Chapter 4 : Summary

- Study summary
- References

THE SUMMARY

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%.

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.

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 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.

fuzzy matching?

Rather than flagging records as a 'match' or 'non-match', fuzzy matching identifies the likelihood that two records are a true match based on whether they agree or disagree on the various identifiers.
The identifiers or parameters you choose here and the weight you assign forms the basis of fuzzy matching. If the parameters are too broad, you will find more matches, true, but you will also invariably increase the chances of 'false positives. These are pairs that are identified by your algorithm or fuzzy matching software of choice as a match, but upon manual review, you will find that your approach identified a false positive.
Consider the strings "Kent" and "10th". While there is clearly no match here, popular fuzzy matching algorithms still rate these two strings nearly 50% similar, based as character count and phonetic match. Check for yourself.

Damerau-Levenshtein distance

It is an extension to Levenshtein Distance, allowing one extra operation: Transposition of two adjacent characters:

Ex: TSAR to STAR

Web scraping

is the process of gathering information from the Internet. Even copy-pasting the lyrics of your favorite song is a form of web scraping! However, the words "web scraping" usually refer to a process that involves automation. Some websites don't like it when automatic scrapers gather their data, while others don't mind.

If you're scraping a page respectfully for educational purposes, then you're unlikely to have any problems. Still, it's a good idea to do some research on your own and make sure that you're not violating any Terms of Service before you start a large-scale project. To learn more about the legal aspects of web scraping, check out Legal Perspectives on Scraping Data From The Modern Web

Some website providers offer **Application Programming Interfaces (APIs)** that allow you to access their data in a predefined manner. With APIs, you can avoid parsing HTML and instead access the data directly using formats like JSON and XML. HTML is primarily a way to visually present content to users.When you use an API, the process is generally more stable than gathering the data through web scraping. That's because APIs are made to be consumed by programs, rather than by human eyes. If the design of a website changes, then it doesn't mean that the structure of the API has changed.to APIs just as they do to websites. Additionally, it's much harder to inspect the structure of an API by yourself if the provided documentation is lacking in quality.The approach and tools you need to gather information using APIs are outside the scope of this tutorial. To learn more about it, check out API Integration in Python.

Why do you need a framework?

To understand what Django is actually for, we need to take a closer look at the servers. The first thing is that the server needs to know that you want it to serve you a web page.
Imagine a mailbox (port) which is monitored for incoming letters (requests). This is done by a web server. The web server reads the letter and then sends a response with a webpage. But when you want to send something, you need to have some content. And Django is something that helps you create the content.

What does Django code look like?

In a traditional data-driven website, a web application waits for HTTP requests from the web browser or other client. When a request is received the application works out what is needed based on the URL and possibly information in POST data or GET data. Depending on what is required it may then read or write information from a database or perform other tasks required to satisfy the request. The application will then return a response to the web browser, often dynamically creating an HTML page for the browser to display by inserting the retrieved data into placeholders in an HTML template.

REFERENCES

- https://tutorial.djangogirls.org/en/django/
- https://djangostars.com/blog/why-we-use-django-framework/
- https://www.ibm.com/cloud/learn/django-explained
- https://developer.mozilla.org/en-US/docs/Learn/Server-side/Django/Introduction
- https://realpython.com/beautiful-soup-web-scraper-python/
- https://www.tutorialspoint.com/requests/requests_quick_guide.htm
- https://realpython.com/python3-object-oriented-programming/
- https://www.datasciencecentral.com/profiles/blogs/fuzzy-matching-algorithms-to-help-data-scientists-match-similar
- https://blog.couchbase.com/fuzzy-matching/
- https://towardsdatascience.com/natural-language-processing-for-fuzzy-string-matching-with-python-6632b7824c49
- https://dataladder.com/fuzzy-matching-101/
- https://www.bain.com/contentassets/2b078686303045ffa1d1207130ab5d79/bain_report__ecommerce_in_mena.pdf