University of Hradec Králové

Faculty of Informatics and Management

DETECTION OF MANIPULATION IN ECONOMIC DATA USING DATA MINING TECHNIQUES

Master's Thesis

Ahmed Samir Mohamed, Alashrafy

June, 2015

University of Hradec Králové

Faculty of Informatics and Management

Department of Information Technologies

DETECTION OF MANIPULATION IN ECONOMIC DATA USING DATA MINING TECHNIQUES

Master's Thesis

Author: Ahmed Samir Mohamed, Alashrafy
Branch of Study: Information Management

Advisor: Dr. Petr Tucnik

Consultant: Dr Diyaa, Atiya
Center for strategic studies, chief operation officer

Hradec Králové June, 2015

ACKNOWLEDGEMENTS

First and foremost I would like to thank my advisor, Dr. Tučník Petr, for his guidance, patience, and support. My sincere thanks also goes to Dr. Diyaa Atiya, my professor in the Cairo University for his Fruitful Idea about applying data mining in the bid rigging detection field. I would also like to thank my parents Samir Alashrafy and Fatma Gamal for their support throughout my education. Finally, I would like to thank my new baby Yassine Alashrafy for pushing me to finish the desert earlier than I think.

Abstract

DETECTION OF MANIPULATION IN ECONOMIC DATA USING DATA MINING TECHNIQUES

Following the increasing use of public procurement all over the world and most of governments started to store the related data in a database, the subject of detecting fraud in procurement bidding has become more researchable and important topic. In this study, we will provide an overview of the bid rigging detection efforts from both public and private entities, then we will introduce an overview of the most important data mining techniques and algorithms which we will use to figure out how we can detect some bid rigging strategies using these techniques. The new approach in this research that we will use a real data from both private entity and Egyptian governments to show how it works better than the existence system for detect the bid rigging which almost depend on the hard evidence to detect it without using some advanced technical techniques for detection.

LIST OF FIGURES

LIST OF EQUATIONS

LIST OF TABLES

Chapter 1

Introduction

This proposal for the degree of Master in Science at Hradec Kralove University was made with the collaboration of the Egyptian government and the Egyptian International School. This particular topic about bid rigging detection using data mining techniques was selected after a discussion with delegates from both organization about their current needs to face these problems and their efforts so far to limit it. In this case several techniques was investigated to detect the bid rigging either automatically or manually , with the new techniques were proposed to prove it on the data that has been provided.

Background

Data can be collected from various resources with different ways but what made this data useful is the value in it. To extract this value it needs some tools for visualization and analysis in order to grasp the useful information. For this purpose, the early scientists used statistics and stochastic techniques and methods, over the years. These methods were improved to include other branches of science such as artificial intelligence, statistical analysis and computer science. This combination of methods and techniques has evolved to become what is now known as the Data mining field. Data mining provides us techniques that can combine different fields of science, in order to extract the useful, accurate and helpful information. These techniques were successfully used in both business and academic area, creating some useful results. For instance, financial field is one of the major application areas for the data mining techniques, the financial data in banking and financial industry generally relies on it to improve their performance. Here are the few typical cases[1]:

- Loan payment prediction and customer credit policy analysis.
- Classification and clustering of customers for targeted marketing.

[1] http://www.tutorialspoint.com/data_mining/dm_applications_trends.htm - 30 Jan 2015

- Detection of money laundering and other financial crimes.
- Visualization tools in genetic data analysis.

This work is primarily focused on the area of bid rigging in the public procurement procedures. Public procurement is the act of purchase to goods services by a public sector entity for achieving certain specified and identified objectives.

Bids are form of trade where one or more entities compete with each other to provide the best quality with minimum prices. So we can find many public and private sectors offering a lot of bids to get the best service for a good price. Also, there is another notion related to the bids - the auction. It is the opposite of the bid where the entity is offering goods/services and seeking the highest price. Both auctions and bids rigging can be realized online or offline even more can happen from single person, groups, companies or governments. When the bidders or competitors in the same market agree in advance on who is going to win which contract then we generally refer to this as bid rigging. Bid rigging has following common forms [69][70]:

- Bid suppression: i.e. agreement to refrain from bidding, or to withdraw a previously submitted bid, so that the designated winner's bid will be accepted.
- Bid rotation: i.e. agreements between firms to take turns as the winning bidder. Bid rotation schemes can be implemented in a variety of ways, e.g. based on tenders' values, or according to the market shares of the firms.
- Market allocation: i.e. agreement not to compete for certain customers or in certain geographic areas.
- Cover Bidding: i.e. higher, or unacceptable, bids that are only submitted to give the appearance of competition between bidders.

However, regardless of the strategy adopted, there are always signs that could expose the existence of bid rigging schemes. These signs include (according to [71][69]):

- The same company is often the lowest bidder.
- The same suppliers submit bids and each company seems to take a turn being the successful bidder.

- Some firms submit tenders that win in only certain geographic areas.
- Regular suppliers fail to bid on a tender they would normally be expected to bid for, but have continued to bid for other tenders.
- Fewer than expected numbers of competitors submit bids.
- Certain companies always submit bids but never win.
- A company appears to be bidding substantially higher on some bids than on other bids, with no apparent cost differences to account for the disparity.

Motivation

There is a lot of ways that can be used to detect the bid rigging but none of these has been efficient enough to reduce the loss from bid rigging as it leads to raising prices or lowering the quality of goods or services for the purchasing entity. Also, the bid rigging can be very costly to the national economy for any country, as tenders account for a large volume of economic output. In OECD countries, for example, public procurement accounts for about 15% of the Gross Domestic Products (GDP); the ratio is even higher in non-OECD countries [69]. Both public and private firms often rely upon a large number of suppliers for providing products and services. In addition to that, the same techniques used to detect bid rigging activities could also be applied to other forms of cartel agreements and anticompetitive practices. As it will be showed in discussions below, findings of this research are applicable to other domains as well, e.g. Regulatory Bodies, Auditors, and the Stock Exchange.

Research Structure and methodology

Some data mining techniques are suited for processing specific types of data, say categorical, while others are better used with other types of data, say numerical. Economic data, as in bid rigging conspiracies, further complicates the process by including a mix of different data types, e.g. names of firms (categorical) and prices (numerical). There is very few limited research on the use of data mining techniques in the realm of fighting bid rigging practices. Thus, to achieve this objective, we shall conduct an extensive survey on data mining techniques in the bid rigging domain as well as other related economic fields, e.g. stock markets and frauds in online bidding which will be introduced in **Chapter 2**.

In Chapter 3, will be for the bid rigging, forms and strategies and the contributions efforts to detect it. Then we list some screening technique that is used by some competition authorities to detect the bid rigging. In addition to this, we will introduce the systems to detect the bid rigging using some data analysis techniques which were produced by the Brazilian competition authority and South Korean competition authority.

Chapter 4, will be for introducing a case study from E-bay website to detect the fraud in the online auctions**.**

Chapters 5 and 6, will be dedicated to show our experiments and application for both of association rules and decision tree to detect the bid rigging. We will use the data that we get from both the Egyptian government and the Egyptian international school.

Chapter 2

Bid Rigging Detection

Bid Rigging Introduction and Overview

Both public and private sectors frequently depend on an extensive number of suppliers for provision of products or services. To keep suppliers net revenue competitive, large number of firms have their own procurement process, where the bid is rewarded to the supplier with the best esteem for the cash offer. Bid rigging is a form of manipulation in which the suppliers furtively plot to raise costs or bring down the nature of products or administrations for the buying firm.

In this chapter we will explore the bid rigging definition according to law and some international organization. In addition to that, we will explore the various forms of bid rigging and how we can detect them. In the next section we will show that effort has been made by authorities to detect bid rigging using advanced techniques.

Bid Rigging Definition

According to Organization for Economic Co-operation and Development (OECD), Bid Rigging can be defined as “An unlawful conspiracy, referred to as bid rigging (also collusive tendering), is a procedure whereby opponent parties while bidding, gather together to falsely raise the amount or quality of offered good / services to future buyers”. Within the competition law/ rules, this conspiring attitude not only has to be investigated but also sanctioned. Indeed, it is a criminal activity.

Bid Rigging – A Critical Criminal Act

Where experts are trying to educate a lot of people about bidders’ behavioral pattern as well as bidding strategies to give them the ability to prevent the bid rigging, a huge effort has also been made to reveal various forms and signs of bid rigging. Bid rigging is now deemed as a criminal offence over the world with individuals or firms found guilty are liable to under severe legal actions. Therefore, in this work we introduce this criminal act in much depth and detail to let the prospective bidders' gain proper understanding and not to indulge in any such activity causing them huge damages/ penalties later on.

Bid Rigging – Its Various Forms

OECD [47] has highlighted various forms with regards to bid rigging strategies:

- **Cover bidding:** Being the most commonly used form of bid rigging, cover/ complementary bidding is crafted in such a way to look like actual competition. It takes place when individuals or firms are in consensus to place either an unacceptable bid or a higher one in comparison to the bid placed by designated winner.

- **Bid Suppression:** This form of bid includes agreements whereby all or one of the competitors refrain from or takes back an already placed bid to let the designated winner's bid become acceptable.
- **Bid Rotation:** As the names says, in this bid rigging form, conspiring bidding firms decide their turns among themselves to be the winning bidder.
- **Market Allocation:** Here, the rivalry firms come into an agreement to not engage in a competing activity in selected geographical boundaries or for selected buyers.

Signs of Bid Rigging

Now we will introduce the different indications and warning signs of Bid Rigging

There are multiple ways to determine and avoid the risk of bid manipulation that can be used by procurement officials to eliminate or reduce the incidence of Bid Rigging. Below is a complete list of pointers which can be used by the officials in charge of accepting and supervising the conduct of bidding.

i.Warning Signs When Businesses Are Submitting Bids

i1. There are no changes to the names of the lowest bidders

i2. Some bidders only win in selected areas

i3. Regular suppliers fail to bid on tenders where they are normally expected to bid but are shown to continue bidding on other tenders

i4. Suppliers immediately withdrawing their bids

i5. The same company names submit bids but never win any

i6. Companies seemingly winning bids alternately

i7. Two businesses submitting a joint bid though one could have bid on its own

i8. Losing bidders joining the winning bidder on the project as sub-contractors

i9. The winning bidder does not accept the contract but ends up becoming a sub-contractor

i10. Meeting of competitors before the tender is awarded

ii.Warning Signs On Documents Submitted

ii0. The same typographical errors in submitted documents by different companies

ii1. Bid documents from different companies contain similar handwriting or typing fonts or use identical forms and stationary

ii2. Different companies submitting documents containing the same telephone numbers, fax numbers or glaring coincidences

ii3. When you see the same mistakes or same errors in calculation in the documents of different companies

ii4. Bids from different companies show a significant number of identical estimates for the cost of certain items

ii5. The letters from different companies have similar postmarks or post metering machine marks

ii6. Bid documents from different companies indicate similar last minute adjustments such as erasures or physical alterations

ii7. When you can see different company documents that have less details than expected or required

ii8. Competitors submitting identical bids in regular instances

iii.Warning Signs In Price Quotations And Adjustments

iii1. Some bidders submit overpriced bids without reasons or justifications

iii2. Estimated discounts or rebates unexpectedly disappear

iii3. Some competitors delay the submission of price quotations, or give quotations which are not compatible with the published cost

iii4. Large differences between the price of the winning bid and the losing bids

iii5. A suppliers bid matches a different suppliers bid in a similar tender

iii6. Significant reductions in project cost after a new company and infrequent bidder enters the tender, the resulting changes may show a disruption of an existing bidding alliance

iii7. Local suppliers giving unusually higher bids on delivery than outside suppliers

iii8. Similarities in transportation costs indicated by both local and outside suppliers

iii9. A bidder contacts wholesalers for pricing information prior to submission of bids

iii10. Glaring mistakes such as writing one hundred instead of one thousand.

iv.Bidders Statements Warning Signs

iv1. Spoken or written references that lead to agreements between bidders

iv2. Statements from bidders justifying their bids as based on standard market prices or suggested retail prices

iv3. Statements indicating that certain companies do not sell in particular areas or to particular customers

iv4. Statements indicating that some customers actually come from the same zone as the supplier

iv5. Statements indicating knowledge of competitors pricing or bid details or knowing in advance whether a company will win or lose the bidding even before publication of results

iv6. Statements indicating that a supplier has given courtesy, compliments, tokens, or placed a symbolic or cover bid

iv7. Different suppliers using the same expressions when explaining the increase in pricing

v.Warning Signs In Bidders Behavior

v1. Any meetings with competitors before the bidding process

v2. When you see any relations between competitors

v3. Another company requesting a bid package for a competitor

v4. Companies submitting their documents along with the documents of another bidder

v5. Companies that do not complete projects

v6. A company brings many bid forms, then opens and chooses one

v7. Competitors asking the same questions, do the same things and anything else in between

Tips and Tricks to Track out Bid Rigging

Experts like Bajari and Summers [48] have strongly suggested tests for Collusion,

Correlating bid decision and bid values: To find out correlations among bidders with regards to their entering *bid decisions or bid values*, it can be achieved via constructing econometric models, cost and market power parameters (inclusive of raw material cost, transportation expenses, closest rival's proximity and backlog etc.). Here, high correlation reflects greater chances of collusive actions as found in cover/ complementary bids.

Low covariance among costs and prices: Competitive markets claiming rationality and sensibility, bids must be showcasing costs. Comparing alongside collusion action, would no longer support this bid cost connectivity to let the conspirators get huge profits exceeding a usual competitive rate.

Bids' ranking in comparison to bidders' costs ranking: Bid ranking must be a reflection of what exactly is the bidders' costs, in order to eradicate the chances of a collusive agreement.

Scrutinizing the Win/ Lose Situation: Suspicion arises especially when a particular firm always gets leads for a specified procurement. Similarly, if few specific firms keep on entering unsuccessful bids for long, it be deemed as suspicious.

Circular Patterns: When the bids are placed by the same old group of suppliers in a way that their turns to become successful bidders appear as predefined.

Excessive Prices: Sometimes, a few bids are found exceeding the published price listing, either prior bids entered by same very firms or engineering cost assessments. Also, it may be noticed that a firm seems to place substantially higher on some bids in comparison to other, without any significant differences in costs due to disparity concerns.

Participants' Numbers Drop In: In this particular technique, only a few competitors place their bids, which is not a normal bidding behavior.

Effect of New Entrants: Here, as soon as a new bidder enters into the bidding activity, bid prices instantly decline.

Subsequent Contractors: It is observed that the one becoming successful bidder gives away the work to rivals who placed unsuccessful bids on the very same project. In another way, a firm takes back its successful bids to get the work subcontracted later by the new successful contractor.

Screens and Tests Options to Track Bid Rigging:

Through the statistical test referred to as "screen", industries having competition related issues are discovered to know about which of the conspiring firms are working on it [51] [52]. Competition Authorities like of which is the US Department of Justice and Federal Trade make use of screens to work as conspiracy detection tool by Commission (FTC) [52]. Mentioned below are some of the major uses of amazingly powerful screens:

- Defendants blamed for the act of collusion in order to help out creating the non-presence of either a conspiracy or its immateriality [53];
- To help and guide concerning decisions to provide room or not for leniency;
- While being at the class classification stage where genuine claims are supposed to be one and the same among class members; and
- To assess overprices and to calculate damages

In order to figure out any traces of collusion, various types of screens are applied. Some of the screen examples in this regard include:

Bid-rigging screens contingent upon improbable events*:* Suspicions for collusion become multifold when the bids are found in high correlation with one another to be the outcome of bidders' independent act/ conduct [54].

Bid-rigging screens depending on control groups: This particular screen highlights how good a bid is to reflect costs (the collusion lessens the connectivity among that of bids and costs in order to let the conspiring firms make extraordinary profits exceeding a normal competitive rate).

Screens Depending on Price-Cost Information*:* According to the US Department of Justice, the below mentioned patterns might highlight collusive behavior [55].

- Similar prices may specify a price-fixing conspiracy, exclusively when the prices stay same in the longer run; prices were dissimilar before; increased costs do not provide any reason for a rise in price.
- Element of discount is diminished particularly in a market having a significant discount history.
- Local customers have to buy from the sellers at rates comparatively higher than those charged to distant buyers.

Bid Rigging and Data Mining

Through various techniques, data mining enables to examine a company's information to find out patterns, associations and signs concerning deceitful acts. It allows government agencies as well as corporations to scrutinize huge data chunks much faster and cheaper.
Data mining is used by the competition agencies in a number of ways. As per agencies, most frequently used data mining efforts include [27]:

- Enhancing performance / services;
- Tracking out cheat, misuse and wastage
- Examining technical and investigative data
- Altering the tender procedure
- Handling human capital
- Discovering and tracking illegal activities/ patterns/ trends

South Korea System and the Federal Trade Commission

South Korea's Federal Trade Commission (KFTC) initiated using data mining techniques and statistical tool named as screen in the year 2006 to figure out bid rigging. Named as Bid Rigging Indicator Analysis System (BRIAS), its screening program investigates the collusive bidding chances in accordance with the bidding information obtained concerning public development projects run by the government, its financial institutions as well as the local authorities [57]. In the year 2008, success of this initiative made KFTC carry forward this screening program's application with every public institution giving information for BRIAS. KFTC has been aided by BRIAS in several ways including [58] :

- To effectively reveal conspiracy concerning bid rigging via allowing it to chronologically observe public sector tenders along with carrying out on-site examination of the ones having considerable signs of bid rigging
- Saved country's finances which previously used to go waste due to bid rigging
- Properly assisted in creating just competitive preferences

- By conveying signals to the firms about KFTC's watchful role over every single bid of public related tasks, made them willfully refrain from bid rigging

Federal Government System of Brazil

A powerful and effective government e-procurement system has been run by the Brazilian Federal Government. Referred to as COMPRASNET [57], it’s a virtual procurement structure. Under the system of COMPRASNET, firms involved in public purchasing get themselves registered by putting forward their demands concerning procurement. Resultantly, the system automatically generates and sends email messages to the already registered suppliers. The suppliers are also provided with the facility to download bidding documents at their end [59]. This virtually enabled procurement system ensures not only effective but also clear procurement without any traces of embezzlement. Besides, it helps saving time otherwise required to run the complex procedure. Last but not the least, it allows micro-level businesses to maximize their participation in government tenders.

Bid Rigging – Impact

Particularly in government procurement cases, bid rigging is spreading its wings like an epidemic leading towards high rise in prices in comparison to ordinary price-fixing [60]. It gears towards uncompetitive tender processes with companies agreeing to pay more price than usual or to compromise at poor quality good/services. Resultantly the businesses suffering from bid rigging are expected to pass on their ill fate forward to their buyers either in the form of low quality or extra price [61].

Bid Rigging – A Criminal Offence Globally

Referred as *private restrictions to competition,* bid rigging frequently paves the way towards criminal enforcement [62]. In the recent years, it has been declared as a criminal act almost all across the globe. Besides US and UK, strict legal actions (penalties and punishments) have been imposed against this critical concern in Canada, Australia, Italy, South Africa and Germany etc. In the United Kingdom, criminal prosecution against those involved in bid rigging is taken under the Enterprise Act 2002. In the United States, under Sherman Act Section 1, it is deemed as a delinquent criminal offence and Canada's marks it as a chargeable criminal offence under Competition Act's section 47. The US Antitrust Division, soon after August 2012, declared several bid rigging related guilty pleas, convictions and sentencing in multiple industries included within which are contractual bids for municipal bond proceeds, public foreclosure auctions and municipal tax lien auctions etc. For the purpose of investigating and prosecuting bid-rigging conspiracies, all three i.e. the Antitrust Division, US attorneys' Office and FBI have to join hands frequently [62]. Bid- rigging in Germany is accounted as a criminal act whereby the guilty one can be imprisoned for 5 years and/or charged with heavy fine [60].

Last year in April, the Australian Federal Court provided the bid rigging victims with damages worth Aus $22.4 million and in July, the same year, Canada's Competition Bureau pronounced JTEKT (Japanese bearings producer) as guilty to two bid rigging counts under the Competition Act [63].

Chapter 3

E-bay Case Study

E-Bay's Bidders Behaviors and Strategies- Case Study:

E-Bay, like other auction websites, is an amazing gateway for the masses to engage in virtual auctions. With the help of e-Bay videogame console auctions, it gets easier to figure out various bidding patterns. The data gathered this way shows which bidding behaviors are more popular, old or new. In this case, we therefore, suggest new attributes to get involved in bidding while identifying strategic rules of the game along with economic factors that can be a motivation to identify a new bid rigging behavior.

With the passage of time, internet auctions have gained considerable value with hundreds of thousands people actively participating via Yahoo, eBay and uBid, etc. Gaining high rankings due to heavy traffic, bidders' activity and action with regards to auction's life need to be clearly understood. The information obtained this way is beneficial for both the bidders and sellers where the former learns to properly use bidding strategies and crafting powerful software agents while the later gets valuable insight for revenue generation. The information gathered not only helps economists to analyze these social auction platforms but also assists in tracking down fraudulent tendencies, if any via normal and abnormal bidding attitudes.

Mining bidding data, though with multiple advantage, has to cope with various challenges. The very first in the series is that of bidder and auction system interaction having multiple attributes including time, value, product description and available bids, measurable either imprecisely or implicitly. Next is temporal data dependency, connected with both, auction time period and economy as a whole. Last but not the least is about increasingly sparse data. All these domain based challenges heavily impact data mining techniques.

This study is aimed to analyze eBay's bidders' behavior via seeking response for below mentioned queries. In this regard, data was gathered from around 12,000 auctions that took place at eBay.

1. Is it workable to categorize individual bidder behavior towards bidding?
2. If yes, then what are the mostly frequently used strategies on eBay?
3. Can considerable number of bidders be identified to achieve desired results?
4. Can fraudulent behavioral practices be discovered?

EBay - Model and Mechanism

EBay auctions, using an ascending bid format, allow bidders to respond within seller's defined fixed end time. Among the four standard auction choices are [18]:

1. Standard Auction: Most frequently used, this type offers only one item (or group of items) to the one placing highest bid.
2. Reserve Price Auction: Having a hidden reserve price with the seller (that needs to be exceeded before selling), as soon the highest bid becomes more or equal to that, the current item price increases to that reserve price.
3. Buy it Now Price: Through this option, a bidder can instantly achieve that item.
4. Dutch Auction: With seller offering items in quantity exceeding one, the bidder has the option to enter the desired number of items along with desired price per item. Eventually, the winners have to pay lowest winning bid price.

EBay makes use of proxy mechanism referred as bidder's proxy bid where bidders enter maximum bid, guaranteed by eBay to automatically highlight their active offer till reaching maximum bid value. This acclaimed auction site also enforces a minimum bid increment via a table showcasing a hike alongside the increased current ask price.

Data Collection and Interpretation

The data for auctions, we selected were of Sony PlayStation 2 console (PS2) and Nintendo Gameboy Advanced consoles (GBA). For PS2, anonymized data was gathered for two-days in October 2000 as well three weeks in January 2001 whereas for GBA, data collection was made from May 31 till July 29, 2001. The data reflecting a liquid secondary market, was a reasonable choice as a normal user would be interested to get just one of the consoles.

A spider was written to carryout historical data search for selected product category. It constructs the URLs to put forward the request for individual auction data as well as the history pages. These history pages have information regarding all the submitted bids to the auction which are then cached by the spider. To avoid overloading eBay's server, all requests are staggered eventually.

Data interpretation is subtle though. The bid history pages give no count of proxy bids due to which ask-price is unknown at the site. Since eBay's bid history page doesn't record reserve price, so in case of this type of auctions, the analysis becomes rather complex. The remaining two auction types also pose same like challenges. Therefore, our study focus encompasses only standard auction

Analysis and Results

We discovered that to get familiarized with some features of bidder's strategy, graphical examination would be of considerable value besides going through bid value charted over time. Figure 1 shows excess increment values of a bidder named 62013's bids for all-3-day auctions. A dotted line connects all specific auction bids with each auction marked by a different color. An "h" above a marked line represents the high bidder by the time bid was processed. For an accurate understanding of bidding behavior and auction context, it demands going back and forth between different views.

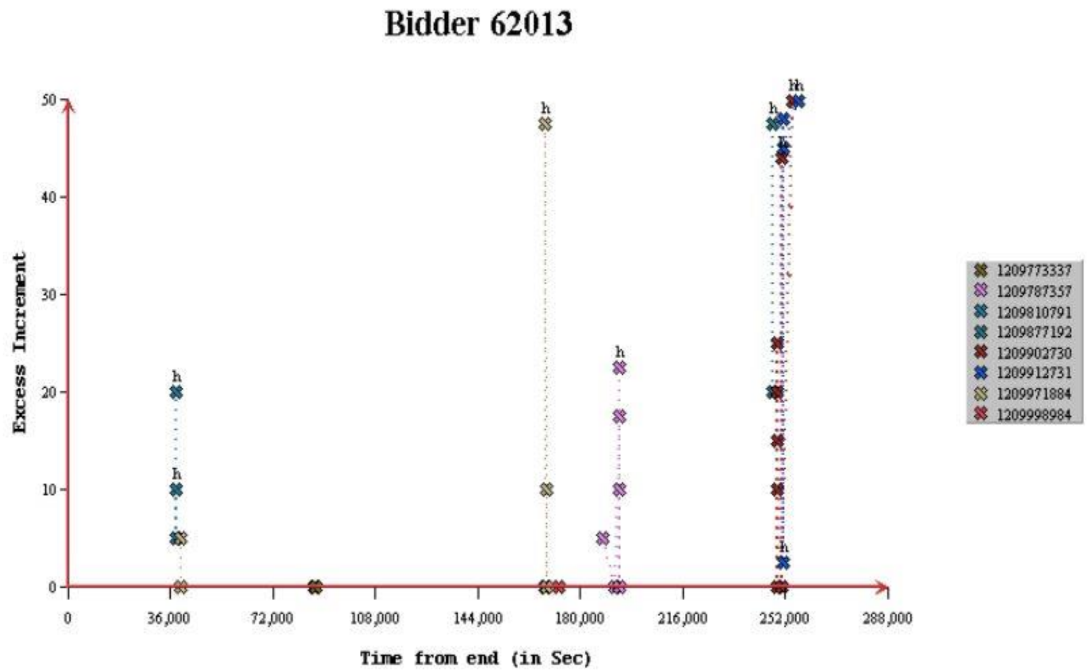


Figure 1 Bidder 62013 behavior according to time

The figure shows how bidder 62013 became the highest bidder. This is one example out of several patterns appearing in the data. Thus, for measuring these patterns' actual frequency within restricted dataset *Dr*, a test was made for labeling individual engagements (all bids placed by a single bidder in an auction). Here, Figure 2represents the distribution of engagement sizes.

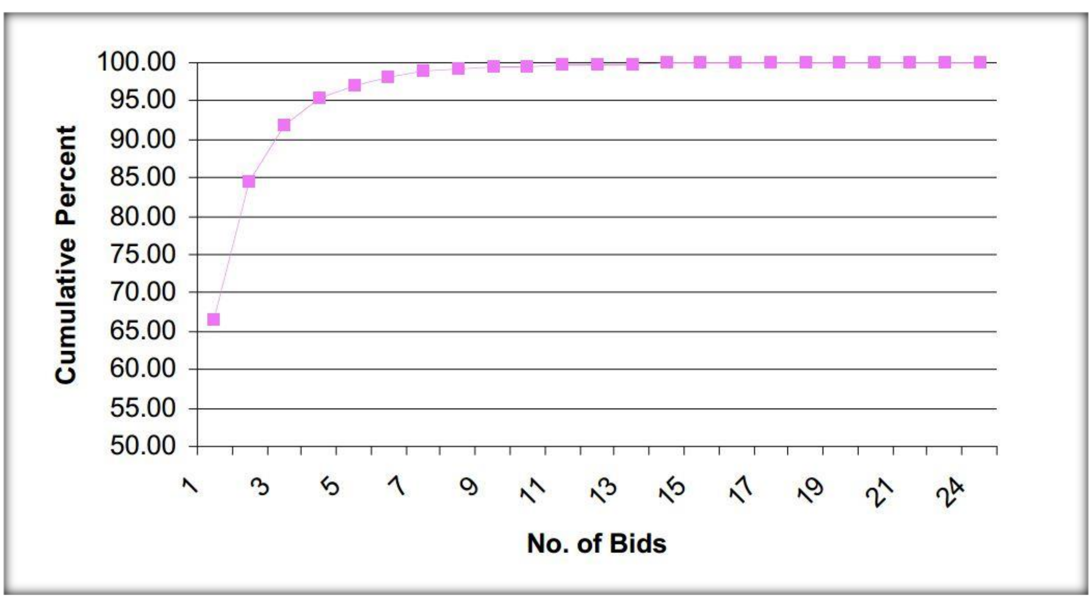


Figure 2 No. of bids engagement

Single Bid Engagements (E_1)

Regardless of the auction duration, most of the single-bid engagements reflect late bidding behavior i.e. about 58%. The practice of submitting the bids in the last minute is called sniping - an irrational approach indeed, is quite obvious which is against the eBay's advice as well as auction theory that demands placing bid at the earliest.

Experts have highlighted another bidder type i.e. Evaluators. Having clear understanding of valuation, they have following two major features.

1. Bid early and once with high value
2. Place significant bids in comparison to minimum bid demanded at that very hour.

A clear differentiation between snipers and evaluators is possible if we focus only on the engagements where single bid was placed a day prior to auction end time.   Yet, it’s not a task as easy we may be thinking off; primarily due to the nature of eBay game console auctions and secondly because of changes in bidder's evaluation alongside market trends.

Multiple-bid Engagements

One type for multiple bid engagements E_1, is skeptic behavior i.e. submitting multiple bids having zero excess increment, is quite common. A reason for so may be that the bidder is naive or skeptical concerning eBay's proxy system so would always like to go for minimum acceptable bid.

One commonly found behavior in E_2 is known as unmasking. It has a pattern of closely placed bids with variable excess increment. A significant reason behind this may be bidder's effort to reveal the maximum bid of another party's proxy bid. It is also possible that behind unmasking, the bidder's intention is to become highest bidder till reaching his willingness to pay limit or to move further with another auction.

Fraud Detection

Shilling is the most common form of online auction frauds by the seller where he attempts to raise the prices either through buy bids under aliases or associates. Very difficult to detect, it has various attributes especially workable at eBay. However, through association analysis between buyers and sellers (we adopted SAS Enterprise Miner for this purpose), it can be detected eventually.

Related Work to the e-bay case

Considerable amount of work and efforts have been made on online auctions. Researches look forward to pick eBay as the most popular auction platform. David Lucking-Reiley, et al [64] study eBay features' impacting final auction price, Houser and Wooders [65] go for checking feedback ratings effect on auction price, Roth and Ockenfels [66] look for late bidding and related strategic issues and last but not the least Unver [67] examines the evolutionary tendencies related to strategic multiple and last minute bidding through artificial agents.

Study Conclusion

This case serves as a useful guide for understanding online auction behaviors common at eBay like sniping, skeptic, and evaluator and unmasking. It also highlights bidding strategies helpful in achieving increased results, if adopted. It gives a detailed account of bid rigging, a legal offence in almost every court of law. Above all, it is focused on classifying engagements with shill detection as an important extension. It is expected that the data collection approach as well as bidding strategy models would assist in making online market place safer and viable for the users in the years to come.

Chapter 4

Data Mining Techniques and Algorithms

In this chapter we will introduce the different type of data structure and various methodologies to collect and clean the data.

We examine number of different data repositories on which mining can be performed. In principle, data mining should be applicable to any kind of data repository, as well as to transient data, such as data streams. Thus the scope of our examination of data repositories will include relational databases, data warehouses, transactional databases, advanced database systems, flat files and data streams.

We will introduce the various techniques to analyze the structured data and see which of them would be proposed to be our techniques. Also we will introduce the latest application and development to each of them.

Machine learning

Machine learning is an artificial intelligence technique for learning the machine how to deal making smart decisions and how to make a conclusions based on variables and properties of the input data in addition to finding a rule that can connect the data to each other to produce a new information.

The Machine learning aims to find a solution for a problem without specified algorithm so that the machine can evaluate new solutions according to the change in the requirements. We can say machine learning is not an algorithm, it’s a way to simulate a human thinking or brains by analyzing the data to detect the patterns associated rules then the machine can make a conclusions after sometimes these conclusion will be a knowledgebase where the machine can use as inputs to predict or analyze the new situations. These techniques can be divided into two categories; one is supervised learning and the other is unsupervised learning.

Supervised learning is the task of inferring a function from training dataset. The data consist of a set of examples to train the machine on it. Each example consists of inputs and outputs for a specific situation where the machine will be able to draw the function between them to use this function in the future. The function should be used to predict the output value from any input value similar to those were in training data. If the output was continues this function will be called a regression but if the output was discrete it will be called classifier.

Unsupervised learning, is the task of learning the system how to define the output without labeling or defining any rules for the system. We don't define anything. In such a case there is no explicit target, the algorithm will be able to cluster the input and try to figure out some relations without any predefined functions between inputs and outputs. Unsupervised learning is more popular than supervised as it seems like the real brain, in real world you don't tell the brain how to behave just it learn itself during the life cycle and each human has his behavior.

Data mining

According to IBM research, there are 2.7 Zetabytes of data exist in the digital universe today. This huge amount of data needs a lot of effort to analyze and need more powerful and efficient techniques to be used. The need for powerful analyzing tools and techniques raised up since 80's as more date produced and started to increase dramatically till now. Nowadays Google processing more than 20,000 terabytes of data (20 petabytes) a day.

That led the company to invest a lot of money in this area. Nowadays, Google, Facebook and Twitter have their own data centers to analyze the huge amount of collected data from various web services. Data Mining is referred to as a set comprising computer-aided expertise and techniques crafted in a way to mine bulks of assimilated data automatically so as to reach out for latest, concealed or unpredictable facts or unique patterns. According to Fayyad, et al [1] the term Data Mining is actually to explore out hidden information stored in huge chunks of data and can be valued as stepping towards the process of knowledge discovery. Also, Han and Kamber [2] have mentioned a variety of names used to describe this process such as knowledge extraction, data archaeology, data dredging and data/pattern analysis etc.

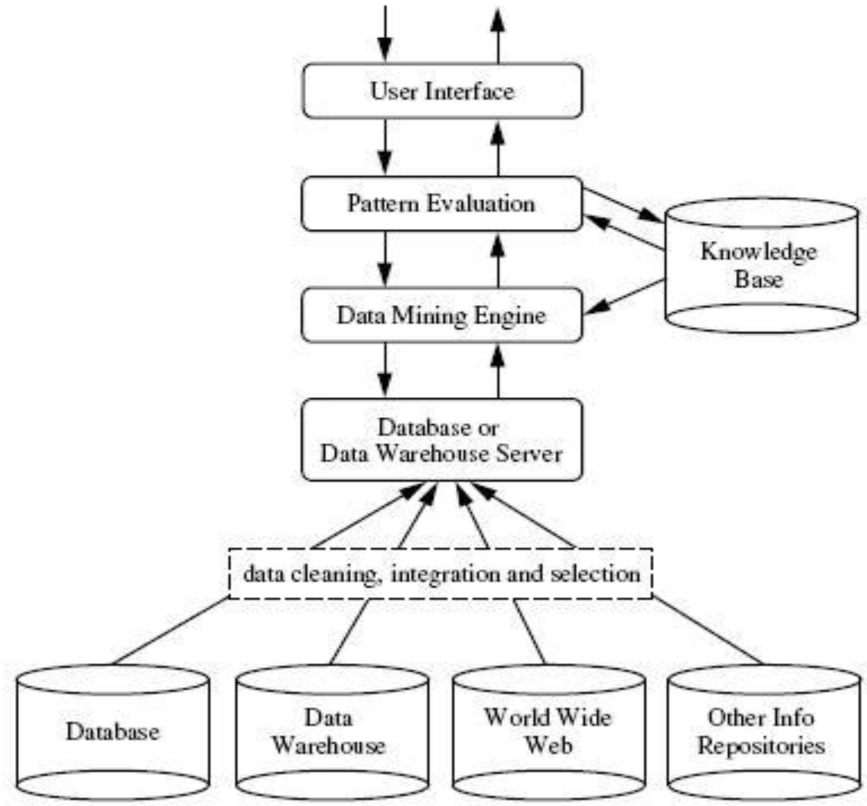


Figure 3 Architecture of Typical Data Mining System [33]

Small data sets can be extremely helpful in making use of traditional statistical analysis. Edelstein [3] highlights that the foremost simple analytical step with regards to data mining involves identifying the data, summarizing its statistical features i.e. means/ standard deviations, examining it through graphs/charts as well as finding out expected relationships between variables (among the values arising out simultaneously).

Declared as one of the top 10 latest technologies by the MIT Technology Review[4], the process of data mining implies various data analysis tools for finding out data related patterns and linkages for accurate estimation[5]. Emerged almost two decades ago, it has rightly proved to be the need of current times and as per ZDNET News [6] a remarkable technical advancement for the years to come. However, it is still believed and applied by a great majority as a step necessary in carrying out the procedure of knowledge discovery in databases (KDD) [7].

Data Mining makes use of a number of scientific fields, namely Database Systems, Statistics, Artificial Intelligence and Visualization along with Pattern Recognition. Significantly unique patterns are identified from datasets by implying a blend of various techniques from all these scientific areas [8]. The powerful analytical process of data mining involves the critical concepts namely bagging (averaging), boosting, Cross Industry Standard Process for Data Mining (CRISP) , data preparation, data reduction, feature selection, text mining etc, to be discussed in detail in this work. So we can define the data mining or knowledge mining as a process of exploration and analysis of a large quantities of data in order to either extract or predict meaningful information that can help to understand human behavior. To do that, data mining used different field of science and it increases with new challenges appears in these fields included: Statistics, Database systems, Pattern Recognition, Artificial Intelligence, and Machine Learning. Data mining can take place not just numerically but also textually and graphically. For the same reason, utilization of data mining can been seen in a multitude of simple and advanced databases; former include transactional database, relational database and object-oriented database while later enlist time-series database, legacy database, spatial database, text multimedia database, as well as World Wide Web[9].

In fact, for a human brain it may be a lengthy activity to explore meaningful information from data available in raw form. In this context, data mining technique comes in handy as it helps extracting out factual details from data which is not very obvious. Mainly, data mining is implied to aid data analysts in figuring out significant behaviors, observations and patterns interwoven within data.

Research on data mining and its applications is going on a massive scale[10] [11]. Effective application of data mining has also been installed by varied sort of organizations in practice. To gain an edge like cost reduction, reinforcing research, and up surging sales, professions like medicine, banking, retailing, and insurance companies are generally and irrefutably using data mining [12][13]. Techniques and skills of data mining are in persistent use with varied domains of astronomy, medicine, molecular biology, health care management, geology, tax fraud detection, sports and even in monitoring money laundering. Significance of data mining is due to its implication in numerous fields like optical character recognition, medicine, radar, speech recognition, multimedia, agriculture, vision and sonar.

University of Texas's Doug Alexander describes the data mining process as one aided by computer to dig huge data sets in depth and detail, access them and finally discover the hidden meanings. It is important to forecast future behavioral patterns and trends so as to assist businesses in making wisely proactive decisions. Through data mining tools business professionals can answer business related queries in a timely fashion. While searching through databases, these tools detect hidden patterns and thus by discovering probable future trends aid professionals to not miss out anything on account of being beyond their expectations [14]. Above all, it helps gaining a definite edge over rival firms/ industry.

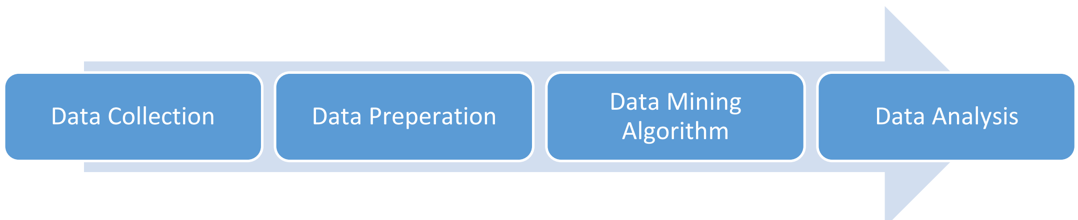


Figure 4 Data Mining Process

Traditional data mining is experiencing serious challenges in accommodating user's requirements as well as adjusting to the business needs Data-Driven Decision Management [DDDM]. Therefore, to ensure the achievement of commendable outcomes, data mining process model is formed that guides the user through series of steps. Fayyad, et al. [15] is anticipated to have been generated the first actual process model in the mid-1990's. Afterwards numerous venders and consulting firms have indicated several process models. A design called CRISP-DM — Cross-Industry Standard Process for Data Mining [16] has been developed by a team of venders and consumers in current times. The CRISP-DM Process Model comprises of six phases. Demonstrated in the Figure 3 given below, there is a cyclical procedure in which the six phases of data mining fit together. The likely iteration paths or feedback loops are signified by the arrows in the figure.

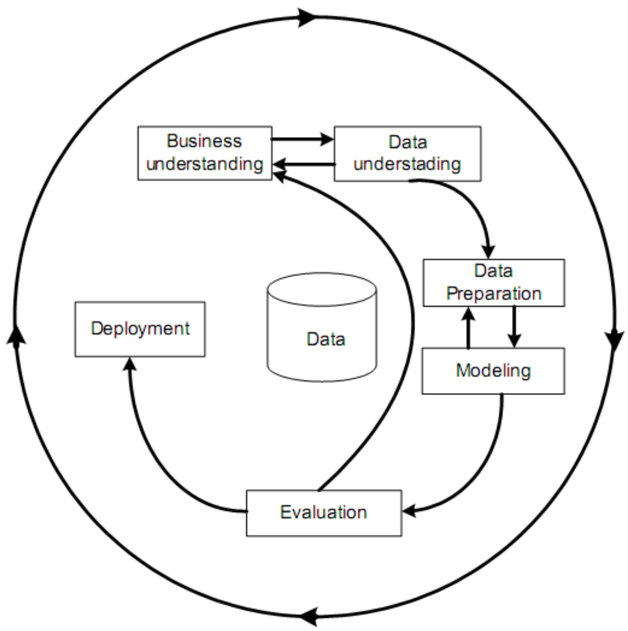


Figure 5 Phases of the CRISP-DM Reference Model [34]

1. Understanding Data: Gathers the data, searches through it to know about unique data subsets and highlights preparation requirements pertaining to data and identifies data quality concerns, if any. This phase is responsible for six major tasks concerning data i.e., to organize, gather/collect, describe, analyze, explore and verify available data.

2. Understanding Business: To comprehend and understand the business objectives and rules of a firm. Impractical and useless consequences might generate by misapprehending the problem resulting in the failure of the whole process, even if the algorithms selected yield precise and accurate results.

3. Preparing Data: In a non-sequential order, the process includes various tasks such as Outlier Handling, Transformation, Transferring, Binning, Missing Values Handling and Loading. Finally, this phase provides with prepared datasets along with reports explaining the entire series of processes.

4. Modeling: Methodology selection, test compilation for model quality authentication alongside purpose in hand validation, model and assessment development (making a comparison of test criteria and success criteria to the actual outcome achieved. To get full command on data mining and its complexities, both mathematical modelling view along with computation algorithm view should be deeply understood. Here the techniques being used include decision tree, association analysis, clustering, regression and time sequence analysis etc [17].

5. Evaluation: Makes an attempt to estimate the successful achievement of business objectives by implying selected algorithm or if this chosen algorithm proves defective due to some constraints.

6. Deployment: Outcomes achieved by data mining technique are deployed into business process. Tasks involved in this phase are plan deployment, monitoring, maintenance, final report making and last but not the least project review.

As per Two Crows Corporation, process model derives some insightful benefits from CRISP-DM [5]. Yet, it is dissimilar from CRISP-DM where the collection for data mining must be in a different database, due to following reasons:

- In case of using corporate data warehouse, data mining will let you become an efficient user of it while leading towards resource allocation issues.

- In case you are interested to bring some changes in data, the data warehouse administration may not let you do so easily.

- In case you are in need of understanding the data, the corporate data warehouse structure may refuse to provide supportive data exploration.

- In case you are willing to store data in a separate DBMS having unique physical design than the one supported by corporate data warehouse, then it may not workout.

Now we will introduce our model of using a data mining in our research. As illustrated in Figure 4, the data mining process comprises four main steps:

Data Collection

Data is collected from different data sources and later on this data is refined so that data redundancy can be minimized. Internal and external data is gathered from different sources either hard sources like the documents or soft sources like web, images, spread sheets or databases. Any kind of data can be collected either it was structured or not.

Construction of data set from this randomly collected data is the next stage of data preparation and perhaps this is the most important part of data preparation.

The best possible practice is to start with a single initial dataset, this would not only help in getting familiarize with the data but also would assist in extracting good useful data. Data preparation is considered to be one of the most time consuming process of data mining [31] and since it is hectic process so there are a lot of chances of errors. Therefore, a database expert needs to be very vigilant while preparing the data set. Extra care needs to be taken for the data set as there is great chance of redundancy and invalid values. If the data is not screened properly then this data can lead to results that are highly invalid. After all the success of the entire data mining process depends on the data preparation process. For Data Mining, while using SAS (Statistical Analysis System) , in order to eliminate the chances of bottleneck during data preparation process, the size of the data is of paramount importance. It is especially to be taken care off while using transformation in case of scoring process [32].

Once the data is collected from all the available sources the next step is to clean data and this process in the language of computer science is known as cleansing data. Data is analyzed carefully and the tables of the database are searched for all sorts of data that is invalid or would repeatedly give null values. Main step in data mining is to remove all the data repetition and this process can only be achieved through quality data cleansing process.

Tables of database need to be looked up and down and repeated and test queries are made in this process to make sure that result obtained from these queries is correct and valid. Unclean and invalid data in the database tables need to be removed from the tables as quickly as possible, because if this is not done then the result of the queries would not be accurate. So in a nutshell one can say that data preparation and data cleaning processes are the backbones of the data mining projects.

Data Types and Structure

In our analysis we can deal with different type of data either it is structured data or unstructured data. Structured data can be defined as the data in the fixed field that is located in a file. This type of file can include spread sheets, database file or text files etc. the way of structuring the data depends on the business model or the application that you will use the data in it and how we will use this data in the future that would specify the way of storing, processing the data. This includes fields' definition starting form the name of the fields, data type and constrains on these fields. The name of data should be related and understandable to the business community that the data deal with them. Data types could be numbers, text, symbols. The main advantage of structured data that we can analyze, store, enter and querying on it with reasonable cost and with efficient way. Sometimes we can face some obstacles due to the limitation in the capacity of the huge amount of structured data that we need to work with it. These days we overcome most of this problems with different technologies to handle the big data. There are unlimited format of structured data included CSV, text files, spread sheets, database etc.

Anything that can't be stored in a structure way then it should be stored on documents and this kind of what we can call unstructured data. Unstructured data is that type of data that don't have a unique way of reading where the info is hardly extracted without human help. We can say everything which is not a structured format we can say it is unstructured data like the PDF files, webpages or PowerPoint presentation also the images and photo categorized as unstructured data. There are a lot of techniques to convert it to structured data we will deal with some of these techniques within next section to prepare the data to be analyzed. Each type of these data can be analyzed using different techniques suit for it.

Data Sources and gathering tools

We have many sources that we can collect the data from it either it was structured or not. Public datasets is an example of data that can be used for analysis. Also we can collect the data from different servers all over the world like if you need to collect twits from twitter website. Another interesting example of collecting data is to collect the audio files from Youtube website to analyze it and detect the relations between it for a specific users. So we need to extract data from data warehouse into database. There is a lot of tools which are able to extract the data from different sources as a lot of companies invested in this area the most famous tools was produced by Microsoft, IBM and Oracle but because of the cost of this software there are also open source tools that can be comparable in the quality with this produced from a big tools like: Weka, Rapid miner, R-Language[2].

Data Preparation and cleaning

Relevant data is selected, cleaned, and preprocessed under the guidance and knowledge of domain experts who capture and integrate both the internal and external data into a comprehensive view that encompasses the whole organization. In this stage we are going to detect and evaluate the business problem and try to figure out the solution and what is the data format and features appropriate for our problem. After preparing the data we then go for removing and detecting the errors and inappropriate values from the data like:

- Negative values in sales
- Empty fields for the name of the product field
- Misspellings in the data entry
- Redundant data

All these kind of errors can be removed or not depend on the business model. Sometimes we need to keep redundant data or null values for some purpose. In this stage we can make advanced cleaning of data like remove the outlier from the data.

After data cleaning stage, we should meet the following requirements:

[2] http://www.r-project.org/

- Remove noisy and errors in the data either it was collected form single source or from multiple sources. We should use any of the mentioned tools to ensure the quality of the produced data and to ease the process of integration of multiple sources data.
- High degree of validity which ensure that every record in the data reflect a valid data not fake or incorrect data in reality like ensure that the streets' names in a specific city are already exist in this cities.
- No confliction in the data where there is no two records conflict each other.
- Furthermore, data cleaning should not be performed in isolation but together with schema-related data transformations based on comprehensive metadata. Mapping functions for data cleaning and other data transformations should be specified in a declarative way and be reusable for other data sources as well as for query processing. Especially for data warehouses, a workflow infrastructure should be supported to execute all data transformation steps for multiple sources and large data sets in a reliable and efficient way.

The data mining algorithms

The analyst chooses the appropriate data mining technique to apply. There is a wealth of literature on data mining techniques and tool support. The choice of a specific data mining techniques depends primarily on the problem investigated, the type of data considered (numeric, categorical, etc.), and the information/patterns to elicit from the data.

Data mining techniques are implied in order to identify the kind of patterns present in data mining tasks which are of two types namely descriptive and predictive. The former one illustrates basic properties of data in the database while the later one executes inference on the latest available data for making predictions [2]. In general data mining tasks can be divided into two main parts one is descriptive data mining and another is predictive data mining. Figure 4 describe the data mining tasks as in the following sections we will introduce the main data mining techniques under the descriptive and predictive techniques and brief introduction to show how to use it.

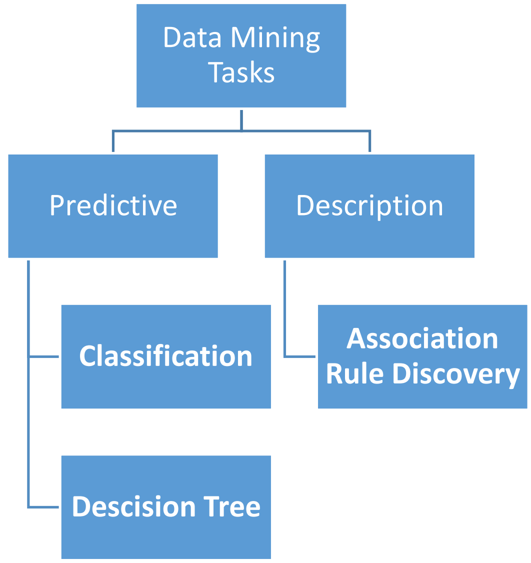


Figure 6 Data Mining Tasks

1. **Classification:** With the help of this technique, database records are classified into various predefined classes as per certain criteria.
2. **Prediction:** In data mining, prediction technique is implied to find out an event's possible future outcome. Predictive analytics ranging a number of algorithms so as to use them for characterizing past facts and figures [18].
3. **Regression:** Regression, being the most commonly used statistical function in data mining [19] is used to predict a number i.e. distance, weight, income, age etc. It starts in data set when the 'target value' knowledge already exists [20].
4. **Time Series:** This model or technique of data mining explains a series of an event's values over a specified time. It helps in expressing the component features of dataset in series to be perfectly categorized.
5. **Clustering:** Clustering helps in discovering not only dense but also sparse regions in a dataset. It aids in exploring data distribution and patterns. The underlying clustering principle is based on the concepts of similarity metric or distance metric. For numerical data clustering algorithm are classified as partition and hierarchical [21].
6. **Association:** The association model is implied to figure out affinities referred to as rules within the data collected. Preferably, in Market Basket Analysis, this model is applied.

7. Sequencing: To highlight patterns over a period of time, this model is adopted. It is employed as a generative device for almost all sequences under study [22]. Most frequently used by catalog firms its major application can also be found in the field of finance.

8. Characterization: To summarize basic features of a target class within data, characterization model is implied.

9. Comparison/ Discrimination: In order to compare the general characteristics/ attributes of target class data objects with that of contrasting class data objects, comparison model is adopted. User, here, has the choice to select both target and contrasting classes. Resultantly, SQL queries are used to recover relevant data objects.

10. Outlier Analysis: Dealing with what is referred as ' data objects not the ones in harmony with basic attributes of the data', the outlier analysis model is mostly adopted in tracking out fraudulent activities. Outlier analysis helps revealing credit card's illegal usage by making purchase detection of huge amounts in comparison with normal purchases made through the same account. Outlier values may be tracked with regards to location, purchase type and purchase frequency [2].

11. Evolution Analysis: It describes data object's trends and regularities undergoing behavioral change with time. Among its exclusive features are included sequence/ periodicity pattern matching, similarity-based data analysis as well as time-series data analysis [2].

Predictive Data Mining

Just like statistics, predictive analytics revolves around discovering considerable relationship between variables and representing them via models. Included among the variables are response variables (which we try to predict) and explanatory variables (which we observe, control or alter) to be related to the response later [23].

Classification

Classification is categorized under predictive analysis technique. It makes use of a training data set to discover a model/ function which in turn explains and differentiates among data classes. This model/ function can be further used to forecast the class of other data objects having unidentified class label. Classification model, thus achieved, may be reflected by a variety of ways. Famously used forms of classification model include Decision Tree induction, Bayesian classifier, IF-THEN Rules, K-nearest neighbor, Support Vector Machine and Neural Networks.

Classification by Logistic Regression

A lot of statistics is concerned with predicting the value of a continuous variable, e.g. price, cost, etc. This kind of statistics dominates economic analysis. But what do we do if the dependent variable is binary? What if, for example, we are running a study where we want to predict whether or not a market is free from anticompetitive conducts based on based on various factors, e.g. number of players, entry barriers, etc. Or, what if we are running a study where we want to predict whether bid rigging exists based on number of participants, costs, prices, historical participation, etc. In this case, the dependent variable can only have two values: Yes and No. One way around this problem is to simply use standard linear regression, and treat the dependent variable as if it was binary. If the two values were coded as 1 and 0, then any value of 0.5 or above would be treated as a 1, and anything below 0.5 would be treated as a zero. However, this approach is not rigorous and can lead to many problematic results, most notably violating the assumptions of linear regression that the error variances (residuals) are normally distributed.

Like ordinary regression, logistic regression is used for describing the relationship between one or more independent variables (X) and a binary response variable (Y). However, the goal of logistic regression is a bit different, because we are predicting the *odds* that Y is equal to 1 (rather than 0) given certain values of *X*. So, instead of the normal regression equation

$$Y = B_0 + B_1 X + error$$

We use the equation

$$ln\left(\frac{P}{1-P}\right) = B_0 + B_1 X + error$$

With simple algebraic conversion, the above formula comes to the following logistic function:

$$P = \frac{e^{B_0 + B_1 X + error}}{e^{B_0 + B_1 X + error} + 1} = \frac{1}{1 + e^{-(B_0 + B_1 X + error)}}$$

Where P is the probability that Y is equal to 1 and e is the base of the natural logarithm (≅ 2.718)

The logistic function takes as an input any value from negative infinity to positive infinity, but confines the output to values between 0 and 1.

Because the relation between X and P is nonlinear, the regression coefficients are now interpreted in differently. The value of B_0 yields the *odds* of Y equal to 1 when X is zero. The value of B_1 adjusts how quickly the probability changes with changing X a single unit. So, if we take the exponent constant (e) and raise it to the power of B_1, we get the *odds ratio*. For example, if $B_1 = 0.75$, the odds ratio is $e^{0.75} \cong 2.12$. This means that the probability that Y equals 1 is twice as likely as the value of X is increased one unit. An odds ratio of .5 indicates that Y=1 is half as likely with an increase of X by one unit; that is, there is a negative relationship between X and Y. An odds ratio of 1.0 indicates there is no relationship between X and Y.

There are many advantages to using logistic regression. It is more robust; the independent variables don't have to be normally distributed, or have equal variance in each group. It does not assume a linear relationship between the independent variable and the dependent variable. There is no assumption on homogeneity of variance or normal distribution of errors.

The advantages of logistic regression come at a cost. Logistic regression requires large amount of data to achieve stable, meaningful results. With standard regression, typically 20 data points per predictor is considered as a reasonable lower bound. For logistic regression, at least 50 data points per predictor is necessary to achieve stable results. Another cost is that there is no R^2 to measure the variance accounted for in the overall model. Instead, a chi-square test is used to indicate how well the logistic regression model fits the data.

Classification of K Nearest Neighbors

K nearest neighbors are mostly used as a simple algorithm containing the combination of storing of all variables and also used for the searching of new cases/variables related to the similarity of old and new methods or formulas. Within the 1970's, KNN was developed to estimate statistical and other related methods considering non-parametric technique. Algorigthm related to the above mention theory, here, a case is discussed where K nearest neighours is calculated by a distance function. If K=1, then the case is considered as the class of its nearest neighbours [38]

$$\sqrt{\sum_{i=1}^{k}(x_i - y_i)^2} \tag{1}$$

Equation 1 Euclidean Distance

$$\sum_{i=1}^{k}|x_i - y_i| \tag{2}$$

Equation 2 Manhattan Distance

$$\left(\sum_{i=1}^{k}(|x_i - y_i|)^q\right)^{1/q} \tag{3}$$

Equation 3 Minkowiski distance

As we can easily see by these equations, these three distance formulas are representative of continuous variables. As categorical variables are used within these equations, therefore Hamming distance is used. As these categorical variables are used, therefore an issue occurs containing standardization of numerical variables within the set of 0 and 1 [38]

$$D_H = \sum_{i=1}^{k} |x_i - y_i| \tag{4}$$

Equation 4 Hamming Distance

$$x = y \rightarrow D = 0$$

$$X \neq y \rightarrow D = 1$$

With the inspection of the data and overall situation, choosing the optimal value of K is important. Using of large value of K would reduce the noise problem but there is no guarantee in it. Another way of setting up a best value for K is through Cross-Validation. Throughout different equations, the value of optimal datasets for K is between 3-10. Through this way, a much better result could be gained instead of 1NN [37]

Example:

Here, credit default is shown. Age and Loan are used as two variables. Whereas, default is used as a target.

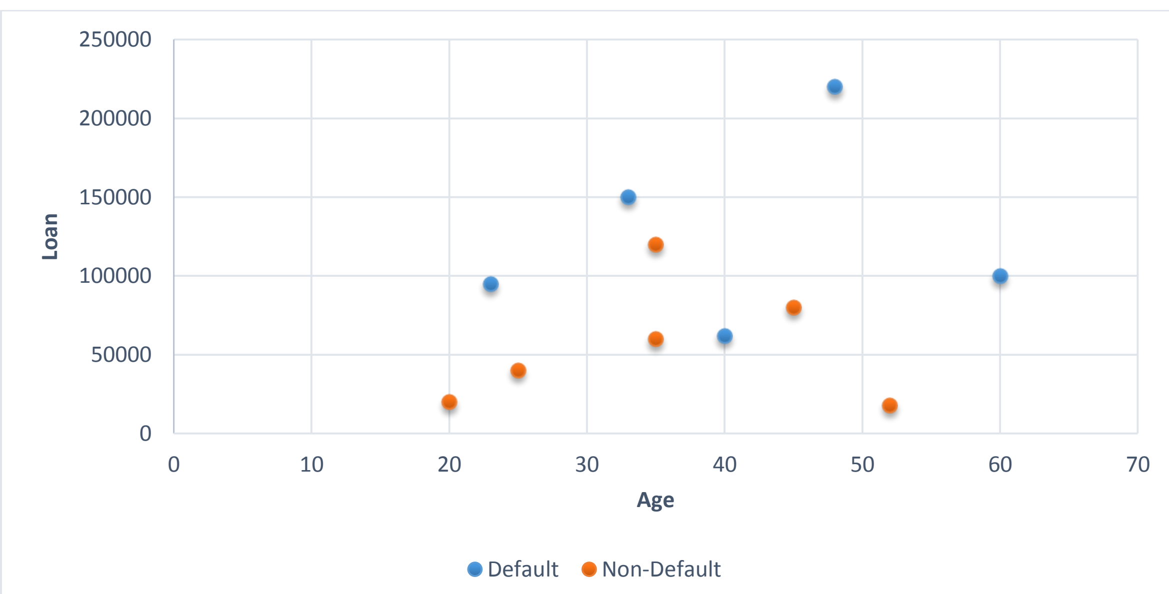


Figure 7 Credit Default Classification

In Figure 7 Considering Age=48 and Loan=$142,000. If we take K=1, then the value for NN containing Default=Y would be

$$D = \sqrt{((48 - 33)^2 + (142000 - 150000)^2)} = 8000.01$$

Equation 5 Disctance Calulation for Age = 48 and Loan =142000

Age	Loan	Default	Distance
25	40000	N	102000
35	60000	N	82000
45	80000	N	62000
20	20000	N	122000
35	120000	N	22000
52	18000	N	124000
23	95000	Y	47000.01
40	62000	Y	80000
60	100000	Y	42000
48	220000	Y	78000
33	150000	Y	8000.014

Table 1Calculation for the distance for Age = 48 and loan =142000

According to calculation in Table 1 Then if K=3 the default would be Y.

Classification of SVM

In order to gain values for SVM, such arrangements are conducted that have less amount of errors in them.

$$\frac{1}{2}w^T w + c\sum \varepsilon_i$$

$$y_i\,(w^T\varphi(x_i) + b) \geq 1 - \varepsilon_i, i = 1,2,\ldots,N$$

Here, C is used as a capacity constant, w is vector of coefficients, b is another constant and ε_i shows that parameter that is used for the calculation and handling of non-separable data (inputs). The index i is mainly used to show N training cases. x_i Represents independent variables. The kernel φ is mainly used to transfer data from input. Greater the value of C shows there would be large number of errors that would occur and that would be caught. Therefore, C must be used and chosen with great care and concern

So to produce the regression on SVM:

Y = f(x) + noise

Hence, in here the major task is to find out such functional form of f that is used to predict such new cases that were not used before for the representation of SVM. Sample test could be used as an example or as a practical work for SVM in order to take out our desired outcome. Example : set of training, it must be composed of a process having optimization on the basis of sequential system for the determination of an error function. Clarifying this error function, two types of SVM models could be taken out from it. [37]

The basic error function that is calculated for this type of error function is:

$$\frac{1}{2}w^T w + c\sum_{i=1}^{N}\varepsilon_i + c\sum_{i=1}^{N}\varepsilon_i^* \tag{6}$$

Equation 6 Error Function for SVM

Large number of kernel could be used in order to give Support Vector machine models. These models are of different types and styles. Some of them are: Linear, Polynomial, Radial basis function (RBF) and also sigmoid [37]

Information of Kernel Functions:

$$K\left(\mathbf{X}_{i}, \mathbf{X}_{j}\right) = \begin{array}{cc} x_i x_i & \mathit{Linear} \\ (\gamma x_i x_i + C)^2 & \mathit{Ploynomial} \\ \tanh(\gamma x_i x_i + C) & \mathit{Sigmoid} \end{array} \quad (6)$$

Equation 7 Kernel Functions

Where K (X_i, X_j) = $\phi(X_i) \cdot \phi(X_j)$

This equation shows a kernel function. It shows a dot product of input data that has been drawn on higher location feature and is spaces by the transformation ϕ. Gamma (γ) is mainly used as an adjustable parameter or tool of different kernel functions. RBF is mainly used throughout the world for the solution of these kernel functions and also provide a better service of catching different error functions [35]

Selection of best tools and equipment are very important for a better performance and functionality. There are large number of tools and equipment within the study of statistics that mainly die or their working capability mainly ends up with the passage of time. The expiration of any tool, equipment, formula, equation is due to the fact that there occurs a slight error within the functionality or the handling/controlling processing of that particular tool or formula. Different terms, tools etc. are used for a better working and better calculation of any outcome. Such an example is R environment. It is mainly used to carry out manipulation of data, for the purpose of calculation and other processes through error freeway. Collection of best and the efficient one is very important as far as calculation or the processing is concerned. With different calculations and precise measurements, a single work could benefit a human being, a lot [35]

Bayesian Classification

Bayesian classifier is a measurable classifier. It is focused around Bayes' guideline of restrictive likelihood. Bayes' tenet says that on the off chance that you have a speculation H and proof X that bears on that theory, then

$$P\left(\frac{H}{X}\right) = P\left(\frac{X}{H}\right) P(H) \div p\ (X)$$

The basic manifestation of Bayesian classifier is known as the Naïve Bayesian (NB) classifier. NB model performs well in examination with the famous choice tree calculation and displays high exactness and rate when connected to huge databases.

Bayesian conviction systems are graphical models, which permit the representation of conditions among subsets of properties.

Classification by Naïve Bayesian (NB) Classification

Naïve Bayesian classifiers expect that the impact of an attribute value on a given class reflects independence from the values of other attributes (class conditional independence).

The performance of this classifier goes as follows [2]:

- D to be taken as tuples' training set and class labels linked with them.
- Every single tuple is shown by an n-dimensional attribute vector as X: $(x_1,x_2,x_3,..,x_n)$.
- Classes as m: $C_1,C_2,C_3\ldots C_m$

According to NB Classifier estimates "X belongs to Ci (Class)" in the following situation:

$$P(Ci|X) > P\ (Cj|X) for\ 1 \le j \le m; j \ne i$$

Thus P((C_i |X) is maximized. As per Bayes' theorem:

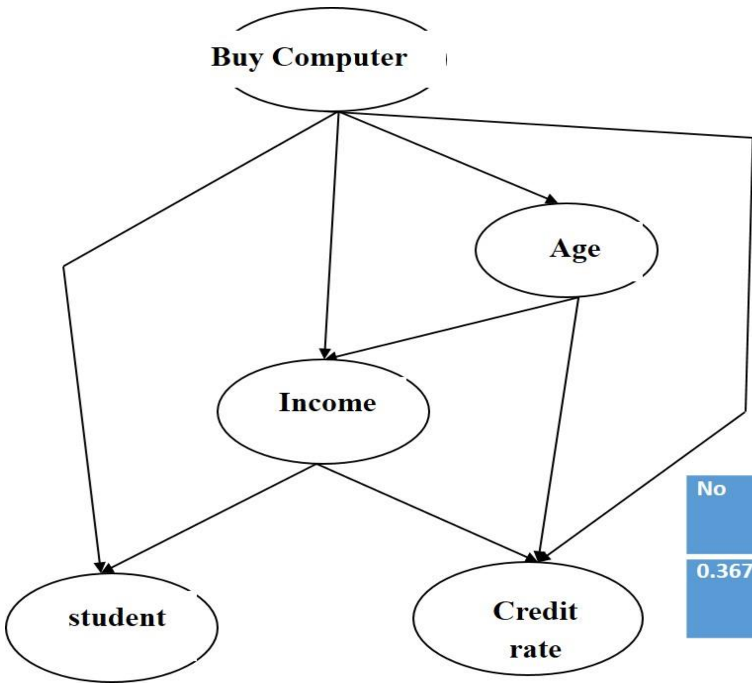
$$P\left(\frac{Ci}{X}\right) = \ P\left(\frac{X}{Ci}\right) P(Ci) \div P(X)$$

With P(X) constant, the aforementioned equation equals to maximizing. Where:

$$p(X / C_i) = \prod_{1}^{n} p(X_k / C_i)$$

Classification by Bayesian Belief Networks

Bayesian conviction system is characterized by two segments [27] an administered a-cyclic chart and a set of restrictive likelihood tables Figure 8. Every hub in the diagram speaks to an arbitrary variable. Edges speak to contingent dependencies. It permits the representation of conditions among subsets of properties. Bayesian conviction system can deal with missing values and can be utilized to addition seeing around an issue space and to foresee the outcomes of intercession. Moreover, Bayesian measurable strategies in conjunction with Bayesian systems offer a productive and principled methodology for escaping the over-fitting of data.



No	Yes
0.367	0.633

Figure 8 Classification by Bayesian Belief Networks

Classification by Rule Based Classification

An IF-THEN rule is an articulation of the structure (IF condition THEN conclusion). A rule-based classifier utilizes a set of IF-THEN rules for classification. A rule R can be evaluated by its scope and precision.

Given a tuple, X, from a class named information set D let n-covers be the quantity of tuples secured by R, n-correct be the quantity of tuples effectively grouped by R, and |d| be the aggregate number of tuples in D. We can characterize the scope and precision of R as:

$$Coverage(R) = \frac{n_{covers}}{\| D \|}$$

$$Accuracy(R) = \frac{n_{correct}}{n_{covers}}$$

That is, a rule's scope is the rate of tuples that are secured by the rule; i.e., whose attributes are taken as valid by the values for the forerunning purpose of rule. However, to ensure correctness of the rule we look at the tuples whether it covers and perceives the exact percentage being accurately classified.

Decision Tree Induction

Decision tree induction [2] is the taking in of decision trees from class-marked preparing tuples. A decision tree is a flowchart-like tree structure, where every inward hub (non-leaf hub) means a test on a characteristic. Each one extension speaks to a result of the test, and each one leaf hub (or terminal hub) holds a class mark. The highest hub in a tree is the root hub. Among the commonly used types of decision tree under classification model are Option decision tree, oblivious decision tree, Lazy tree and Oblique decision tree [24]. Decision trees (DT) are so poplar classifier in information mining on the grounds that they require less information readiness, not delicate to anomalies and skewed circulations, can deal with high dimensional information, have great precision and easily understood by human brain. According to Rokach, and Maimon algorithms as Decision tree inducers automatically buildup decision tree from an available dataset. By reducing generalization error, the basic aim is to look for the optimal decision tree [24].

There are numerous distinctive calculations that are utilized for building decision trees, including CHAID (Chi-squared Automatic Interaction Detection), ID3, CART (Classification and Regression Trees), Quest, and C5.0. [25].

Decision tree calculation begins with a preparation set of tuples and their related class names. The preparation set is recursively part (discover the best part utilizing entropy, Gini or Chi square) into more diminutive subsets as the tree is constantly manufactured. At each one succeeding level of the tree, the subsets made by the previous part are as per whatever guideline works best for them. The tree keeps on growing until it is no more conceivable to discover better approaches to part up the incoming records [26].

Trees left to develop without bound take more time to fabricate and appear as over fit the available information. In order to control the tree size, either means of ceasing rules to stop its growth/ development are used or techniques to prune the tree such as Cost Complexity Pruning, Pessimist Pruning, Reduced Error Pruning and Optimal Pruning etc. (whichever is suitable) are applied [25] (the tree is pruned over to the smallest size that does not trade off exactness).

However the decision tree has many advantages but it has also disadvantages, one of them is that the decision with regards to splitting of variable is not concerned about the impact of current split on future ones. Also every future split is based on the first split i.e. a change in first split may result in final solution to be totally different. At one time, algorithms implied for splitting are deemed as only one predictive variable.

Descriptive Data Mining

Descriptive data mining handle the data and analyze it in the way of analyzing historical events and try to figure out insights to show how to approach the future. Descriptive data mining looks at past events and understands these events by mining past data looking for the causes behind it either it was perform good or bad in the past. The uses of this type of analysis can be used in a lot of management reporting like marketing, finance, sales and operations.

Descriptive analysis evaluate relationships in data in a way that is most of times used to classify inputs into groups with the same behavior. Descriptive models often used to identify many different relationships between customers or products.

For example descriptive data mining can be used to categorize European customers verses African Customers by their product preferences and life style the model can identify the factors that differentiate between those two classes. The produced model can be used as a tools itself to describe the data or can be used as an input in another model to produce more sophisticated model which can produce deeper analysis.

We will introduce the main algorithms that have been used to apply the descriptive analysis with focusing on the Association Rules techniques and clustering.

Association Rules

Association rules overview

Data mining has many techniques for discovering important information from databases. It is a technique, which currently has become profitably used by many industries. A number of algorithm tools have been developed, and used to retrieve information and to discover knowledge patterns which are useful in support of decision making. Several data techniques are clustering, patterns recognition, classification and association on classification it has been identified as the important challenge in the data mining field. The Association rule and Apriori algorithm are among the knowledge pattern technique discovered or developed. They are discussed below.

Association rules implementation

Association rule has become the leading approach for discovering information, which it reveals the relationships among the different items. This approach is use to analyze large database contain medical record-data. The aim here is to obtain the association rules, and to indicate the relationships between the preformed procedures on the patient and reported diagnoses. This concept of associated with data mining, it also help in the discussion for identifying the association rules and to report on generated rules.

Association rule finds the interesting correlation relationships or associations among the large set of data. The rules show the attribute value which occurs frequently in given set of data. The most used example of this rule is market basket analysis. For instance, the collected data using a bar code scanners in supermarkets. Such market basket databases consist of large transactions records. Managers are interested in knowing which certain groups are consistently purchased. They used this data items for adjusting stores layouts, for cross selling and for promotions.

This rule also provide information of this type in form of if then statements. Association rule are completed from the data and unlike if statement rules of logic, therefore they are probabilistic in nature. The rules have two numbers which express the degree of uncertainty of the rule. In the analysis of association rule, the antecedent and consequent are sets of data which are disjoint. The database containing transaction that consists of set items and the transaction identifier such as market basket. The association rule has the implication from X->Y, such that

X and Y are the two disjoint subsets for items available. X is name or called the antecedent or Left hand side and Y is called the consequent or right hand side. This rule have to satisfy the constrains on measures of interesting and significance.

The principle has two step approaches, frequent generation and rule generation. The figure represents the association rule performances.

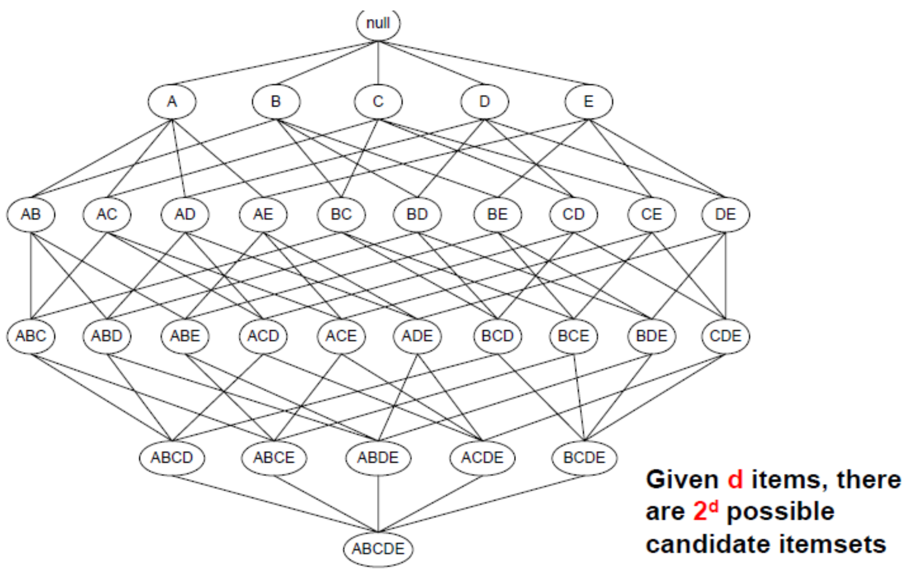


Figure 9 Association Rule Technique

The association rule is the implication of if-then-rule that supported by the data. The development of association rule is the "market basket analysis" that deals with the contents of the point of sale transaction in retailers. This association rule generates the analysis of market basket and it is used in other areas such as text data analysis. The results in large collections of basket market provide the information about, that the items are sold. The example is in the figure below:

Market Basket Id	**Market basket content**

1	Orange juice, soda
2	Milk, bread, orange juice
3	Orange juice, butter
4	Orange juice, bread, soda water
5	Bread

Table 2 Five Grocery Basket Market

This rule is simple for the association rule. The efficiency of the rule will depend on the features of the data sets. And the important feature of this rule is that many retailers data sets is the average market basket contains only small subsets.

Apriori algorithm

This rule uses a prior knowledge about the important property for frequent set of data, thus its name. Apriori property of a set of items says that all the nonempty subsets of frequent set item must be frequent. However, if the given set item is not frequent, then the superset of the set items will not be frequent also, because it doesn't occur frequently than the original set items.

Generating itemset efficiently, the Apriori algorithm, performs iterative search through the set items, starting with one-itemset, through two-set items, and through three-set items and the process continues. Generally, Apriori algorithm finds and processes K-set items based on exploration of k-1 set items. Using Apriori property, Apriori algorithm performs the following: finds all the 1-setitmes, next it finds among the set of frequent 1-setitems, next it extends 1-setitems to generate 2-set items, next it finds among the 2-set items set of frequent 2-setitems, and finally repeats the process in order to obtain 3-setitems and 4-setitems.

Basing on Apriori algorithm, in each iteration of k- set items which do not satisfy a minimum support will be removed and the remaining k-items set are used in generating set items for the next k+1 iteration. The process reduces substantially the number of set items that must be check if they are frequent.

The principle of Apriori

That if the itemset are frequent, then all of its subsets must be frequent or if the itemset are frequent then all the supersets also must be infrequent. The principle is holds due to the property of the support measure: that the support of an itemset never exceeds the subsets support, also this is known as the anti-monotone support property. It is illustrated in the below figure.

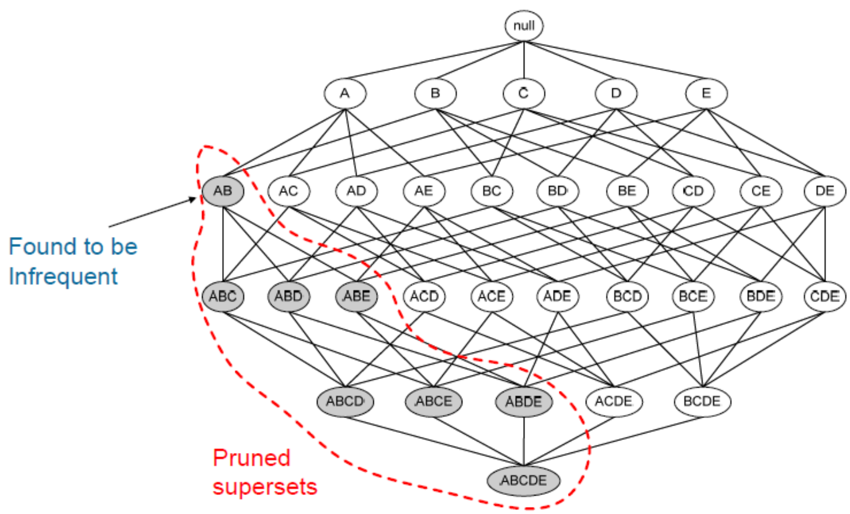


Figure 10 Apriori Principle Found Infrequent

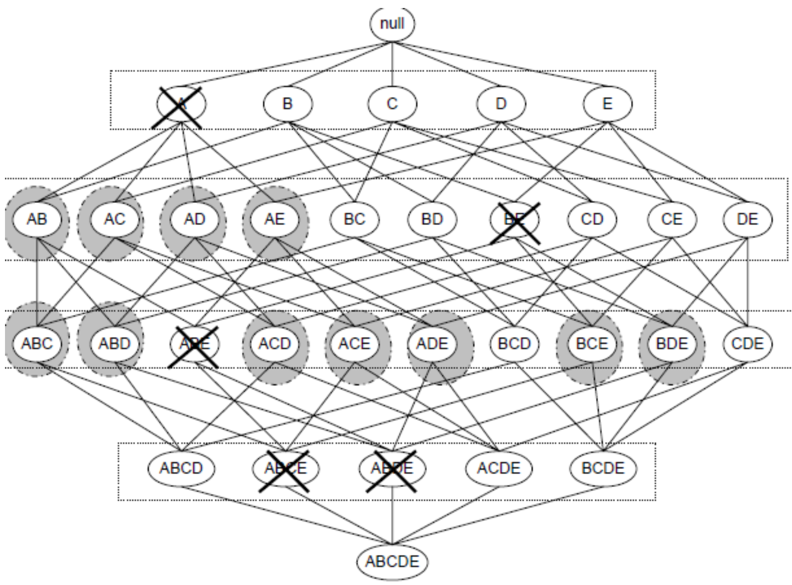


Figure 11 Apriori Principle Found Frequent

The use of Apriori algorithm is applied in the cases of mining of the influenza treated by the contemporary world. The computational tools are needed to extract information from the databases and the manpower is also required to apply such tools in order to diverse the community to settle in new areas.

Data Analysis Phase

Data mining output going under evaluation process to interpret the output and see if additional knowledge has been discovered and to add the relevant discovered important facts which was produced by the mining algorithms to the current knowledge.

R Language

We will rely on R programming language to produce our analysis. R programming language became the single most important tool for data analysis, visualization and science. All over the world a lot of statisticians and data scientists use R to solve their most challenging problems. R has become the most popular language for data science and an essential tool for Finance and analytics-driven companies such as Google, Facebook, and LinkedIn.

Chapter 5

Detection Of Manipulation in the Egyptian Market

Introduction

Regardless of the technique that can be used or applied to the problem at hand, i.e. bid rigging detection, the best practice of the data mining is that one can predict the future events using the past data with detecting some patterns or repeated action in the past. In this chapter the application of the data mining techniques will be introduced to help in detection of the bid rigging. We will use it to identify bid rigging signals which can be used as red flags either governments or private entities to indicate the anti-competitive behavior.

In the following sections we will try to detect some specific bid rigging defined strategy from the OECD[3] which are:

Cover bidding: Cover (also called complementary, courtesy, token, or symbolic) bidding is the most frequent way in which bid-rigging schemes are implemented. It occurs when individuals or firms agree to submit bids that involve at least one of the following: (1) a competitor agrees to submit a bid that is higher than the bid of the designated winner, (2) a competitor submits a bid that is known to be too high to be accepted, or (3) a competitor submits a bid that contains special terms that are known to be unacceptable to the purchaser. Cover bidding is designed to give the appearance of genuine competition.

Bid suppression: Bid-suppression schemes involve agreements among competitors in which one or more companies agree to refrain from bidding or to withdraw a previously submitted bid

[3] Organisation for Economic Co-operation and Development

so that the designated winner's bid will be accepted. In essence, bid suppression means that a company does not submit a bid for final consideration.

Bid rotation: In bid-rotation schemes, conspiring firms continue to bid, but they agree to take turns being the winning (i.e., lowest qualifying) bidder. The way in which bid-rotation agreements are implemented can vary. For example, conspirators might choose to allocate approximately equal monetary values from a certain group of contracts to each firm or to allocate volumes that correspond to the size of each company.

Market allocation: Competitors carve up the market and agree not to compete for certain customers or in certain geographic areas. Competing firms may, for example, allocate specific customers or types of customers to different firms, so that competitors will not bid (or will submit only a cover bid) on contracts offered by a certain class of potential customers which are allocated to a specific firm. In return, that competitor will not competitively bid to a designated group of customers allocated to other firms in the agreement.

Egyptian public procurement

Public procurement in Egypt is regulated by Tender Law No. 89 / 1998 public procurement law (PPL) and its executive regulations (secondary laws) issued pursuant to the decision of the Minister of Finance, no. 7 1368/1998. Since 2008, the PPL has been subject to two amendments which have had no significant impact on the regulations. The most recent amendment, decree no. 33/2010 by the Prime Minister, introduced electronic means for tender notification and established a government website where contracting entities should publish contract notices in addition to traditional means (a public tender board or newspaper) of publication. [45]

In Egypt, there is no single autonomous power with administrative forces in charge of creating procurement policies. Also, there is additionally no free devoted cures body to handle complaints identified with public procurement and monitor agreeability of contracting elements. The survey showed that the Egyptian public procurement institutional framework is intricate, with various powers government divisions included in conveying open acquirement capacities.

According to OECD [46], the value of the public procurement market in Egypt reached EGP 28.7 billion in the fiscal year 2010/2011:

- EGP 5.7 billion pounds (13.2%) from the total budgetary resources was allocated to the Central Public Utilities
- EGP 6.8 billion (12.6%) from the total budgetary resources allocated to the Local and Municipal Public Utilities
- EGP 16.2 billion pounds from the total budgetary resources of the administrative authorities of the Egyptian Government

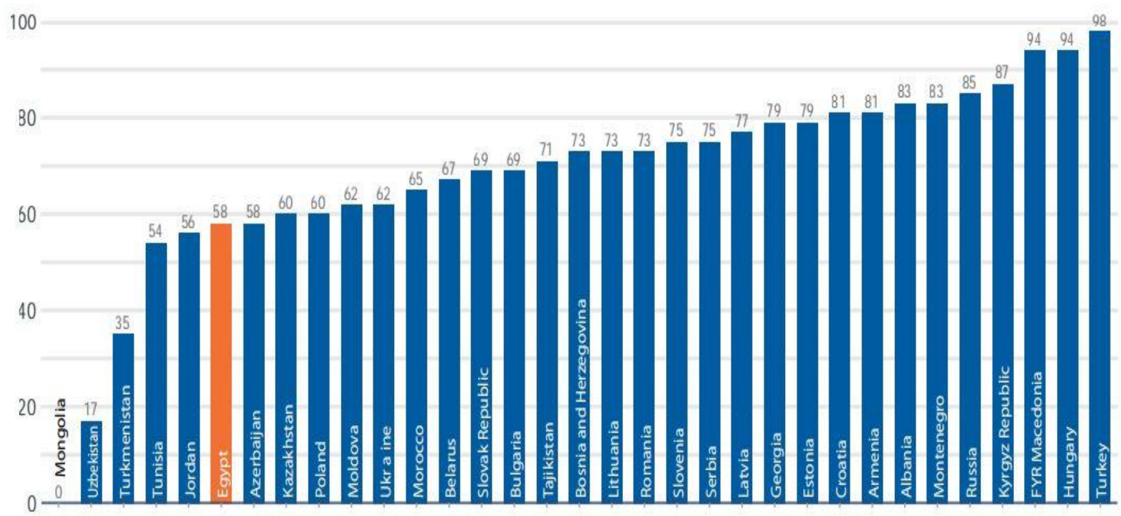


Figure 12 The Quality Of Public Procurement In Egypt Comparing To Other Countries [45]

Figure 12 represent the score for the quality of the process if the bidding in Egypt comparing to other European Bank for reconstruction and development region (ERBD). Total scores are presented as a percentage with 100 per cent representing the highest performance in the legal efficiency concept benchmark indicators.

So we can say according to previous discussions that the Egyptian public procurement is not running efficiently.

Data Description

According to previous definitions for some bid rigging strategies, we think there are minimum requirement to apply the proposed techniques to any bidding system as any data should contain some essential variable:

1- Location of bid
2- Bidders details (Name, Contacts, …)
3- Products (Name, Type)
4- The winner
5- The bid value for each bidder

So we contacted the Egyptian government and a private school which is the Egyptian International School to provide us with the real data to prove the concept of how can data mining detect the bid rigging.

Case 1: The Egyptian International Schools (EIS) Bids

In this case, the data was in CSV format and it was consist of six columns and 31 rows. The product is the paper of size A4 and size A3. Market consists of two suppliers as they were the only supplier for this type of paper our data variables as the following:

1- Supplier: company name that will provide the school with the required paper, we will hide the name of the two provider according to the data provider request.
2- Winner bid Value: the value that has been paid to the winner of the bid
3- Quantity: the number of blocks of the paper
4- Location: the school own many location in different governorates but the bid were going on in Cairo.
5- Date: the date of the decision of declaring the winner

The full details of the bids are available in Appendix A.

Table 3 EIS Bids Data

Supplier	Winner Bid Value (EGP)	Quantity	Location	Product	Date
C2	12690	705	Cairo	Paper A3	Sunday, July 9, 1995
C1	5757	303	Cairo	Paper A4	Tuesday, November 7, 1995
C2	14520	726	Cairo	Paper A4	Monday, July 15, 1996
C2	14605	635	Cairo	Paper A4	Wednesday, November 20, 1996
C2	16399	713	Cairo	Paper A4	Tuesday, October 7, 1997
C2	13368	557	Cairo	Paper A3	Wednesday, November 5, 1997
C1	11316	492	Cairo	Paper A3	Monday, July 13, 1998
C2	17520	730	Cairo	Paper A4	Sunday, November 1, 1998
C2	12744	708	Cairo	Paper A4	Thursday, July 1, 1999

We will apply two data mining techniques trying to figure out the bid rigging indicators.

Case 2: The Egyptian Competition Authority (ECA) Oil Market

Table 4 ECA Oil Bids Data

Comp Id	Bid Id	Comp Price	Win/ Not	Price	Bid Value	Oil Type 1	Quantity
1	66	928.85	TRUE	920	0	Sun flower	10000
1	69	893	TRUE	893	0	Sun flower	6000
1	70	888	TRUE	0	0	Sun flower	10000
1	71	865	TRUE	865	8650000	Sun flower	10000
1	73	859	FALSE	859	0	Soya	15000
1	73	859	TRUE	859	0	Sun flower	10000
1	74	982	FALSE	982		Soya	20000

We have used the R-Language to examine the bids data from real market to detect anomalous patterns. R-language is a developed suite of software facilities for data manipulation, calculation and graphical display .

Our choice for the R-language as a tool among other tools for evaluate our modules and data mining techniques rely on many reasons:

- An effective data handling and storage facility
- A suite of operators for calculations on arrays, in particular matrices
- A large, coherent, integrated collection of intermediate tools for data analysis, graphical facilities for data analysis and display either directly at the computer or on hard- copy, and
- A well-developed, simple and effective programming language (called ‘S’) which includes conditionals, loops, user defined recursive functions and input and output facilities. (Indeed most of the system supplied functions are themselves written in the S language.) [36]

For the first case, we generated a table listing the following attributes: the *supplier Name*, the *product* procured either it was paper of type A3 or A4 and the *quantity* required.

For the second case which related to the public procurement of the oil market in Egypt, we generated a table listing the following attributes: the *company ID*, the *Bid ID*, the *company price* for the bidders prices, *final price* which is the winner price, and the *quantity* required

Then, for each of the two cases concerned, we applied the following two data mining techniques:

Classification trees: Given a set of points and a set of labels, the technique assign labels to points so that similar objects are labeled by similar labels and a point is labeled by a more likely label.

Association rules: The goal of association analysis is to extract significant patterns, in the form of rules or sets of events that will predict the occurrence of certain events based on the occurrence of other events. For example, the association rule A=>B suggests that event B is expected to occur whenever event A is observed. A standard association analysis algorithm, such as *Apriori* [4] is applied to extract rules from the transformed data sets. The rules extracted by the *Apriori* algorithm are evaluated using the well-known support and confidence measures

Rules with low support and low confidence tend to be statistically insignificant, and are pruned automatically by the *Apriori* algorithm [4].

Experimental Results

Now we will introduce the results of applying both of decision tree and association rules to our two cases and show how we can use these results either it was rules or classification to identify the bid rigging technique(s).

Results for Case 1

Applying the decision tree (DT) classification algorithm to the case one produced Figure 13 and Figure 14. The first figure clearly shows that the market is divided by product, and identifies how many times each supplier won with which product.

The first number is the total number of instances which reached to the leaf. The second number is the number instances that were misclassified

For example, the figure shows that Paper (A3) was always the product when for all 10 bid that supplied C1 win in them. And there was three bid misclassified.

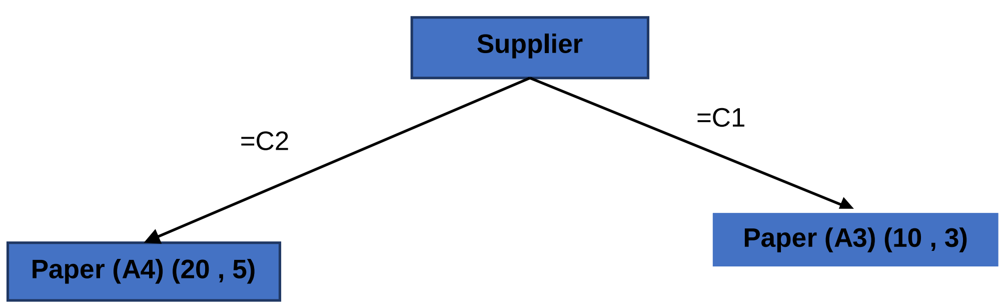


Figure 13 First Decision Tree for case One

The second figure shows that is divided by quantity such that if the quantity is less than or equal 495 paper package then C1 always win and if it was greater than 495 then C2 win.

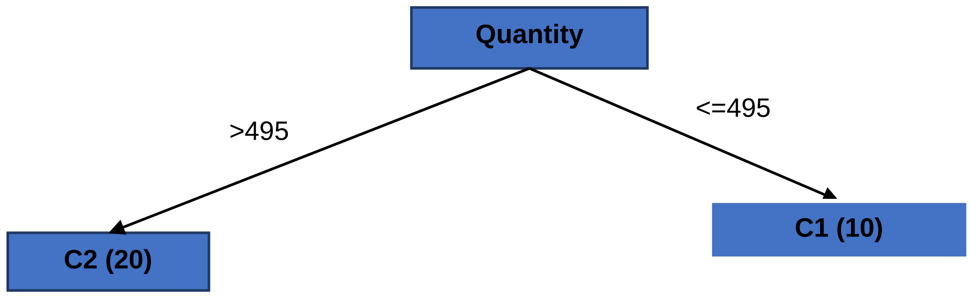


Figure 14 Second Decision Tree for Case one

The classification which has been identified by the decision tree was confirmed by the results from the association rules algorithm. But before I applied the association rules, I converted the quantities to be categorized values instead of continues values because association rules can't be applied to continues values so instead of removing it from the variables that we will Apriori to them, I converted it to categorical variable. To do that I relied on the previous classification of the decision tree which has classified the quantities into categories either greater than 495 or less than or equal 495. So I converted the quantity values to be either greater than 495 or less than or equal. So the data will be like that and the full details of the data in Appendix

Supplier	Winner Bid Value (EGP)	Quantity	Quantity (Categorized)	Location	Product	Date
C2	12690	705	Greater Than > 495	Cairo	Paper A3	Sunday, July 9, 1995
C1	5757	303	Less Than or Equal 495	Cairo	Paper A4	Tuesday, November 7, 1995
C2	14520	726	Greater Than > 495	Cairo	Paper A4	Monday, July 15, 1996
C2	14605	635	Greater Than > 495	Cairo	Paper A4	Wednesday, November 20, 1996
C2	16399	713	Greater Than > 495	Cairo	Paper A4	Tuesday, October 7, 1997

Table 5 Modified EIS Data (Conversion of Quanitity to Categorical)

The results from the association rules algorithm were as the following:

Left Hand Side Rule	Right Hand Side Rule	Confidence
Quantity (Categorized)=Greater Than > 495 20	Supplier=C2	100%
Supplier=C2 20	Quantity (Categorized)=Greater Than > 495	100%
Supplier=C2 20	Location=Cairo	100%
Quantity (Categorized)=Greater Than > 495 20	Location=Cairo	100%
Quantity (Categorized)=Greater Than > 495 Location=Cairo 20	Supplier=C2	100%
Supplier=C2 Location=Cairo 20	Quantity (Categorized)=Greater Than > 495	100%
Supplier=C2 Quantity (Categorized)=Greater Than > 495 20	Location=Cairo	100%
Quantity (Categorized)=Greater Than > 495 20	Location=Cairo	100%
Supplier=C2 20	Location=Cairo	100%
Product =Paper A4 18	Location=Cairo	100%

So from the previous rules and the decisions tree we can confirm there is a big signs with high confidence that there is a bid rigging by dividing the market according to the quantity and the product which is prohibited by law.

Results for Case 2

Applying the decision tree (DT) classification algorithm to the case one produced Figure 15 and because the tree was too big to be included in one graph I decided to divide it to many graphs which connected to each other.

The main result indicate that there are market allocation according to product ant prices. We can interpret the results as the followings:

1. If the oil was of type Soya and the company price for the bid was less than or equal 851.85 EGP then Company C14 will be the winner
2. If the oil type neither Soya nor Sun flowers and the company price for the bid was less than or equal 737 then Company C2 will be the winner
3. If the oil type was Sun flowers and the bidder price was less than or equal 1063 then the company C3 will be the winner

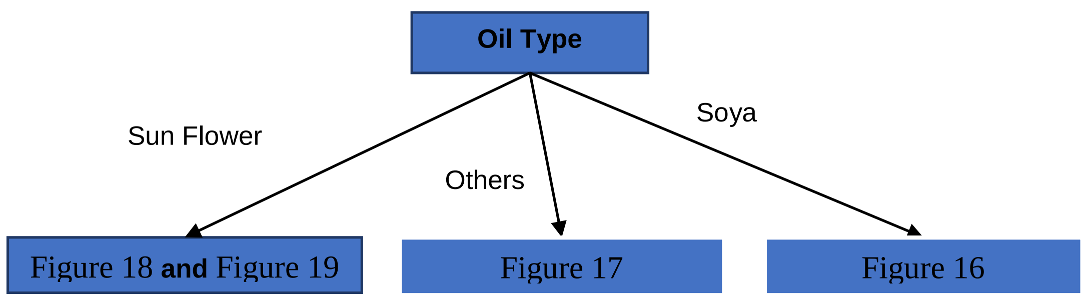


Figure 15 Case Two Decision Tree

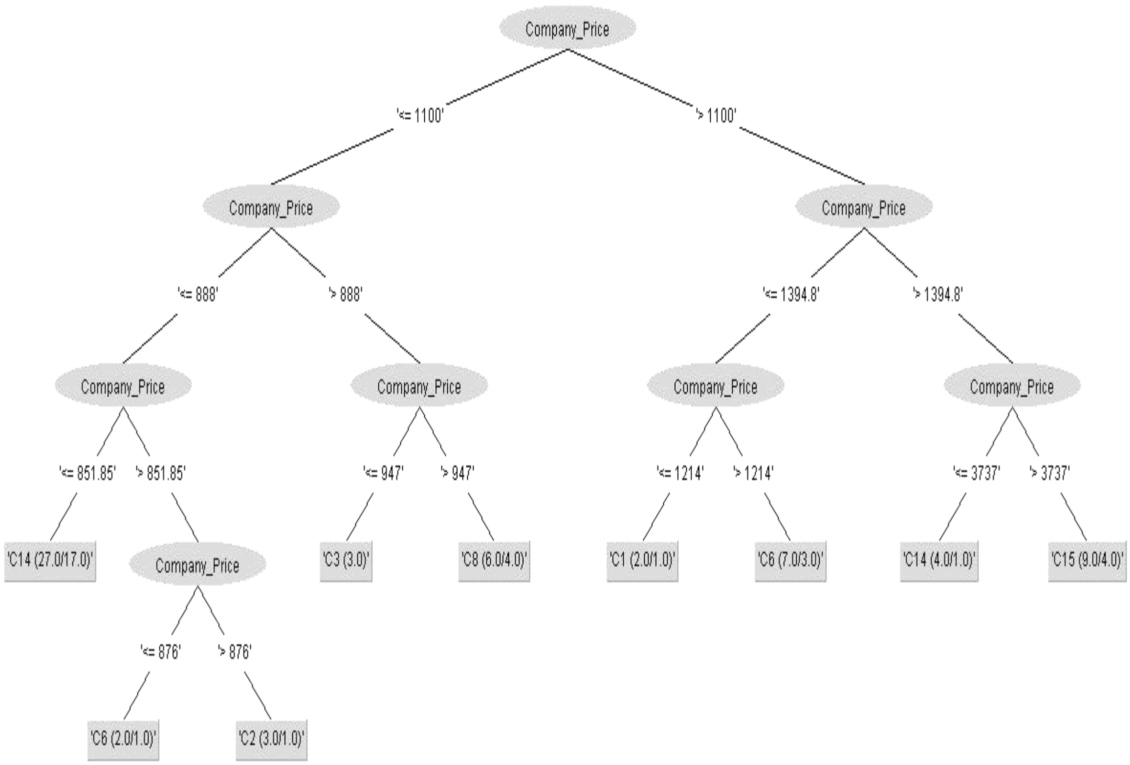


Figure 16 Soya Oil Decision Tree

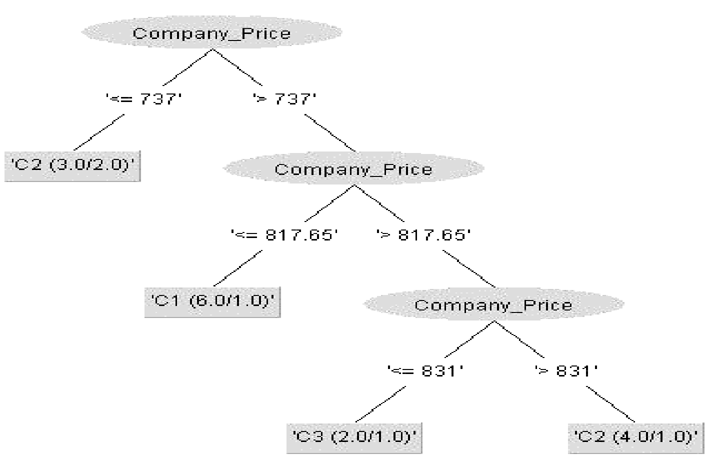
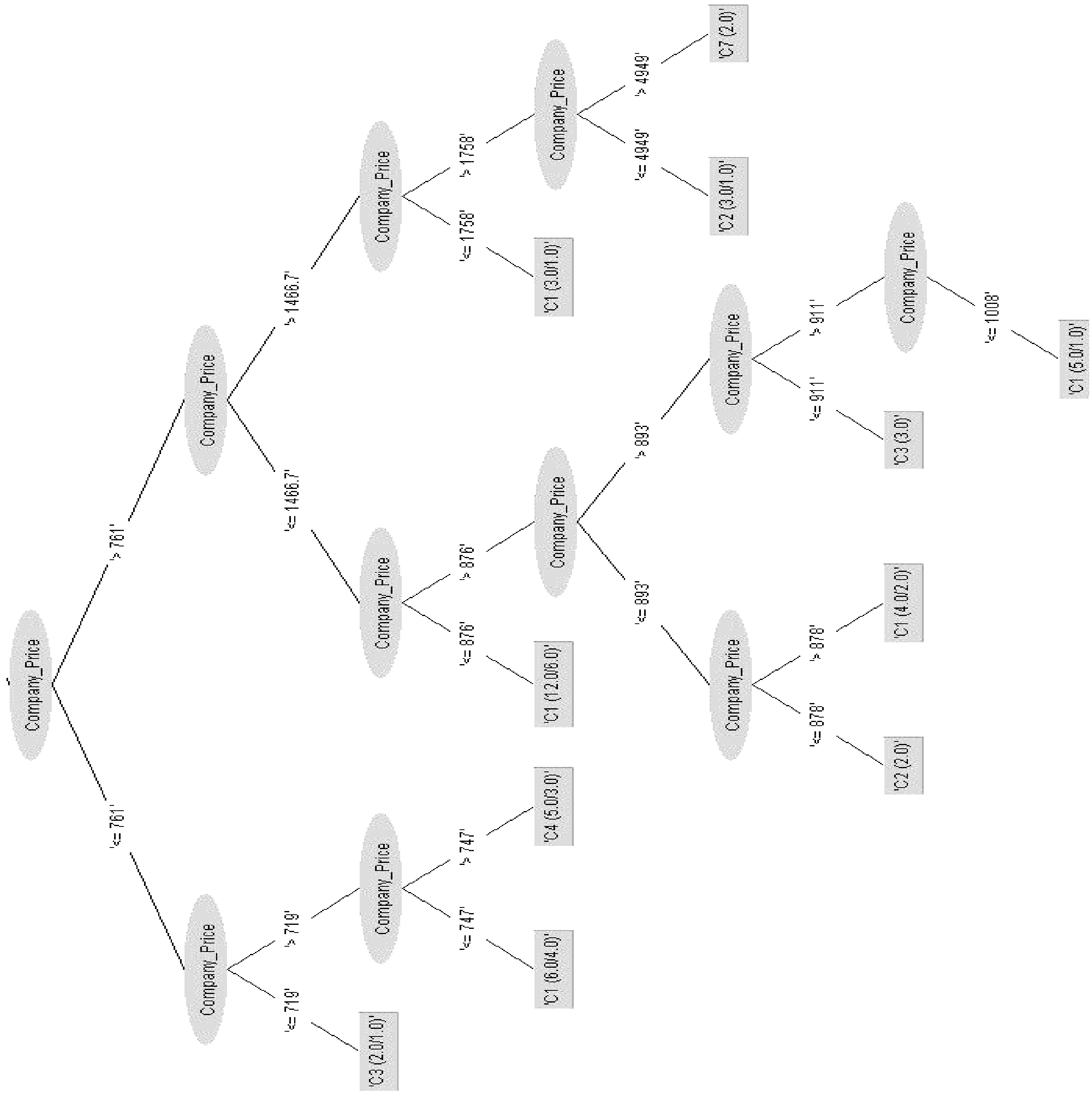


Figure 17 Other Type of Oil Decision Tree



Company_Price
'<= 761'
'> 761'
Company_Price
'<= 719'
'> 719'
'C3 (2.0/1.0)'
Company_Price
'<= 747'
'> 747'
'C1 (6.0/4.0)'
'C4 (5.0/3.0)'
Company_Price
'<= 1466.7'
'> 1466.7'
Company_Price
'<= 876'
'> 876'
'C1 (12.0/6.0)'
Company_Price
'<= 893'
'> 893'
Company_Price
'<= 878'
'> 878'
'C2 (2.0)'
'C1 (4.0/2.0)'
Company_Price
'<= 911'
'> 911'
'C3 (3.0)'
Company_Price
'<= 1008'
'C1 (5.0/1.0)'
Company_Price
'<= 1758'
'> 1758'
'C1 (3.0/1.0)'
Company_Price
'<= 4949'
'> 4949'
'C2 (3.0/1.0)'
'C7 (2.0)'

Figure 18 Sun Flower Oil Decision Tree Part 1

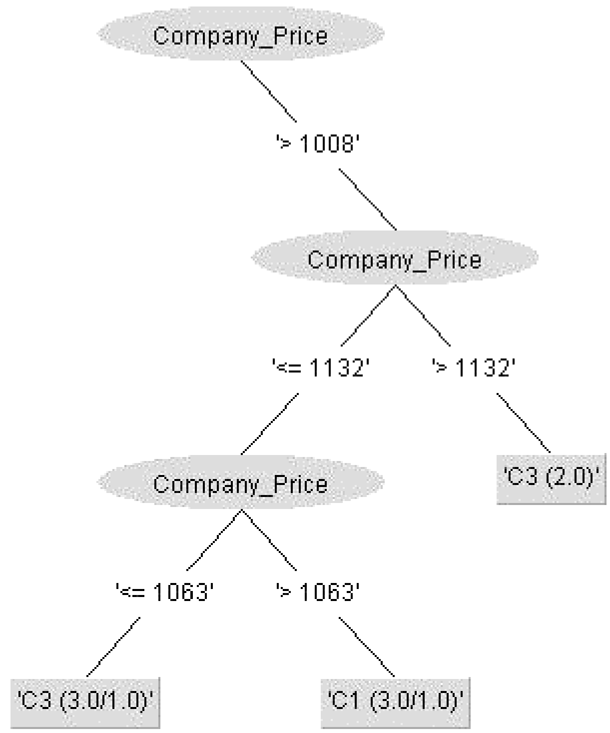


Figure 19 Sun Flower Decision Tree Part 2

In addition to the previous result we applied the association rules algorithm which produced the following rules:

1. Company_Id=C14 19 ==> Oil Type 1=Soya 19   conf:(100%)

2. Company_Id=C15 15 ==> Oil Type 1=Soya 15   conf:(100%)

However we only get two rules from association rules but these two rules supports the previous results from decision tree.

So we encourage applying the most techniques that we can use then choose the best techniques that give use best accuracy.

Chapter 6

Conclusion and Future Work

This thesis provided an introduction to the process of data mining and its application. A detailed literature survey of the data mining techniques. Of all the surveyed techniques we identified Decision Trees Classification and association rules (Apriori) as potential candidates for detecting bid rigging conducts in the market. We have introduced the problem of detecting bid rigging of procurement processes, and applied the data mining to the bids data. We also showed that these techniques could be applied in both private and public sectors, i.e. for detecting bid rigging conducts that already existing in the market, or ones that are just about to happen. We then used a data mining tool, e.g. R language to identify suspicious or unusual patterns between bidders. For various libraries of R language we applied Apriori algorithm to produce association rules and J48 to produce the decision tree. We run two experiments on a real data from the private sector which has been provided from the Egyptian international school and data from public sector which provided from the Egyptian government with two techniques classification and association rules on detecting bid rigging with great success.

This theses proposed a new application for data mining techniques and proved their efficiency to detect some bid rigging strategies which come from the bidders.

The application of data mining to detect bid rigging conspiracies is work still in its early stages. However the above results show a high potential for this new domain of data mining applications. We will try to create a new system that can be used to automatic detect the bid rigging by raising a red flags that can be reliable. These system will use mainly the decision tree and association rules in addition to more data mining techniques. Also, we will try with these and other data mining techniques to detect the other types of bid rigging conducts. For example, we envisage a high potential for using the “pattern recognition” data mining technique for detecting the “bid rotation” schemes. The goal is to eventually identify which data mining are more effective in detecting which types of bid rigging schemes.

References:

[1]Fayyad, U., Shapiro, G. And Smyth P. , Advances in Knowledge Discovery and Data Mining, MIT Press, 1996

[2]Han, J., and Kamber, M., *Data Mining: Concept and Techniques*, Morgan Kufmann publishers, 2006.

[3]Edelstein, H., *Introduction to Data Mining and Knowledge Discovery 2nd Edition*, Two Crows Corporation, ISBN 1892095009

[4]MIT, *The Technology Review Ten*, MIT Technology Review, January/February 2001.

[5]Two Crows Corporation, *Introduction to Data Mining and Knowledge Discovery 3rd*, two Crows Corporation, 2005

[6]Konrad, R., *Data mining: Digging user info for gold*, ZDNET News, February 7, 2001

[7]Fayyad, U., Piatetsky-Shapiro, G. and Smyth, B., *From Data Mining to Knowledge Discovery in Databases*, American Association for Artificial Intelligence, 1996

[8]Tan, Pang-Ning Steinbach, M. Kumar V., *Introduction to Data Mining*, Pearson, 2006

[9]Cheeseman, P. and Stutz, J., *Bayesian Classification (AutoClass): theory and results*, 1995

[10]Jin, H., *Scalable Model-based Clustering Algorithms for Large Databases and Their Applications*, the Chinese University of Hong Kong, August, 2002

[11]Zurada, J. and Kantardzic, M., *New Generation of Data Mining Applications*, Wiley Interscience Publication, 2003

[12]Setty D., Rangaswamy, T. and Subramanya, K., *A Review on Data Mining Applications to the Performance of Stock Marketing*, International Journal of Computer Applications, Vol 1, No. 3, 2010.

[13]Sumathi, S. and Sivanandam, S.N., *Introduction to Data Mining and its Applications*, Springer-Verlag Berlin Heidelberg, 2006.

[14]http://observationbaltimore.com/blog/2011/09/data-mining%E2%80%94why-is-it-important/

[15]Fayyad, U., Piatetsky-Shapiro, G. and Smyth, B., *From Data Mining to Knowledge Discovery in Databases*, American Association for Artificial Intelligence, 1996

[16]Chapman, P., Clinton, J., Khabaza, T., Kerber, R., Reinartz, T., Shearer T. &Wirth, R., *CRISP-DM 1.0 Step-by-step data mining guide*, http://www.crisp-dm.org/, 2000

[17]Williams, Graham J., and Simeon J. Simoff., *Data mining theory, methodology, techniques, and applications*. Berlin New York: Springer, 2006

[18]McCue, Colleen. *Data mining and predictive analysis intelligence gathering and crime analysis*. Amsterdam Boston: Butterworth-Heinemann, 2007

[19]Vaidya, Jaideep, Michael Zhu, and Christopher W. Clifton, *Privacy preserving data mining*, New York: Springer, 2006

[20]http://www.comp.dit.ie/btierney/oracle11gdoc/datamine.111/b28129/regress.htm

[21]Pujari, Arun K., *Data mining techniques*. Hyderabad Great Britain: Universities Press, 2001

[22]Dong, Guozhu, and Jian Pei. *Sequence data mining*. New York: Springer, 2007

[23]Miller, T. H., *Modeling Techniques in Predictive Analytics: Business Problems and Solutions with R, Revised and Expanded Edition*, FT Press, US. 2014.

[24]Rokach, Lior, and Oded Maimon, *Data mining with decision trees theory and applications*, Singapore: World Scientific, 2008

[25] MURTHY, S., K., *Automatic Construction of Decision Trees from Data: A Multi-Disciplinary Survey*, Kluwer Academic Publishers, Boston, 2009

[26]Bramer, M., *Principles of Data Mining*, Springer-Verlag London Limited, 2007

[27]GAO, *DATA MINING Federal Efforts Cover a Wide Range of Uses*, United States General Accounting Office, 2004

[28]Crat, Bogdan, ZhaoHui Tang, and Jamie MacLennan, *Data Mining with Microsoft SQL Server 2008*, N.J: Wiley, 2013.

[29]Cao, *Longbing. Domain driven data mining*, Springer, 2010

[30]Thuraisingham, Bhavani M., *Web data mining and applications in business intelligence and counter-terrorism*, Boca Raton Fla: CRC Press, 2003

[31]http://www.saedsayad.com/data_preparation.htm

[32]Refaat, Mamdouh, *Data preparation for data mining using SAS*, Amsterdam Boston: Morgan Kaufmann Publishers, 2007

[33]Han, J., and Kamber, M, Data Mining: Concepts and Techniques, 3rd, 2006

[34]http://the-modeling-agency.com/crisp-dm.pdf

[35]M.J. Crawley, *STATISTICS: AN INTRODUCTION USING R*, Paperback, 2005

[36]W. N. Venables, D. M. Smith and the R Core Team, *An Introduction to R*, *Notes on R: A Programming Environment for Data Analysis and Graphics* Version 3.1.1, 2014

[37]Ian H.Witten & Eibe Frank, *Data Mining Practical Machine Learning Tools and Techniques*, 2005

[38]William J. O'Neil, *How to Make Money in Stocks: A Winning System in Good Times or Bad*, 1995

[39]Agrawal. , *Mining association rules between sets of items in large databases*, ACM SIGMOD International Conference on Management of Data, 1995

[40]Cios, K. j., *Data Mining: Knowledge discovery approach*, Springer, 2007

[41]Hegland., The Apriori Algorithm, 2005

[42]M.kamber, J. H., *Data Mining: concept and Techniques.* Morgan Kauman, 2000

[43]Nigel, M. C. , *New Ranking algorithm for parsing and tagging.* Philadelphia, 2002

[44]Sahoo, Y. K., *Analysis Of Bayes,Neural Network and Tree classifier of classifcation Technique in Dats Mining using WEKA*, Computer Science & Information Technology, 2012

[45]Public procurment Sector Assesment, Review of law and practise in the SEMED Reigon, 2013

[46]Organization for Economic Cooperation and Development, *Detecting Bid Rigging in Puplic Procurement: helping governments to obtain best value for money*, 2010

[47]OECD, *Guidelines for Fighting Bid Rigging in Public Procurement*, 2010

[48]Bajari, P. and Summers, G., Detecting collusion in procurement auctions, Antitrust Law Journal, Vol 145, 2003

[49]Porter, R and Zona , D., *Detection of bid rigging in procurement auctions*, NBER, 1992.

[50]Porter, R. and Zona, D., *Ohio School Milk Markets: An Analysis of Bidding*, RAND Journal of Economics, 1999

[51] Joseph E., *Behavioral Screening and the Detection of Cartels*, EU Competition Law and Policy Workshop/Proceedings, 2006

[52]Rosa Abrantes- Metz and Luke Froeb, *Competition Authorities Are Screening for Conspiracies: What Are They Likely to Find?*, ABA Section of Antitrust Law Economics Committee Newsletter, 2008

[53]Rosa M. Abrantes-Metz and Sumanth Addanki, *Is the Market Being Fooled? An Error-Based Test for Manipulation*, 2007

[54]Porter, R. and Zona, D., *Ohio School Milk Markets: An Analysis of Bidding*, RAND Journal of Economics, 30, 263-88, 1999.

[55]Antitrust Div., U.S. Dep’t of Justice, *Price Fixing, Bid Rigging, and Market Allocation Schemes: What They Are and What to Look For*, 2000

[56]GAO, *DATA MINING Federal Efforts Cover a Wide Range of Uses*, United States General Accounting Office, 2004

[57]OECD, *Increasing competition by reducing Bid Rigging in Latin America*, Fifth meeting of the Latin American Competition forum, 2007

[58]Korean Government, *Collusion and Corruption in public procurement*, globule forum on competition, 2010

[59] http://www.comprasnet.gov.br/

[60]Jones, A. and Sufrin, B. E., *EU competition law: text, cases, and materials*, Oxford Publisher, 2014

[61]https://www.accc.gov.au/business/anti-competitive-behaviour/cartels/bid-rigging

[62]Murray, S.B., Cummis, S. and Gross, P.C., *Bid-Rigging Remains Focus Of DOJ Antitrust Criminal Enforcement: Businesses Need To Ensure Their Compliance*, The Metropolitan Corporate Counsel, 2013

[63]http://www.mondaq.com/australia/x/269072/Antitrust+Competition/

[64]D. Lucking-Reiley, D. Bryan, N. Prasad and D. Reeves, *Pennies from eBay: The determinants of price in online auctions*, 2000

[65]D. Houser and J. Wooders, *Reputation in auctions: Theory, and evidence from eBay*, 2000.

[66]A.E. Roth and A. Ockenfels, *Last minute bidding and the rules for ending second-price auctions: Theory and evidence from a natural experiment on the Internet*, 2001

[67]M. Utku Ünver, *Internet auctions with artificial adaptive agents: Evolution of late and multiple bidding*, 2001

[68] R. Bapna, P. Goes and A. Gupta, *A theoretical and empirical investigation of multi-item on-line auctions. Information Technology and Management*, 2000

[69]Organization for Economic Cooperation and Development, *Detecting Bid Rigging in Public Procurement: helping governments to obtain best value for money*, 2010

[70]Organization for Economic Cooperation and Development, *Guidelines for Fighting Bid Rigging in Public Procurement*, 2010.

[71]Antitrust Division, U.S. Department of Justice, *Price Fixing, Bid Rigging, and Market Allocation Schemes: What They Are and What to Look For*, 2011

Appendices

Appendix A: The Egyptian International School Bids

The following table lists the various bidding suppliers in their won bids, together with the location and dates of those bids.

Supplier	Winner Bid Value (EGP)	Quantity	Location	Date
C2	12690	705	Cairo	Sunday, July 9, 1995
C1	5757	303	Cairo	Tuesday, November 7, 1995
C2	14520	726	Cairo	Monday, July 15, 1996
C2	14605	635	Cairo	Wednesday, November 20, 1996
C2	16399	713	Cairo	Tuesday, October 7, 1997
C2	13368	557	Cairo	Wednesday, November 5, 1997
C1	11316	492	Cairo	Monday, July 13, 1998
C2	17520	730	Cairo	Sunday, November 1, 1998
C2	12744	708	Cairo	Thursday, July 1, 1999
C2	14364	798	Cairo	Wednesday, November 3, 1999
C1	3045	145	Cairo	Wednesday, July 5, 2000
C2	11080	554	Cairo	Thursday, November 9, 2000
C2	15168	632	Cairo	Sunday, July 8, 2001
C2	16390	745	Cairo	Sunday, November 11, 2001
C1	8280	345	Cairo	Sunday, July 7, 2002
C1	8910	495	Cairo	Sunday, November 3, 2002
C2	12180	609	Cairo	Wednesday, July 9, 2003
C2	17273	751	Cairo	Saturday, July 10, 2004
C2	15180	690	Cairo	Tuesday, July 5, 2005
C1	6308	332	Cairo	Sunday, July 9, 2006
C2	16896	768	Cairo	Sunday, November 11, 2007
C1	4669	203	Cairo	Wednesday, November 12, 2008
C2	12906	717	Cairo	Tuesday, July 7, 2009
C2	16742	761	Cairo	Saturday, July 11, 2009
C1	3969	189	Cairo	Monday, November 15, 2010
C2	15260	763	Cairo	Saturday, November 12, 2011
C1	4400	200	Cairo	Thursday, July 12, 2012

C2	15709	683	Cairo	Thursday, December 13, 2012
C2	9954	553	Cairo	Sunday, May 12, 2013
C1	10186	463	Cairo	Friday, May 9, 2014

Table 6 The Egyptian International School Bids

Appendix B: The public procurement Bids for oil market

The following illustrates the bids, together with their identification numbers and assigned geographic region.

Bid_Id	Bid_Date	Place
92	Thursday, January 03, 2008	Place 1
31	Thursday, February 07, 2008	Place 2
93	Wednesday, February 13, 2008	Place 1
30	Thursday, February 28, 2008	Place 2
104	Wednesday, March 12, 2008	Place 2
94	Tuesday, April 08, 2008	Place 1
32	Tuesday, April 15, 2008	Place 2
36	Thursday, April 17, 2008	Place 2
35	Wednesday, April 30, 2008	Place 2
95	Tuesday, May 06, 2008	Place 1
34	Tuesday, May 13, 2008	Place 2
29	Monday, May 26, 2008	Place 2
107	Thursday, May 29, 2008	Place 2
96	Tuesday, June 17, 2008	Place 1
27	Thursday, June 19, 2008	Place 2
97	Thursday, July 03, 2008	Place 1
26	Tuesday, July 08, 2008	Place 2
33	Tuesday, July 29, 2008	Place 2
98	Tuesday, August 05, 2008	Place 1
99	Tuesday, August 12, 2008	Place 1
25	Thursday, September 04, 2008	Place 2
24	Monday, September 08, 2008	Place 2
101	Tuesday, September 09, 2008	Place 1
22	Wednesday, September 24, 2008	Place 2
23	Monday, September 29, 2008	Place 2
102	Tuesday, October 14, 2008	Place 1
21	Tuesday, November 25, 2008	Place 2
20	Tuesday, December 02, 2008	Place 2
103	Wednesday, December 03, 2008	Place 1
100	Friday, December 12, 2008	Place 1

19	Tuesday, December 16, 2008	Place 2
18	Tuesday, December 23, 2008	Place 2
78	Thursday, January 15, 2009	Place 1
79	Thursday, January 22, 2009	Place 1

The public procurement Bids for oil market

81	Thursday, February 05, 2009	Place 1
17	Tuesday, February 10, 2009	Place 2
51	Thursday, February 12, 2009	Place 2
1	Thursday, March 05, 2009	Place 2
82	Thursday, March 05, 2009	Place 1
2	Thursday, March 12, 2009	Place 2
3	Thursday, April 02, 2009	Place 2
4	Wednesday, April 08, 2009	Place 2
50	Wednesday, April 08, 2009	Place 2
84	Thursday, April 09, 2009	Place 1
5	Tuesday, May 12, 2009	Place 2
85	Thursday, May 14, 2009	Place 1
6	Tuesday, May 26, 2009	Place 2
86	Thursday, June 11, 2009	Place 1
7	Thursday, June 18, 2009	Place 2
8	Wednesday, July 15, 2009	Place 2
87	Tuesday, July 28, 2009	Place 1
14	Wednesday, August 05, 2009	Place 2
41	Wednesday, August 05, 2009	Place 2
45	Wednesday, August 05, 2009	Place 2
106	Wednesday, August 05, 2009	Place 2
88	Tuesday, September 08, 2009	Place 1
10	Tuesday, September 15, 2009	Place 2
11	Wednesday, September 23, 2009	Place 2
89	Wednesday, September 30, 2009	Place 1
43	Friday, October 02, 2009	Place 2
37	Monday, October 12, 2009	Place 2
12	Wednesday, November 04, 2009	Place 2
38	Wednesday, November 04, 2009	Place 2
90	Thursday, November 05, 2009	Place 1
53	Saturday, December 05, 2009	Place 2
13	Thursday, December 10, 2009	Place 2
105	Thursday, December 10, 2009	Place 2
91	Tuesday, December 15, 2009	Place 1
72	Wednesday, January 06, 2010	Place 1

66	Tuesday, January 12, 2010	Place 2
68	Tuesday, January 19, 2010	Place 2
109	Tuesday, January 19, 2010	Place 1
65	Wednesday, February 10, 2010	Place 1

The public procurement Bids for oil market

69	Tuesday, February 16, 2010	Place 1
57	Saturday, February 27, 2010	Place 2
64	Wednesday, March 03, 2010	Place 2
63	Monday, March 15, 2010	Place 2
70	Tuesday, March 16, 2010	Place 1
62	Thursday, April 01, 2010	Place 2
71	Tuesday, April 20, 2010	Place 1
61	Tuesday, May 04, 2010	Place 2
110	Tuesday, June 01, 2010	Place 1
60	Wednesday, June 16, 2010	Place 2
59	Wednesday, June 23, 2010	Place 2
73	Tuesday, July 13, 2010	Place 1
58	Tuesday, July 27, 2010	Place 2
74	Tuesday, August 10, 2010	Place 1
56	Wednesday, August 18, 2010	Place 2
75	Tuesday, August 24, 2010	Place 1
55	Wednesday, September 15, 2010	Place 2
76	Wednesday, September 22, 2010	Place 1
54	Tuesday, October 05, 2010	Place 2
77	Wednesday, October 13, 2010	Place 1

Table 7 Egyptian Market Bids Details

The following table provides the data for bids on the Oil Market, including the participating companies, their prices, and whether they won the bid or not.

Company_Id	Bid_Id	Company_Price	WinOr Not	Final_Company_Price	Bid_Value	Oil Type 1	Quantity
1	19	672.35	FALSE	0	0	Soya	0
1	20	734.75	FALSE		0	Sun flower	0
1	21	743.85	FALSE	0	0	Soya	0
1	22	1014.5	TRUE	1013.5	0	Sun flower	15000
1	23	1046.35	FALSE	0	0	Soya	0
1	26	1441.75	FALSE	0	0	Soya	0
1	27	1476.75	FALSE	0	0	Soya	0
1	31	1774.85	FALSE	0	0	Sun flower	0
1	32	1753	TRUE	1753	0	Sun flower	15000
1	33	1336.55	FALSE	0	0	Soya	0
1	34	1398.45	FALSE	0	0	Soya	0
1	35	1343.6	FALSE	0	0	Soya	0
1	36	1466.7	TRUE	1466.7	0	Soya	15000
1	54	1022	FALSE	0	0	Soya	0
1	55	1068.65	FALSE	0	0	Sun flower	0
1	56	964.41	FALSE	0	0	Soya	0
1	58	935	TRUE	911	0	Sun flower	6000
1	59	838	FALSE	0	0	Sun flower	0
1	60	833	FALSE	0	0	Sun flower	0
1	61	873.65	TRUE	873.5	0	Sun flower	20000
1	62	874.25	FALSE	0	0	Sun flower	0
1	63	869.85	FALSE	0	0	Soya	0
1	65	884	FALSE	0	0	Soya	0
1	66	928.85	TRUE	920	0	Sun flower	10000
1	69	893	TRUE	893	0	Sun flower	6000
1	70	888	TRUE	0	0	Sun flower	10000

1	71	865	TRUE	865	865000 0	Sun flower	10000
1	73	859	FALSE	859	0	Soya	15000
1	73	859	TRUE	859	0	Sun flower	10000
1	74	982	FALSE	982		Soya	20000
1	74	1100	TRUE	1100	0	Sun flower	6000
1	75	1008	TRUE	1008	0	Sun flower	15000
1	76	1132	FALSE	1132	0	Soya	6000
1	76	1132	TRUE	1132	0	Sun flower	6000
1	77	1233	FALSE	1233		None	0
1	78	724	FALSE	724		Soya	25000
1	79	747	FALSE	0	0	Soya	25000
1	79	744	FALSE	0	0	Sun flower	6000
1	81	723	FALSE			Soya	25000
1	81	729	TRUE	729	0	Sun flower	6000
1	82	685	FALSE	0	0	Soya	20000
1	82	709	FALSE	0	0	Sun flower	6000
1	84	765	FALSE		0	Sun flower	2000
1	85	896	FALSE	0	0	Soya	15000
1	85	895	FALSE		0	Sun flower	5000
1	86	884	FALSE	0	0	Soya	20000
1	86	857	TRUE	857	0	Sun flower	35000
1	87	779	FALSE	0	0	Soya	15000
1	87	747	TRUE		0	Sun flower	20000
1	88	812	FALSE		0	Soya	20000
1	88	764	FALSE	764		Sun flower	3000
1	89	824	FALSE		0	Soya	20000
1	89	764	TRUE	761	0	Sun flower	15000
1	90	0	FALSE	0	0	None	0
1	90	858	FALSE	0	0	Soya	18000

1	90	815	TRUE	815	0	Sun flower	25000
1	91	909	FALSE		0	Soya	20000
1	91	912	TRUE	912	0	Sun flower	20000
1	92	1164	TRUE	0	0	Soya	15000
1	93	1375	FALSE	1350	0	Soya	15000
1	93	1758	TRUE	1756	0	Sun flower	6000
1	94	1349	FALSE	0	0	Soya	15000
1	95	1338	FALSE	0	0	Soya	15000
1	95	1890	FALSE	0	0	Sun flower	6000
1	96	1474	FALSE	0	0	Soya	15000
1	96	1816	FALSE			Sun flower	6000
1	109	894	FALSE	0	0	Soya	20000
1	109	917	FALSE	0	0	Sun flower	6000
1	110	843	TRUE	843	0	Soya	18000
2	20	741	FALSE	0	0	Sun flower	0
2	21	748.88	FALSE	0	0	Soya	0
2	22	1035	FALSE	0	0	Sun flower	0
2	23	1051	FALSE	0	0	Soya	0
2	26	1433.69	FALSE	0	0	Soya	0
2	27	1487.88	FALSE	0	0	Soya	0
2	31	1762	TRUE	1762	0	Sun flower	3000
2	33	1330.88	FALSE	0	0	Soya	0
2	34	1409.22	FALSE	0	0	Soya	0
2	35	1344.98	FALSE	0	0	Soya	0
2	36	1469.69	FALSE	0	0	Soya	0
2	54	1024.42	FALSE	0	0	Soya	0
2	55	1063.75	TRUE	1063	0	Sun flower	9000
2	59	832.32	FALSE	0	0	Sun flower	0
2	60	823.88	FALSE	0	0	Sun flower	0
2	61	892.49	FALSE	0	0	Sun flower	0

2	62	877.78	TRUE	868.7	0	Sun flower	10000
2	63	867.92	FALSE	0	0	Soya	0
2	65	882.72	FALSE	0	0	Soya	0
2	66	929.8	FALSE	0	0	Sun flower	0
2	68	878	TRUE	878	0	Soya	20000
2	69	878	FALSE	0	0	Soya	0
2	69	909	FALSE	0	0	Sun flower	15000
2	70	854	FALSE	854	854	Soya	20000
2	70	892	FALSE	0	0	Sun flower	6000
2	71	851	FALSE	0	0	Soya	20000
2	71	865	FALSE	0	0	Sun flower	10000
2	73	855	FALSE	855	0	Soya	8000
2	73	885	FALSE	885	0	Sun flower	10000
2	74	990	FALSE	990	0	Soya	20000
2	75	1044	FALSE		0	Sun flower	9000
2	76	1016	FALSE	0	0	Soya	20000
2	76	1147	FALSE	0	0	Sun flower	6000
2	77	1098	FALSE	1098	0	Soya	20000
2	77	1236	FALSE	1236	0	Sun flower	6000
2	78	747	FALSE	747	0	Soya	20000
2	78	752	FALSE	752	0	Sun flower	6000
2	79	729	FALSE	0	0	Soya	25000
2	79	760	FALSE	0	0	Sun flower	6000
2	81	723	TRUE	720	0	Soya	25000
2	81	729	TRUE	729	0	Sun flower	6000
2	82	688	FALSE	0	0	Soya	25000
2	82	713	FALSE	0	0	Sun flower	6000
2	84	779	FALSE	0	0	Soya	25000
2	84	750	FALSE		0	Sun flower	9000
2	85	893	FALSE		0	Soya	20000

2	85	878	TRUE	876	0	Sun flower	15000
2	86	882	FALSE	0	0	Soya	20000
2	86	874	FALSE	0	0	Sun flower	9000
2	87	797	FALSE	0	0	Soya	15000
2	87	777	FALSE	0	0	Sun flower	9000
2	88	759	TRUE	758	0	Sun flower	9000
2	89	813	FALSE	0	0	Soya	20000
2	89	774	FALSE	0	0	Sun flower	6000
2	90		FALSE	0	0	None	0
2	90	864	FALSE	0	0	Soya	15000
2	90	827	FALSE	0	0	Sun flower	3000
2	91	915	FALSE	0	0	Soya	20000
2	91	916	FALSE	0	0	Sun flower	3500
2	92	1177	FALSE	0	0	Soya	15000
2	93	1369	FALSE	0	0	Soya	15000
2	93	1769	FALSE	0	0	Sun flower	6000
2	94	1342	FALSE	0	0	Soya	15000
2	96	1498	FALSE	0	0	Soya	20000
2	96	1790	TRUE	1790	0	Sun flower	3000
2	109	878	TRUE	878	0	Soya	20000
2	110	829	FALSE	829	0	Soya	20000
2	110	848	FALSE	0	0	Sun flower	15000
3	18	668.95	TRUE	668.95		Soya	20000
3	19	669.95	FALSE	0	0	Soya	
3	20	719	TRUE	718	0	Sun flower	10000
3	21	747.75	FALSE	0	0	Soya	0
3	22	1042	FALSE	0	0	Sun flower	0
3	23	1049.75	FALSE	0	0	Soya	0
3	26	1444.35	FALSE	0	0	Soya	0
3	27	1476.75	FALSE	0	0	Soya	0
3	31	1770	FALSE	0	0	Sun flower	0

3	32	1767.2	FALSE	0	0	Sun flower	0
3	33	1318.87	FALSE	0	0	Soya	0
3	35	1367.75	FALSE	0	0	Soya	0
3	36	1472.75	FALSE	0	0	Soya	0
3	54	1011	FALSE	0	0	Soya	0
3	55	1063	TRUE	1063	0	Sun flower	6000
3	56	947	TRUE	945	0	Soya	15000
3	58	911	TRUE	911	0	Sun flower	6000
3	59	831	TRUE	830.5	0	Sun flower	6000
3	60	819	TRUE	819	0	Sun flower	6000
3	61	876	FALSE		0	Sun flower	
3	62	878	FALSE		0	Sun flower	20000
3	63	873	FALSE	0	0	Soya	0
3	65	889	FALSE	0	0	Soya	0
3	66	937	FALSE	0	0	Sun flower	0
3	68	900	TRUE	900	0	Sun flower	15000
3	69	891	FALSE	891	0	Soya	15000
3	69	899	FALSE	899	0	Sun flower	9000
3	70	864	FALSE	0	0	Soya	20000
3	70	897	FALSE	0	0	Sun flower	6000
3	71	856	FALSE	0	0	Sun flower	9000
3	73	852	FALSE	852	0	Soya	15000
3	73	859	TRUE	859	0	Sun flower	6000
3	74	965	FALSE	965	0	Soya	20000
3	75	1011	TRUE	1007	0	Sun flower	6000
3	76	1012	FALSE		0	Soya	20000
3	76	1135	TRUE	1132		Sun flower	6000
3	77	1097	FALSE	0	0	Soya	20000

3	77	1214	TRUE	1214	0	Sun flower	6000
3	78	729	FALSE	729		Soya	20000
3	78	727	FALSE	727	0	Sun flower	9000
3	79	727	FALSE	0	0	Soya	20000
3	79	747	FALSE	0	0	Sun flower	6000
3	81	717	TRUE	717	0	Soya	20000
3	81	732	FALSE		0	Sun flower	9000
3	82	693	FALSE		0	Soya	20000
3	82	708	FALSE	0	0	Sun flower	9000
3	84	775	TRUE	0	0	Soya	20000
3	84	732	FALSE			Sun flower	9000
3	85	887	FALSE		0	Soya	9000
3	85	876	TRUE	876	0	Sun flower	9000
3	86	891	FALSE	0	0	Soya	20000
3	86	874	FALSE			Sun flower	9000
3	87	790	FALSE	0	0	Soya	15000
3	87	773	FALSE	0	0	Sun flower	6000
3	88	802	FALSE	802		Soya	20000
3	88	768	FALSE	768	0	Sun flower	9000
3	89	804	FALSE	0	0	Soya	15000
3	89	761	TRUE	761	0	Sun flower	9000
3	90	857	FALSE		0	Soya	20000
3	90	824	FALSE		0	Sun flower	9000
3	91	899	TRUE	899	0	Soya	20000
3	91	929	FALSE	0	0	Sun flower	9000
3	92	1182	FALSE	0	0	Soya	15000
3	93	1375	FALSE	0	0	Soya	10000
3	93	1767	FALSE	0	0	Sun flower	3000
3	94	1342	TRUE	1333	0	Soya	20000
3	95	1355	FALSE	0	0	Soya	15000

3	96	1817	FALSE	0	0	Sun flower	6000
3	109	900	TRUE	900	0	Soya	15000
3	109	900	TRUE	900	0	Sun flower	5000
3	110	828	FALSE	0	0	Soya	20000
3	110	843	TRUE	0	0	Sun flower	9000
4	22	1070	FALSE	0	0	Sun flower	0
4	31	1780	FALSE	0	0	Sun flower	0
4	66	934.75	FALSE	0	0	Sun flower	0
4	70	896	FALSE	0	0	Sun flower	2000
4	78	720	TRUE	725		Sun flower	6000
4	79	769	FALSE	769	0	Sun flower	6000
4	86	887	FALSE		0	Sun flower	3000
4	88	748	TRUE	748		Sun flower	15000
4	89	761	TRUE	761	0	Sun flower	6000
4	96	1830	FALSE	0	0	Sun flower	6000
5	19	669	FALSE	0	0	Soya	0
5	20	721.5	FALSE	0	0	Sun flower	0
5	21	745.2	FALSE	0	0	Soya	0
5	22	1025	FALSE	0	0	Sun flower	0
5	23	1065	FALSE	0	0	Soya	0
5	26	1436.5	FALSE	0	0	Soya	0
5	34	1444	FALSE	0	0	Soya	0
5	35	1387	FALSE	0	0	Soya	0
5	36	1503	FALSE	0	0	Soya	0
5	54	1035	FALSE	0	0	Soya	0
5	62	884.7	FALSE	0	0	Sun flower	0
5	63	867.9	FALSE	0	0	Soya	0

5	66	951	FALSE	0	0	Sun flower	0
5	69	893	TRUE	893	0	Sun flower	18000
5	70	845	TRUE	0	0	Soya	20000
5	71	878	FALSE	0	0	Sun flower	6000
5	75	1067	FALSE	0	0	Sun flower	10000
5	76	1019	FALSE		0	Soya	20000
5	77	1102	FALSE	1102	0	Soya	20000
5	78	747	FALSE	747	0	Soya	10000
5	78	762	FALSE		0	Sun flower	6000
5	79	737	TRUE	728	0	Soya	10000
5	79	766	FALSE	0	0	Sun flower	2000
5	81	729	FALSE	0	0	Soya	20000
5	81	742	FALSE			Sun flower	6000
5	82	692	FALSE	0	0	Soya	20000
5	82	692	FALSE	0	0	Sun flower	9000
5	84	790	FALSE		0	Soya	20000
5	85	908	FALSE		0	Soya	20000
5	85	895	FALSE			Sun flower	6000
5	86	895	FALSE	0	0	Soya	20000
5	86	874	FALSE	0	0	Sun flower	9000
5	87	801	FALSE		0	Soya	15000
5	88	812	FALSE		0	Soya	15000
5	88	789	FALSE	0	0	Sun flower	15000
5	89	810	FALSE	0	0	Soya	15000
5	89	779	FALSE	779	0	Sun flower	15000
5	90	876	FALSE	0	0	Soya	18000
5	91	943	FALSE	0	0	Sun flower	15000
5	95	1373	FALSE	0	0	Soya	15000
5	96	1528	FALSE	0	0	Soya	20000
5	109	898	FALSE	0	0	Soya	15000

5	109	947	FALSE		0	Sun flower	18000
6	23	1046	FALSE	1043	0	Soya	25000
6	26	1439.4	FALSE	0	0	Soya	0
6	33	1312.7	TRUE	1305	0	Soya	20000
6	34	1394.8	TRUE	1390	0	Soya	25000
6	35	1343	TRUE	1340	0	Soya	25000
6	54	1022.4	FALSE	0	0	Soya	0
6	56	957	FALSE	0	0	Soya	0
6	63	863	TRUE	863	0	Soya	20000
6	65	884	TRUE	881	0	Soya	15000
6	69	879	FALSE	879	0	Soya	20000
6	70	888	TRUE	888	0	Sun flower	18000
6	71	852	FALSE	0	0	Soya	20000
6	73	864	FALSE	864	0	Soya	15000
6	74	964	FALSE	964	0	Soya	20000
6	76	1009	FALSE	1009	0	Soya	20000
6	77	1089	FALSE	1089	0	Soya	20000
6	78	729	FALSE	0	0	Soya	25000
6	79	755	FALSE	0	0	Soya	18000
6	81	732	FALSE	0	0	Soya	25000
6	82	677	TRUE	677	0	Soya	25000
6	82	677	TRUE	677	0	Sun flower	6000
6	84	785	FALSE	0	0	Soya	20000
6	85	886	FALSE	0	0	Soya	25000
6	85	886	FALSE		0	Sun flower	5000
6	86	892	FALSE		0	Soya	20000
6	86	888	FALSE	892	0	Sun flower	10000
6	87	797	FALSE	0	0	Soya	15000
6	88	799	FALSE	0	0	Soya	25000
6	89	812	FALSE	812	0	Soya	18500
6	90	860	FALSE	0	0	Soya	20000
6	91	913	FALSE	0	0	Soya	20000
6	95	1334	TRUE	1330	0	Soya	15000
6	110	820	FALSE	0	0	Soya	20000
7	14	805	FALSE	0	0	None	
7	21	4000	FALSE	0	0	Soya	0
7	24	6200	TRUE	6200	0	Soya	8000

7	25	6570	FALSE	0	0	Soya	0
7	29	7860	FALSE	0	0	Soya	0
7	31	9825	TRUE	9785	0	Sun flower	2500
7	32	10375	FALSE	0	0	Sun flower	0
7	33	7150	FALSE	0	0	Soya	0
7	88	774	FALSE		0	Sun flower	5000
7	95	1729	TRUE	1729	0	Sun flower	6000
7	96	1377	TRUE	1374	0	Soya	4000
7	96	1780	TRUE	1701	0	Sun flower	4000
7	107	10175	TRUE	10175	0	Sun flower	2000
8	14	830	FALSE	0	0	None	
8	21	752	FALSE	0	0	Soya	0
8	26	1455	FALSE	0	0	Soya	0
8	27	1497.43	FALSE	0	0	Soya	0
8	33	1336	FALSE	0	0	Soya	0
8	34	1435	FALSE	0	0	Soya	0
8	35	1359.65	FALSE	0	0	Soya	0
8	36	1472.31	FALSE	0	0	Soya	0
8	54	1004.34	TRUE	1004.34	0	Soya	20000
8	56	952.69	FALSE	0	0	Soya	0
8	63	874.63	FALSE	0	0	Soya	0
8	65	890.95	FALSE	0	0	Soya	0
8	66	930	FALSE	0	0	Sun flower	0
8	69	879	FALSE			Soya	20000
8	70	849	FALSE	0	0	Soya	20000
8	71	841	TRUE		0	Soya	20000
8	73	858	FALSE	858	0	Soya	15000
8	74	965	FALSE	965	0	Soya	20000
8	76	1005	TRUE	1005	0	Soya	20000
8	77	1075	FALSE		0	Soya	20000
8	78	727	FALSE	0	0	Soya	12000
8	78	730	TRUE	725		Sun flower	6000
8	79	744	FALSE	0	0	Soya	6000
8	81	725	FALSE		0	Soya	25000
8	82	691	FALSE	0	0	Soya	20000

8	85	889	FALSE		0	Soya	20000
8	85	880	FALSE	880	0	Sun flower	6000
8	86	882	FALSE	882	0	Soya	20000
8	88	803	FALSE	803	0	Soya	15000
8	89	808	FALSE	0	0	Soya	15000
8	90	856	FALSE	0	0	Soya	20000
8	91	904	FALSE	0	0	Soya	15000
8	92	1177	FALSE	0	0	Soya	15000
8	93	1376	FALSE	0	0	Soya	15000
8	94	1357	FALSE	0	0	Soya	20000
8	95	1351	FALSE	0	0	Soya	15000
8	96	1509	FALSE	0	0	Soya	15000
8	109	908	FALSE	0	0	Sun flower	6000
9	61	884.4	FALSE	0	0	Sun flower	0
9	62	868.7	TRUE	868.7	0	Sun flower	6000
9	65	882.45	FALSE	0	0	Soya	0
9	68	50	FALSE	440	40	None	54
9	89	0	FALSE		0	None	0
9	92	1169	TRUE	1169	0	Soya	5000
10	81	749	FALSE			Sun flower	2000
10	82	728	FALSE		0	Sun flower	6000
10	84	732	TRUE	732	0	Sun flower	6000
10	86	864	FALSE		0	Sun flower	3000
11	66	920	TRUE	920	0	Sun flower	5000
11	91	920	FALSE	0	0	Sun flower	5000
14	19	3666	TRUE	3666	0	Soya	5000
14	21	3737	TRUE	3737	0	Soya	15000
14	25	6215	TRUE	6215	0	Soya	4000
14	26	7654	FALSE	0	0	Soya	0
14	27	7970	FALSE	0	0	Soya	0
14	29	7654	FALSE	0	0	Soya	0
14	30	8299	TRUE	0	0	Soya	15000
14	33	7130	FALSE	0	0	Soya	0

14	36	8080	FALSE	0	0	Soya	0
14	54	5858	FALSE	0	0	Soya	0
14	64	4949	TRUE	4850	0	Soya	10000
14	69	850	TRUE	0	0	Soya	13000
14	70	821	TRUE	821	0	Soya	10000
14	71	841	FALSE	841	0	Soya	20000
14	73	850	TRUE	850		Soya	15000
14	74	965	TRUE	964	0	Soya	8000
14	76	1042	FALSE	1042	0	Soya	5000
14	77	1071	TRUE	1071	0	Soya	13000
14	78	708	TRUE	708	0	Soya	10000
14	82	682	TRUE	666	0	Soya	5000
14	84	759	TRUE	759	0	Soya	20000
14	85	897	FALSE		0	Soya	15000
14	86	864	FALSE		0	Soya	12000
14	88	783	TRUE	771	0	Soya	10000
14	89	780	TRUE	783	0	Soya	8000
14	90	829	TRUE	829	0	Soya	10000
14	91	876	TRUE	875		Soya	10000
14	93	1328	FALSE	0	0	Soya	15000
14	95	1289	FALSE	0	0	Soya	7500
14	96	1417	TRUE	1394	0	Soya	16000
14	110	828	TRUE	828	0	Soya	28000
15	21	4100	FALSE	0	0	Soya	0
15	24	6300	TRUE	6200	0	Soya	2500
15	25	6350	FALSE	0	0	Soya	0
15	26	7950	FALSE	0	0	Soya	0
15	27	7900	TRUE	7850	0	Soya	6000
15	29	7350	TRUE	7350	0	Soya	5000
15	33	7200	FALSE	0	0	Soya	0
15	36	8150	FALSE	0	0	Soya	0
15	54	5950	FALSE	0	0	Soya	0
15	64	4850	TRUE	4850	0	Soya	5000
15	69	850	TRUE	850	0	Soya	13000
15	71	824	TRUE	824		Soya	5000
15	73	850	TRUE	850	0	Soya	5000
15	74	966	TRUE	964	0	Soya	10000
15	76	1008	TRUE	1008	0	Soya	7000
15	77	6110	TRUE	6110	0	Soya	8180
15	87	728	TRUE	0	0	Soya	8000
15	88	763	TRUE	763	0	Soya	5000

15	89	771	TRUE	771	0	Soya	5000
15	90	831	TRUE	829	0	Soya	7000
15	110	828	TRUE	828	0	Soya	10000
17	22	1059	FALSE	0	0	Sun flower	0
18	25	6500	FALSE	0	0	Soya	0
18	29	8250	FALSE	0	0	Soya	0
19	26	7950	FALSE	0	0	Soya	0
19	54	1027	FALSE	0	0	Soya	0
19	56	972.5	FALSE	0	0	Soya	0
19	59	840.3	FALSE	0	0	Sun flower	0
19	60	825.3	FALSE	0	0	Sun flower	0
19	61	881.3	FALSE	0	0	Sun flower	0
19	62	882.6	FALSE	0	0	Sun flower	0
19	63	880	FALSE	0	0	Soya	0
19	65	896	FALSE	0	0	Soya	0
19	68	906	FALSE	0	0	Sun flower	10000
19	69	904	FALSE	904	0	Soya	20000
19	69	915	FALSE	915	0	Sun flower	20000
19	70	865	FALSE	0	0	Soya	20000
19	70	909	FALSE	0	0	Sun flower	6000
19	71	856	FALSE	0	0	Soya	20000
19	71	869	FALSE	0	0	Sun flower	6000
19	73	868	FALSE	868	0	Soya	15000
19	76	1023	FALSE	1023	0	Soya	20000
19	77	1097	FALSE	0	0	Soya	20000
19	109	892	FALSE	892	0	Soya	15000
19	109	906	FALSE	906	0	Sun flower	10000
19	110	824	FALSE		0	Soya	20000
19	110	861	FALSE	861	0	Sun flower	6000
20	27	8330	FALSE	0	0	Soya	0
20	33	7436	FALSE	0	0	Soya	0
21	107	10225	FALSE	0	0	Sun flower	0

22	29	8244	FALSE	0	0	Soya	0
22	36	8187	FALSE	0	0	Soya	0
23	88	758	TRUE	758	0	Sun flower	6000
23	89	767	FALSE	0	0	Sun flower	6000
23	91	918	FALSE	0	0	Sun flower	5800
23	110	868	FALSE	868	0	Sun flower	6000
24	79	811	FALSE	811	0	Sun flower	2000
25	79	729	FALSE	0	0	Soya	25000
25	88	907	FALSE	0	0	Soya	20000
26	93	1328	TRUE	1328	0	Soya	3000
26	93	1728	FALSE	0	0	Sun flower	4000

Table 8 Egyptian Market Companies Bids

Finally, the following table lists the data for the Egyptian International Schools bids, including the participating companies, their bid values, and the products they competed for.

Supplier	Winner Bid Value (EGP)	Quantity	Quantity (Categorized)	Location	Product	Date
C2	12690	705	Greater Than > 495	Cairo	Paper A3	Sunday, July 9, 1995
C1	5757	303	Less Than or Equal 495	Cairo	Paper A4	Tuesday, November 7, 1995
C2	14520	726	Greater Than > 495	Cairo	Paper A4	Monday, July 15, 1996
C2	14605	635	Greater Than > 495	Cairo	Paper A4	Wednesday, November 20, 1996
C2	16399	713	Greater Than > 495	Cairo	Paper A4	Tuesday, October 7, 1997
C2	13368	557	Greater Than > 495	Cairo	Paper A3	Wednesday, November 5, 1997
C1	11316	492	Less Than or Equal 495	Cairo	Paper A3	Monday, July 13, 1998
C2	17520	730	Greater Than > 495	Cairo	Paper A4	Sunday, November 1, 1998
C2	12744	708	Greater Than > 495	Cairo	Paper A4	Thursday, July 1, 1999
C2	14364	798	Greater Than > 495	Cairo	Paper A4	Wednesday, November 3, 1999
C1	3045	145	Less Than or Equal 495	Cairo	Paper A3	Wednesday, July 5, 2000
C2	11080	554	Greater Than > 495	Cairo	Paper A4	Thursday, November 9, 2000
C2	15168	632	Greater Than > 495	Cairo	Paper A4	Sunday, July 8, 2001

C2	16390	745	Greater Than > 495	Cairo	Paper A3	Sunday, November 11, 2001
C1	8280	345	Less Than or Equal 495	Cairo	Paper A4	Sunday, July 7, 2002
C1	8910	495	Less Than or Equal 495	Cairo	Paper A3	Sunday, November 3, 2002
C2	12180	609	Greater Than > 495	Cairo	Paper A4	Wednesday, July 9, 2003
C2	17273	751	Greater Than > 495	Cairo	Paper A4	Saturday, July 10, 2004
C2	15180	690	Greater Than > 495	Cairo	Paper A4	Tuesday, July 5, 2005
C1	6308	332	Less Than or Equal 495	Cairo	Paper A3	Sunday, July 9, 2006
C2	16896	768	Greater Than > 495	Cairo	Paper A4	Sunday, November 11, 2007
C1	4669	203	Less Than or Equal 495	Cairo	Paper A3	Wednesday, November 12, 2008
C2	12906	717	Greater Than > 495	Cairo	Paper A3	Tuesday, July 7, 2009
C2	16742	761	Greater Than > 495	Cairo	Paper A4	Saturday, July 11, 2009
C1	3969	189	Less Than or Equal 495	Cairo	Paper A3	Monday, November 15, 2010
C2	15260	763	Greater Than > 495	Cairo	Paper A4	Saturday, November 12, 2011
C1	4400	200	Less Than or Equal 495	Cairo	Paper A4	Thursday, July 12, 2012
C2	15709	683	Greater Than > 495	Cairo	Paper A3	Thursday, December 13, 2012
C2	9954	553	Greater Than > 495	Cairo	Paper A4	Sunday, May 12, 2013
C1	10186	463	Less Than or Equal 495	Cairo	Paper A3	Friday, May 9, 2014

Table 9 The Egyptian International Schools Bids (Modified)

University of Hradec Králové
Faculty of Informatics and Management

MASTER'S THESIS ASSIGNMENT FORM

Student's Surname, Forename(s):	**Alashrafy Ahmed Samir Mohamed AbdElAziz**
Branch of Study:	**Information Management**
Advisor's Name:	**Tučník Petr**
Thesis Title (80 characters max):	**Detection of Manipulation in Economic Data Using Data Mining Techniques**
Thesis Title in Czech:	Detekce manipulace s ekonomickými daty s využitím data miningu
Subtitle:	
Objectives (3 lines max.):	develop a module for fraud detection in procurement bidding using data mining techniques
Contents:	1.Introduction 2.Bid Rigging Detection 3.E-bay Case Study 4.Data Mining Techniques and Algorithms 5.Detection Of Manipulation in the Egyptian Market 6.Conclusion And Future Work

Board of Examiners:

Dated of:

Student's Signature: Ahmed Alashrafy

Advisor's Signature: ....

One copy to be retained by the student, one copy at the Department, and one copy is attached to the clean copy of the Thesis. See "Student's Guide to Bachelor's/Master's Thesis.