III. BIG DATA ANALYTICS IMPLEMENTATION, TYPES, TECHNIQUES, AND TOOLS

Due to its characteristics, Big Data requires techniques and even tools to extract its value. This section will explain types, techniques, tools in processing Big Data, and also the implementation of Big Data.

3.1. Big Data Analytics Types

Big Data Analytics are categorized into 4 general types, namely descriptive, predictive, diagnostic, and prescriptive analytics (Zetino & Mendoza, 2019).

3.1.1. Descriptive Analytics

Descriptive analytics are used to generate information of how, who, where something is happening so what and might be happening can be visualized (Zetino & Mendoza, 2019). Descriptive analysis is used to analyze past data and identifies patterns for trend commentary (Deka, 2016). For example, describing information about the geographical area, gender, age, and visitation patterns to develop a predictive model (Zetino & Mendoza, 2019).

3.1.2. Predictive Analytics

Predictive analytics are the development of models that reveal what might happen in the future by extracting data to establish a predictive model (Zetino & Mendoza, 2019). Predictive analytics is used to reveal relationships in data that can't be viewed using conventional analysis and predict the trends and probabilities of the future (Deka, 2016). For example, identify predictive factors to develop a predictive model about revenue (Zetino & Mendoza, 2019). The most widely used modeling techniques in predictive analytics are clustering, regression, decision trees, SVM, NN, association rules, and scorecards (Vashisht & Gupta, 2015).

3.1.3. Diagnostic Analytics

Diagnostic analytics are used to explain why something happened by diagnosing problems, identifying what is wrong, investigating data, and determining variables that are linked to the answer (Zetino & Mendoza, 2019). For example, diagnosis of someone's behavior (Zetino & Mendoza, 2019).

3.1.4. Prescriptive Analytics

Prescriptive analytics are used to expertly determine the best action from uncovered information through the other forms of analysis, not only focusing on How, Why, What, and When (Zetino & Mendoza, 2019; Deka, 2016). Prescriptive analytics can be used to evaluate and make decisions to operate in new ways, and also targets business aims by balancing all barriers (Deka, 2016).

3.2. Big Data Implementation

According to Hurwitz et al. (2013), the first way in implementing Big Data to solve a problem is to acquire the needed data from various sources. The data then are organized, integrated, analyzed to get the answer (Hurwitz et al., 2013). Necessary resources are also needed to efficiently implement Big Data for a purpose (Wielki, 2013). Industries can use various proper techniques, technologies, tools, and people to collect, manipulate, analyze, and visualize Big Data (Wielki, 2013). Krishnan (2013) mentioned that there are four stages to process Big Data: gather, analyze, process, and distribute. Data is first gathered from multiple sources (real-time, near-real-time, and batch systems/ applications). Gathered data are analyzed by tagging, classifying, and categorizing data to group the data based on attributes, subject, classification, and type of data. The contexts of data and its occurrence then are explored. From the context and occurrence, the pattern in the data can be discovered. After the data are processed, each data can be linked and integrated with a data warehouse or other systems. Last, Big Data can be distributed to systems by processing it using analytical or other applications/ platforms (Krishnan, 2013).

3.3. Big Data Techniques

Big Data is a term assigned to large-size datasets that hide any information in its massive volume and can be explored using algorithms or data mining techniques (Arun & Jabasheela, 2014). Meanwhile, Yaqoob et al. (2016) stated that important Big Data techniques consist of data mining, web mining, visualization methods, machine learning, optimization methods, and social network analysis. (Chen & Zhang, 2014) also state that statistics, data mining, machine learning, neural networks, social network analysis, signal processing, pattern recognition, optimization methods, and visualization approaches are several disciplines related to Big Data techniques. Other techniques that can be used to extract data from Big data are topic modeling and sentiment analysis.

3.3.1. Data Mining

Data mining is a whole Big Data Analytic process, from collecting, extracting, analyzing, and statistics (Arun & Jabasheela, 2014). Data mining is used to find interesting/unknown patterns, unusual records, or dependencies (Arun & Jabasheela, 2014). Applications of data mining are for artificial neural networks, business applications (database marketing, stock analysis, credit approval, and retail data analysis), science application (astronomy, molecular biology, medicine, geology), tax fraud detection, and health care management (Gupta, 2014). Data mining is a set technique of extracting specific information from data (Chen & Zhang, 2014; Yaqoob et al., 2016). The type of data mining includes clustering analysis, classification, regression, and association rule learning (Chen & Zhang, 2014; Yaqoob et al., 2016; Arun & Jabasheela, 2014). Trend analysis is also a type of data mining based on the type of patterns to be mined (Gupta, 2014). The type of data mining used depends on the objective or problem to be resolved (Arun & Jabasheela, 2014). Conversely, different articles state that data mining techniques consist of association, classification, clustering, decision trees, and sequential patterns (Unnisabegum et al., 2019).

Anomaly or outlier detection is the search for data items that do not fit the projected or expected patterns/ behavior in order to detect defective or improper procedures or actions (Arun & Jabasheela, 2014). Association rule learning is the process of discovering interesting relation/interdependencies between different variables in large databases that reveal hidden patterns in the data and can be used to identify variables in the data as well as the co-occurrences of different variables that occur with the greatest frequency (Arun & Jabasheela, 2014). Association is the process of finding a correlation to identify patterns between variables (Unnisabegum et al., 2019). For example, finding a correlation to identify patterns between purchases of whipped cream and strawberries: whipped cream is often purchased when customers buy strawberries and vice versa (Unnisabegum et al., 2019).

Clustering analysis is the process of identifying and differentiating data sets based on their similarity (Arun & Jabasheela, 2014). For example, clustering based on attributes of each cluster of customer buying patterns or other subjects for marketing strategy purposes (Unnisabegum et al., 2019). Classification analysis is a systematic process of obtaining important and relevant information about data or data about data that helps to identify different categories of data (Arun & Jabasheela, 2014). Classification is used to identify a particular class of variable using various attributes (Unnisabegum et al., 2019). For example, classify social groups and the age of target audiences for the use of marketing campaigns (Unnisabegum et al., 2019). Regression analysis is the process to determine the dependence between variables and is usually used (Arun & Jabasheela, 2014). For example, determine the level of customer satisfaction and its effect on customer loyalty (Arun & Jabasheela, 2014).

Decision tree analysis can categorize or predict data (Unnisabegum et al., 2019). Decision tree analysis uses a simple question with two or more

answers for the starter and then each answer leads to a further question that is used to make a prediction, identify and classified data that can be categorized (Unnisabegum et al., 2019). Sequential patterns allow trends or regular occurrences of a specific event to be identified (Unnisabegum et al., 2019). Trend analysis is the process of discovering patterns in a period of time (Gupta, 2014).

3.3.2. Machine Learning

Machine learning can develop computers' behaviors based on empirical data by designing algorithms (Chen & Yang, 2014). Machine learning aimed to make intelligent decisions automatically by exploiting knowledge (Chen & Zhang, 2014). Machine learning techniques are required to scale up to cope with Big Data, such as using frameworks Map/Reduce and DryadLINQ (Yaqoob et al., 2016).

3.3.3. Social Network Analysis

Social network analysis is a modern technique that presents social relations within the framework of a network (Chen & Zhang, 2014; Yaqoob et al., 2016). Social network analysis has better performance when the amount of data is not huge (Yaqoob et al., 2016). Social network analysis has been applied in communication, organizational, and development studies, information science, also as a consumer tool (Chen & Zhang, 2014).

3.3.4. Optimization Methods

Optimization methods are used to solve quantitative problems (Chen & Zhang, 2014; Yaqoob et al., 2016). These methods are highly efficient, provide optimization, have high complexity, and are used in multidisciplinary fields (Yaqoob et al., 2016).

3.3.5. Web Mining

Web mining uses data mining techniques to discover usage patterns from huge web repositories (Gupta, 2014). Web mining will find unknown knowledge that then is used for further analysis (Gupta, 2014; Yaqoob et al., 2016). Based on the type of data, web mining is classified into three, namely web content, web structure, and web usage mining (Gupta, 2014). Web content mining enables useful information to be extracted from web content in the form of text, images, audio, video, or structured notes (lists and tables) (Gupta, 2014; Yaqoob et al., 2016). Web structure mining makes the node and connection structure of websites can be analyzed through graph theory (Yaqoob et al., 2016). Web usage mining can be used to reveal useful knowledge from web user sessions data (Gupta, 2014). Web mining can be applied for e-commerce, e-services, e-learning, e-government use, self-organizing websites, security, and crime investigation (Gupta, 2014).

3.3.6. Artificial Neural Networks

Artificial neural network techniques have various applications that can be found in adaptive control, image analysis, pattern recognition, and other fields (Chen & Zhang, 2014). Artificial neural networks are time-consuming for the learning process (Yaqoob et al., 2016).

3.4. Big Data Analytics Tools

Platform as a processing technology is needed to easily provide the value from Big Data. Big Data tools are categorized into three classes based on how the data is processed, namely, batch processing tools, stream processing tools, and interactive analysis tools (Chen & Zhang, 2014). The interactive analysis enables users to conduct their real-time information analysis so that the data is processed in an interactive environment (Chen & Zhang, 2014). The tools of each class can be seen in Table 2.

Table 2. Big Data Tools (Chen & Zhang, 2014; Yaqoob et al., 2016)

Batch Processing tools	Steam Processing Tools	Interactive Analysis

		Tools
Hadoop	Storm	Google's Dremel
Skytree Server	Splunk	Apache drill
Talend Open Studio	S4	
Jaspersoft	SAP Hana	
Dryad	SQLstream s-Server	
Pentaho	Apache Kafka	
Tableau		
Karmasphere		
Apache mahout		