Big Data Analysis For Business In Social Media Networks: Existing And New Approaches

Abbas Fadhil Aljuboori¹, Saddam Hamdan Ahmed²

¹University of Information Technology and Communications, Baghdad, Iraq.
²University of Babylon, Information Technology Collage, Iraq.

Abstract
Business intelligence is often confused with smart reports, as many people consider any dashboard to be business intelligence. This is not true. The report that contains a set of graphs that interact with each other is nothing but smart reports. They are nothing but reports that people have been writing for decades in statistical books, but the technology has been added to them to interact with each other in a coherent manner. Business intelligence is linking these reports with the goals of the company or the department, where the goals and indicators that we are looking for an answer to are determined and their performance is monitored so that the machine replaces the human being in monitoring performance and tracking patterns from the data, studying and analyzing it to reach the best possible performance through the available data. The quality of the data in terms of structuring the data forces us to determine the technology used. Data warehouses are built for structured data, which is data that is stored in databases or files such as Excel in the form of tables. Unstructured data is stored on file distribution systems (such as internet data, social networks, and text files). The concept of a data lake is a scientific concept that means that you have both structured and unstructured data in the company. You can shorten the topic and say we want solutions to a data lake. The researcher will conduct a survey of the current and new methods of analyzing big data for business in social networks and others.

Keywords Social Media, Business, Big Data, Data Analysis, Decisions Making.

Introduction
Big data has several characteristics that must be present in order for organizations and governments to determine if they are suffering from a data inflation problem, referred to as 3Vs (Volume,
Velocity, Variety) and these are considered essential characteristics. And later (Vacity, Value) accuracy and value were added to it, so we have 5Vs, which are as follows: (1)

**Data Team**
Weak training and lack of experience are a major factor in poor products, and reliance on a group of poorly trained recent graduates without the availability of an expert or a professional advisory body inevitably leads to poor projects. WEhope he does not understand that WEam against hiring fresh graduates, but WEbelieve that there must be an expert manager or group of experts with these young people to guide them. What we see as a trend to train young people is a beautiful thing, but reducing the specialization of data analysis to courses on a data display product and marketing this product as data analysis does more harm than good.(2) Analyzing data without learning many principles of statistics and the skill of dealing with databases and some programming languages produces an employee who is not ready to deal with data. In my opinion, this is the weakest belief, although there are skills that must be taken into account such as systems analysis, strategy science, project management ability, and data quality. (2) Of course, not every data team has to be familiar with it, but at the very least there should be someone on the team who is fluent in such skills. And WEhope that educational institutions adopt full paths, and not what is happening now from a narrow view of data specialization. The solution, in my opinion, is only to hire someone who has the experience and knowledge to lead this dangerous team for the facility, or to seek the assistance of the expertise of professional companies in this field, or both. (1) And at the very least provide good training and invest in the data team. It is also worth noting the importance of project management science in improving the performance and path of the work team, as most of the current projects are managed as if they were university experiments with limited goals. (3)

**Methodology**
Minimizing data projects and reducing them to data presentation programs such as Tableau or Power BI, and the organization believing that it has a data project without using statistics, data science, or setting goals from the project is unfortunately a common practice. (4) As well as building data projects as a one-stage project. And ignoring that data projects are built in several stages and require time, effort and financial support, which are attempts that will lead to the weakness of the final product. The solution is defining objectives, building strategies, direct and permanent interest from decision makers, and choosing appropriate technical solutions for the facility. In this field, there is a wide sea that cannot be mentioned. (2) It can be reduced to the data strategy, and it is one of the topics that needs further research and study.

**Decision makers**
Kevin Quinn, a data management expert with more than 25 years of experience, says: A lot of the reports that data experts make are ignored and looked at once they are received and never returned! It is a common practice. (4) WEassert that there is no one working in the field of data but his reports have been ignored after a while. Kevin Quinn says the reason is very simple: what they say
and what they mean are two different things. He believes that the decision-maker must rely on a data expert who explains to him the problem he is looking for a solution to, and then the data expert determines what fits the data and how the data will be presented. (1) It was added to him that it is preferable for the decision maker to increase his skill in dealing with data and understanding its problems and needs. In this aspect, WEhope that courses dedicated to decision-makers in dealing with data will appear, such as courses in finance and accounting for non-specialists. The prevailing illusion that the availability of data and reports solves many problems controls many decision makers. (4) Many of them dream of creating platforms for him through which he can review and follow up data without being familiar with the details of the administrative systems in his facility, causing him to drown in the details and lose confidence in his organization and his employees, and think that everyone is a failure or that they want to mislead him. For example, he is not aware of all the policies and procedures followed in his personnel affairs system, which makes him think that the numbers appearing to him are full of errors until their reasons are explained to him, which would be a shock. (2) Then he tries to solve the problems he thinks by wearing the hat of the personnel manager without experience in this field to start chaos in the facility. (5) The solution is to rely on data experts and specialists in the target field (such as personnel affairs) to evaluate the data and ensure its logicality and satisfaction with the results of this team, and then determines the goals and problems that it seeks to solve. (3) Leave it to the experts to solve it. In other words, it makes room for data led by the organization via what is known as Data Driven. WEguess WEdidn't cover all the problems with the data projects, but this is what WEthink most influences the success of the project or not. WEalso hope that the specialists will cover this aspect with further study and research.

Historical glance and previous studies
In a 1958 article, IBM researcher Hans-Peter Len used the term business intelligence. Intelligence has been defined as follows: (6) "The ability to capture the interrelationships of facts presented in such a way as to direct action towards a desired goal". Howard Dresner in 1989, later a Gartner Group analyst, proposed business intelligence as an umbrella term to describe "concepts and methods for improving decision-making using reality-based support systems." (7) It was not until the late 1990s that this use became widespread. The researcher Saddam and others published an article using some artificial intelligence algorithms to find the optimum accuracy for analyzing medical data and predicting potential heart diseases entitled The Heart Disease Prediction by Using Random Forest Algorithm, (8), the article concluded that Applying Data mining techniques for chronic diseases can support decision making, reduce the time and the cost in hence reduce the chronic disease death.

Research Aims & Objectives
The research aims to identify the current and modern methods of analyzing big data in the field of business on social networking sites and what are the methods used to follow the interests of users.
on those sites to analyze and categorize them and thus display future publications that correspond to the desires of the browsers to ensure the greatest success rate and profits for business operations.

Research procedure
Social media platforms provide data through which people's opinions and feelings can be known about various issues (social, political, economic, sports, etc.), or about products, companies, advertising campaigns, or others, without the need to survey people's opinions on these issues. However, the methods of obtaining and analyzing this data may be a challenge, especially for non-specialists in computer and information sciences because some of these tools require the use of software codes to obtain and analyze the data. (9) This article discusses the most famous methods used to obtain data from social media and then analyze it without going into technical details in some of the methods.

Data analysis methods
There are several ways to analyze social media data, the most important of which are:

Content Analysis
A technique/method for understanding and inferring meaning from a large amount of text by following a systematic classification process. Content analysis may be quantitative or qualitative. (10) The quantitative method is concerned with counting the amount or number of repetitions of a particular word or term in the text, while the qualitative method is concerned with classifying the data into topics or categories; It may be a mixture between the quantitative and qualitative method, and this is what prevails. (11) Content analysis is a subjective analysis that may produce different results when the same data is analyzed by different people. Therefore, when analyzing any content, it is recommended to set goals to be focused on during the data analysis process and questions to be answered through the process of data analysis and auditing the results by using inter-coder reliability documentation. Content analysis technology is mainly used to analyze text, but it can also be used to analyze images, audio clips, and video.(12)

There are one or two approaches to content analysis:

a- Deductive analysis: where the analysis process is carried out with the presence of previous sections or topics that were deduced from the analysis of similar previous data (through previous studies, for example), and the current data is analyzed by using the previous topics or sections, where the researcher matches the current data with the results Previous research. (13)

b- Inductive analysis: where the analysis process begins without the presence of topics or sections expected to be present in the data, where the sections or topics are reached by using the current data completely.
The basic steps of data analysis by content analysis technique are as follows:

1- Data collection and processing: In this stage, data is collected and processed for the analysis process. You may have a very large number of data that you cannot analyze all of it manually, so you must at this stage choose the data to be analyzed. You may have certain reasons in the selection process so that the selected data sample will help you answer your questions and reach your goals of analysis[^14] (for example: choosing certain times/dates for tweets, the sample containing certain words, targeting specific types of tweet accounts, etc.) The selection process is completely random.

2- Data cleaning: from unwanted items. Whether it is unrelated comments and data such as ads, etc., or unwanted elements such as: the tweeter's geographical location, profile picture, tweet time, and others. This information may be important to your analysis process and may be undesirable. If this data is unwanted, remove it with this step.

3- Getting to know the data: It is very important to read the data and know its content before starting the analysis process to form an initial picture of what the data is and what is the content exchanged between users in that sample.

4- Determining the main units: when you adapt and get to know the data in the previous step, you will notice the main units/topics that were traded in that sample. For example, when analyzing data on different ad campaigns, you may find comments about actors, advertising music and directing, and content or product; These may be considered as different units or topics.

5- Building the Unit Chart: As you read and immerse yourself in the content, you will find many units and the relationships between them. And some topics or units may contain several sub-topics of the main topics, so you build a complete chart of the topics and sub-topics and the relationships between them. In this step, you name and define each section and specify what is included or not included in each section.

6- Experiment with a sample of data on the unit chart: This step enables you to know whether the unit chart covers all the data and topics with the data or not. If you find that the unit chart does not include some data, you modify it to include the data.
7- Analyze all the data: After we have ensured that the units chart fits all the data, we analyze the data.
8- Check the consistency of the analysis: In this step we check that we have included the same type of data for the same section. For example, if there are Tweets that talk about how ads are being output, we check that all of those Tweets are categorized under the same section.
9- Drawing conclusions from the analysis: At this stage and after we have completed the analysis process, we understand the whole picture of the data that we analyzed and link the results with other information (for example: previous studies, other information and statistics) to link the results to the surrounding social world.
10- Writing the final report: If one of the objectives of the analysis process is to write a final report, at this stage we write the final report with the results and the steps that were followed to reach the results. (15)

For example; when analyzing the following tweets about drifting, we have collected tweets containing the word 'drift' and will analyze the content:

![Fig. 2: Analyzing the tweets about drifting](image)

The text in the first tweet suggests that teenage driving is one of the most important causes of drifting (from Muhammad's point of view). So WEput together a section called "Teen Driving" and it contains all the similar tweets that mention teen driving. From the point of view of “Asmaa” in the second tweet, drifting is a major cause of traffic accidents, so WEput a section called “Traffic Accidents” containing all the tweets that refer to traffic accidents. The end result of the content analysis is similar to the following table:
Table 1: Result of the contents analysis

<table>
<thead>
<tr>
<th>% Repetition</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>154 (6.7)</td>
<td>laws and regulations</td>
</tr>
<tr>
<td>412 (18.3)</td>
<td>Teen driving</td>
</tr>
<tr>
<td>819 (38.9)</td>
<td>traffic accidents</td>
</tr>
<tr>
<td>295 (13.8)</td>
<td>awareness and education</td>
</tr>
</tbody>
</table>

Table 2: Result of the contents classification

<table>
<thead>
<tr>
<th>TP Rate</th>
<th>FP Rate</th>
<th>Precision</th>
<th>Recall</th>
<th>F-Measure</th>
<th>MCC</th>
<th>ROC Area</th>
<th>IPCC Area</th>
<th>Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.455</td>
<td>0.214</td>
<td>0.625</td>
<td>0.455</td>
<td>0.526</td>
<td>0.256</td>
<td>0.679</td>
<td>0.756</td>
<td>no</td>
</tr>
<tr>
<td>0.583</td>
<td>0.769</td>
<td></td>
<td>0.583</td>
<td>0.667</td>
<td>0.195</td>
<td>0.641</td>
<td>0.602</td>
<td>yes</td>
</tr>
<tr>
<td>0.000</td>
<td>0.000</td>
<td>?</td>
<td>0.000</td>
<td>?</td>
<td>?</td>
<td>0.250</td>
<td>0.040</td>
<td>?</td>
</tr>
</tbody>
</table>

| Weighted Avg. | 0.600 | 0.356 | 0.600 | ? | ? | 0.642 | 0.648 |

After using classification method for twits types, the classifier shows 60% as accuracy rate and shows that most car accidents caused by Traffic accidents.

**Sentiment Analysis**

This type of analysis means analyzing people's feelings about issues and topics, organizations, products, companies, campaigns, and others. There are four main sections when analyzing sentiment:

Section 1: Positive: This type of data shows positive feelings about the topic we are analyzing the data about. Section Two: Natural: This type shows neither positive nor negative feelings about the subject.

Section 3: Negative: This type shows negative feelings. Fourth Section: Mixed: Contains a mixture of positive and negative feelings.

An example of an analysis of people's feelings about an advertising campaign for a new car:
Table 3: Analysis of people's feelings about an advertising campaign for a new car.

<table>
<thead>
<tr>
<th></th>
<th>Positive</th>
<th>Natural</th>
<th>Negative</th>
<th>Mixed</th>
</tr>
</thead>
<tbody>
<tr>
<td>😊 Wow, I love the ad!</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I just noticed a new car ad</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The car ad is annoying, the direction and the sound is bad</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I loved the idea, but I didn't like the actors' performance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similarly, people's opinions of certain issues can also be analyzed, which is known as (Stance Identification) where the data is analyzed either for or against a particular issue. For example, when compiling a number of tweets about store closings, prayer times, women driving, the VAT law, or other topics, the data analyst can sort the tweets into three main sections:

1- Positive: with the issue at hand (example: with closing shops, women driving or value-added tax).
2- Negative: Against the issue at hand (eg, against store closures, women driving or value added tax).
3- Others: He did not show any clear or strong feelings for either party.

Social Network Analysis
Or network analysis, which is a method of depicting a particular social network or circle by determining the relationships between users in the circle, the knowledge of influencers in that circle, the types and forms of data spread in each circle or relationship, and the nature of communication between users.\(^{(17)}\) This type of data analysis is used when you want to highlight the social network as a whole and its dynamism and not just the content. Several tools can be used to aid in network analysis such as Gephi and NodeXL.

The role of data analysis techniques in solving business problems
Despite the continuous progress that the world is witnessing in all fields (scientific and technical), it is facing a real problem, which is the shortage of material and natural resources. In order to overcome this problem, it is necessary to achieve the optimal use of these resources, invest them in the right way, and save them from loss and waste.\(^{(18)}\) Data analysis techniques can solve this problem by helping businesses make optimal use of available resources by rationalizing the decision-making process to achieve the maximum returns possible for them. Where it can help departments in carrying out their work efficiently and dealing with complex cases to reach optimal solutions to their problems, in order to raise their efficiency in solving problems, by helping them to know the best way to perform something, and keeping these solutions as new knowledge added to the knowledge stock of the organization to contribute to Improving the capabilities and expertise of business institutions to solve problems and face challenges in the future. Because of the expansion and diversification of business activities and the expansion of the scope of the work of...
business enterprises than before, as well as the complex and complex nature of business problems and the overlap of their activities, activities such as production, warehousing, shipping and distribution have become overlapping and affected by one another. In addition to the increasing standards of competitiveness among business organizations.\(19\) So, business organizations must not only solve business activities’ problems, but also achieve efficiency in the process of solving these problems in an agile and practical manner to resolve conflicts or problems that arise during the implementation of business activities in a manner that ensures achieving the greatest return at the lowest costs, as well as helping to meet the entire demand for the product. Hence, the possibility of converting this solution into added knowledge that helps organizations to develop their capabilities in dealing with similar cases that they may face in the future. In order to transform the weaknesses facing organizations into strengths that enable business organizations to have a competitive advantage in carrying out their work and activities in their field of competence. To achieve this, there must be specialized experts who can do this task. However, due to the scarcity of these experts and the increase in demand for them in the future, it is difficult to retain them for a long time due to the high demand for them in the business world and the monopoly of major institutions due to the availability of the great potential of these institutions that help attract these experts. Consequently, many institutions are deprived of the services and expertise of these experts. Here, the tools and techniques of smart data analysis and modeling can provide support for business organizations of all kinds and sizes, through their ability to use the available information to simplify the problem and to determine the degree of correlation of the variables affecting them, as these techniques are characterized by their ability to process data, mix solutions and model problems, and not only provide information ready, which helps business organizations to achieve their goals in the best way.\(14\) As well as the accumulation of experiences and knowledge resulting from these solutions leads to creativity in solving problems, achieving integration with internal processes, and enriching the experiences of the organization to build its capabilities to face problems and similar situations in the future, with the aim of improving the core capabilities of the organization\(15\) Where intelligent data analysis techniques provide assistance to business organizations to solve problems that there is no typical way to solve them, by studying the logical thinking processes of the human element in solving problems, and enabling computers to carry out specialized tasks based on understanding how human experts perform the same tasks.\(2\) In addition, these technologies not only help business organizations achieve physical growth by increasing returns and reducing costs, but also help in developing human capital, which is the most valuable asset of business organizations. Where solutions resulting from the use of intelligence techniques help individuals expand their awareness by dealing with cases that direct thinking cannot address and solve, they help increase their learning opportunities, and improve their abilities to deal with complex situations in the work environment.\(17\)

**Conclusions**

There are many characteristics and factors that may be useful in the process of analyzing social media data, such as gender, geographic location, age group and others, and the importance of these
and other factors is determined by our goals of data analysis. Automation of some data analysis processes (such as sentiment analysis or knowing the gender of the user, for example) is one of the most prominent current research areas by specialists in the field of natural language processing, and the difference and change of dialects and different spelling methods are among the most important challenges in this field. Business intelligence as a process is a continuous process, meaning that the organization continuously collects, records and analyzes data according to certain goals and uses the results in the decision-making process to improve the performance of the organization. Business intelligence as a technology is a group of technologies in information technology that give business intelligence content and form in organizations. In general, it is the processes, strategies, culture, tools, structure, standards and techniques that result in smarter performance of organizations. Business intelligence systems and data warehousing as organizations around the world face real challenges in using the available data in decision support. Business intelligence applications frequently use data collected from a data warehouse. However, not all data warehouses are used for business intelligence and not all business intelligence applications require a data warehouse. The most important of these challenges in extracting trade indicators are as follows:

- Store data in multiple files and databases that are not interconnected with each other.
- Link multiple data sources to provide a unified and integrated view of the organization's data.
- The ability to follow up on the implementation of the strategies and plan developed from the reality of data and indicators.
- The quality of the data for the purposes of analysis - especially historical ones.
- Not using the appropriate presentation and design for each indicator and statistic through graphs and linking that to databases.

Business Intelligence and Analytics, Thomas Davenport argued that business intelligence should be subdivided into inquiry, reporting, and related analytical processing, the "alerts" and business analysis tool. Business analysis is the branch of knowledge that defines business needs and identifies solutions to its problems. Solutions often include a system development component, and can consist of process improvements or organizational changes. The person who performs this task is called a business analyst. Business analysts who work solely on developing software systems may be called business IT analysts, technical business analysts, or systems analysts.

References


Joshua Eckroth, 2018, Python Artificial Intelligence Projects for Beginners: Get up and running with Artificial Intelligence.
Zacharias Voulgaris PhD & Yunus Emrah Bulut [Voulgaris & Zacharias PhD & Bulut & Yunus Emrah], 2018, AI for Data Science: Artificial Intelligence Frameworks and Functionality for Deep Learning.

Wolfgang Ertel & Nathanael T. Black, 2018, Introduction to Artificial Intelligence.


Peter Gentsch, 2018, AI in Marketing, Sales and Service: How Marketers without a Data Science Degree can use AI, Big Data.

Stephanie Diamond, 2019, Digital marketing all-in-one for dummies. Francesco Corea, 2019, An Introduction to Data: Everything You Need to Know about AI, Big Data and Data Science.

Mark Skilton & Felix Hovsepian (auth.), 2018, the 4th Industrial Revolution: Responding to the Impact of Artificial Intelligence on Business.


Ossama Embarak, 2018, Data Analysis and Visualization Using Python Analyze Data to Create Visualizations for BI Systems.

Thomas Mailund (auth.), 2017, Beginning Data Science in R: Data Analysis, Visualization, and Modelling for the Data Scientist.


Stephanie Diamond, 2019, Digital marketing all-in-one for dummies.

Muhammad Mashhood Alam, 2019, Transforming an Idea Into a Business with Design Thinking: The Structured Approach from Silicon.