

RESEARCH ARTICLE



Received: 18.04.2020

Accepted: 20.11.2020

Published: 19.02.2021

Citation: Abbas I, Ahmad M, Imran M, Khan MN, Niazi ZZK, Raza SMU, Naveed A, Ali Z (2021) Role of data mining techniques in business. Indian Journal of Science and Technology 14(5): 508-518. <http://doi.org/10.17485/IJST/v14i5.315>

* **Corresponding author.**

abbasirfan440@gmail.com

Funding: None

Competing Interests: None

Copyright: © 2021 Abbas et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.indst.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Role of data mining techniques in business

Irfan Abbas^{1,2*}, Munib Ahmad², Muhammad Imran²,
Muhammad Nadeem Khan², Zohaib Zafar Khan Niazi²,
Syed Muhammad Usman Raza², Aamir Naveed², Zulfiqar Ali²

¹ Lecturer, Department of Computer Science and IT, University of Central Punjab, Natt Road, 50700, Gujrat, Pakistan

² Department of Computer Science and IT, Minhaj University Lahore, Punjab, Pakistan

Abstract

Objective: The main objective is to elaborate and discuss different techniques used in data mining, to analyze different strategies of data mining to make improvements, and to find more powerful mining techniques for the betterment of the business. **Methods:** Multiple techniques and strategies of data mining are used to improve the business. We employed the data warehouse methods for the improvements of the business using Business Intelligence (BI) and Business Analytics (BA) along with their types and instruments. We are also discussing some tools used for data mining or ordering organizational data. **Findings:** We employed Business Intelligence-(BI) and Business Analytics-(BA) techniques for the improvement of the business. Earlier, there were only four (Regression, Classification, Association, and Clustering) techniques that were used for business improvements. It is found that Crawler is the best tool for BI or BA data mining. **Novelty :** This study analyzed that, BI and BA are the best ways used for data mining, data ordering, or format of data in business. Earlier, these ways were not in use for data mining. Data mining may be the best approach to improve the business.

Keywords: Business Intelligence (BI); Business Analytics Data Mining (BADM); Data Warehouse (DW); Knowledge Discovery (KDD) Customer Relationship Management (CRM); Enterprise Resource Planning (ERP)

1 Introduction

Data management is an important component for every organization and individual. In the business, data and conversion with the client are very important. Clients discuss their ideas with organizational officials for the best-featured product. Institutes use data mining techniques for the formatting of the data and the business Data ordering is very important for the betterment of business performance and collecting information about the process. It is also used for scoring unauthorized information and data. In the business community, Regression, Classification, Association, and Clustering techniques are generally used for data ordering. Their merits and limitations are elaborated below:

1.1 Regression Data Mining-(RDM)

Regression is used for the Prediction Range-(PR) of required data. In Data Science-(DS), regression uses continuous values which also known as numeric values from the dataset and predicts results. In business and marketing planning strategies, regression is used for the prediction of expected output and expected profit. It is also used for Financial Forecasting (FF) and Environmental Modeling (EM). There are some regression types and Linear Regression (LR) is one of the oldest types. The LR is defined as a relationship between two different variables X and Y or M and N. The other one is Multiple Regression (MR) used for prediction from more than two variables. In data mining, regression is used for the prediction. The predicted data provides help for future planning and cost optimization but it can take a lot of time for the optimization.

1.2 Classification Data Mining (CDM)

The Classification Data Mining (CDM) technique is also used for the prediction of data. The CDM uses discrete values for the expected results. Classification is employed to classify each item during a bunch of data into one predefined class or set of classes or groups. The task analysis information classification is where a model or classifier is made to predict the class label attributes. It may be a data processing function that assigns items during a collection to focus on categories. The objective is to accurately predict the target class for every case within the data. For instance, a classification model may be used to identify loan applicants as low, medium, or high credit risks. A classification task begins with a knowledge set during the class. But, it is very difficult as a class generation a time-consuming process.

1.3 Association Data Mining (ADM)

An Association Data mining (ADM) is a rule-based procedure that has the aim to detect regularly occurring patterns or relations from a database. It is also used to find the kinds of relationships of data from a dataset. The different association rules are used in data mining for the improvement of the business like antecedent and consequent. In other words, two different rules of associations are 'if' and 'then'. Both associations use for business improvements. The 'if' association says to find something on the base of the condition and the 'then' association says that output in the form of a false condition. All the association rules build up after a deep analysis of data. It is a time-consuming process and required resources for data analysis.

1.4 Clustering Data Mining-(CLDM)

Clustering is a technique that comes from the CDM. In this technique, we use group data and classes of group data. All the data classified according to the different labels or categories. Organizations use Cluster Analysis-(CA) from the dataset which is based on the similarity report of the data.

2 Suggested Techniques

It assumes an important job in all sorts of businesses and ventures that data mining is extremely helpful. Since it is important to gather the information that might be in crude certainties, graphs, and figures, at that point all the procedures are unadulterated information and data. The data assists in basic leadership to improve our business quality and reliability. Nowadays, it is difficult to gather real information about business and business strategies. Everything should be possible through manual working and finding algorithms for better performance, but it becomes costlier. In all cases, the data mining devices, techniques, methods, and strategies are difficult to get genuine/ actual data issues, arrangements, control, and true picture of your business. We extricate information from different sources and information distribution centers.⁽¹⁻³⁾

Business Intelligence-(BI) and Business Analytics-(BA) are the suggested techniques use for data mining in business for better product features and product development. These techniques used to improve the business through CRM, ERP, OLAP, business checks, and reports for information mining. The information is extracted from inside and outside sources which may be from information bases, distribution centers, or information mining. We test all the data by the operational framework and the ETL handling framework for data mining. The detailed description of these techniques areas

2.1 Business Intelligence (BI)

Business Intelligence (BI) is a technique that we use for better decision-making, advanced analysis, business query, business competition, CRM, and Reports making. Better decision-making techniques are used for better results for our business and product. The advanced analysis method is used to get accurate results and accurate information after analyzing the results or reports for our organization's improvements. Same as business query, business competition, dashboard, and ERP reports are

used for the better and improvements of our quality business.^(4,5) We take a dataset from any source and apply a BI approach to it. The BI gets data from data centers and clouds. The data is provided to one of the types of BI like CRM or Advanced Analysis-(AA). These types perform some operations on the data according to the requirements of the data. The Customer Relationship Management-(CRM) Dashboard gets data from BI and performs advanced analysis on data using a different tool of BI. We are using Crawler or I Crawler tools to perform BI and generate Enterprise Resource Planning-(ERP) reports. The ERP reports defined the user reports or product developer plaining reports, results, and cost of the product. The CRM dashboard must obtain the Business Query-(BQ) as an input from the product developer. The BQ is a problem that businessman wants to solve against the product. The BQ defined the Business Competition-(BC) in the market and develop extra features in the product against the competitors. After that, Better Decision Making-(BDM) approach defines the decision (results) of data. It is a time and cost-saving approach to analyze the whole data and generate results for the business improvements.



Fig 1. Business Intelligence

2.2 Business Analytics (BA)

Business Analytics is a method for investigating and converting information into a significant and usable structure or format. First of all, we gather the unstructured information from the clients. After that, we process it and get filtered and important data called information that helps for business improvement. We are using the “Breakdown” technique for preparing or dissecting the information. By dividing the important information into specific parts, we get accurate data for the different tasks. It covers all the issues, procedures, and testing for the business improvements. The Data enhancement, Pattern, Statistical model, Relationship, and questions are found by BA and use these data for further process. It helps us to focus on the results and remain

with the contestant by delivering expectations about the future. Business Analysis produces better outcomes, significant data, and control on business activities. BA depends on relevant information which sees the present and past circumstance, results, administrations, client conduct, client choice, and Sales.⁽⁶⁾ BA is a mix of innovation, applications, instruments, programming projects, calculations, and processes utilized by organizations to accomplish objectives, goals, and business positioning. It may be utilized in any office from deals, generation, import, export, and customer service also. After examining, it will be as certainty-based information which helps to measure past execution to guide and association business arranging. Such a key focus is important to settle on better business, customer satisfaction, and basic leadership.⁽⁷⁾ (Figure 2)



Fig 2. Business Analytics using Data Mining

2.3 Tools and Techniques of Business Intelligence/ Business Analytics

Different BI applications, instruments, and strategies are utilized in BI to process or concentrate information into useful data. Data Visualization, Reporting (Ad-hoc and constant), Predictive Modeling (Foresee future results and Patterns), Descriptive Data Mining (Describe occasions that previously occurred), Pattern Analysis, Statistical Analysis-(SA), Sales Intelligence-(SI), Decision Trees-(DT), Clusters, SAP business target undertaking and SAP Net Weaver Business Intelligence-(BI), Microsoft Business Intelligence stage (MSBIS), Online Analytical Processing (OLAP), Web Focus, Quick View-(QV), Micro methodology and Board Management Intelligent-(BMI) Toolkit are the tools that we used for the business intelligence and business analytics.

2.4 Business Intelligence Software

We use BI software to perform different operations on data, i.e. extraction of information from the sample data. Business intelligence programming utilizes numerous diagnostic elements like measurements, data mining, content mining, and prescient information to assess and provide results. There are numerous products for examining the information. Some are open-source like cloud facilitated and restrictive based and some are the⁽⁸⁾ Looker, Microsoft Power BI, Databox, Sisense, Logi Analytics, Exago BI, Grow BI Dashboard Software, Yellowfin BI, SAP Analytical Cloud, Corporates, J Report, Cloudera Enterprise, BIME Analytics, and KNIME are the suggested business intelligence software's that we used for business improvements in different organizations. We are using all these software's one by one, and we check that every software produces the same results in the same range (Time Period).

3 Data Warehouse

Data Warehouse is the platform where sample and use data stored by the organizations. It is also known as Information Distribution Centers-(IDC). Another name of IDC is Endeavor Information Stockroom/Enterprise Data Warehouse (EDW) and It is a significant part of the BI. The information distribution center is an arranged, incorporated, non-volatile, and time variation assortment of information for the support of the business. The DW's are the databases that stored data. The database has different types. Some are the operational database for the information distribution center and the information stockroom. It is also known as a vital database.⁽⁹⁾ In which, we can include, erase, addition, evacuate, or update information. For instance, we have 50 product branches and every branch have 50 databases. In every database for each branch there will be an item that we buy, deal, erase, alter, or update. However, then it is difficult to maintain and get data, final result, and time squandering method. So, the data Warehouse house takes care of this issue where information is stored with the vital database. In a vital database, all databases will be put away in the information distribution center. Where no information will be erased or updated. It is a useful and efficient technique to accomplish data. This technique is useful for settling on choices to make business more beneficial. Information might be gathered from this database, ERP, Websites, CRMs, Excel, or from different applications. It is useful for removing, recovering, and putting away information. It is a quicker and more consistent technique.⁽¹⁰⁾ (Figure 3)

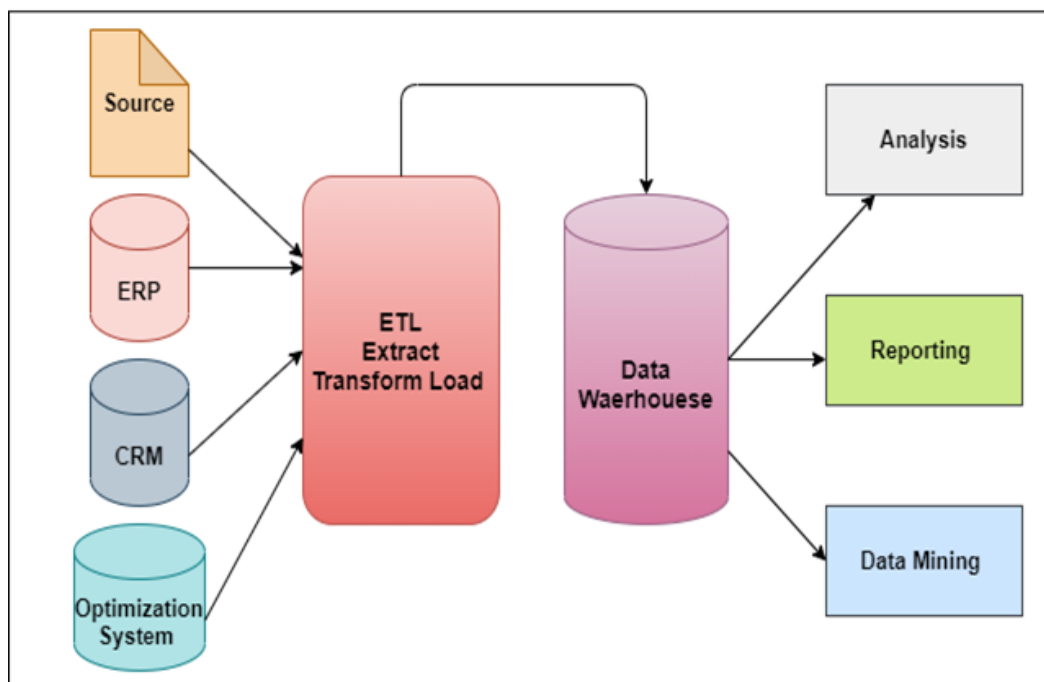


Fig 3. Data Warehouse in Data Mining

3.1 Data Warehouse Design Approach-(DWDA)

The Data Warehouse Design-(DWDA) approach is a fundamental and significant piece of information stockroom for building, picking, and overseeing with efficient procedure. With business classification and time, the correct methodology was identified. There are three information distribution center methodologies which we are going to talk about it. ⁽¹¹⁾

3.1.1 Top-Down Information Distribution Center Structure Approach-(TDIDCSS)

The TDIDCSS was introduced by the researcher named “Bill Inman” and known as the father of data Warehouse. In this methodology, information is efficiently put away in the Third Normal Form-(NF3) because this stockpiling method makes the retrieval and capacity of information from a value-based framework. Right now, the stockroom is structured, and information must be made on the upper side of the information distribution center. Information can be extracted from different parts, sources, or databases. At last, ETL substantial is utilized for checking the information correctness. Essentially information will be stacked and approve to demonstrate the exact structure of information. ⁽¹²⁾ It is important to utilize ETL for handling and transformation of the information distribution center into an information bazaar for business utilization. Information can be removed on a routine basis for better information stockroom results. ⁽¹³⁾ (Figure 4)

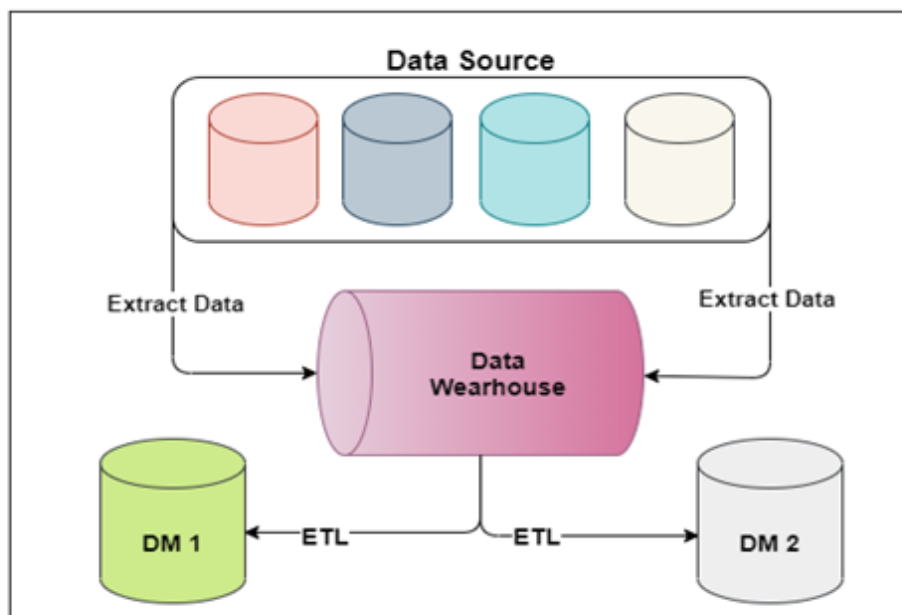


Fig 4. Top-down information Distribution Center Plan Approach

3.1.2 Base-up Information Distribution Center Structure Approach-(BIDCSS)

The BIDCSS was invented by “Ralph Kimball”. This methodology is called Dimensional Strategy or Kimball Approach. ⁽¹⁴⁾ Essentially this is the turn-around strategy for the Bill Inman Approach. Information Bazaars are stacked from different sources, with the assistance of ETL strategy, these are stacked into an information distribution center. The flow of information starts with the extraction of information and the various segment, sources, or databases. After refreshing current information, it is separated into different stages, and changes are applied to make information store structure. After that, the information will be outlined, amassed, and stacked and will be used for end clients. This methodology is known as a dimensional information distribution center. Since it is more technical and direct opposite of Inman’s philosophy which is known as the top-down approach or information distribution center plan approach. ⁽¹⁵⁾ Information shops consist of various wellsprings of information that are changed from the third typical structure to the dimensional models and these models are comprised of realities and measurements. These realities are a measure for a specified purpose of time. ⁽¹⁶⁾ (Figure 5)

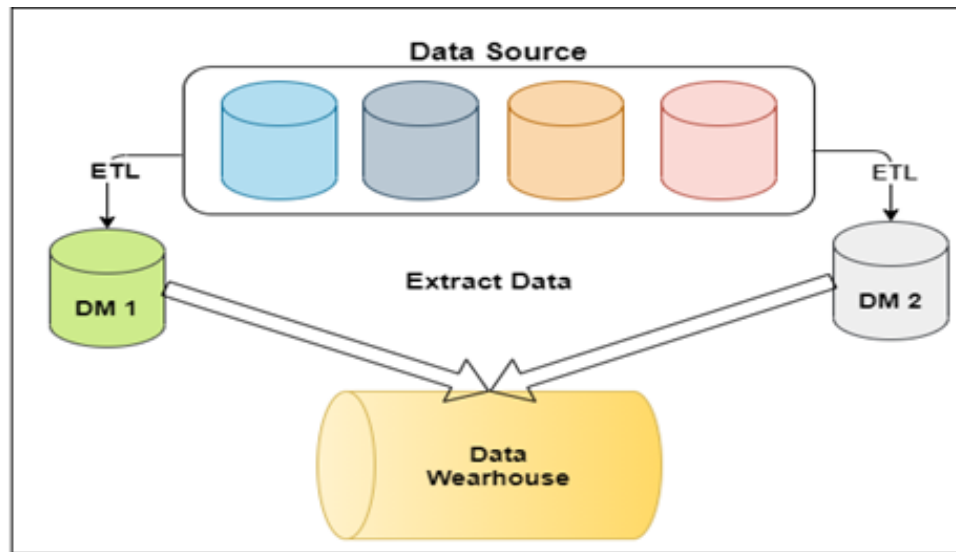


Fig 5. Bottom-up Data Warehouse Design Approach

3.1.3 Hybrid Data Warehouse Design Approach- (HDWDA)

This is the third and last structure approach of information stockroom. The mixture information stockroom approach is including angles from the top-down and base up approach. It is the combination of two strategies and arrangements. Crossover information stockroom is kept on the third ordinary structure to lessen the excess of information. ⁽¹⁷⁾ (Figure 6)

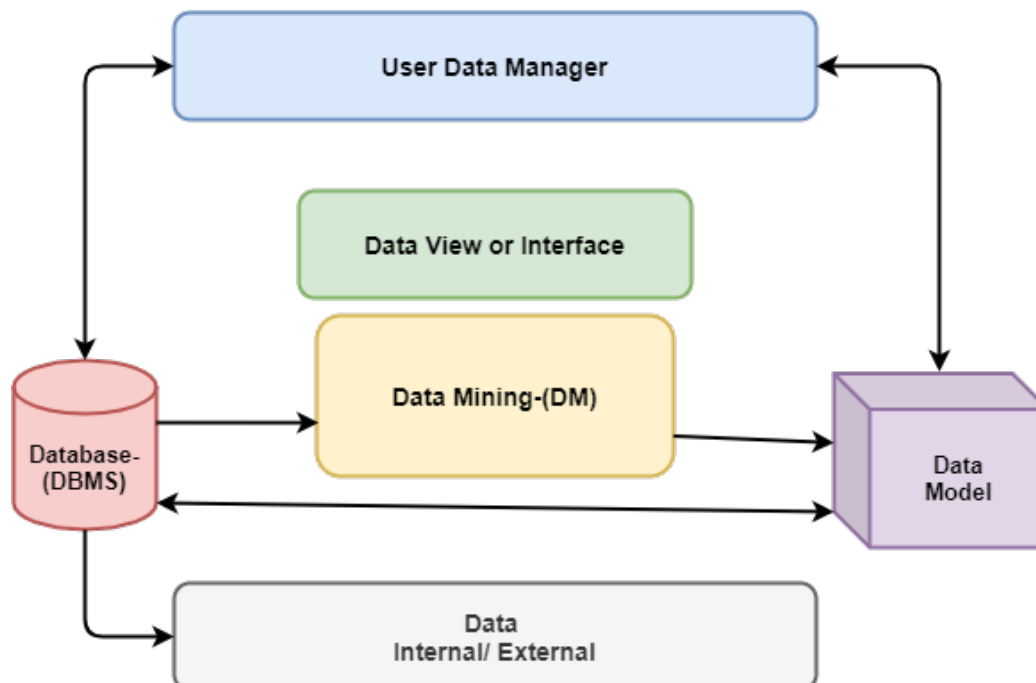


Fig 6. Hybrid Data Warehouse Design Approach

The information stockroom gives a private method for data where little information Bazaars can be made. The crossover configuration approach permits a data distribution center to supplant with an ace information shop, where operational and

non-static information is kept. Information can keep up by utilizing SQOOP.⁽¹⁸⁾ It includes various frameworks. Putting the information away into Hadoop is called a crossbreed information distribution center. On the off chance that information is put away first in Half Breed information distribution center, at that point it may be utilized by end clients using HIVE ODBC or Microsoft Power BI, Decision Trees, Board Management Intelligent Toolkit, or Qlik view or different devices of information mining and information stockroom or by the source clients.⁽¹⁹⁾

4 Web data crawling by data mining

E-commerce plays an important role in business communication between different buyers and sellers. Nowadays, Organizations can get the users' data using e-commerce. Due to the fast data contamination and use of the internet, people should feel to be made a site for business data and even online buying and selling sites. This is the best approach to advancing a business. In many cases, the all-related contender is making the same sort of item or business site. Then how we expect that our site is clicking or visiting by the client. Perhaps our rival site is on the principal page of any program, it will visit the client and this is an approach to draw in and pull the client. We are defining the site ordering method. The crawler is called boot or insect, which is the pivot around the web via e-commerce. The files of our sites are right to state that the primary reason for the crawler is ordering the sites. One thing is critically realized that the crawler doesn't creep or file in each site.⁽²⁰⁾ The crawler has a few systems, rules, and guidelines to finish the checking on sites. It checks the trustworthiness, consistency, copyright, language, and duplication of information. If a web is recorded by slithering, at that point it implies this web getting traffic and earned. Then again if the web isn't recorded by creep, at that point there is no traffic. The reality about through the Index, a web is spared in the database through the storehouse technique in the wake of slithering.⁽²¹⁾

Web Data Crawling is a technique that display multiple results on the screen when we search something on the web. The principal motivation behind the Crawler is to go ahead of sites and watch users' recent searches. On the other ways, a web has no excess, reiteration, or duplication of information then Crawler duplicates and every one of its connections and URL that spares it on to XML, XLS, or on spreadsheet file. Typically, Crawler comes into a web without authorization. We permit ourselves to visit our site and list it. Crawler visit is very beneficial for us when something new and significant on our web. In any case, definitely Crawler visit the web because of solicitation.⁽²²⁾

There are several apparatuses and programming languages to creep information from the clients. We can hide the crucial parts of the web page from clients or crawlers then we would use automated files. The automated files are relying upon web security approach that is show and escape clients. At some point, we can't show the piece of page which is inadequate. The crawler doesn't check the only URL of the web, it can also check the title, photographs, watchwords, and different hyperlinks which are saved money on the database. In this way, we should mind that there are no reiteration, duplication, or copyright issues.^(15,21,23) (Figure 7)

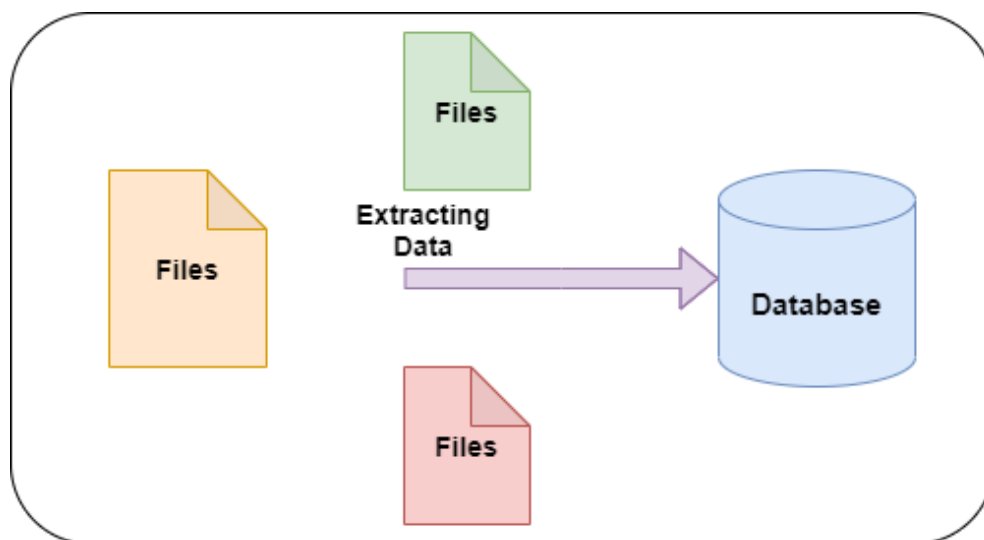


Fig 7. Web data crawling

4.1 Tools and techniques for data crawling and data mining

We are using the hypothetical procedure and the technique of Crawling information. Few applications and devices are accessible for data mining. Several instruments are composed and made by various kinds of dialects. This is the rundown of certain devices with composed language and stage which are utilizing for slithering the information.⁽²⁴⁾ The tools are Crawler, Web Eater, I Crawler, Distributed Web Data Crawler-(DWDC), ASP Seeks-(ASPS), Scraper Wiki-(SW), Parse Hub-(PH), Web Scraper Chrome Extension-(WSCE), HT Track-(HTT), Netspeak Spider-(NS) and Content Grabber-(CG).

4.1.1 HT track

HT Track is a free and open-source programming framework that permits downloading the entire whole site from the web to the PC. At least two sites can be downloading all the while. HT Track works with the strategy for mirrors and concentrates information. It can continuously refresh the whole site. It gets all pictures, archives, figures, HTML, catalogs, and different files from the server to your PC. It can stop and resume the whole site which you are downloading. There is a difference between crept/removed or downloaded web and on the server-side web. A broad contrast between crept/download/concentrate and server-side/on the web. The downloaded webpage is a method which is costly because you have to visit different locales and use your memory for data storage for different websites.⁽²⁵⁾

4.1.2 Scrappy

Scrappy is a permitted and exposed source net crawler that is developed in python language. The primary capacity of scrappy is packets of information and additional is permitting engineers to reuse their code. It is utilized for extraction of information with the assistance of APIs or as a broadly useful and as a web creeping. It is one of the independent crawlers which is assembled around bugs. An engineer can test their web working. They can reuse the code and check the site behavior.⁽²⁶⁾

4.1.3 Data miner

Data Miner (DMr) is an information excavator for the program augmentation. It assists with sparing information that you are seeing on your program page. it is also used to remove information. It spans information on spreadsheets or exceeds expectations. It is additionally useful to utilize web offline mode.⁽²⁷⁾

4.1.4 BIO web mining toolkit

The Bio-Web Mining (BWM) toolbox is composed of java as an open-source. This is a cross-stage web crawler having an Apache server. The toolbox runs on Hadoop with the arrangement of falling funnels. The BWM is effectively running on the rebuild devices for specific needs. The workflow module is used for gathering and making activities stronger. The URL used for data searching. Scoring Algorithms are utilized for the data ordering. It is useful for Business Intelligence-(BI), Social media marketing, shopping baskets and diagrams.^(28,29)

5 Results and Conclusion

This study focused on the importance of data management for the improvement of the business wherein we also discussed the different data mining methods that were used to format the data in various ways. Different organizations use Regression, Classification, Association, and Clustering for data management and analysis. These techniques have some pros and cons. All these techniques produce good results and predict better ways but time-consuming. We use BI and BA for the format and decision of the data. We analyze that BI and BA consume lesser time but contribute for better results to improve business.

The article additionally portrays the devices and strategies of BI and BA. Through BI we examine, analyze and extract information that is helping in basic leadership and improving a wide range of organizations. The information stockroom examined the significant piece of BI and information mining. Data Warehouse (DW) helps to examine, analyze and glean from databases and ERP sources. The BI and BA are better ways for the improvements of the business But the Hybrid, the combination of both, performed still better.

We also examined the web information slithering through information mining and strategy working procedure. We use different applications and strategies for data mining using BI and BA. Different DWD (Top-down, Base-up, and Hybrid) approaches are discussed in detail for business improvement. A hybrid approach is the best for data mining for product data. This study also presented the application of BI, BA, and programming projects. The web information crawler and web information mining can illuminate the issues, change from crude information to others which supply the specialized feasibility. The most perfect programming languages are C, C++, Java and Python that most effectively used by the organizations. After the discussion of different types, tools, methods, and data warehouse types, we analyze that, there is a lot of shortcomings in prevailing methods

that can be improved by our approach. The new hybrid way (BI and BA) can be an effective approach of data mining to gain business objectives and goals because it takes lesser time and cost as compared to the others. We also found that crawling is the best way for business development and improvement of ordering/ formatting data.

Future work

In the future, we will develop a model using the BI method or hybrid method that can improve organizational business, produce a featured product and save organization time and cost during the format of the data.

References

- 1) Tan PN, Steinbach M, Kumar V. Introduction to data mining. and others, editor; Pearson Education India. 2016. Available from: https://books.google.com.pk/books?id=v1qqDwAAQBAJ&sitesec=buy&source=gbs_atb.
- 2) Aggarwal CC. Data mining: the textbook. and others, editor; Springer. 2015. Available from: https://books.google.com.pk/books?id=cfNICAQBAJ&sitesec=buy&source=gbs_vpt_read.
- 3) Han J, Pei J, Kamber M. Data mining: concepts and techniques. and others, editor; Elsevier. 2011. Available from: <https://books.google.com.pk/books?id=pQws07tdpjoC&printsec=frontcover#v=onepage&q&f=false>.
- 4) Mishra BK, Hazra D, Tarannum K, Kumar M. Business intelligence using data mining techniques and business analytics. In: and others, editor. 2016 International Conference System Modeling & Advancement in Research Trends (SMART). 2016;p. 84–89. Available from: <https://ieeexplore.ieee.org/abstract/document/7894496/>.
- 5) Böhringer M, Gluchowski P, Kurze C, Schieder C. A Business Intelligence Perspective on the Future Internet. *AMCIS*. 2010;p. 267–267. Available from: https://www.researchgate.net/profile/Christian_Schieder/publication/220894233_A_Business_Intelligence_Perspective_on_the_Future_Internet/links/5658ab2708ae1ef9297df54c.pdf.
- 6) Seah BK, Selan NE. Design and implementation of data warehouse with data model using survey-based services data. In: and others, editor. Fourth edition of the International Conference on the Innovative Computing Technology. INTECH. 2014;p. 58–64. Available from: <https://ieeexplore.ieee.org/abstract/document/6927748>.
- 7) Hussain R, Lee J, Zeadally S. Trust in VANET: A Survey of Current Solutions and Future Research Opportunities. *IEEE Transactions on Intelligent Transportation Systems*. 2020;p. 1–19. Available from: <https://dx.doi.org/10.1109/tits.2020.2973715>.
- 8) Zhou L. Continuous Authentication and Lightweight Implementation of Elliptic-Curve Cryptography for the Internet of Things. 2019.
- 9) Taherkordi A, Zahid F, Verginadis Y, Horn G. Future Cloud Systems Design: Challenges and Research Directions. *IEEE Access*. 2018;6:74120–74150. Available from: <https://dx.doi.org/10.1109/access.2018.2883149>.
- 10) Mishra BK, Hazra D, Tarannum K, Kumar M. Business intelligence using data mining techniques and business analytics. In: and others, editor. 2016 International Conference System Modeling & Advancement in Research Trends (SMART). 2016;p. 84–89. Available from: <https://ieeexplore.ieee.org/abstract/document/7894496>.
- 11) Apte C, Dietrich B, Fleming M. Business leadership through analytics. *IBM Journal of Research and Development*. 2012;56(6):1–5. Available from: <https://dx.doi.org/10.1147/jrd.2012.2214555>.
- 12) Kim SM, Ha YG. Automated discovery of small business domain knowledge using web crawling and data mining. *2016 International Conference on Big Data and Smart Computing (BigComp)*. 2016;p. 481–484. Available from: <https://ieeexplore.ieee.org/abstract/document/7425974>.
- 13) Surya A, Sharma DK. A comparative analysis of clickstream as web page importance metric. In: and others, editor. 2013 IEEE Conference on Information & Communication Technologies. 2013;p. 776–781. Available from: <https://ieeexplore.ieee.org/abstract/document/6558199>.
- 14) Taherdoost H, Sahibuddin S, Jalaliyoon N. Perceived barriers and benefits of Web based services. In: and others, editor. 2014 International Conference on Computational Science and Computational Intelligence; vol. 2. IEEE. 2014;p. 34–39. Available from: <https://ieeexplore.ieee.org/abstract/document/6822300>.
- 15) Eickhoff C, Kim Y, White RW. Overview of the health search and data mining (hsdm 2020) workshop. In: and others, editor. Proceedings of the 13th International Conference on Web Search and Data Mining. 2020;p. 901–902. Available from: <https://dl.acm.org/doi/abs/10.1145/3336191.3371879>.
- 16) Leskovec J, Rajaraman A, Ullman JD. Mining of massive data sets. and others, editor; Cambridge university press. 2020. Available from: https://books.google.com.pk/books?hl=en&lr=&id=S4HCDwAAQBAJ&oi=fnd&pg=PR9&dq=19.%09Leskovec+J,+Rajaraman+A,+Ullman+JD.+Mining+of+massive+data+sets.+Cambridge+university+press%3B+2020+Jan+9.&ots=iQlfgoiAPp&sig=5sewB3ooUiOh3tZ89hRIVGXd4Ko&redir_esc=y#v=onepage&q&f=false.
- 17) Vashani H, Sullivan J, Asmar ME. DB 2020: Analyzing and Forecasting Design-Build Market Trends. *Journal of Construction Engineering and Management*. 2016;142(6). Available from: [https://dx.doi.org/10.1061/\(asce\)co.1943-7862.0001113](https://dx.doi.org/10.1061/(asce)co.1943-7862.0001113).
- 18) Schoier G, Borruso G. A methodology for dealing with spatial big data. *International Journal of Business Intelligence and Data Mining*. 2017;12(1):1–1. Available from: <https://dx.doi.org/10.1504/ijbidm.2017.082705>.
- 19) Rajeswari S, Suthendran K, Rajakumar K. A smart agricultural model by integrating IoT, mobile and cloud-based big data analytics. In: and others, editor. 2017 international conference on intelligent computing and control (I2C2). 2017;p. 1–5. Available from: <https://ieeexplore.ieee.org/abstract/document/8321902>.
- 20) Kraus M, Feuerriegel S, Oztekin A. Deep learning in business analytics and operations research: Models, applications and managerial implications. *European Journal of Operational Research*. 2020;281(3):628–641. Available from: <https://dx.doi.org/10.1016/j.ejor.2019.09.018>.
- 21) Neves TFM. A data mining approach to predict probabilities of football matches. 2019. Available from: <https://repositorio-aberto.up.pt/handle/10216/121217>.
- 22) MIS S. Evaluation of the Adequacy of the Vidas® SET2 Kit in the Detection of Staphylococcal Enterotoxins in Cheese of Portuguese Origin. *Nutrition & Food Science International Journal*. 2018;5(4):50–60. Available from: <https://dx.doi.org/10.19080/nfsij.2018.05.555669>.
- 23) Kumar A. Implementing real time recommendation systems using graph algorithms & exploring graph analytics in a graph database platform (Neo4j) (Doctoral dissertation. 2019. Available from: <https://esource.dbs.ie/handle/10788/3708>.

- 24) Abbas I, Muneer U. The Prediction of death causes using regression models and moving averages. *International Journal of Data Science and Advanced Analytics*. 2019;1(1):39–46. Available from: <http://ijdsaa.com/index.php/welcome/article/view/69>.
- 25) Mazimpaka JD, Timpf S. Trajectory data mining: A review of methods and applications. *Journal of Spatial Information Science*. 2016;(13):61–99. Available from: <https://dx.doi.org/10.5311/josis.2016.13.263>.
- 26) Thomas DM, Mathur S. Data analysis by web scraping using python. *2019 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA)*. 2019;p. 450–454. Available from: <https://ieeexplore.ieee.org/abstract/document/8822022>.
- 27) Saurkar AV, Pathare KG, Gode SA. An Overview On Web Scraping Techniques And Tools. *International Journal on Future Revolution in Computer Science & Communication Engineering*. 2018;4(4):363–367.
- 28) Santos PI. Ressignificando o papel da CPA: possibilidades e perspectivas estratégicas das ações de autoavaliação institucional. .
- 29) Rajavel R. Trace Event–A Web Mining based Application for the Collection of Online Web Events. . Available from: <https://cutt.ly/Dk5oYa0>.