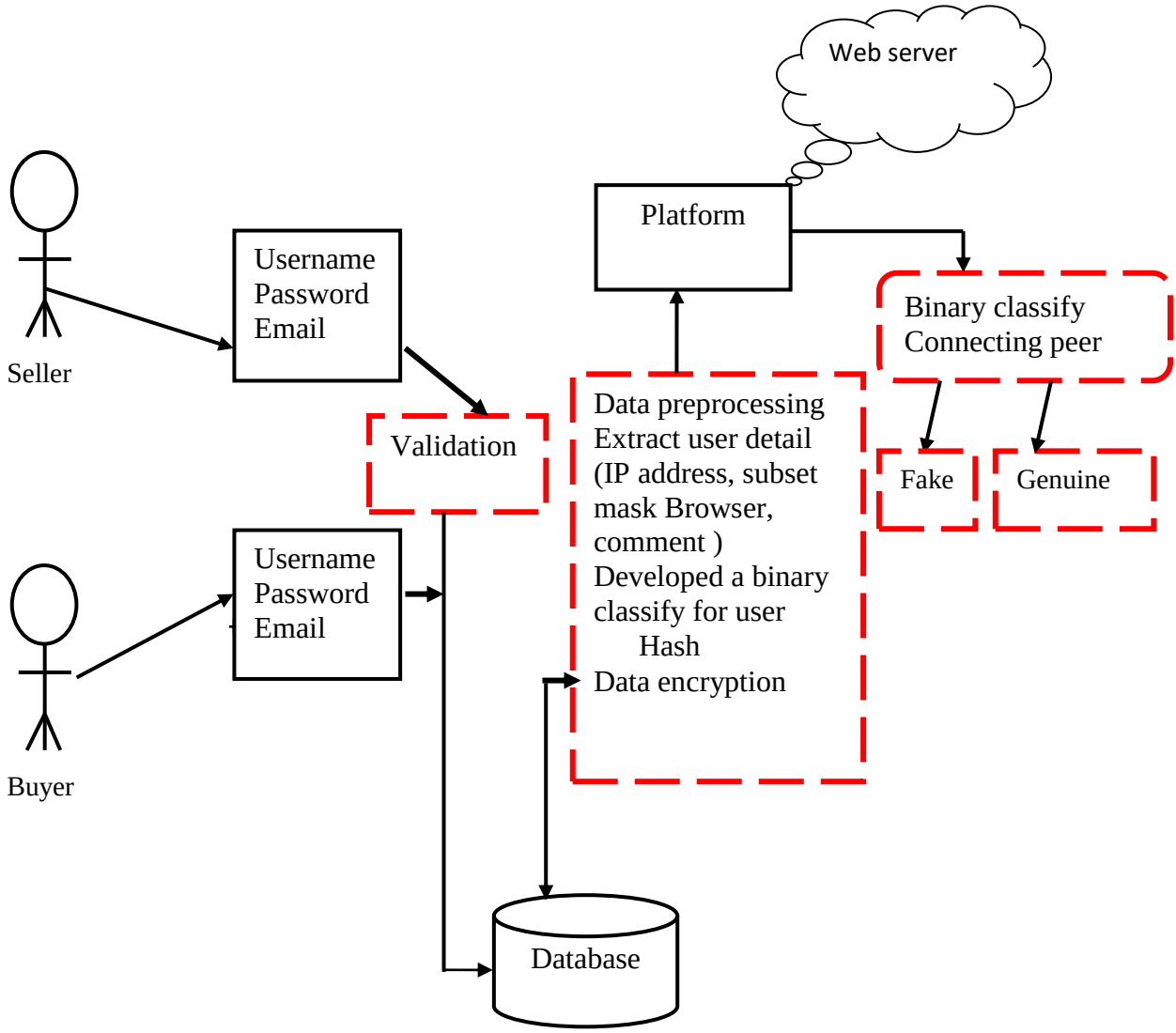


Figure 3.2: Architecture of the Proposed System

Binary classification has been employed to categorize the proposed system into two groups. Within the dataset, there exist two types of variables: input variables (genuine) and output variables (fake). A supervised learning algorithm constructs a model that uses the response variable across a known dataset to assess the model's accuracy. As a facet of supervised machine learning, classification has gained notable prominence. The K-means clustering algorithm, a form of supervised learning, is utilized to identify patterns in a dataset without predefined categories, distinguishing both similar and dissimilar behaviors.

Patterns that exhibit consistency from the initial stages and persist over time, such as specific IP addresses, usernames, passwords, and software tokens, are trained for legitimate users. Whenever such users seek access using their login details, the incoming information is compared to the trained data. If similarity is identified, the entry/activity is considered normal and non-malicious, granting access. Conversely, if dissimilarity emerges upon comparison with the trained data, the entry/activity is deemed malicious, leading to access denial. In the

subsequent discussion, we introduce mathematical notations to facilitate discourse, followed by the introduction of word-level, semantic, and structural features. Transactions contravening this validation are categorized as suspicious, attributed a user-defined weight Wa.

A product T_{jp}^{Ck} is detected as an outlier if it does not belong to any cluster in the set C Such an observation gives evidence that the product could be fake. We measure the extent of deviation of an incoming transaction by its degree of outlierness. If the average distance of the amount p of an outlier transaction T_{jp}^{Ck} from the set of existing clusters in C is v_{avg}, then its degree of outlierness $d_{outlier}$ is given by:

$$d_{outlier} = \begin{cases} 1 - \frac{\mathcal{E}}{v_{avg}} if\ |N_{\mathcal{E}}(p)| < MinPts \\ 0 \end{cases} \quad \text{…….. (equ 3. 1)}$$

where

MinPts is the minimum number of points required in the e-neighborhood of each point to form a cluster:

Ɛ: maximum radius of the neighborhood

$N_{\mathcal{E}}(p) = \{q \in D | dist(pq) \leq \mathcal{E}\}$ ……..… (equ3. 2)

Thus, The ‘‘Internet protocol” rule as R1 and the ‘‘ browser” rule as R2

However, Computing the comment (CTH) and product (PTH) is given as:

R2 = comment > product max specified limit ……. (equ 3. 3)

$C = \{c_1, \ldots, c_n\}$ ……..….. (equ 3. 4)

(denote the clusters in a database D for a particular card $C_{k)}$

$A = \{a_1, a_2, \ldots, a_n\}$ …….. (equ 3.5).

The function y(t) is described as the modified version of x(t) weighed by w(a). In a neural network context, x(t) is the input, w(a) the kernel, and y(t) the output or feature map. Since algorithms only ɑ $_{+\infty}$ discrete steps, the convolution operation has to be discretized:

$y(t) = (x \star w)(t) = \sum$ x(a)w(t−a) (3.6)

Here, t is an inte $_{-\infty}$ volution can be generalized for multiple dimensions, where functions x and w are defined on a set t of integers. With a 2-dimensional input x(i, j), for example, images with width and height coordinates i and j, the convolution algorithm is written in the following equation below:

Algorithm for K-Means

This involves first looping for a fixed number of epochs and within each epoch updating the network for each row in the training dataset. Because updates are made for each training pattern, If errors were accumulated across an epoch before updating the weights, this is called batch learning or batch gradient descent.

Input:

Given T_1, T_2, ... …T_i//set of elements

k // number of desired clusters

Output:

K //set of clusters

K-means Algorithm:

randomly initialize values for mean c_1, c_2, ...c_k;

repeat

assign each element T_i to centroid with the closest mean;

recalculate to find new mean to each cluster;

process is repeated;

until merging target is met;

Back propagation is used to train the sensor. The back propagation algorithm is frequently used training technique in the feed-forward networks. Back propagation algorithm is used demonstrate on how to change the weights and bias the cost function in a network.

Initialize all weights with small random numbers. typically between 1 and 0

for every pattern in the training set

Present the pattern to the network

// Propagate the input forward through the network:

for each layer in the network

for every node in the layer

1. Calculate the weight sum of the inputs to the node
2. Add the threshold to the sum
3. Calculate the activation for the node

end

end

// Propagate the errors backward through the network

for every node in the output layer

calculate the error signal

end

```
        for all hidden layers
            for every node in the layer
                1. Calculate the node's signal error
                2. Update each node's weight in the network
            end
        end
//      Calculate  Error
        Calculate the Error Function
    end
while ((maximum  number of iterations < than specified) AND
        (Error Function is > than specified))
```

Training the Network

To validate the network at regular intervals during training, validation data should be specified by choosing the validation frequency value so that the network is validated about once per epoch. This involves first looping for a fixed number of epochs and within each epoch updating the network for each row in the training dataset. Because updates are made for each training pattern, If errors were accumulated across an epoch before updating the weights, this is called batch learning or batch gradient descent.

Table 3.1: Key data structures

Data	Description	Type
Training samples	Number of Training samples	Integer
Epoch's	iterations	Integer
Input Unit	Data input	String
Learning Rate	Used for improving	double
	Prediction output	

Epochs 100

Leaning rate 0.01

Data Blocks:

This block initializes the data sequence in a list, the time and the dataset variables. The dataset variable is very important as it contains the necessary input context data for the predictive modeling and simulations. This dataset is stored in a suitable file in the computer system and is accessed either through conventional means or programmatically through the use of appropriate file commands in the main program. During the training process, backward propagation was used for repetition of the process at the same time for checking error. Below is the training error graph.

An error could occur if the total number of iterations set in the training process is not

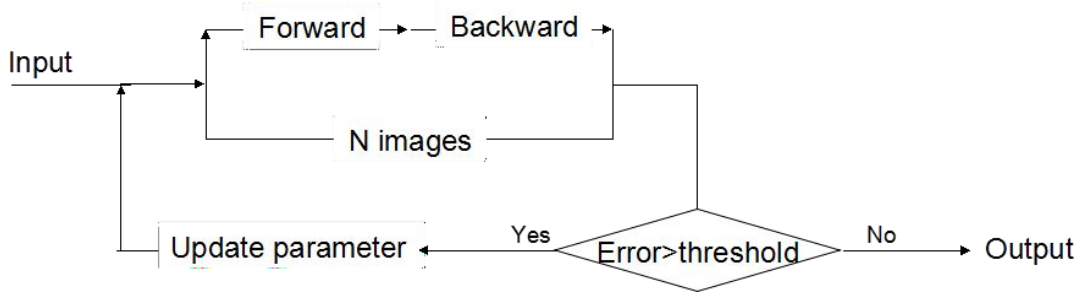


Figure 3.3 : Error Output of the System

completed. Sometime the network has a specific number of input and output units. The number of hidden units was increased as more data were added. The data is divided into training and validation data sets.

Predictive System Specifications

The predictive system is centred on the provision of neural network which shows number of hidden neurons (hidden unit size), number of training epochs, letter size, maximum number of trials.

Table 3.2: Key predictive System Network Default Parameters

System Parameter	Value
Hidden unit size	40 by 40
Epochs	100
Learning rate	0.01
Soft-max sample temperature	0.10
Letter (character) size	5
Maximum trials	3

Dataset for the Trained Model

Raw data of a sufficient no. of auctions is used which contains the minimum required attributes like (bidder id, item id, item type value of bids, no. of bids) for each auction. Using the definition of K-means we formulate the initial required parameters necessary for applying the first K-means clustering algorithm to the system like (no. of distinct observation symbols M). Using the parameters identified in step we apply K-means algorithm and later on identify the bidding behavior of users. After this the identification of other tuple parameters• required by K-means is calculated using other algorithm named as Baum Welch algorithm. After the calculation phase, the next phase starts i.e. Detection where each new bid placed by any respective user will be compared with its bidding habit and decisions are taken if the new bid is false bid or not

Table 3.3 Parameter dataset

Purchase id	**Weight**
Company name	
Product id	
Product name	
Product rate	
Ip address	
Login details	
Browser type	
Description	
Complaint	
Number of bidders	
Average bid	
Winning bid	
Initial bid	
Auction ID	
Total	

Training Phase

It is necessary to analyze this huge amount of data and extract useful information from it. So I split the time series into training and validation sets with ratios of 80% and 20% respectively. The experiment was repeated again several times but the dataset was truncated to consider historical data. Figure 3.4 trained system model

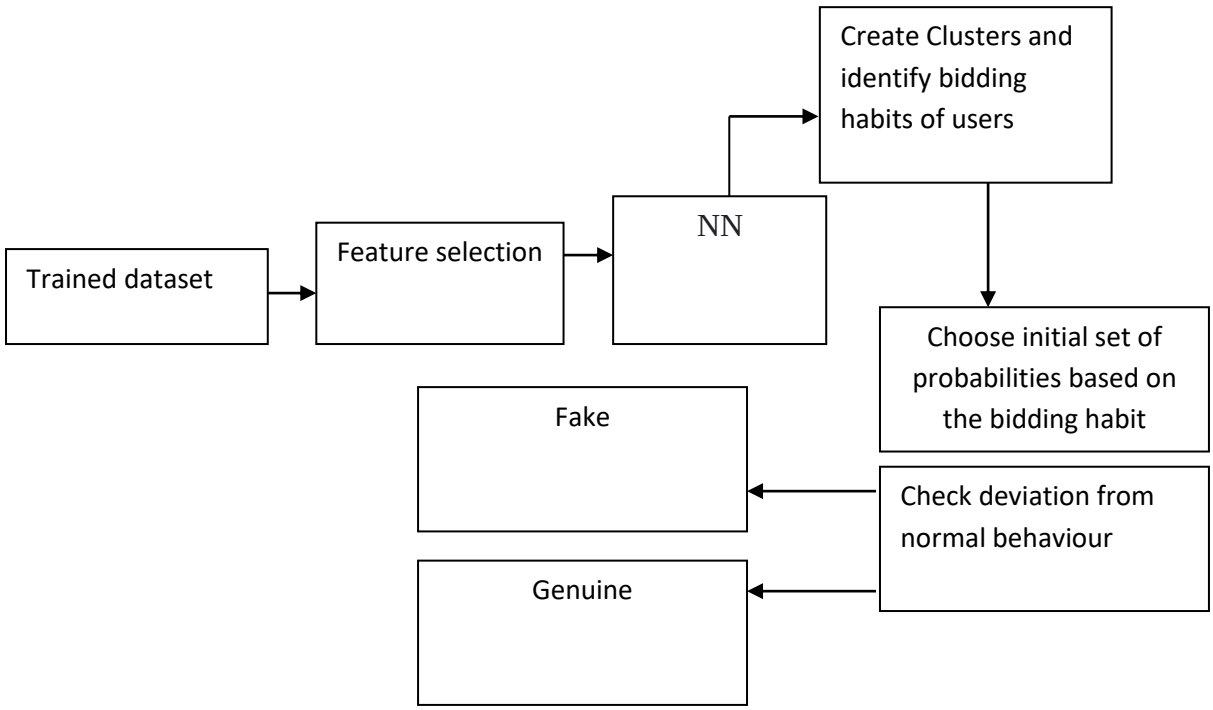


Figure 3.4 : Training Phase

3.4 High Level Input Model.

This is about the interface that the user will use to interact with the program. The users must verify their identities to reduce the fraud make by users and producing higher level of confidently.. The verification of user's identities is done by asking user to input his/her email, Ip address and submet mask time. Any device can be use to capture seller or buyer details into the database. figure 3.5 Input design of the proposed system.

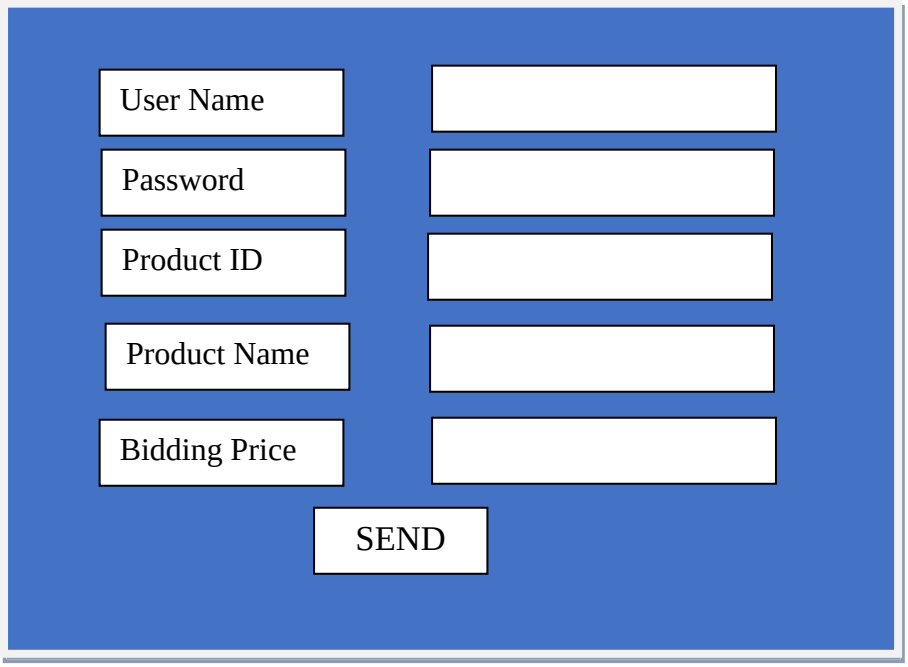


Figure 3.5: Input Design of the Proposed System

3.5 High Level Process Design

The high level design are the different component of the proposed system use in designing the application. The following are the component

USE-CASE diagram

.The USE-CASE consists of a set of possible sequences of interactions between systems and users in a particular environment for achieving a particular goal. The seller that makes a bid is asked to login if not already logged. If the seller is accepted by the system, the item is listed in the user personal page. Bids can only be placed during the auction interval and they must be at least one minimum increment bid above the current price. Post an auction: From his/her personal page, the user can. post an auction from a specific form, providing the characteristics of the item he/she is willing to sell. If the auction is accepted by the system, the item is listed in the user personal page and other users can place a bid for it. Figure 3.6 consist of the USE-CASE diagram of the proposed system

1

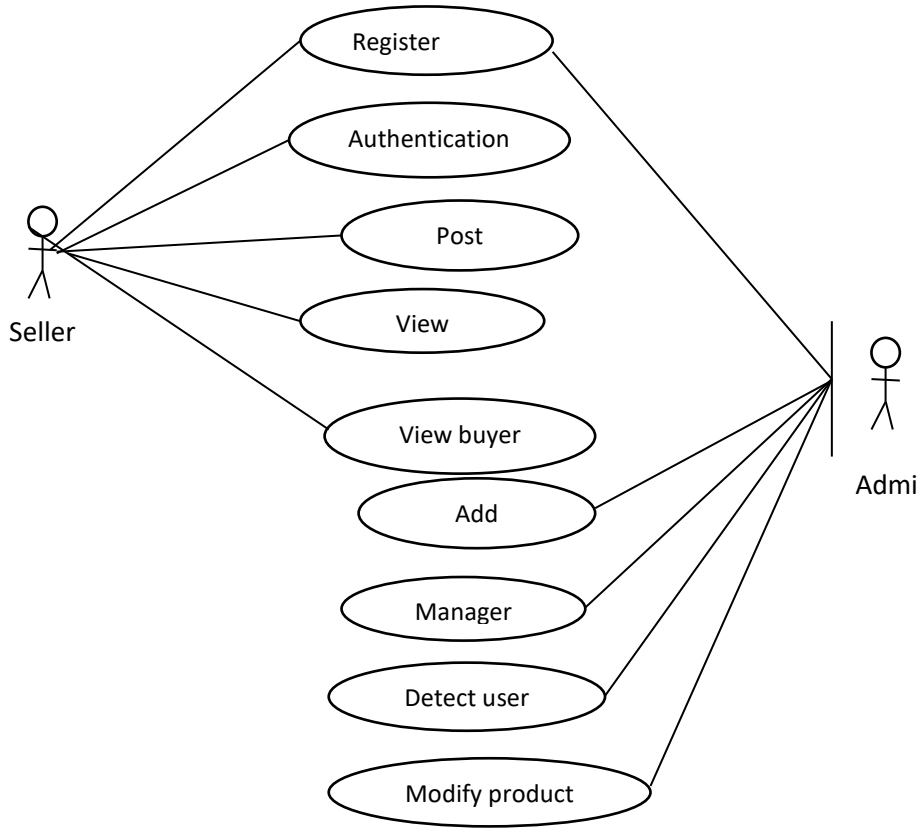


Figure 3.6 : Admin Use Case Diagram of the Proposed System

Sequence Diagrams

Sequence diagram show how object interact with each other via messages in the execution of an operation. After having drawn the Sequence Diagrams for the On Line Auction System, it is clear what kind of data should be stored in the database. Since PostgreSQL is a relational

database, the Sequence Diagrams approach is very useful to design the database schema since it maps well to the relational model and the constructs used in the sequence Diagram model can easily be transformed into relational tables. Here is the sequence Diagram for the database of the system. A sequence diagram depicts in figure 3.7.

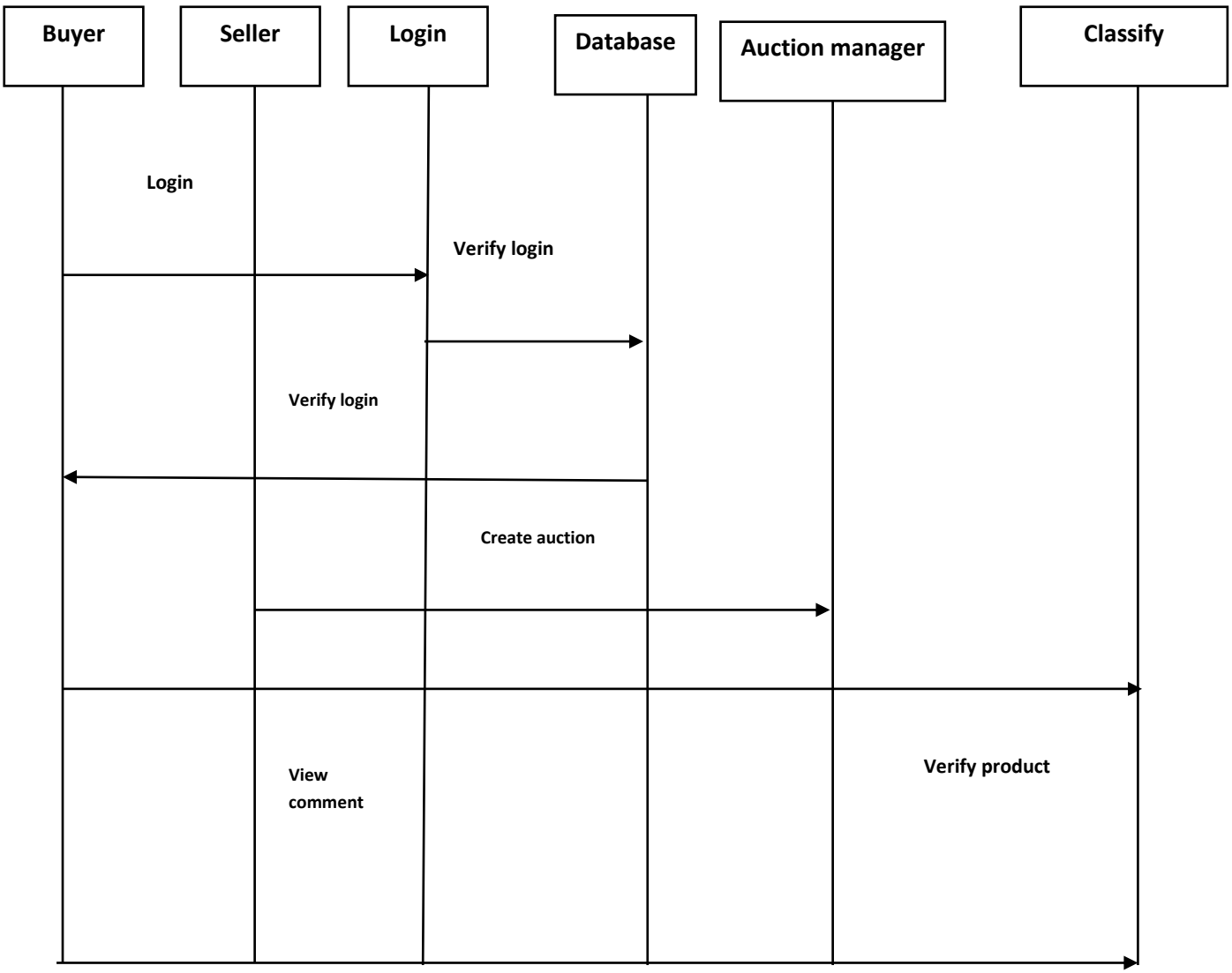


Figure 3.7: Sequence Diagram of the Proposed System.

Class Diagram

The bid must be higher than the start or the current bid. Administrative actor, which can login to the system with a particular username and password. Administrative performs many actions in the system that contain managing category, removing users, editing categories, etc. The user capable of adding an item on auction for sale, adjusting it and remove.It by login to the auction site and accessing his/her personal page. figure 3.8 Class Diagram of the proposed system

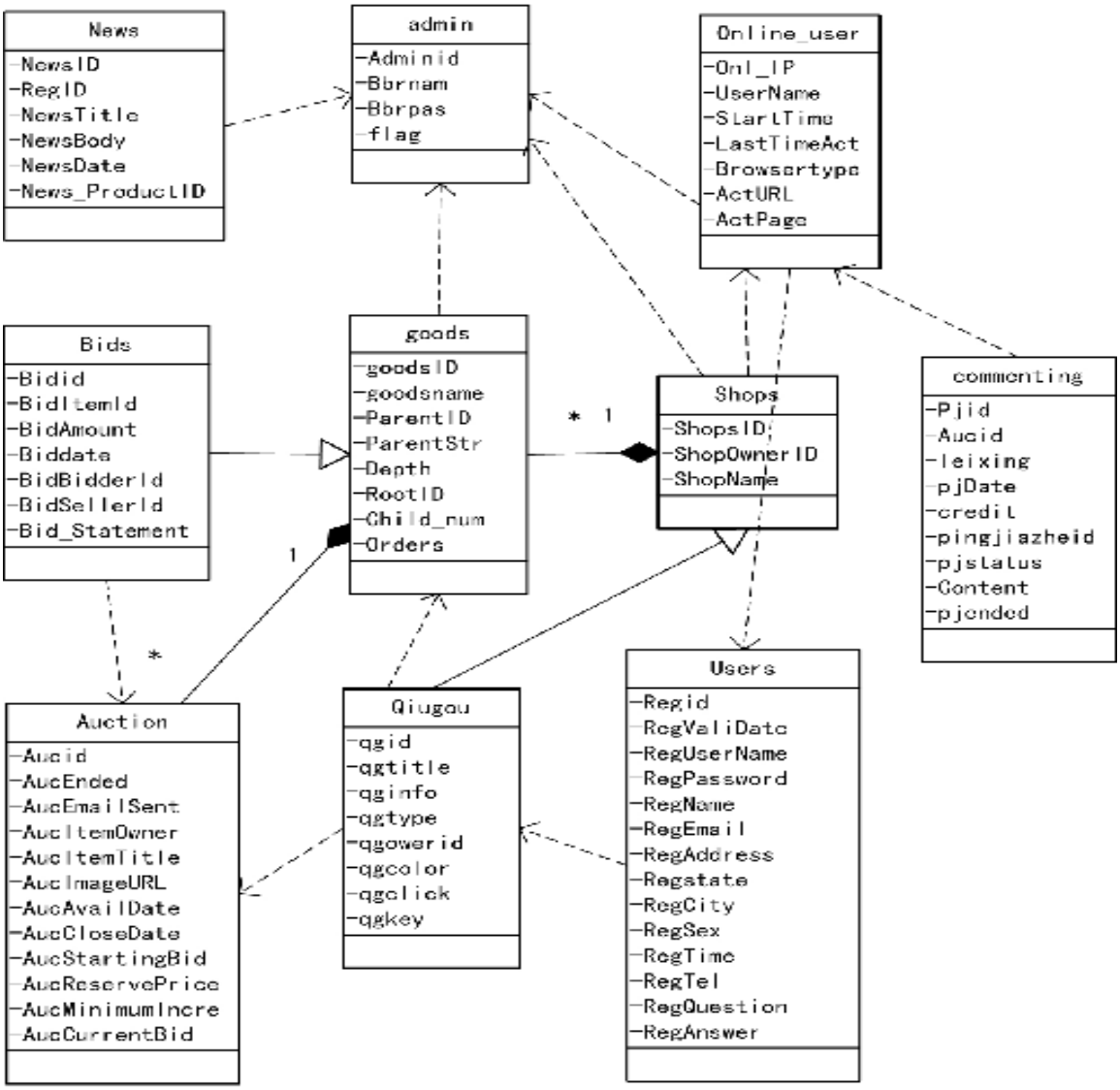


Figure 3.8: Class Diagram of the Proposed System

CHAPTER FOUR

IMPLEMENTATION AND DISCUSSION

4.1 Choice of Programming Language

One should consider a programming language in which the programs can be developed and executed rapidly. The programming language for the adoption of the proposed e-auction shill bidding fraud detection model is Python.

4.2 Reasons for Choosing Python for Implementation

The following are some of the reasons Python is used for implementing the proposed Machine Learning model for e-auction fraud detection.

i. **User-friendly and Uniform:** Python presents concise and coherent code. While intricate algorithms and versatile workflows underpin machine learning and AI, Python's straightforwardness empowers developers to craft dependable systems. This allows developers to channel their efforts into addressing ML challenges rather than grappling with the technical intricacies of the language. Moreover, Python's learnability appeals to many programmers. Python code is comprehensible to humans, streamlining the construction of machine learning models. Several programmers affirm that Python exhibits greater intuitiveness compared to other programming languages. Others highlight the abundance of frameworks, libraries, and extensions that streamline the incorporation of diverse functionalities. It's widely acknowledged that Python suits collaborative implementation, especially when multiple developers are involved. As a versatile language, Python adeptly handles complex machine learning tasks and enables rapid prototyping for rigorous testing of your machine learning products.

ii. **Comprehensive array of libraries and frameworks:** The implementation of AI and ML algorithms can prove intricate and time-consuming. A well-organized and thoroughly tested environment is indispensable to facilitate developers in devising optimal coding solutions. In order to streamline development timelines, programmers frequently turn to a range of Python frameworks and libraries. A software library constitutes pre-authored code that developers utilize to tackle common programming tasks. Python, with its extensive technological toolkit, boasts a wide spectrum of libraries tailored for artificial intelligence and machine learning endeavors. Here are a few notable ones:

   a) Keras, TensorFlow, and Scikit-learn for machine learning

   b) NumPy for high-performance scientific computation and data analysis

c) SciPy for advanced computation

d) Pandas for versatile data analysis

e) Seaborn for data visualization

Scikit-learn encompasses diverse classification, regression, and clustering algorithms, including support vector machines, random forests, gradient boosting, k-means, and DBSCAN. It is designed to synergize seamlessly with Python's numerical and scientific libraries, namely NumPy and SciPy.

Moreover, a wide array of Python integrated development environments (IDEs) furnishes an all-encompassing toolset encompassing testing, debugging, refactoring, and localized build automation within a unified interface. These solutions expedite product development, obviating the need for your development team to reinvent the wheel. Instead, they can capitalize on existing libraries to incorporate requisite functionalities.

iii. **Platform independence:** A programming language or framework that is platform independent enables developers to implement things on one system and utilize them on another machine without any (or with very little) modification. The fact that Python is a platform independent language is one factor in its success. Numerous operating systems, including Linux, Windows, and macOS, support Python. The majority of popular operating systems may be used to produce standalone executable programs using Python code, making it simple to distribute and utilize Python software on those operating systems without a Python interpreter. What's more, developers usually use services such as Google or Amazon for their computing needs. However, you can often find companies and data scientists who use their own machines with powerful Graphics Processing Units (GPUs) to train their ML models. And the fact that Python is platform independent makes this training a lot cheaper and easier.

iv. **Great community and popularity:** In the Developer Survey 2020 by Stack Overflow, Python was among the top 5 most popular programming languages, which ultimately means that you can find and hire a development company with the necessary skill set to build your AI-based project. In the Python Developers Survey 2020, it was observed that Python is commonly used for web development. At first glance, web development prevails, accounting for over 26% of the use cases shown in the image below. However, if you combine data science and machine learning, they make up a stunning 27%.

4.3 System Requirements

To implement the e-auction fraud detection model, the following hardware and software specifications will be requires:

i) A laptop with at least 8 GB of RAM and 50 GB of free disk space

ii) Python 3.7 or higher installed: The following Python libraries will be installed: pandas, numpy, sklearn, imblearn, gensim, nltk, matplotlib, seaborn, shap

iii) VS Code or any other code editor of your choice

4.4 Steps in Creating An E-auction Fraud Detection using Python

To create an e-auction fraud detector with Python, we utilized a combination of supervised and unsupervised learning algorithms, as well as some data preprocessing and feature engineering techniques. The main steps are:

1. Load and Investigate the Data: Our initial step involves utilizing a sample dataset extracted from eBay transactions. This dataset encompasses details about the merchandise, the seller, the purchaser, and the bidding procedure. Additionally, a labeled column is incorporated, signifying whether the transaction was fraudulent.
2. Process and Modify the Data: Our subsequent phase entails data refinement, incorporating actions like addressing missing values and outliers, along with data modification such as encoding categorical variables and scaling numerical ones.
3. Divide the Data into Training and Testing Sets: We'll divide our data utilizing a 70/30 ratio, segregating it into a training set and a test set. This division will facilitate employing the training set for model training and the test set for evaluating model performance.
4. Train a Supervised Learning Model: The following stage involves training a logistic regression model, selected as our baseline model for supervised learning. This model will be instrumental in predicting the likelihood of a transaction being fraudulent or legitimate, grounded in the extracted data features.
5. Train an Unsupervised Learning Model: In this step, a DBSCAN model will be employed as our unsupervised learning baseline. This model will assist in clustering transactions into distinct groups according to their resemblances, subsequently recognizing groups prone to containing anomalies or outliers.
6. Evaluate and Merge Models: Our focus will shift to assessing model performance through various metrics such as accuracy, precision, recall, F1-score, ROC curve, and

AUC score. Furthermore, we'll explore techniques to merge models, employing either a voting classifier or an ensemble approach, aiming to enhance overall outcomes.

7. Visualize and Interpret Outcomes: Visualization tools like matplotlib and seaborn will be utilized to graphically present results, garnering insights into both models and data. We'll also employ interpretation tools like SHAP values and feature importance scores to gain an understanding of the rationale behind model predictions and the pivotal features impacting fraud detection.

4.5 Program Installation

i) Installing virtualenv

To install virtualenv, you need to have Python installed on your system. You can check if you have Python by running the following command in your terminal:

```bash
python --version
```

If you see a Python version number, such as 3.9.5, then you have Python installed. If not, you need to install Python first.

You can install virtualenv using pip, which is the Python package manager. To install pip, you can follow the instructions on this page: https://pip.pypa.io/en/stable/installing/

Once you have pip installed, you can run the following command to install virtualenv:

```bash
pip install virtualenv
```

This will download and install the virtualenv package on your system.

ii) Creating a virtual environment: To create a virtual environment using virtualenv, you need to specify a name and a location for the environment. For example, if you want to create a virtual environment called venv in your current directory, you can run the following command:

```bash
virtualenv venv
```

This will create a subdirectory called venv in your current directory, and copy or link the Python interpreter and the necessary files to run it. It will also create a subdirectory called lib

that will contain the site-packages directory, where the Python packages for the virtual environment will be installed. Additionally, it will create a subdirectory called bin (or Scripts on Windows) that will contain the scripts to activate and deactivate the virtual environment.

iii) **Activating and deactivating a virtual environment:** To use a virtual environment, you need to activate it first. This will modify your shell's environment variables and path settings so that you can use the Python interpreter and the packages installed in the virtual environment.

To activate a virtual environment, you need to run the activate script that is located in the bin (or Scripts on Windows) subdirectory of the virtual environment. For example, if you are using a Unix-like system and your virtual environment is called venv, you can run the following command:

```bash
source venv/bin/activate
```

If you are using Windows, you can run the following command:

```bash
venv\Scripts\activate.bat
```

After running this command, you should see a (venv) prefix before your shell prompt, indicating that you are in the virtual environment.

To deactivate a virtual environment, you need to run the deactivate script that is also located in the bin (or Scripts on Windows) subdirectory of the virtual environment. For example, if you are using a Unix-like system and your virtual environment is called venv, you can run the following command:

```bash
deactivate
```

If you are using Windows, you can run the following command:

```bash
venv\Scripts\deactivate.bat
```

After running this command, you should see that the (venv) prefix is gone from your shell prompt, indicating that you are out of the virtual environment.

iv) Installing and managing packages in a virtual environment : Once you have activated a virtual environment, you can use pip to install and manage Python packages for your project. For example, if you want to install Flask, which is a web framework for Python, you can run the following command:

```bash
pip install Flask
```

This will download and install Flask and its dependencies in the site-packages directory of your virtual environment. You can check what packages are installed in your virtual environment by running the following command:

```bash
pip list
```

This will show you a list of packages and their versions that are available in your virtual environment.

You can also use pip to uninstall or upgrade packages in your virtual environment. For example, if you want to uninstall Flask from your virtual environment, you can run the following command:

```bash
pip uninstall Flask
```

This will remove Flask and its dependencies from your site-packages directory.

If you want to upgrade Flask to the latest version available on PyPI (the Python Package Index), you can run the following command:

```bash
pip install --upgrade Flask
```

This will download and install the latest

4.6 Implementation and Discussion

Figure 4.1 depicts the installation of visual studio code. Visual Studio Code is a freeware which can be downloaded from the web and installed in Windows or Linux environment. In Windows environment the exe installer can be downloaded and installed by double clicking on the downloaded file. The installation wizard then completes the installation process. Step i. Search for and launch Visual Studio Code. The search for Visual Studio Code is as shown in figure 4.1 after locating vs code, click on the hyperlink and control was transferred to vs code code editor.

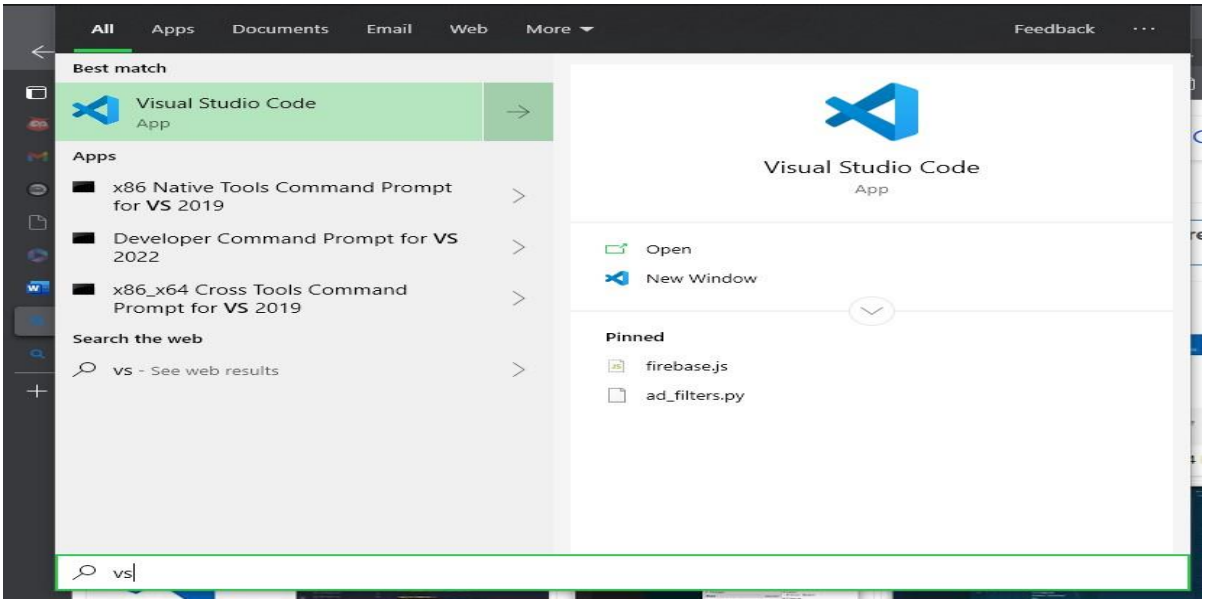


Figure 4.1:Location of Visual Studio Code

In figure 4.2, the visual Studio Code is shown. Visual Studio Code is a 64-bit IDE that makes it easier to work with bigger projects and complex workloads, boasting a fluid and responsive experience for users. The IDE features IntelliCode, its automatic code completion tools that understand code context and that can complete up to a whole line at once to drive accurate coding. Visual Studio Code comes fairly complete out of the box, but there are many plug-ins available to extend its functionality.

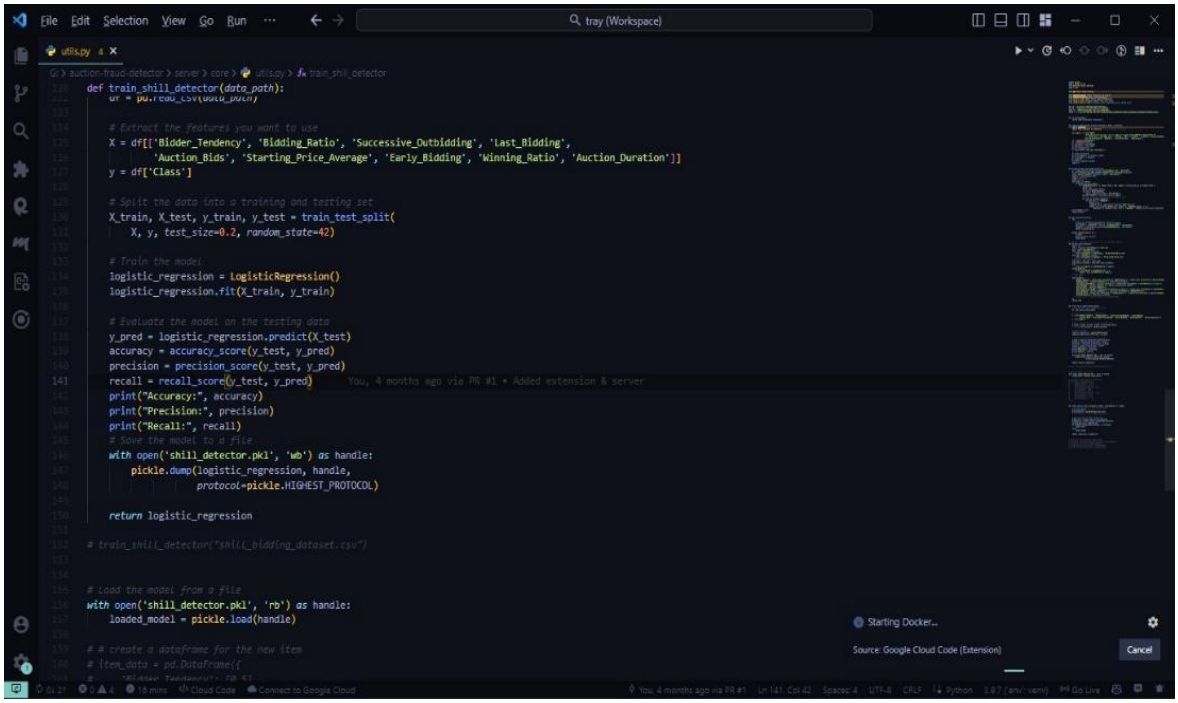


Figure 4.2:Virtual Studio Code

Figure 4.3 shows the create a virtual environment in Python using the package virtualenv. A virtual environment is a way of isolating a Python project from the system-wide Python installation and other projects. It allows you to install and manage Python packages without affecting or being affected by other packages on your system. One of the tools that you can use to create and manage virtual environments in Python is virtualenv. In this study, we explained how to create, install and use virtualenv to create a virtual environment for Python project.

Figure 4.3 shows the bidders ID and the threshold probability. If the threshold exceeds 10, then it will be flag shill bid fraud.

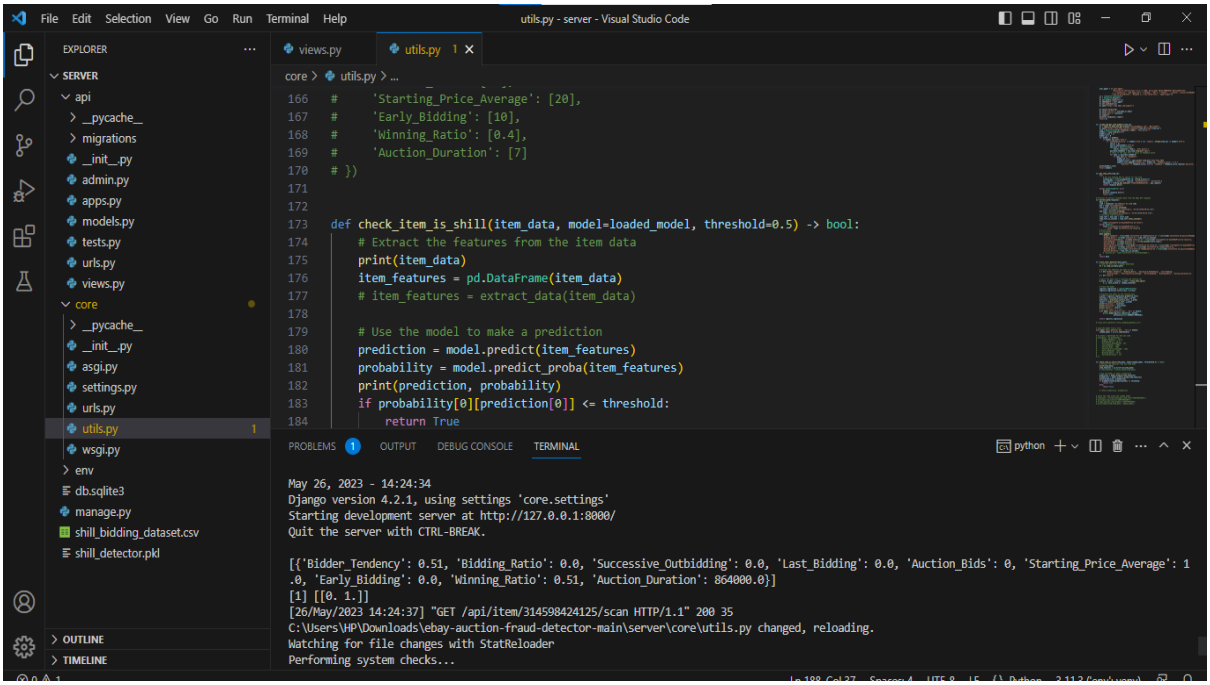


Figure 4.3:Bidders Threshold

Figure 4.4 reveals the sample output of the shill detection probabilities. This depicts the source code and the possibility of the occurrence of a shill bid when auction is done by the bidder.



Figure 4.4: Shill Bidding Probabilities Output

Figure 4.5 depicts the source code where the model is being trained using logistic regression. The proposed system uses logistic regression to build the classifier to prevent shill_bidding fraud in online auction transactions



```
utils.py 8 ×
server > core > utils.py > ...
        return data


    def train_shill_detector(data_path):
        # Load the data into a pandas DataFrame
        df = pd.read_csv(data_path)

        # Extract the features you want to use
        X = df[['Bidder_Tendency', 'Bidding_Ratio', 'Successive_Outbidding', 'Last_Bidding',
                'Auction_Bids', 'Starting_Price_Average', 'Early_Bidding', 'Winning_Ratio', 'Auction_Duration']]
        y = df['Class']

        # Split the data into a training and testing set
        X_train, X_test, y_train, y_test = train_test_split(
            X, y, test_size=0.2, random_state=42)

        # Train the model
        logistic_regression = LogisticRegression()
        logistic_regression.fit(X_train, y_train)

        # Evaluate the model on the testing data
        y_pred = logistic_regression.predict(X_test)
        accuracy = accuracy_score(y_test, y_pred)
        precision = precision_score(y_test, y_pred)
        recall = recall_score(y_test, y_pred)
        print("Accuracy:", accuracy)
        print("Precision:", precision)
        print("Recall:", recall)
        # Save the model to a file
        with open('shill_detector.pkl', 'wb') as handle:
            pickle.dump(logistic_regression, handle,
                        protocol=pickle.HIGHEST_PROTOCOL)

        return logistic_regression
```

Figure 4.4: Source Code showing the trained model

The performance evaluation metrics such as accuracy, precision, recall and F1 Scores of the proposed shill-bidding detection model is shown in figure . 70 % of the dataset was used for training and 20% (200) was used for testing the algorithm. The result of the performance predicted by machine learning algorithm such as logistic regression is as follows for over 100 e-auction bidders:

Prediction = [78.81559325 92.41931081 92.57981572 92.40782513 93.74550495
98.07158033 99.64497992 101.0729118 100.95194109 102.73447463
93.74550495 80.73372616 104.71377705 101.0729118 97.29425866
68.03186809 101.44234013 101.76991467 111.20184943 102.13756159
97.54168545 86.018984 93.52447982 102.18681372 85.07838286
92.41931081 94.27028428 107.20965062 68.03186809 94.80376267
71.60972286 96.97642976 105.84309096 91.7349983 82.62661917
101.0729118 98.52081378 100.48099604 100.63562115 93.27810562
97.17932152 80.59667339 100.65269864 102.46326967 88.29632349
100.63562115 97.74291238 81.12862731 103.65814732 78.79710386
101.0729118 82.38107109 79.81801047 98.20973176 103.9759703
78.81559325 86.018984 98.83229802 94.16380588 86.018984
85.07838286 104.48308356 102.90438284 101.0729118 97.38169679
90.07586223 102.73447463 97.74291238 95.90256134 101.86076371
95.93227797 89.06204128 101.68809341 103.97159811 95.5454376
95.93227797 103.95881661 102.73447463 95.17995555 89.48445267
96.97642976 94.26311106 92.50467535 97.38169679 106.23504818
97.43758072 85.07838286 100.90114888 94.20721401 78.79710386
94.45679122 91.7349983 80.73372616 88.23442164 95.93215884
98.20973176 75.88768819 80.73372616 87.64912 101.44234013
95.90256134 95.5454376 101.57058302 83.81864544 101.0729118
80.60638369 93.21584938 82.19252462 93.27810562 102.34056163
101.26965245 102.46326967 86.018984 86.89347084 97.58023043

97.58023043 85.07872315 86.89347084 86.03260665 95.90256134
86.018984 79.75326031 94.80376267 101.44234013 104.62002803
85.0939143 95.908746 94.20721401 101.15297423 97.58023043
106.68503023 97.17932152 75.88768819 103.95881661 101.15757748
82.23032151 111.33490764 91.38709254 99.59271385 97.17932152
111.33490764 95.17995555 97.17932152 98.20973176 79.81801047
84.05912245 85.19727099 82.79043188 85.9571095 95.93215884
92.75471557 103.33895261 101.31902652 104.62002803 94.80376267
90.81674658 89.06204128 93.74550495 99.59271385 100.65269864
93.48006561 99.59271385 82.79043188 101.0729118 82.62661917
99.78315021 77.50715086 98.20973176 95.93215884 97.17932152
99.59271385 98.71592205 81.12862731 101.73679213 93.74550495
95.93215884 91.38709254 91.7349983 95.00253332 97.17932152
107.20965062 88.29632349 85.2861698 98.20973176 94.20721401
102.73447463 100.63562115 92.50467535 78.79710386 92.75471557
95.48295896 91.65184899 104.85950855 93.52447982 78.97842285
92.40782513 80.52355603 97.38169679 98.81562706 96.10621687]
Mean Absolute Error scores:
[0.13459795 0.11948379 0.12353664 0.1218063 0.12229716]

Mean Absolute Error (MAE), as mentioned above, is a regression metric that is frequently used in Python Sklearn. Without taking into account their direction, it is a measurement of the average size of errors in a group of forecasts. It is the mean of the absolute differences between forecast and actual observation across the test sample. Since MAE is a linear score, the average is weighted equally for each unique difference. **The average error between the predictions and actuals in this dataset is 0.1 This is a very good result since the variation is 1 out of 100.**

4.7 Performance Evaluation

In our study, we are interested in detecting fraudsters more than in identifying normal bidders. So, we choose the most common quality metrics for the fraud
class:

1. Precision:

T P/T P+FP (1)

It calculates the ratio of correctly predicted fraudsters to the total predicted true and false fraud.

2. Recall:

T P/T P+FN (2)

It calculates the ratio of fraudsters that are correctly classified and fraudsters that are misclassified.

3. F1-Score:

$$2 * (\text{Recall} * \text{Precision})/(\text{Recall} + \text{Precision}) \quad (3)$$

It calculates the weighted average of Precision and Recall.

4. The Area Under the ROC Curve (AUC):

It tells us how much a model can distinguish between normal and fraud classes. The closer value is to 1 the better.

5. False Negative Rate (FNR):

$$1 - TP \quad (4)$$

It measures the ratio of fraudsters that are classified as normal bidders.

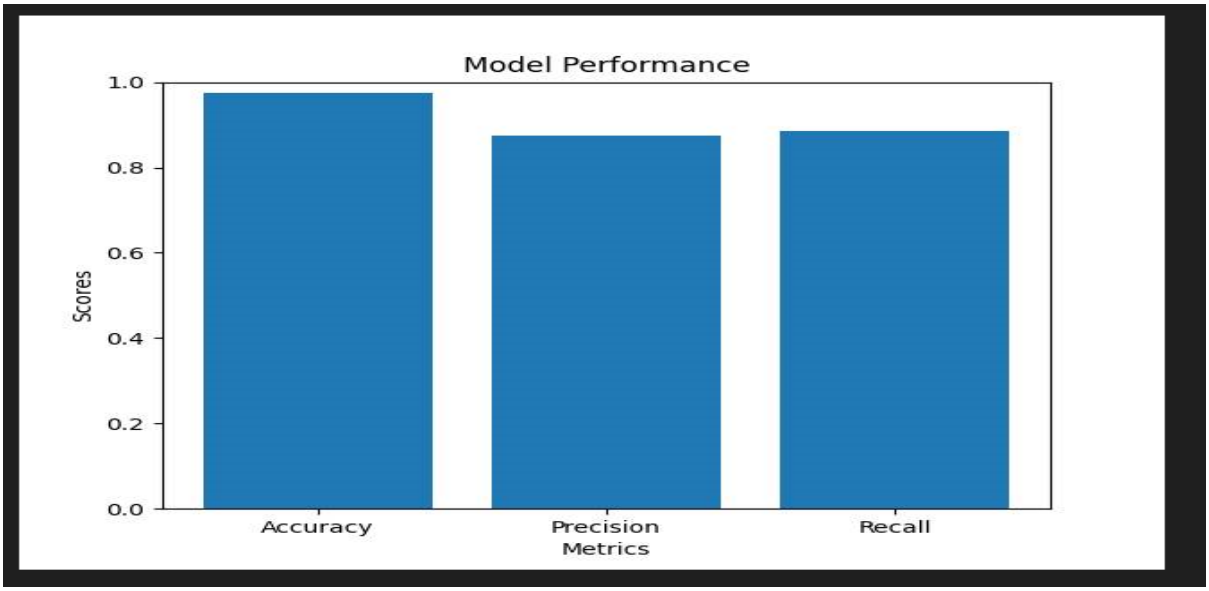


Figure 4.5: shill-bidding detector performance evaluation

CHAPTER FIVE

SUMMARY, CONCLUSION AND RECOMMENDATION

5.1 Summary

Auction fraud has become a major issue in the world of e-commerce due to the quick development of online auctions. Shill bidding, one of the many types of auction fraud, is particularly difficult to spot. As a result, the negative effects of shill bidding frequently go unreported, causing victims to endure unjustified suffering. These actions damage the reputation of auction houses in addition to hurting auction participants. The development of user confidence in online auctions has been hampered by the absence of real-time monitoring systems. Participants have the chance to hide their identities by creating fictitious accounts and acting dishonestly because of this void. Auction houses and unintentional buyers could suffer if shill bidding is not swiftly identified and stopped through rigorous monitoring. Shill bidding occurs when sellers add fake bidders to fraudulently raise the winning bid. Due to its ostensibly genuine bidding patterns, this fraudulent behavior takes place during e-auctions and is difficult to detect. As a result, bidders overpay and suffer financial losses, while sellers benefit from shill bidding techniques. Training and validation, which are divided into 70% and 30% of the proposed model, respectively. The technique looks for instances of shill bidding on eBay by using a binary classification method. The suggested model is trained using a dataset taken from the Kaggle website, and it is then evaluated.

5.2 Conclusion

Engaging in online auctions has gained immense popularity and practicality as a means of conducting business transactions. However, the substantial surge in consumer participation in online auctions has unfortunately led to the proliferation of auction fraud. Among various forms of auction fraud, shill bidding stands out as the most prevalent. Shill bidding entails a seller placing a bid on their own auction with the intention of artificially inflating the final bid amount. This manipulation can be executed directly by the seller or indirectly through collaboration with a third party who submits fictitious bids on behalf of the seller. Given the sheer volume of auctions and bids, manually detecting instances of shill bidding becomes a daunting task. The most effective approach to mitigate the negative impact of auction fraud is to promptly identify shill bidding in real-time. To minimize the losses stemming from shill bidding, researchers have explored various strategies.

5.3 Recommendation

The findings in the study are recommended to be used by:

i) Researchers and students who are interested in learning about and advancing the development of machine learning models for fraud detection in e-commerce website.

ii) Various e-commerce shops in detecting shill bidding fraud

iii) Organizations where real-time fraudulent activities are prevalent and the need for a predictive model in predicting the possibility of fraud and machine learning algorithms in detecting such fraud when it occurs.

iv) Both bidders and the e-Auction companies that face the yearly millions of dollars loss and fraud reports.

REFERENCES

Adabi, S., Farhadinasab, H. and Jahanbani, P.R. 2022. A genetic algorithm-based approach to create a safe and profitable marketplace for cloud customers. Journal of Ambient Intelligence and Humanized Computing.13. 2381–2413. https://link.springer.com/article/10.1007/s12652-021-03682-z Metrics.

Aishawarya, V. (2020). Binary Classification model for fraudulent credit card transactions. *IOSR Journal of Computer Engineering (IOSR-JCE).* 22(3). 38-45.

Aljaf, B. (2016). Online Auction System. Tempre University of Applied Sciences. 50-65.

Anowar, F. and S. Sadaoui(2020). Detection of Auction Fraud in commercial sites. *Journal of Theoretical and Applied Electronic Commerce Research (JTAER).* 15(1). 81-98.

Benjamin, J. F.,H. Xu and I.Valova (2013).A Real-Time Self-Adaptive Classifier for Identifying Suspicious Bidders in Online Auctions. The Computer Journal. 56(5). 22-29.

Bhamare, M., A. Charme, G. More and P.A. Rindhe (2019). Online Auction. *International Journal for Scientific Research & Development (IJSRD).*7(4). 18-21.

Borle, S., Boatwright, P., & Kadane, J.B. (2006). The timing of bid placement and extent of multiple bidding: An empirical investigation using ebay online auctions. Statistical Science, 1,194-205.

Carta, S., G. Fenu, D. F.,Recupero and R. Saia (2019). Fraud detection for E-commerce transactions by employing a prudential Multiple Consensus model. Journal of Information Security and Applications. 46 (2019). 13–22.

Chang, W-S. and Jau-Shien Chang (2010). Using Clustering Techniques to Analyze Fraudulent Behavior Changes in Online Auctions. *International Conference on Networking and Information Technology.*

Chang, W.-H., & Chang, J.-S. (2012). An effective early fraud detection method for online auctions. Electronic Commerce Research and Applications. 11(4). 346–360. doi:10.1016/j.elerap.2012.02.005.

Chau, D., S. Pandit, C. Faloutsos(2009).Detecting Fraudulent Personalities in Networks of Online Auctioneers. Berlin. DOI: 10.1007/11871637_1

Chen, Y.-J., Liou, W.-C., Chen, Y.-M., and Wu, J.-H. (2018).Fraud detection for financial statements of businessgroups. International Journal of Accounting InformationSystems. doi:10.1016/j.accinf.2018.11.004.

Chothani, R.G, N.A. Patel, A. H. Dekavadiya and R.P. Punit (2015). A review of Online Auction and its Pros and Cons. *International Journal of Advance Engineering & Research.* 2(1). 8- 11.

Chua, C. and J. Wareham (2004). Fighting Internet Auction Fraud: An Assessment and Proposal. IEEE. 37(10).31-37.

Chua, C., D.W. Straub, H.M. Khoo, S. Kadiyala, and D. Kuechler (2018). The Evolution of e-CommerceResearch: A Stakeholder Perspective. Journal of Electronic Commerce Research. 6(4).262-280.

Deorukhakar, S., N. Khabiya, A. Kulkarni and A. Thorat (2014). Online Auction Fraud Detection. *International Journal of Engineering Research & Technology (IJERT).* 2(7). 57-63.

Didwana, M. and S. Chattaraj (2015). Design of an improved System. *International Journal of Innovative Research in Science, Engineering and Technology.* 4(9). 69-74.

Dong, F.I., S.M. Shatz1 and H. Xu (2009). Inference of Online Auction Shills Using Dempster-Shafer Theory. *Sixth International Conference on Information Technology: New Generations.*

Donga, F. S., M. Shatza and H. Xub (2017). Combating Online In-Auction Fraud: Clues, Techniques and Challenges. Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607, USA.

Dornadula, V. N and S. Geetha (2019). Credit card fraud detection using machine learning algorithms. *International Conference on recent trends in advance computing (ICRTAC, 2019).* Procedia Computer Science. 165 (2019). 631–641
eBay(2008). Shill Bidding Examples. eBay Inc.. Retrieved 2/02/2022 http://pages.ebay.com/help/policies/seller-Shill-bidding.html

Elshaar, S. and S. Sadaoui (2020). Detecting Bidding Fraud using a Few Labeled Data. *In Proceedings of the 12th International Conference on Agents and Artificial Intelligence (ICAART 2020).*2. 17-25.

EunjiKima, JehyukLeea, Hunsik Shina, Hoseong Yanga,SungzoonChoa, Seung-Kwan Namb, Youngmi Song b,Jeong-a Yoonb, Jong-il Kim, ―Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning‖, Expert Systems With Applications 128(2019) 214–224.

Ford, B., H. Xu and I. Valova (2012). A real-time self-adaptive classifier for identifying suspicious bidders in online auctions. The Computer Journal. 56(5). 646-663

Gadade, H.D., B.P. Manoj and B.D. Patil. (2014). Online Auction fraud detection techniques and security measures using proactive moderation system. *International Journal of Engineering Research & Technology (IJERT).* 3(2). 1158-1163.

Ganguly, S. (2016). Auction Shill Detection framework based on SVM. University of Regina, Saskatchewan. 1-123.

Gavish, B. and C.L. Tucci (2008). Reducing Internet Auction Fraud," Communications of the ACM (ACM). 51(5).89-97.

Goel,A (2010). A Multi-state Bayesian Network for Shill Verification in Online Auctions. San Francisco Bay,USA: Redwood City.

Gregg, D.G., and J.E. Scott (2008). Typology of Complaints about eBay Sellers. Communications of the ACM (CACM). 51(4). 69-74.

Gregg, D.G. and J.E. Scott (2006). The Role of Reputation Systems in Reducing On-Line Auction Fraud. *International Journal of Electronic Commerce.*10 (3).95-120.

Gupta, P. and A. Mudra (2015). Online in-auction fraud detection using online hybrid model. IEEE Proceedings on International Conference on Computing, Communication & Automation (ICCCA). India. 901-907.

Han, J. and M. Kamber (2001). Data Mining: Concepts and Techniques. Morgan Kaufmann Publishers. 2-6.

Jurjees, A. ., S. Kumar, K. Satosh., S. Anwar and B. Manikandan (2018). Developing a web-based Android Application for Auction Based E-Commerce System. *IJESC.* 8(4). 16859- 16863.

Kauffman, R.J., and C.A. Wood (2003). Running up the Bid: Detecting, Predicting, and Preventing Reserve Price Shilling in Online Auctions. Proceeding of the 5th International Conference on Electronic Commerce.

Kauffman, R.J., and Wood, C.A. (2005). The effects of shilling on final bid prices in online auctions. Electronic commerce research and Applications, 4(1), 21-34.

Khan, H.U., M. A. Al-Faifi and D.M. Diab (2012). Theoretical model for . *International Journal of Computer Science and Network Security (IJCSNS).* 12(4). 116-120.

Kim, A., Yang, Y., Lessmann, S., Ma, T., Sung, M.-C., & Johnson, J. E. V. (2019). Can Deep Learning Predict Risky Retail Investors? A Case Study in Financial Risk Behavior Forecasting. *European Journal of Operational* Research. Doi:10.1016/j.ejor.2019.11.007.

Kumar, I., K. Dogra, C. Utreja and P. Yadav (2018). A Comparative Study of Supervised Machine Learning Algorithms for Stock Market Trend Prediction, *Second International Conference on Inventive Communication and Computational Technologies (ICICCT). Coimbatore.* 1003-1007. *Doi:10.1109/ICICCT.2018.8473214*

Lin, J-L and L. Khomnotai (2017). Online Auction Detection in privacy-aware reputation systems. *International Journal of Engineering Research and Technology (IJERT). 5*(9). 24-29.

Logeswari, G. and K. Sangeetha (2015). Mobile Auction System using internet. *International Journal of Engineering Research and Technology (IJERT).* 3(15). 1-3.

Madhuri, H., J. Rukmini, P. Sachin., B. Somnath, O.S. Ghotap (2020). Analysis, Design and Implementation of a secure online Auction System using detaining technique. *International Research Journal of Engineering and Technology (IRJET).* 7(6). 2285-2288.

Majadi, N., J. Trevathan, H. Gray, V. Estivill-Castro and N. Bergmann (2017). Real-time detection of shill bidding in online auctions: A literature review. Computer Science Review. 25(2017). 15-32.

Majadi, N., Trevathan, J., and Bergmann, N. (2016). Analysis on Bidding Behaviors for Detecting Shill Bidders in Online Auctions. InComputer and Information Technology (CIT). International Conference on 2016 Dec 8. 383-390.IEEE.

Mamun (2015). Combating Shill Bidding in Real Time: Prevention, Detection and Response. Computer and Information Science. 8(2). 24-36.

McMillan, R. (2008). auction Fraud Tops Net Crime. PC World. Retrieved 2/02/2022. http://www.pcworld.com/article/129940/auction_fraud_tops_net_crime.html

Mubalaike, A.M. and E. Adali (2018). Deep Learning Approach for Intelligent Financial Fraud Detection System. *3rd International Conference on Computer Science and Engineering (UBMK).* 598-603.

Pandit, S., Chau, D. H., Wang, S., and Faloutsos (2007). C. Net Probe :A fast and scalable system for fraud detection in online auction networks. *In Proceedings of the 16th International Conference on the World Wide Web.* Banff, Canada. 8(12). 201–210

Pondhu, L.N and G. Kummari (2018). Performance Analysis of Machine Learning Algorithms for Gender Classification. Second International Conference on Inventive Communication and Computational Technologies (ICICCT). 1626-1628.

Porter, R. and Y. Shoham (2003). On Cheating in Sealed-Bid Auctions. Proc. of the ACM Conference on Electronic Commerce. 76–84.

Qadeer, M.A., N. Akhtar and F. Khan (2009). Comparison of Tools for Data Mining and Retrieval in High Volume Data Stream. Second International Workshop on Knowledge Discovery and Data Mining. Moscow. 252-255.

Read, J.(2009). Detecting Shill Bidding in Online English Auctions. sHandbook of Research on Social and Organizational Liabilities in Information. 446-70.

Romahniyah, D.S. Musyarofah and D. Tarjo (2020). Dissect Fraud of Planning and Bidding for Procurement of Goods and Service. *International journal of Managerial Studies and Research (IJMSR)*. 8(11). 26-33.

\Roy, A. J. Sun, R. Mahoney, L. Alonzi, S. Adams and P.Beling, "Deep learning detecting fraud in credit cardtransactions," 2018 Systems and Information EngineeringDesign Symposium (SIEDS), Charlottesville, VA, 2018. 29-134. doi: 10.1109/SIEDS.2018.8374722.

Salmon, D., Callie W.B, Gregory A.B and Christopher E.W. 2021. A Framework for Modeling Fraud in E-Waste Management. https://www.sciencedirect.com/science/article/pii/S0921344921002226.

Salvatore Carta, Gianni Fenu, Diego Reforgiato Recupero, and Roberto Saia (2019). Fraud detection for Ecommerce transactions by employing a prudential Multiple Consensus model‖, Journal of Information Security andApplications 46 (2019). 13–22.

Shah, H.S., Joshi, N.R., and Wurman, P.R. (2002). Mining for bidding strategies on eBay. InSIGKDD'2002 Workshop on Web Mining for Usage Patterns and User Profiles

Shah, H.S., N.R. Joshi, A. Sureka, and P.R. Wurman (2003). Mining eBay: Bidding Strategies and Shill Detection. Proceeding 4th International Workshop on Mining Web Data for Discovering Usage Patterns and Profiles.

Singh, P. and S.C. Jat (2020). A survey to detect financial fraud using deep learning approaches. *International Journal of Scientific & Technology.* 9(3). 16-20.

Tan, F. T. C., Guo, Z., Cahalane, M., and Cheng, D. (2016).Developing business analytic capabilities for combating ecommerce identity fraud: A study of Trustev's digitalverification solution. Information & Management, 53(7),878–891. doi:10.1016/j.im.2016.07.002.

Temponeras, G.S., S. N. Alexandropoulos, S. B.Kotsiantis, and M. N. Vrahatis, "Financial FraudulentStatements Detection through a Deep Dense Artificial Neural Network," 2019 10th International Conference onInformation, Intelligence, Systems and Applications (IISA),PATRAS, Greece. 1-5. Doi:10.1109/IISA.2019.8900741.

Thiagu, K., B. Maheswari, D. S. Kharthik (2014). An effective method for identifying online auction fraud using Traingulation detection techniques. *International Journal of Software and Hardware Research in Engineering.* 2(5). 15-21.

Trevathan, J., and W. Read (2018). Investigating shill bidding behavior involving colluding bidders. Journal of Computer. 2 (10). 63–75.

Tsang, S., Y. Koh, G. Dobbie, and S. Alam (2014). Detecting online auction shilling frauds using supervised learning. Expert Systems with Applications. 41(6). 3027-3040.

Tyagi and Sharma (2021). Implementation of fraudulent sellers detection system of online market places using machine techniques. *International Journal of Recent Technology and Engineerin (IJRTE).* 10(2). 194-198.

Vitor, P., E. Lughofer and A. J. Guimaraes (2019). Evolving Fuzzy Neural Network Based on Uni-null neuron to Identify Auction Fraud. Joint Proceedings of the 19th World Congress of the International Fuzzy Systems Association (IFSA), the 12th Conference of the European Society for Fuzzy Logic and Technology (EUSFLAT), and the 11th International Summer School on Aggregation Operators (AGOP). 3. 314-321.

Xu, H. and Y.-T. Cheng (2007). Model Checking Bidding Behaviors in Internet Concurrent Auctions *International Journal of Computer Systems Science & Engineering.* 22(4). 179-191.

Yoshida, T. and H. Ohwada (2010). Shill bidder detection for online auctions. Proceedings Pacific Rim International Conference on Artificial Intelligence. Berlin, Heidelberg. 351-358.

Zhang, X., Han, Y., Xu, W., and Wang, Q. (2019). HOBA: A Novel Feature Engineering Methodology for Credit CardFraud Detection with a Deep Learning Architecture.Information Sciences. Doi:10.1016/j.ins.2019.05.023.10.1109/UBMK.2018.8566574.

APPENDIX A
SOURCE CODES

```
Import pickle
import pandas as pd
from datetime import datetime
import time

from ebaysdk.finding import Connection as Finding
from ebaysdk.exception import ConnectionError
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, precision_score, recall_score

APP_ID = "DivineIk-s-PRD-59c7a1d25-67a7167e"
DEV_ID = "f3547dca-f775-4599-bb9f-671721a836a9"
CERT_ID = "PRD-9c7a1d252d9c-ce6c-47c1-b962-69f7"
TOKEN =
"v^1.1#i^1#I^3#p^3#r^1#f^0#t^Ul4xMF8wOjRBOEIzQUVDNkFDOTkxMUE1QzQ4MDgz
MjEzREFBOUQ3XzNfMSNFXjI2MA=="

def get_item_info(item_id):
try:
# Use the Finding API to search for the item
        finding_api = Finding(appid=APP_ID, config_file=None)
        api_request = {'keywords': item_id, 'outputSelector': 'SellerInfo'}
        response = finding_api.execute('findItemsByKeywords', api_request)
return response.dict()



except ConnectionError as e:
        print(e)
        print(e.response.dict())
return None

# Function to extract relevant data from the eBay API response
def extract_data(response):
    data = []
    item = response['searchResult']['item'][0]
    bid = item['sellingStatus']
    start_time = datetime.strptime(
        item['listingInfo']['startTime'], "%Y-%m-%dT%H:%M:%S.%fZ")
    end_time = datetime.strptime(
        item['listingInfo']['endTime'], "%Y-%m-%dT%H:%M:%S.%fZ")
# Calculate time difference
    time_left = end_time - start_time
    time_left_in_seconds = time_left.total_seconds()
try:
        item['listingInfo']['buyItNowPrice']['value']
except KeyError:
        item['listingInfo']['buyItNowPrice'] = {
            'value': bid['currentPrice']['value']}
```

```
# listingInfo
    data.append({
        'Bidder_Tendency': float(item['sellerInfo']['feedbackScore']) /
float(item['sellerInfo']['positiveFeedbackPercent']),
        'Bidding_Ratio': int(bid['bidCount']) / time_left_in_seconds,
        'Successive_Outbidding': int(bid['bidCount']) /
float(item['listingInfo']['buyItNowPrice']['value']),
        'Last_Bidding': int(bid['bidCount']) / int(to_seconds(start_time)),
        'Auction_Bids': int(bid['bidCount']),
        'Starting_Price_Average': float(bid['currentPrice']['value']) /
float(item['listingInfo']['buyItNowPrice']['value']),
        'Early_Bidding': int(bid['bidCount']) / int(item['sellerInfo']['feedbackScore']),
        'Winning_Ratio': float(item['sellerInfo']['feedbackScore']) /
float(item['sellerInfo']['positiveFeedbackPercent']),
        'Auction_Duration': time_left_in_seconds,
    })
return data

def train_shill_detector(data_path):
# Load the data into a pandas DataFrame
    df = pd.read_csv(data_path)


# Extract the features you want to use
    X = df[['Bidder_Tendency', 'Bidding_Ratio', 'Successive_Outbidding', 'Last_Bidding',
            'Auction_Bids', 'Starting_Price_Average', 'Early_Bidding', 'Winning_Ratio',
'Auction_Duration']]
    y = df['Class']

# Split the data into a training and testing set
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42)

# Train the model
    logistic_regression = LogisticRegression()
    logistic_regression.fit(X_train, y_train)

# Evaluate the model on the testing data
    y_pred = logistic_regression.predict(X_test)
    accuracy = accuracy_score(y_test, y_pred)
    precision = precision_score(y_test, y_pred)
    recall = recall_score(y_test, y_pred)
    print("Accuracy:", accuracy)
    print("Precision:", precision)
    print("Recall:", recall)
# Save the model to a file
with open('shill_detector.pkl', 'wb') as handle:
        pickle.dump(logistic_regression, handle,
protocol=pickle.HIGHEST_PROTOCOL)

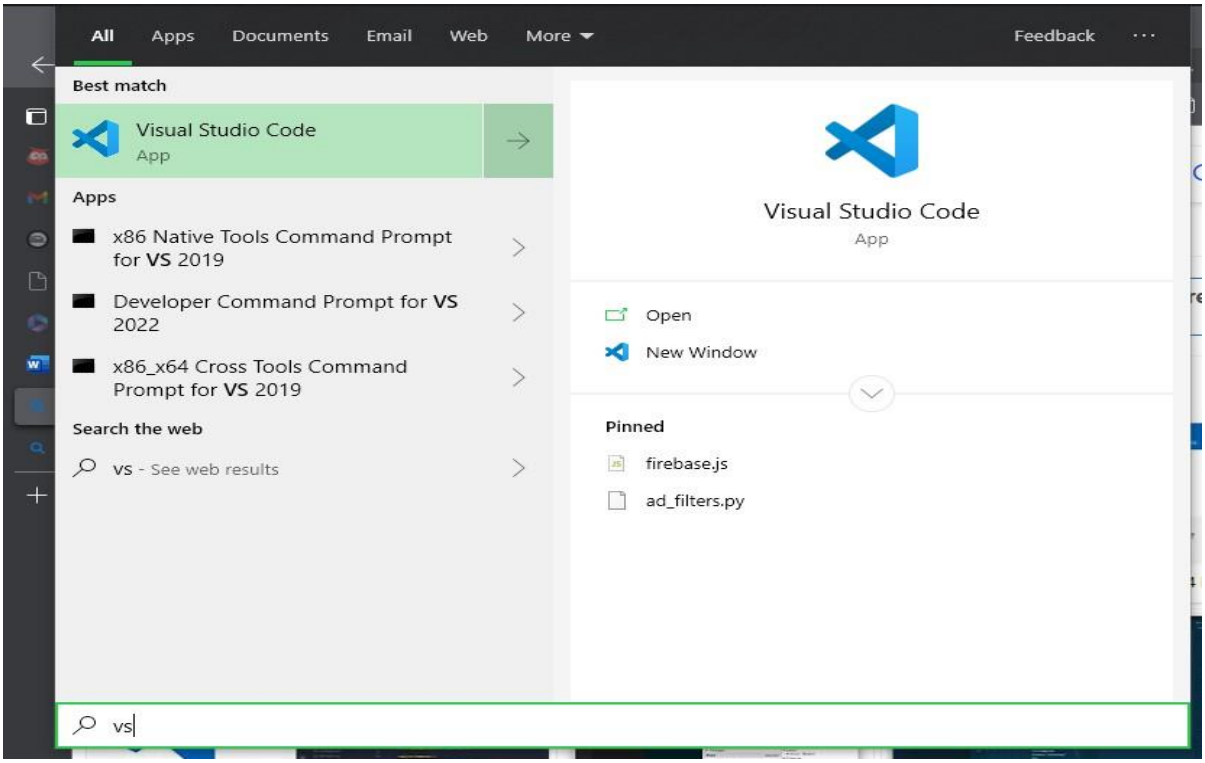
return logistic_regression
```

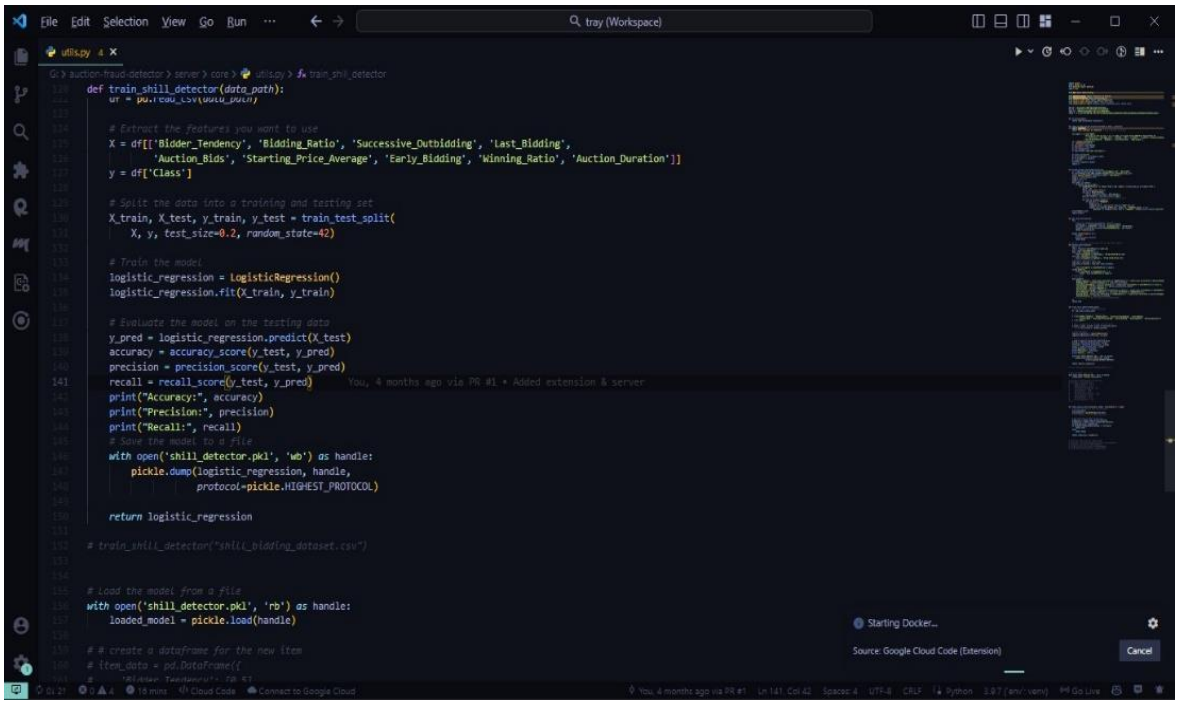
```
# Load the model from a file
with open('shill_detector.pkl', 'rb') as handle:
    loaded_model = pickle.load(handle)

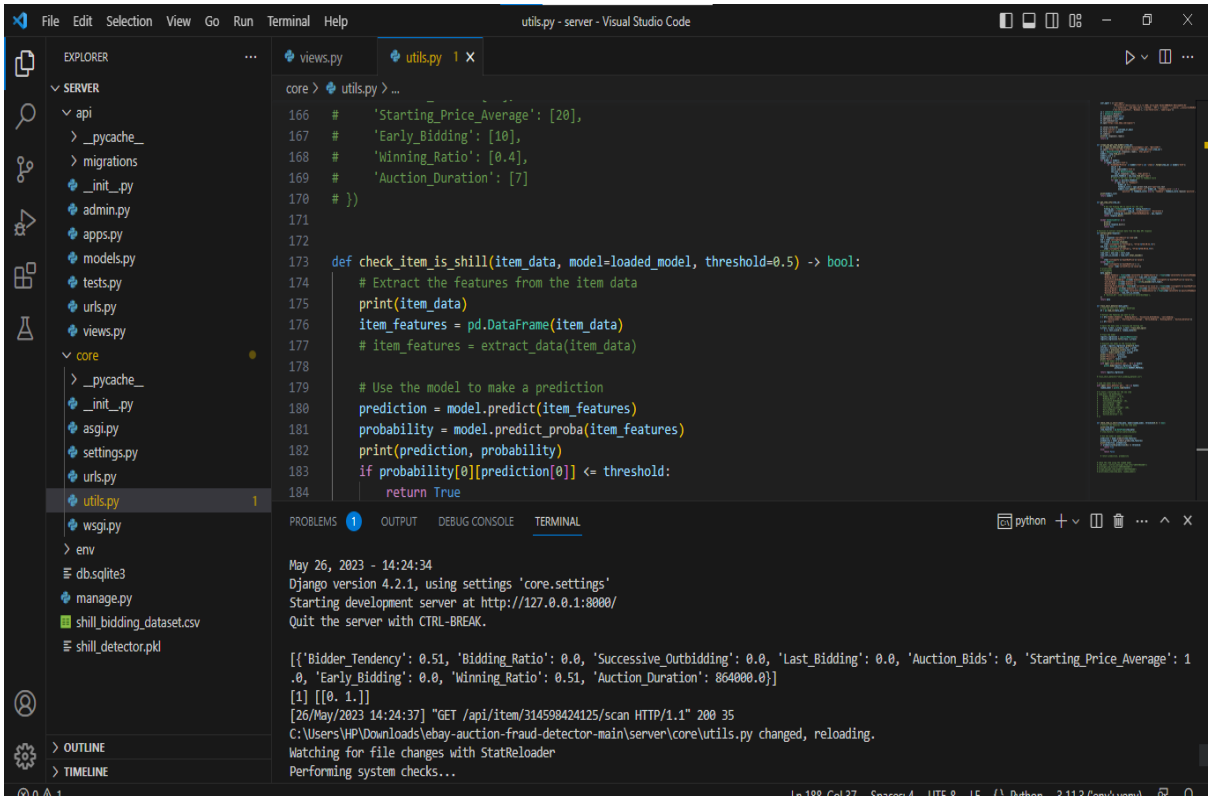

def check_item_is_shill(item_data, model, threshold=0.5) -> bool:
# Extract the features from the item data
    print(item_data)
    item_features = pd.DataFrame(item_data)
# item_features = extract_data(item_data)


# Use the model to make a prediction
    prediction = model.predict(item_features)
    probability = model.predict_proba(item_features)
    print(prediction, probability)
if probability[0][prediction[0]] <= threshold:
return True
else:
return False
```

APPENDIX B
PROGRAM OUTPUT SCREEN






File Edit Selection View Go Run Terminal Help
utils.py - server - Visual Studio Code
EXPLORER
SERVER
api
__pycache__
migrations
__init__.py
admin.py
apps.py
models.py
tests.py
urls.py
views.py
core
__pycache__
__init__.py
asgi.py
settings.py
urls.py
utils.py
wsgi.py
env
db.sqlite3
manage.py
shill_bidding_dataset.csv
shill_detector.pkl
OUTLINE
TIMELINE
views.py
utils.py 1
core > utils.py > ...
166 # 'Starting_Price_Average': [20],
167 # 'Early_Bidding': [10],
168 # 'Winning_Ratio': [0.4],
169 # 'Auction_Duration': [7]
170 # })
171
172
173 def check_item_is_shill(item_data, model=loaded_model, threshold=0.5) -> bool:
174 # Extract the features from the item data
175 print(item_data)
176 item_features = pd.DataFrame(item_data)
177 # item_features = extract_data(item_data)
178
179 # Use the model to make a prediction
180 prediction = model.predict(item_features)
181 probability = model.predict_proba(item_features)
182 print(prediction, probability)
183 if probability[0][prediction[0]] <= threshold:
184 return True
PROBLEMS 1 OUTPUT DEBUG CONSOLE TERMINAL
python
May 26, 2023 - 14:24:34
Django version 4.2.1, using settings 'core.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
[{'Bidder_Tendency': 0.51, 'Bidding_Ratio': 0.0, 'Successive_Outbidding': 0.0, 'Last_Bidding': 0.0, 'Auction_Bids': 0, 'Starting_Price_Average': 1
.0, 'Early_Bidding': 0.0, 'Winning_Ratio': 0.51, 'Auction_Duration': 864000.0}]
[1] [[0. 1.]]
[26/May/2023 14:24:37] "GET /api/item/314598424125/scan HTTP/1.1" 200 35
C:\Users\HP\Downloads\ebay-auction-fraud-detector-main\server\core\utils.py changed, reloading.
Watching for file changes with StatReloader
Performing system checks...

APPENDIX C
DATASET

Record_ID	Auction_ID	Bidder_ID	Bidder_Tendency	Bidding_Ratio	Successive_Outbidding	Last_Bidding	Auction_Bids	Starting_Price_Average	Early_Bidding	Winning_Ratio	Auction_Duration	Class
1	732	_***i	0.2	0.4	0	2.78E-05	0	0.993593	2.78E-05	0.666667	5	0
2	732	g***r	0.02439	0.2	0	0.013123	0	0.993593	0.013123	0.944444	5	0
3	732	t***p	0.142857	0.2	0	0.003042	0	0.993593	0.003042	1	5	0
4	732	7***n	0.1	0.2	0	0.097477	0	0.993593	0.097477	1	5	0
5	900	z***z	0.051282	0.222222	0	0.001318	0	0	0.001242	0.5	7	0
8	900	i***e	0.038462	0.111111	0	0.016844	0	0	0.016844	0.8	7	0
10	900	m***p	0.4	0.222222	0	0.006781	0	0	0.006774	0.75	7	0
12	900	k***a	0.137931	0.444444	1	0.768044	0	0	0.016311	1	7	1
13	2370	g***r	0.121951	0.185185	1	0.035021	0.333333	0.993528	0.023963	0.944444	7	1
27	600	e***t	0.155172	0.346154	0.5	0.570994	0.307692	0.993593	0.413788	0.611111	7	1
37	2172	o***u	0.6	0.5625	1	0.457631	0	0	0.457474	0.6	7	1
38	1370	7***3	0.5	0.105263	0	0.028692	0.052632	0	0.028654	0.666667	7	0
39	1370	i***r	0.017241	0.052632	0	0.057655	0.052632	0	0.057655	0	7	0
40	2236	i***i	0.041322	0.208333	1	0.286045	0.25	0	0.286025	0.818182	7	1
43	2236	_***r	0.142857	0.041667	0	0.387348	0.25	0	0.387348	0	7	0
44	2236	e***e	0.007246	0.041667	0	0.547427	0.25	0	0.547427	0	7	0
45	2236	a***n	0.01	0.041667	0	0.589286	0.25	0	0.589286	0	7	0
46	2236	t***s	0.025641	0.041667	0	0.912639	0.25	0	0.912639	0	7	0
53	2370	i***m	0.172414	0.185185	1	0.133783	0.333333	0.993528	0.048748	0.933333	7	1
54	2370	s***i	0.010309	0.037037	0	0.130995	0.333333	0.993528	0.130995	0	7	0
59	181	p***b	0.25	0.111111	0	0.994641	0	0	0.646852	1	1	0
65	181	i***0	0.023256	0.055556	0	0.917951	0	0	0.917951	0	1	0
67	432	g***s	0.041667	0.142857	0	0.107607	0	0	0.107607	0.842105	7	0
68	432	0***0	0.111111	0.714286	0.5	0.055501	0	0	0.053763	0.818182	7	1
71	1370	o***-	0.25	0.052632	0	0.333924	0.052632	0	0.333924	0	7	0
72	1370	0***e	0.052632	0.157895	0	0.460966	0.052632	0	0.31975	0.793103	7	0
73	1968	9***r	1	0.6	0.5	0.000201	0	0	4.86E-05	0	5	0

85	522	u***0	0.142857	0.041667	0	0.000316	0.25	0.993593	0.000316	0	3	0
88	522	r***a	0.017391	0.083333	0	0.662307	0.25	0.993593	0.000212	0	3	0
90	1847	9***m	1	0.090909	0.5	0.930026	0.454545	0	0.92999	0	7	0
93	1847	t***o	0.045455	0.030303	0	0.974153	0.454545	0	0.974153	0	7	0
103	522	0***e	0.070175	0.166667	0	0.347921	0.25	0.993593	0.347882	0.793103	3	0
107	522	a***_	0.066667	0.041667	0	0.765764	0.25	0.993593	0.765764	0	3	0
108	522	n***u	0.222222	0.083333	0	0.884132	0.25	0.993593	0.765652	0	3	0
109	522	r***0	0.052632	0.041667	0	0.827562	0.25	0.993593	0.827562	0	3	0
111	547	t***c	0.043478	0.166667	0	0.056414	0	0	0.056414	0.8	7	0
113	143	h***m	0.157895	0.375	0	0.523162	0	0	0.000306	0.727273	5	0
114	547	w***n	0.571429	0.666667	0.5	0.001976	0	0	0.001976	1	7	1
116	143	c***r	0.052632	0.125	0	0.024542	0	0	0.024542	1	5	0
117	2419	o***o	0.009804	0.02381	0	0.991195	0.571429	0.993528	0.991195	0	7	0
129	2076	r***b	0.210526	0.235294	0	3.8E-05	0	0	8.27E-06	0.833333	7	0
130	2165	e***z	0.481481	0.26	1	0.930847	0.64	0.993593	0.927181	0.666667	5	1
132	461	i***4	0.103448	0.136364	0	0.690207	0.181818	0.993593	0.690202	0.8	7	0
133	461	r***n	0.058824	0.045455	0	0.158763	0.181818	0.993593	0.158763	0	7	0
134	461	c***i	0.022727	0.045455	0	0.577358	0.181818	0.993593	0.577358	0	7	0
135	2076	n***n	0.025974	0.117647	0	1.16E-05	0	0	3.31E-06	0.833333	7	0
137	2076	n***s	0.043478	0.058824	0	0.007363	0	0	0.007363	0	7	0
138	368	3***3	0.071429	0.025	0	0.851111	0.55	0.993593	0.851111	0	5	0
141	1165	h***h	0.020833	0.017241	0	0.317963	0.689655	0.993593	0.317963	0	1	0
142	1165	n***n	0.025974	0.034483	0	0.319051	0.689655	0.993593	0.316725	0	1	0
143	1165	s***n	0.011765	0.017241	0	0.422697	0.689655	0.993593	0.422697	0	1	0
144	1165	0***h	0.034483	0.017241	0	0.492894	0.689655	0.993593	0.492894	0	1	0
149	368	r***s	0.142857	0.1	0.5	0.867056	0.55	0.993593	0.866942	0.9	5	1
151	979	s***c	0.111111	0.2	0	0.002004	0	0	0.001991	0.777778	7	0
153	979	s***-	0.2	0.3	0	0.619858	0	0	0.244773	0.7	7	0
154	368	a***9	0.083333	0.05	0	0.899944	0.55	0.993593	0.899907	0	5	0
155	979	e***e	0.014493	0.2	0	0.169709	0	0	0.142793	0.865672	7	0

159	368	c***o	0.181818	0.05	0	0.914979	0.55	0.993593	0.914968	0	5	0
160	979	d***u	0.105263	0.2	0	0.263722	0	0	0.263715	0.909091	7	0
161	979	c***9	0.125	0.1	0	0.426615	0	0	0.426615	1	7	0
165	1269	a***g	0.4	0.076923	0	0.949882	0.307692	0	0.949866	0	5	0
166	1269	_***i	0.1	0.038462	0	0.956627	0.307692	0	0.956627	0	5	0
169	2086	w***s	0.142857	0.030303	0	4.96E-06	0.454545	0.993593	4.96E-06	0	7	0
170	2086	s***s	0.010204	0.030303	0	0.027149	0.454545	0.993593	0.027149	0	7	0
171	2086	i***b	0.5	0.151515	0	0.727521	0.454545	0.993593	0.713036	0.75	7	0
176	550	1***i	0.04	0.2	0	0.000313	0	0	0.000313	1	3	0
177	818	n***t	0.010417	0.076923	0	0.964954	0	0.993593	0.964954	0	3	0
180	1001	b***b	0.012987	0.058824	0	0.990347	0	0	0.990347	0	1	0
191	352	y***s	0.027027	0.058824	0	0.914136	0	0.993593	0.914136	0	3	0
192	352	9***3	0	0.058824	0	0.001713	0	0.993593	0.001713	0	3	0
193	995	r***r	0.078261	0.219512	0	0.97282	0.560976	0.993593	0.779151	0.877551	3	0
194	995	p***u	0.75	0.073171	0.5	0.94162	0.560976	0.993593	0.941397	0	3	1
197	352	m***r	0.025	0.058824	0	0.002921	0	0.993593	0.002921	0	3	0
198	438	0***i	0.060606	0.4	0	7.11E-05	0	0	1.65E-05	0.736842	7	0
201	438	y***s	0.054054	0.4	0	0.019026	0	0	0.019021	0.857143	7	0
202	870	a***k	0.026667	0.043478	0	0.795313	0.608696	0.999353	0.795289	0	1	0
203	352	-***w	1	0.235294	0	0.945521	0	0.993593	0.945436	1	3	0
207	738	a***6	0.015152	0.033333	0	1.62E-05	0.4	0	1.62E-05	0	5	0
208	738	e***o	0.028571	0.033333	0	6.94E-06	0.4	0	6.94E-06	0	5	0
209	738	k***2	0.333333	0.033333	0	9.26E-06	0.4	0	9.26E-06	0	5	0
211	738	y***g	0.166667	0.033333	0	0.011752	0.4	0	0.011752	0	5	0
212	1349	r***r	0.008696	0.032258	0	1.16E-05	0.419355	0.993593	1.16E-05	0	5	0
214	1349	a***a	0.012085	0.129032	0	0.054347	0.419355	0.993593	0.035961	0.848214	5	0
217	1349	e***l	0.056604	0.096774	0	0.092199	0.419355	0.993593	0.001597	0	5	0
221	1045	c***2	0.033333	0.142857	0	0.892388	0	0	0.892388	1	7	0
223	870	d***a	0.08	0.086957	0	0.960856	0.608696	0.999353	0.84485	0	1	0
224	870	c***m	0.75	0.065217	0	0.875046	0.608696	0.999353	0.875046	0	1	0

242	1892	u***r	0.014085	0.066667	0	0.332959	0	0	0.332959	0	3	0
245	1349	r***1	0.090909	0.032258	0	0.171134	0.419355	0.993593	0.171134	0	5	0
246	1349	r***i	0.266667	0.129032	0	0.556299	0.419355	0.993593	0.4275	1	5	0
247	1349	d***a	0.18	0.290323	0	0.700718	0.419355	0.993593	0.697171	1	5	0
251	838	n***_	0.090909	0.029412	0	1.32E-05	0.470588	0.999935	1.32E-05	0	7	0
252	838	t***r	0.05	0.029412	0	0.020008	0.470588	0.999935	0.020008	0	7	0
255	1943	a***k	0.04	0.09375	0	0.851632	0.4375	0.993593	0.851632	0	3	0
256	1943	u***m	0.025641	0.03125	0	0.966609	0.4375	0.993593	0.966609	0	3	0
257	1661	r***g	0.034483	0.142857	0	0.698981	0	0	0.698981	0.823529	7	0
258	1943	1***v	0	0.03125	0	0.981138	0.4375	0.993593	0.981138	0	3	0
264	807	e***o	0.028571	0.04	0	0.1424	0.28	0	0.1424	0	3	0
265	807	f***f	0.466667	0.28	0	0.28995	0.28	0	0.28995	0.833333	3	0
266	807	.***.	0.454545	0.2	1	0.782917	0.28	0	0.782859	1	3	1
267	807	n***s	0.043478	0.04	0	0.905783	0.28	0	0.905783	0	3	0
272	807	n***t	0.010417	0.04	0	0.912967	0.28	0	0.912967	0	3	0
273	2079	i***o	0.030303	0.166667	0	8.1E-05	0	0	8.1E-05	0.866667	1	0
274	2079	o***l	0.028571	0.166667	0	0.000127	0	0	0.000127	0.761905	1	0
275	2079	i***v	0.058824	0.166667	0	0.00956	0	0	0.00956	0.75	1	0
278	25	n***l	0.054054	0.111111	0	0.509167	0	0	3.31E-06	0.777778	7	0
282	1644	a***k	0.026667	0.04878	0	0.073955	0.560976	0.993528	0.073943	0	7	0
284	1644	_***t	0.8	0.097561	0	0.341318	0.560976	0.993528	0.163504	0	7	0
288	556	r***r	0.026087	0.054545	0	0.990189	0.672727	0.999353	0.988337	0	3	0
299	1392	t***m	0.25	0.1	0	0.988219	0	0	0.988219	1	7	0
300	1392	t***l	0.043478	0.1	0	0.997488	0	0	0.997488	1	7	0
301	1392	4***b	0.5	0.1	0	0.993137	0	0	0.993137	1	7	0
302	1392	1***i	0.04	0.1	0	0.995761	0	0	0.995761	1	7	0
308	1577	x***i	0.2	0.025	0	0.836227	0.55	0.993593	0.836227	0	1	0
313	1610	i***n	0.025	0.076923	0	0.641578	0	0	0.641578	0	3	0
314	1610	s***l	0.085714	0.230769	0.5	0.6688	0	0	0.668765	1	3	0
316	1141	u***_	0.424242	0.666667	1	0.906998	0.142857	0	0.700131	0.8	3	1

329	1392	0***e	0.052632	0.3	0.5	0.855554	0	0	0.680514	0.793103	7	1
337	166	i***e	0.038462	0.166667	0	0.000127	0	0	0.000127	0.8	1	0
338	166	l***s	0.037037	0.166667	0	0.02956	0	0	0.02956	1	1	0
339	166	r***p	0.2	0.166667	0	0.246609	0	0	0.246609	0.5	1	0
340	166	0***i	0.030303	0.166667	0	0.341181	0	0	0.341181	0.736842	1	0
342	651	o***t	0.030769	0.2	0	0.000135	0	0	9.26E-05	0.782609	3	0
343	651	c***b	0.4	0.2	0	0.000772	0	0	0.000772	1	3	0
346	651	a***a	0.003021	0.1	0	0.475417	0	0	0.475417	0.848214	3	0
349	1370	t***n	0.009259	0.052632	0	0.534491	0.052632	0	0.534491	0	7	0
350	600	a***o	0.067227	0.307692	0	0.435184	0.307692	0.993593	0.424454	0.757576	7	0
354	600	r***h	0.008197	0.038462	0	0.559907	0.307692	0.993593	0.559907	0	7	0
355	600	u***-	1	0.115385	0	0.608808	0.307692	0.993593	0.554413	1	7	0
356	600	h***c	0.055556	0.038462	0	0.658451	0.307692	0.993593	0.658451	0	7	0
358	600	a***j	0.076923	0.038462	0	0.615486	0.307692	0.993593	0.615486	0	7	0
359	600	o***s	0.04	0.038462	0	0.636009	0.307692	0.993593	0.636009	0	7	0
362	181	o***o	0.009804	0.055556	0	0.996806	0	0	0.996806	0	1	0
363	1083	y***b	0.034483	0.083333	0	0.000336	0	0	0.000336	0	1	0
364	1083	u***4	0.25	0.083333	0	0.000463	0	0	0.000463	0	1	0
368	2370	a***r	0.014925	0.037037	0	0.137634	0.333333	0.993528	0.137634	0	7	0
370	2370	n***o	0.073171	0.111111	0	0.264478	0.333333	0.993528	0.264469	1	7	0
371	2370	a***y	0.043478	0.037037	0	0.905258	0.333333	0.993528	0.905258	0	7	0
373	2370	s***h	0.1	0.037037	0	0.311369	0.333333	0.993528	0.311369	0	7	0
374	432	5***_	0.125	0.142857	0	0.490602	0	0	0.490602	0.5	7	0
378	783	l***s	0.222222	0.214286	0	0.964861	0.357143	0.993528	0.964803	1	1	0
379	783	h***a	0.045455	0.035714	0	0.978727	0.357143	0.993528	0.978727	0	1	0
381	1370	6***h	0	0.052632	0	0.983547	0.052632	0	0.983547	0	7	0
382	1185	i***o	0.151515	0.3125	0	0.877667	0	0	0.778978	0.866667	7	0
383	1185	o***h	0.176471	0.1875	0.5	0.298953	0	0	0.298932	1	7	1
386	1185	h***u	0.235294	0.25	1	0.37564	0	0	0.375612	1	7	1
389	715	l***4	0.111111	0.090909	0	4.96E-06	0	0	4.96E-06	0	7	0

390	715	9***9	0.022222	0.090909	0	1.65E-06	0	0	1.65E-06	0	7	0
392	1185	r***0	0.052632	0.0625	0	0.387566	0	0	0.387566	0	7	0
399	2370	7***.	1	0.222222	1	0.762244	0.333333	0.993528	0.762216	1	7	1
400	1370	e***e	0.036232	0.263158	1	0.805121	0.052632	0	0.805116	0.865672	7	1
402	1083	h***h	0.020833	0.083333	0	0.102338	0	0	0.102338	0	1	0
403	1083	y***e	0.041667	0.166667	0	0.045394	0	0	0.045278	1	1	0
406	1264	q***0	0.125	0.166667	0	0.000162	0	0	0.000162	0.5	1	0
407	1264	a***a	0.003021	0.166667	0	0.000637	0	0	0.000637	0.848214	1	0
408	1264	k***w	1	0.333333	0	0.048912	0	0	0.045787	1	1	0
410	1264	s***o	0.038462	0.166667	0	0.145208	0	0	0.145208	0.769231	1	0
411	1264	k***k	0.028571	0.166667	0	0.362303	0	0	0.362303	0.888889	1	0
412	2534	r***o	0.069767	0.272727	0	0.106208	0	0	0.025736	0.866667	5	0
413	2534	c***a	0.035714	0.090909	0	0.058919	0	0	0.058919	0	5	0
420	496	v***t	0.130435	0.065217	0	0.982578	0.608696	0.999935	0.979428	0	7	0
422	1541	a***a	0.003021	0.0625	0	4.63E-06	0	0	4.63E-06	0	5	0
427	2273	m***c	0.2	0.021739	0	0.000139	0.608696	0.993593	0.000139	0	1	0
428	2273	0***c	0.5	0.043478	0	0.040706	0.608696	0.993593	4.63E-05	0	1	0
429	2273	b***n	0.2	0.021739	0	0.004236	0.608696	0.993593	0.004236	0	1	0
430	2273	n***d	0.025	0.021739	0	0.019063	0.608696	0.993593	0.019063	0	1	0
432	1582	m***b	0.066667	0.166667	0	0.608044	0	0.993593	0.608044	0.6	1	0
433	1582	n***0	0.105263	0.333333	0	0.471852	0	0.993593	0.471852	0.75	1	0
442	2273	z***l	0.333333	0.130435	0.5	0.738738	0.608696	0.993593	0.734051	1	1	1
446	276	s***n	0.023529	0.181818	0	0.022121	0	0	0.000336	0.935484	7	0
447	276	t***a	0.055556	0.090909	0	9.92E-06	0	0	9.92E-06	0	7	0