

# *The Importance Of Data Mining In Retail Industry*

Muminov Ibrokhim Botir Ugli

Independent Researcher  
Asan City, Republic of Korea



**Abstract**— Thanks to the continuous advancements in data mining, which is an interdisciplinary subfield of computer science and statistics, it has become significantly easier for the retail industry to improve customer service, acquire and retain new consumers by developing products in demand, and compete successfully in the marketplace. As a reason, data mining allows retail companies to analyze product sales trends, customer buying habits as well as preferences, delivery performance, customer peak traffic, and supplier lead times. Consequently, retail companies are now able to control the costs required for the development of existing products, innovate new targeted brands, and dispose of the products that are no longer in use by their customers. This article will provide background information related to data mining, how it became an inextricable part of retail industry, and the advantages as well as disadvantages of exploiting this powerful tool by retail companies.

**Keywords**— data mining, retail industry/companies, clustering, classification, regression, anomaly detection, summarization, data analysis, customer retention, targeting, Market Basket Analysis (MBA), social media optimization, financial fraud identification, price optimization, disadvantages of data mining.

## I. WHAT IS DATA MINING?

### 1.1. History of Data Mining

According to the best estimates of Cloud Tweeks, which is one of the leading authorities in cloud computing information and resources on the planet, approximately 2.5 quintillion bytes of data is produced every single day. It can be said that such a mountainous amount of data does not have any uses until it is processed and analyzed. The field that helps human beings with this intricate task is called “data mining”. Data mining is an interdisciplinary subfield of computer science and statistics that shares responsibility for discovering and extracting similarities in large datasets using methods at the intersection of machine learning, statistics and probability and turn information into understandable structure for further use. Data mining is not something new in today’s fast-paced world, since it has been evolving for centuries. In 1763, two years after death of an English statistician and philosopher Thomas Bayes, his paper was published posthumously about a theorem for relating current probability to prior probability called the Bayes’ theorem, which is fundamental to data mining and probability. Data mining became widely used from 1990, when computers were already able to handle large amounts of data. It also has to be mentioned that before 1990s, terms like “database mining” and “Knowledge Discovery in Databases” (KDD) coined by a data scientist Gregory Piatetsky-Shapiro were used, however, in the same year they were replaced by a new term - data mining, which first appeared in the database community.

### 1.2. Tasks of Data Mining

Generally, data mining involves six common tasks:

- a) **Clustering** – the process of uncovering similarities between groups and structures in the data.

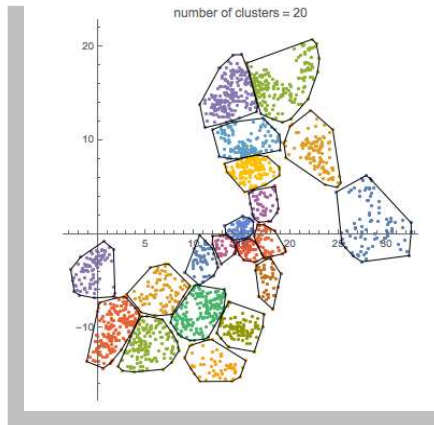


Figure 1: Cluster analysis

Source: WOLFRAM Demonstrations Project

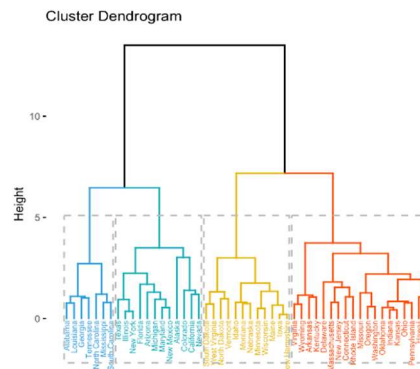


Figure 2: Region Growing and Clustering Segmentation

Source: GitHub

- b) **Association rule learning (dependency modeling)** – searches for relationships of variables. For instance, in order to target more potential customers, a retail company might gather and analyze data related to one or more specific products that are sold better than the other types of products.

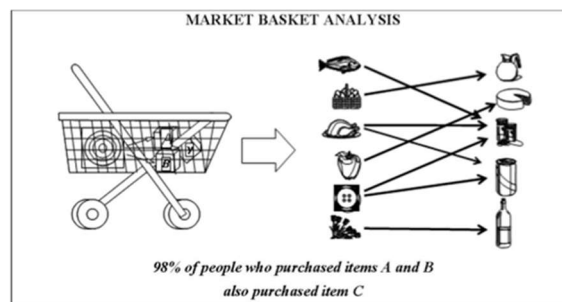


Figure 3: Market Basket Analysis

Source: MATLAB

- c) **Classification** - is the task of generalizing known structure to apply to new data.

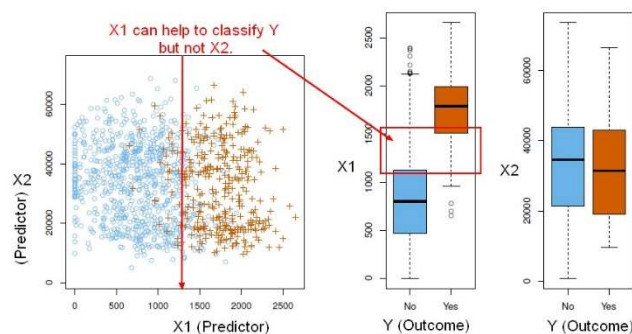


Figure 4: Classifier/Classification Function

Source: Datacadamia

- d) **Regression** - attempts to identify the best function for modeling data with the least amount of inaccuracy, that is, for analyzing the relationship between data or datasets.

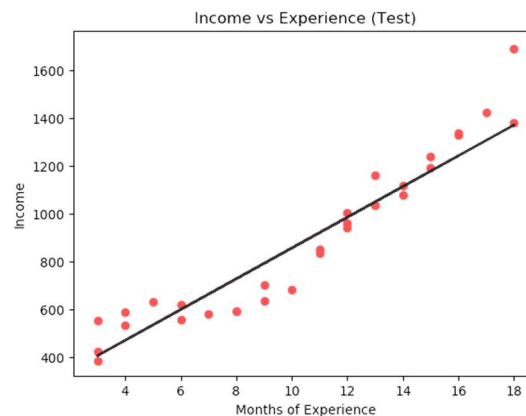


Figure 5: Simple Linear Regression

Source: Elle Knows Machines

- e) **Anomaly detection** – the process of identifying anomalous data records that may be of interest or data mistakes that need to be investigated further.

| timestamp  | value   | anomaly | change point | trend | noise  | 12 hour seasonality | daily seasonality | weekly seasonality |
|------------|---------|---------|--------------|-------|--------|---------------------|-------------------|--------------------|
| 1422237600 | 4333.43 | 0       | 0            | 4599  | 1.81   | -190.95             | -128.86           | 52.44              |
| 1422241200 | 4316.14 | 0       | 0            | 4602  | -14.65 | -220.5              | -105.21           | 54.51              |
| 1422244800 | 4403.20 | 0       | 0            | 4605  | 7.04   | -190.95             | -74.39            | 56.51              |
| 1422248400 | 4531.20 | 0       | 0            | 4608  | 13.52  | -110.25             | -38.51            | 58.43              |
| 1422252000 | 4967.50 | 1       | 0            | 4911  | -3.77  | -6.91               | -2.33             | 60.27              |

Figure 6: Announcing A Benchmark Dataset for Time Series Anomaly Detection

Source: Yahoo!

- f) **Summarization** – presenting the data collection in a more compact form, including visualization and report production.

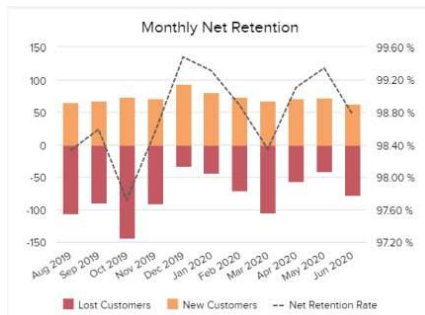
| Invoice Date | Customer | Type         | Country | Amount | Status |
|--------------|----------|--------------|---------|--------|--------|
| 04-05-2018   | Dominoz  | Manufacturer | India   | 54000  | Paid   |
| 08-05-2018   | McDonald | Retailer     | USA     | 14560  | Due    |
| 15-06-2018   | KFC      | Distributor  | Canada  | 13500  | Future |
| 07-05-2018   | Subway   | Manufacturer | USA     | 69500  | Due    |
| 06-02-2018   | KFC      | Distributor  | India   | 12500  | Paid   |
| 07-02-2018   | Dominoz  | Retailer     | Canada  | 15200  | Future |
| 28-02-2018   | KFC      | Distributor  | Canada  | 56000  | Due    |
| 24-06-2018   | Subway   | Manufacturer | USA     | 25200  | Paid   |
| 25-06-2018   | Dominoz  | Distributor  | India   | 32000  | Future |
| 26-06-2018   | KFC      | Retailer     | Canada  | 17100  | Paid   |
| 27-06-2018   | Subway   | Manufacturer | India   | 26800  | Paid   |
| 29-12-2018   | KFC      | Distributor  | Canada  | 12650  | Due    |
| 30-11-2018   | Dominoz  | Manufacturer | USA     | 12200  | Paid   |
| 19-08-2018   | Subway   | Distributor  | India   | 23003  | Paid   |
| 20-08-2018   | Subway   | Retailer     | Canada  | 12400  | Future |
| 14-07-2018   | Dominoz  | Manufacturer | USA     | 51000  | Due    |

Figure 7: Example of summarized data using data mining

Source: Loginworks Softwares

## II. WHY SHOULD RETAIL COMPANIES FOCUS ON DATA MINING?

**2.1. By understanding the behavior of customers, data mining helps retail companies to achieve more potential customers as well as retain them for a long period of time, control the process of manufacturing new products, and stay competitive in the market.**



**Figure 8: Data mining to analyze monthly net retention churn rate**

Source: datapine



**Figure 9: Data mining to analyze monthly customer churn rate**

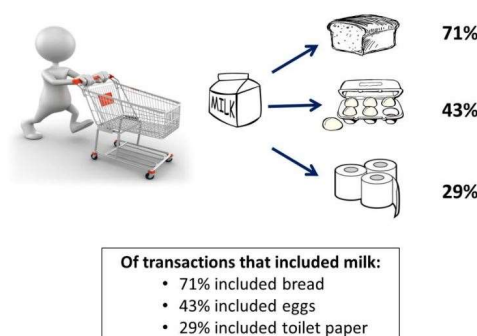
Source: datapine

Analyzing which specific products are frequently chosen by customers makes it easier to create new strategies for targeting new customers and retain previous ones, to bring a significant amount of revenues that can be redirected for the development of well-selling products and, of course, to avoid producing goods that are vulnerable at competition and no longer profitable.

One of the world's leading major home appliance companies Whirlpool Corporation analyzes a huge amount of data collected from 40 different websites to answer vital for every company questions like which features of a product are liked the most by consumers, the competitiveness of the product in the market, customer satisfaction, and what should be improved about a specific product or service.

### 2.2. Marketing methods can be substantially enhanced by using data mining.

One marketing method that is used by many retailers is "Market Basket Analysis", which discovers a particular location where products can be effectively promoted. Using information collected from this method improves promotion effectiveness, the process of determining which products are likely to be sold to customers, which results in an increase in store sales. Most importantly, having exploited data mining for collecting and analyzing information related to Market Basket Analysis, detecting loyal customers becomes significantly faster.



**Figure 10: Market Basket Analysis using data mining**

Source: Fiverr

American Express, a multinational financial services corporation, exploits data mining to analyze and predict consumer behavior. By looking at earlier transactions and integrating more than 100 variables, the company uses complex predictive models

instead of traditional business intelligence-based hindsight reporting. This makes it possible to predict client turnover and loyalty with greater accuracy.

Capital One, an American bank holding company specializing in credit cards, auto loans, banking, and savings accounts, uses data mining for analysis of the demographics and spending habits of customers. Capital One evaluates the best times to give different offers to clients in order to boost conversion rates from their communications.

### 2.3. Data mining is a powerful tool for the optimization of social media websites.

It identifies important social media trends and signals, optimizes brand performance on social media, and helps to analyze customers' satisfaction about a specific product by analyzing positive and negative comments.

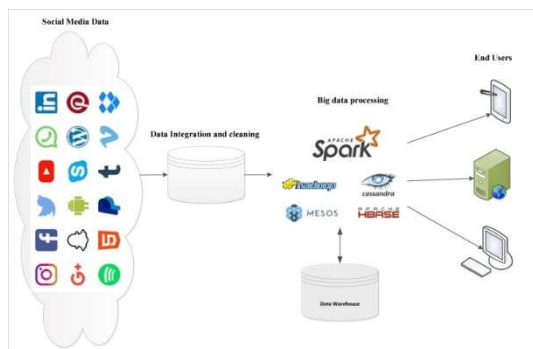


Figure 11: Example of how social media data is processed

Source: N/A

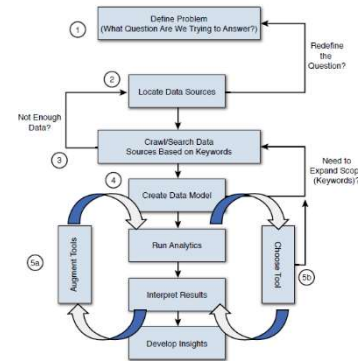


Figure 12: Using social media mining tools

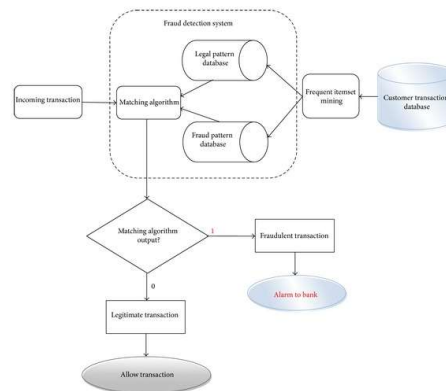
Source: The Digital Media Strategy Blog

For instance, according to Social Media Today, McDonald's Corporation has approximately 70 million followers across all countries, regions, and platforms. Thousands of comments are left on the official social media pages of McDonald's Corporation daily, so the company uses data mining for uncovering valuable insights of customers and to make important decisions about new breakfast menus. Furthermore, thanks to the control over its social media McDonald's can react fast if there is an issue related to the service at a specific location, for example.

### 2.4. Data mining is found to be extremely useful in increasing financial fraud identification.

Financial fraud identification is critical for retail companies, since hundreds of thousands of dollars are stolen by dishonest employees annually. It has to be mentioned that most of these fraud cases happen at Point-of-Sale. To prevent an employee from stealing or to convince for stealing, supermarkets and big shopping malls install CCTV cameras in addition to Point-of-Sales data mining.

Data mining for financial fraud detection relies on pattern analysis to uncover unusual activities and trends. One of the best sources of information for data mining can be big data that includes customer zip codes, income levels, age, and other demographic factors that impacts consumer's financial purchases and decisions.



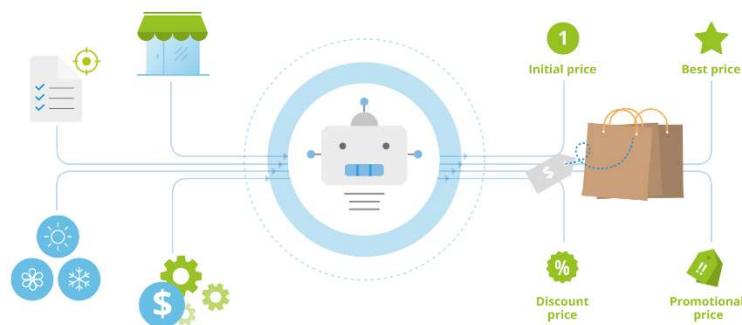
**Figure 13: Proposed credit card fraud detection model**

**Source: ResearchGate**

Even little variations in a consumer's transactions or credit activity can be evaluated and highlighted as possible fraud. As soon as such fraud cases are detected, the suspected transaction history is retrieved from the data mining software as an evidence.

### 2.5. Data mining is extremely useful while pricing products.

By using data mining, companies can anticipate trends before they happen, create a new product, and optimize its price by continually monitoring relevant search words. Furthermore, to define which pricing fits best it is crucial to collect and analyze data about customers from different geographical areas, since knowing what customers expect from the company can increase the chances of winning competition in the marketplace.



**Figure 14: Price optimization using big data and data mining**

**Source: N/A**

Furthermore, it is irrefutable that when prices are controlled correctly, it will increase customers' loyalty to the product and bring more profits to the company, and retailers will always understand why they are successful and can repeat similar method to achieve more customers and increase sales.

Finally, thanks to the advancements in data mining, retail companies now can make pricing decisions faster and change prices for particular products more often.

## III. DISADVANTAGES OF USING DATA MINING

### 3.1. Cost related difficulties and technical skills

Due to the fact that implementation and maintenance of data mining technologies is substantially expensive, not every company can financially afford exploiting this powerful tool. For the better data mining, a mountainous amount of data has to be stored, so the companies who deal with this technology must have their own storage space to save information. Maintaining huge databases is an expensive and intricate task.

Besides, there are different data mining tools with various algorithms and designs, so a skilled software engineer is required to choose the right tool for a specific task. Consequently, one or more specialists who will share responsibility for controlling data mining technologies have to be hired. That is why, only companies like Amazon, Facebook, Microsoft, Netflix, Capital One can afford using this cutting-edge technology nowadays.

### 3.2. Security concerns and information misuse

It is a well-known fact that a significant amount of personal information is collected in data mining. Subsequently, companies that use data mining have to have strong security, since there are a lot of hackers who want to steal vital information from the databases and use it for their own purposes, which can be dangerous.

Not only hackers, but also employees who have access to personal information could use it for personal advantage, or a group of people could be targeted and injured as a result of information misuse.

Therefore, the companies that use data mining have to ensure its security and share responsibility for using collected data for the right purposes.

### 3.3. Accuracy and additional information

Although data mining technologies has eased the burden of collecting and analyzing data from people's shoulders, it might, sometimes, provide inaccurate information, leading companies to wrong conclusions.

Furthermore, as it was mentioned earlier, a huge amount of information has to be collected for data mining, which can be sometimes extraneous or irrelevant, making some people inconvenient to work with this technology. As a reason, working with big datasets is complex and a software engineer who works with data mining has to put massive effort and is responsible for the outcome, which can be both positive and negative.

## IV. CONCLUSION

Having taken into consideration above facts, it may be concluded that data mining has more positives than its negatives. Rightly used in retail companies, and not only, this powerful tool can be extremely beneficial when it comes to collecting and analyzing data. As we live in the century of big data and continuous technological advancements, the value of having proper information will rise significantly, resulting in a substantial use of data mining in the upcoming future. The only thing that has to be done by the companies is to maintain the harmony in using data mining technology and ensure its safety to avoid negative consequences.

## REFERENCES

- [1] Han, Jaiwei; Kamber, Micheline; Pei, Jian (2011). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann. ISBN 978-0-12-381479-1.
- [2] Marketing91 2019, *Data Mining In Retail: Applications and Six Phases in the Life Cycle*, viewed 8 August 2021, <<https://www.marketing91.com/data-mining-in-retail/>>.
- [3] CoolAvenues.com 2013, *Role of Data Mining in retail sector*, viewed 8 August 2021, <<https://www.coolavenues.com/marketing-zone/role-of-data-mining-in-retail-sector>>.
- [4] Intellspot n.d., *7 Real-World Examples Of Data Mining In Business, Marketing, Retail*, viewed 9 August 2021, <<https://www.intellspot.com/data-mining-examples/>>.
- [5] Fayyad, Usama; Piatetsky-Shapiro, Gregory; Smyth, Padhraic (1996). "From Data Mining to Knowledge Discovery in Databases" (PDF). Retrieved 10 August 2021.
- [6] HitechWhizz n.d., *7 Advantages and Disadvantages of Data Mining | Limitations & Benefits of Data Mining*, viewed 10 August 2021, <<https://www.hitechwhizz.com/2021/04/7-advantages-and-disadvantages-limitations-benefits-of-data-mining.html>>.
- [7] KDnuggets 2016, *History of Data Mining*, viewed 11 August 2021, <<https://www.kdnuggets.com/2016/06/rayli-history-data-mining.html>>.
- [8] ICAS 2016, *10 companies that are using big data*, viewed 12 August 2021, <<https://www.icas.com/thought-leadership/technology/10-companies-using-big-data>>.

- [9] Syntelli Solutions 2020, *Using Big Data for Financial Fraud Prevention*, viewed 12 August 2021, <<https://www.syntelli.com/using-big-data-financial-fraud-prevention>>.