Dedication

This study is dedicated to our fathers, Thank you so so much for your passion for our profession and educating. I feel so lucky to have a professor like you, finding the just-right-challenge between pushing us to be better, encouraging us, and inspiring us as Developers. You are appreciated, and hope you feel encouraged as an educator and proud of the work you've done with our class. Thank you so much.

Big thanks for.

- Associated professor/ AMR EL Shora
- Associated professor/ Reham Esaa
- Dr/ AMIRA EL ATTAR
- Dr/ Mona Mowafi

Chapter 1 : An introduction

- Introduction
- Importance of the study
- The problem
- The objectives of the study
- The study samples
- Limitations

INTRODUCTION

progress in digital acquisition and storage technology has resulted in the development of vast database. This has happened in all areas of human attempt from the mundane to

the exotic. Little wonder than that attention has development in the possibility of tapping these data, of demanding from them information that might be of value to the proprietor of the database. The regulation concerned with this assignment has become recognized as data mining. Defining a scientific discipline is always a contentious task; researchers often disagree about the exact range and limits of their field of study.

Bearing this in and tolerant that others might oppose about the details, they shall accept as their functioning definition of data mining. Progress in digital acquisition storage technology has resulted in the development of vast database.

This has happened in all areas of human attempt from the mundane to the exotic. Little wonder then that attention has development in the possibility of tapping these data, of demanding from them information that might be of value to the proprietor of the database. The regulation concerned with this assignment has become recognized as data mining. Defining a scientific discipline is always a contentious task; researchers often disagree about the exact range and limits of their field of study. Bearing this in and tolerant that others might oppose about the details, they shall accept as their functioning definition of data mining. An explosive growth is witnessed in data generation and data collection capabilities across all domains. In the business world, very large databases on commercial transactions have been generated by retailers and ecommerce. Huge amount of scientific data have been generated in various fields as well. One case in point is the human genome project which has aggregated gigabytes of data on the human genetic code.

The World Wide Web provides another example with billions of web pages consisting of textual and multimedia information that are used by millions of people. Analyzing huge bodies of data that can be understood and used efficiently remains a challenging problem. Data mining addresses this problem by providing techniques and software to automate the analysis and exploration of large and complex data sets. Research on data mining is being pursued in a wide variety of fields, including statistics, computer science, machine learning, database management and data visualization. The process of digging through data to discover hidden connections and predict future trends has a long history.

Sometimes referred to as "knowledge discovery in databases," the term "data mining" wasn’t coined until the 1990s. But its foundation comprises three intertwined scientific disciplines: statistics (the numeric study of data relationships), artificial intelligence

(human-like intelligence displayed by software and/or machines) and machine learning (algorithms that can learn from data to make predictions).

What was old is new again, as data mining technology keeps evolving to keep pace with the limitless potential of big data and affordable computing power. Civil libertarians consider some databases held by businesses and governments to be an unwarranted intrusion and an invitation to abuse. For example, the American Civil Liberties Union sued the U.S. National Security Agency (NSA) alleging warrantless spying on American citizens through the acquisition of call records from some American telecommunication companies.

The program, which began in 2001, was not discovered by the public until 2006, when the information began to leak out. Often the risk is not from data mining itself (which usually aims to produce general knowledge rather than to learn information about specific issues) but from misuse or inappropriate disclosure of information in these databases. The growth and future of e-commerce in the Middle East Perhaps the most significant of all recent developments in retail is the growth and progress of e-commerce. According to a joint study by Dubai Economy and Visa, the UAE is currently the most advanced e-commerce market in the Middle East and North Africa, revealing an estimated annual growth of 23% between 2018 and 2022. While customers continue to enjoy the 'destination shopping' aspect of visiting a physical store, COVID-19 has encouraged many people to try e-commerce for the first time. This jump in online users is not expected to reduce.

However, predicting the post-COVID retail landscape is not simple. We anticipate the emergence of a hybrid shopping experience combining online and physical stores, because technology has not yet reached the point where consumers can experience texture in a virtual environment; we cannot touch, feel, smell or interact with products online. It is also worth mentioning the important social role that malls play in the Middle East, especially during the hot summer months. Individuals, couples, families, and groups of friends will continue to visit the malls for shopping, dining, entertainment – a chance to stretch their legs and socialize whilst respecting social distancing in large air-conditioned environments. Digital shopping will never replace the experience of physically interacting with a product, but with more people looking to buy online more often, providing additional context through digital means is a smart way to go. Provided that the supply chain is properly set up to support both online and physical stores, it's a winning strategy.

IMPORTANCE OF THE STUDY

brief introduction The retail industry in the Middle East and North Africa (MENA) is on the verge of a pivotal shift. E-commerce is becoming a reality, reinventing consumers' path to purchase, forming new customer experiences, disrupting business models, and creating growth opportunities for large and small retailers as well as for a new generation of e-commerce pure players.

E-commerce is the engine of global retail growth In 2017, the e-commerce industry came to the fore when it surpassed 10% of all global retail sales. Today, e-commerce is a $2.2 trillion market, and it is expanding at an annual growth rate of 24%, four times faster than the global retail sector as a whole. E-commerce is becoming the key growth engine for retail: Its contribution has risen from 7% in 2012 to 39% in 2017, and we expect it to surpass 50% by 2020.

The MENA e-commerce market is worth $8.3 billion, with significant room for growth In 2017, the MENA e-commerce market reached $8.3 billion. With an average annual growth rate of 25%, e-commerce in the region has been growing slightly ahead of the global average. The GCC and Egypt account for 80% of the e-commerce market, and they have been growing at a 30% annual rate, more than twice as fast as the rest of MENA.

Globally, e-commerce development differs by country. In China and the most advanced Western markets, such as the UK, US, France and Germany, the e-commerce penetration of total retail sales surpassed 10%, reaching nearly 16% in the case of the UK. By contrast, the e-commerce penetration of total retail sales in some other markets, such as Brazil, Turkey and India, is still below the 5% mark.

In MENA, the e-commerce penetration of total retail sales averaged 1.9% in 2017, with the GCC at 3%. The United Arab Emirates (UAE) is the most advanced e-commerce market in the region, with a penetration rate of 4.2%, similar to that of Turkey and Brazil. The Kingdom of Saudi Arabia (KSA) at 3.8% closely follows the UAE. At 2.5%, Egypt's e-commerce penetration of total retail sales is comparable to that of India and Indonesia.

The MENA e-commerce market is worth $8.3 billion, with significant room for growth In 2017, the MENA e-commerce market reached $8.3 billion. With an average annual growth rate of 25%, e-commerce in the region has been growing slightly ahead of the global average. The GCC and Egypt account for 80% of the e-commerce market, and they have been growing at a 30% annual rate, more than twice as fast as the rest of MENA. Globally, e-commerce development differs by country. In China and the most advanced Western markets, such as the UK, US, France and Germany, the e-commerce penetration of total retail sales surpassed 10%, reaching nearly 16% in the case of the UK. By contrast, the e-commerce penetration of total retail sales in some other markets, such as Brazil, Turkey and India, is still below the 5% mark.

In MENA, the e-commerce penetration of total retail sales averaged 1.9% in 2017, with the GCC at 3%. The United Arab Emirates (UAE) is the most advanced e-commerce market in the region, with a penetration rate of 4.2%, similar to that of Turkey and Brazil. The Kingdom of Saudi Arabia (KSA) at 3.8% closely follows the UAE. At 2.5%, Egypt's e-commerce penetration of total retail sales is comparable to that of India and Indonesia.

The problem

MENA is entering the third phase of digital adoption Digital adoption in MENA has not followed the path typically seen in other markets. As consumers move online, businesses typically follow suit, enabling a gradual development of the digital ecosystem in areas such as media and e-commerce. MENA's experience has been different. Mass Internet adoption in MENA, especially in the Gulf Cooperation Council (GCC), took off around the mid-2000s, mostly driven by smartphones and social media and underpinned by faster Internet speeds. We call this the "digital consumer phase." Businesses were late to join the digital party. We finally started to see digital taking a more central role in business strategy after 2010, with media coming to the forefront. And the shift has been dramatic. Within the span of just five years, the share of digital media increased from less than 10% in 2012 to more than 30% by 2017. We call this the "digital media phase." In March 2017, Amazon announced the acquisition of Souq, the leading MENA e-commerce company, for $580 million. Later in the year, Mohamed Alabbar, chairman of Emaar Properties, launched noon, a new e-commerce venture, with $1 billion of funding

backed by the Public Investment Fund of Saudi Arabia. This marked the beginning of a new digital phase in MENA, the "e-commerce phase."

Online product selection is catching up Limited access to a broad and deep selection of products has been one of the main challenges hindering the development of MENA e-commerce. The region's largest e-commerce platform, Souq, offers an estimated 8.4 million products, far fewer than the 550 million offered by Amazon in the US. While the top four e-commerce categories represent half of the US market, these same shopping categories represent 80% of the market in MENA. A small to medium-sized enterprise (SME) marketplace model has been the backbone of product expansion in most markets. In the US, half of all items purchased on Amazon come from more than a million SMEs.

In China, Alibaba managed to enroll SMEs in its marketplace, offering consumers a wide array of online shopping choices. This has not been the case in MENA. While SMEs constitute around 90% of registered companies in the region, they only make up 15% to 30% of MENA's GDP, which contrasts with the 50% share typically seen in developed markets. They have also been slow to move online. Large groups, especially in core categories such as fashion and electronics, historically have dominated the MENA retail industry. Most of these groups have been late in making the vast product selection offered in their brick-and-mortar stores available online. It is estimated that less than 20% of top global fashion brands physically present in MENA have locally based e-commerce platforms that offer delivery to the regional consumer.

This demand–supply imbalance in the region has created an opportunity for cross-border e-commerce, with non-MENA-based e-commerce players capturing a significant share of the rising consumer interest in online shopping. As of 2017, cross-border e-commerce represented nearly 50% of the market in the GCC vs. less than 13% in the UK and 6% in the US. Amazon, eBay and AliExpress are a case in point. These global e-commerce players have been actively marketing to MENA consumers in the general merchandise category. Popular global delivery services, such as Aramex's Shop & Ship, which offers MENA online shoppers a local delivery address in major cities around the world and facilitates shipping to the region, have supported global e-commerce activity in MENA. The fashion category has been particularly favorable to cross-border e-commerce, with players such as Yoox Net-a-Porter (YNAP), Farfetch and Asos actively targeting MENA consumers and offering fast and inexpensive shipping and return options (two-day delivery and free returns in most cases). China-based players have significantly grown in

the high street fashion category by investing in local marketing strategies and customizing their business models with Arabic websites, local warehousing, fast shipping options and convenient return policies. In KSA, the number of Google Search queries for Chinese e-commerce players such as JollyChic and Shein have seen 60% annual growth over the past three years, making them among the most popular e-commerce destinations in the market.

The MENA e-commerce market is relatively fragmented Souq was the first entrant to the MENA e-commerce market and continues to be the market leader. However, the market remains relatively fragmented, with the top two e-commerce players capturing between 25% and 40% of the market. This is in contrast to most other e-commerce markets, where the share of the top two players is typically more than 50%.

As the e-commerce market develops and competition intensifies, some markets have seen a movement toward consolidation. India is a case in point. In 2005, India was a nascent e-commerce market, with a large number of players, including Flipkart and Myntra, competing for market leadership. The market has since consolidated, with Amazon and Flipkart holding more than 60% of the market .

As of 2017, e-commerce pure players controlled more than 90% of the e-commerce market in MENA, cementing their first-mover advantage. In most emerging markets where retail infrastructure is relatively underdeveloped, e-commerce pure players have managed to capture the lion's share of the market.

The case of China is telling. At the turn of the 21st century and in the space of 10 years, China's middle-class population increased by almost 200 million, creating significant demand for retail products that a historically underdeveloped physical retail ecosystem could not meet. E-commerce came to the rescue. Alibaba's efforts to address core e-commerce ecosystem challenges in areas such as product selection, payment and logistics paid off, paving the way for e-commerce to scale quickly and address the growing consumer demand. Alibaba and other e-commerce pure players managed to capture the majority of the market along the way.

By contrast, traditional retailers in developed retail markets, such as the US and UK, enjoy a strong physical footprint and collectively capture about 40% of market share. While Amazon continues to be the market leader, with almost 50% of the e-commerce

share in the US, retailers that moved early on into e-commerce and leveraged their competitive advantage offline into online shopping have been faring particularly well. US retailer Williams Sonoma launched an online shopping presence as early as the turn of the century, and thanks to its catalog heritage, the company was well positioned to collect and use data from customers and other sources in its e-commerce efforts. Williams Sonoma generates nearly 54% of its sales online today. Despite a sluggish start and a less than 4% share of total e-commerce sales, Walmart has recently stepped up its e-commerce investments with the acquisition of Jet.com and a focus on offering a more convenient and integrated online—offline shopping experience, especially in grocery. Walmart is hoping to increase its current share of the e-commerce market.

THE OBJECTIVES OF THE STUDY

Data mining deals with the kind of data to be mined, there are two categories of functions involved are Descriptive and Classification and Prediction. There are many kinds of data mining goals, let us explain all the goals according to different categories.

The high-level primary goals of data mining are as follow:

- ***The descriptive function*** deals with the general properties of data in the database such as Class Description, Frequent Patterns, Associations, Correlations and Clusters as well.
- ***Classification*** is that process for finding a model that describes the classes and concepts of data. This model used to predict the class of objects whose label is unknown. The derived model is based on the analysis of sets of training data with forms such as Classification rules; decision Trees, neural networks and many more.
- ***Prediction*** is used to predict missing and unavailable numerical data values rather than class labels during data mining process. For prediction regression Analysis is used. Prediction can also be used for the identification of trends based on available data in database. In context of Knowledge Discovery process, the description tends to be more sophisticated than prediction. This is in related to pattern recognition and machine learning tasks where prediction is the primary goal of the Knowledge discovery process.

The goals related to prediction and description is achieved by the following primary data mining goals:

- ***Prediction***: Determine how certain attributes will behave in the future.
- ***Identification***: Identify patterns in data.
- ***Classification***: is function to classify the data items into several predefined classes.
- ***Optimization***: Optimize the use of limited resources such as time, space, money or materials.
- ***Regression***: is that function to map data items into real valued prediction variable.
- ***Clustering***: is a descriptive task to identify a finite set of clusters to describe the data. The probability density estimation consists of techniques for estimating, from data and performs joint multi-variate probability density function of all variables in the database.
- ***Summarization***: is a set of methods for finding a description for a subset of data in database.
- ***Dependency***: Modeling is the method of finding a model which describes significant dependencies between variables, such as structural level which specifies the variables that are locally dependent on each other. Other one is quantitative level which specifies the strengths of the dependencies.

According to styles of data mining, predefined goals are as follows:

Directed data mining: it's a top-down approach, used when user knows what they are looking for want to predict. It's a predictive model used to rank the outcomes of future by estimating score of each outcome. This model emerges like black box about the predictions. The main goal is to build a model to apply past outcomes for future predictions.

Undirected data mining: it's a bottom-up approach that finds patterns in data that used to differentiate whether patterns are important or not that are basically used during data exploration.

The other data mining goals are as follows:

Classification: data mining system performs how to classify the data by using classification rules such as classification performed in customer database used in bank.The question arises that is a new customer applying for loan or not?s
And classification rule applied here is:

if STATUS = married **and** INCOME > **10000**

Then INVESTMENT_TYPE = good

Association: These rules associate one attribute to another follows set oriented methods for discovering proper rules such as supermarket database. It contains 62% of records which contains items A,B and C with 62 is confidence factor.

Sequence: the pattern functions analyze the records and delete frequently occurring patterns such as in retailer database. It can be used to evaluate set of purchases that frequently proceeds.
In data mining various techniques are used for analysis of data, finding patterns and set the regularities in data, identifying underlying rules and features of data. Mining extracts patterns that are not previously identified just to perform mining analogy. In this huge volume of data are explored in an attempt to find patterns , low materials or data are sifted to find new value.

THE STUDY SAMPLES

E-commerce is at different levels of development across core retail categories Not all retail categories in the region are at the same level of e-commerce maturity. While electronics and beauty and personal care enjoy penetration levels closer to those seen in developed e-commerce markets, fashion and grocery are still behind. The room for growth is significant across categories.

Electronics Electronics has pioneered the shift to e-commerce in MENA. At $2.9 billion in the GCC and Egypt, it is the largest e-commerce category, and it has been growing 23% annually over the past few years.

On Google Search, electronics-related queries have grown by nearly 30% over the past three years. Electronics is also the most highly penetrated e-commerce category in the region in terms of total retail sales—16% in 2017, up from 10% in 2014. And it has the highest average basket size: $350 online vs. $371 offline. Nearly 70% of online shoppers in the GCC and Egypt have bought electronics online at least once, mainly using a smartphone (56% vs. 31% using a desktop). While mobile phones make up the largest segment within electronics, thanks to an active upgrade culture, the home appliances segment is still lagging behind with challenges related to last-mile delivery. MENA-based e-commerce pure players largely dominate the market. Souq has been the first mover and regional market leader, while Wadi in KSA and Jumia in Egypt are also strong players. Noon is gaining market share as a strong new entrant, and Awok, meanwhile, competes in the value segment. Most brick-and-mortar electronics retailers have launched e-commerce channels with varying levels of success to date.

Fashion Fashion e-commerce is valued at $1.1 billion in the GCC and Egypt, and it has been growing 20% annually over the past three years. The e-commerce share of total retail sales is at 2%, which is still relatively low compared with other developed e-commerce markets, such as the UK (25%) and China

(31%), primarily due to limited product selection. Nearly 50% of online shoppers in the GCC and Egypt have bought fashion items online at least once. At $107, the average basket size online is higher than the $95 in-store cart. With almost $900 million, the high street fashion e-commerce market represented nearly 2% of total retail sales in this category in 2017. Shoes and women's clothing are the biggest subcategories. Competition has been intensifying among regional and international e-commerce players. Namshi has been an early mover in shoes and clothing; the company is now facing significant competition from other regional e-commerce players, such as Souq and noon, especially in KSA. China-based e-commerce players, led by JollyChic and Shein, have built a strong market position over the past few years.

Omnichannel retailers have also increased their presence, with Landmark Group, Chalhoub Group, Al Tayer Group and Apparel Group launching new e-commerce channels and making a dent in the market. With an estimated $200 million market, luxury fashion

e-commerce has been developing at a faster pace than high street fashion. It represents 7% of total luxury retail sales. Traditionally dominated by global luxury e-commerce pure players such as Net-A-Porter and Farfetch, regional omnichannel retailers have been catching up, with Ounass and Level Shoes gaining significant traction. The Chalhoub-Farfetch and Alabbar-YNAP partnerships should fuel market growth and provide more product options for MENA shoppers.

The future looks promising. Can MENA deliver?

Despite being late to take off, the MENA e-commerce market has gained momentum over the past couple of years, and it is set to experience significant growth should elements of the ecosystem continue to fall into place. The region is still in the beginning of the cycle of e-commerce adoption, and the future can take on many forms. The growth trajectory of e-commerce has varied depending on the market. In well-established retail markets, such as the US, the growth of e-commerce has been linear, and it took more than two decades for e-commerce to reach a 12% share of total retail sales. In other markets, such as China, where the physical retail ecosystem has been historically underdeveloped, e-commerce growth has been significantly faster, with the e-commerce share of total retail sales growing to 14% in just eight years.

Besides consumer readiness and the development of the different elements of the ecosystem, the entry of large e-commerce players could help accelerate the shift to e-commerce. India is a case in point. Prior to Amazon's launch, the e-commerce market in India grew by 38% annually between 2009 and 2012. Following Amazon's entry in 2013, India's e-commerce market experienced a significant uplift. Growth in e-commerce more than doubled, averaging 88% annually between 2013 and 2016, while e-commerce penetration grew from 0.4% in 2013 to 1.8% in 2016. The launch of Amazon Prime in 2016 served as another catalyst, with e-commerce penetration growing to 2.5% the following year.

In 2019 bain and company predicted that :

We expect electronics e-commerce to grow to $5.9 billion in the GCC and Egypt by 2022, with a share of total category sales of 25%. With markets such as the UK and US recording more than 40% penetration in this category, the room for growth in the region is significant. An expansion of product selection, beyond mobile phones and into home

appliances, would be key. Omnichannel retailers taking a more active role, especially hypermarkets with well-established physical distribution networks, can help solve some of the logistical issues associated with this category.We forecast that fashion e-commerce will grow by 40% annually and reach $5.1 billion in the GCC and Egypt by 2022, a penetration rate of 10% of total category sales. A broader product selection, faster deliveries and streamlined return policies can help unlock the potential in this category.

In the GCC and Egypt, we expect beauty e-commerce to reach $1.8 billion by 2022, with a share of total category sales of 14%. Growth catalysts include increased customer confidence in product authenticity, global players entering the space and geographical expansion to new cities, especially in KSA.

We see the grocery e-commerce category growing by almost 90% per year and reaching $4.1 billion in the GCC and Egypt by 2022, with a share of total category sales of 4%. A more aggressive expansion strategy by hypermarkets, further transition from the region's drive-through and phone delivery culture to online marketplaces, and a broader geographic expansion to new cities would help facilitate market growth.

In conclusion There has never been a better time to be in e-commerce in MENA. The market is at $8.3 billion today and estimated to reach $28.5 billion by 2022. The region's digitally savvy consumers are hungry for a broader online product selection and new shopping experiences. E-commerce is at the core of retailers' strategies, and e-commerce pure players are expanding into new markets and product categories. The different elements of the e-commerce ecosystem—namely, payments and logistics—have come a long way despite the many remaining challenges.

Building on the strong momentum of the past couple of years, e-commerce enters a pivotal time in the region. The opportunity is significant for consumers, businesses, investors and ecosystem players. Exponential or linear, the pace of growth will depend on how fast the three Ps—product selection, payments and product delivery—come together. When it comes to consumer Internet adoption and digital media investments, exponential growth has been a common theme in the region. Will history repeat itself?

LIMITATIONS

Limitations or Disadvantages Of Data Mining Techniques

Data mining technology is something that helps one person in their decision making and that decision making is a process wherein which all the factors of mining is involved precisely.

And while the involvement of these mining systems, one can come across several disadvantages of data mining and they are as follows :

- violates user privacy:

It is a known fact that data mining collects information about people using some market-based techniques and information technology. And these data mining process involves several numbers of factors.

But while involving those factors, data mining system violates the privacy of its user and that is why it lacks in the matters of safety and security of its users.

- Additional irrelevant information:

The main functions of the data mining systems create a relevant space for beneficial information.

But the main problem with these information collections is that there is a possibility that the collection of information processes can be a little overwhelming for all.

Therefore, it is very much essential to maintain a minimum level of limit for all the data mining techniques.

- Misuse of information:

As it has been explained earlier that in the data mining system the possibility of safety and security measure are really minimal. And that is why some can misuse this information to harm others in their own way.

Therefore, the data mining system needs to change its course of working so that it can reduce the ratio of misuse of information through the mining process.

- Accuracy of data:

Most of the time while collecting information about certain elements one used to seek help from their clients, but nowadays everything has changed. And now the process of information collection made things easy with the mining technology and their methods.

One of the most possible limitations of this data mining system is that it can provide accuracy of data with its own limits.

Conclusion:

Finally, the bottom line is that all the techniques, methods and data mining systems help in the discovery of new creative things. And at the end of this discussion about the data mining methodology, one can clearly understand the feature, elements, purpose, characteristics, and benefits with its own limitations.

Therefore, after reading all the above-mentioned information about the data mining techniques, one can determine its credibility and feasibility even better