

Business Decision Making by Big Data Analytics

Vishal Kumar Goar, Nagendra Singh Yadav

Engineering College Bikaner, Rajasthan, India
dr.vishalgoar@gmail.com, nksyadav100@gmail.com)

Abstract -

Information is the key component towards success when it comes to controlling the decision-makers performance with the quality of a decision. In the modern era, an absolute amount of data is available to organizations for analysis usage. Data is the most important component of the business in the 21st century and a significant number of devices are already equipped with the internet. Based on this the solutions should be studied in order to control and capture the knowledge value pair out of the datasets. Following this, the decision-makers should have access to insightful and valuable data based on the dynamic high volume & velocity using big data analytics. Our research focuses on how to integrate big data analytics into the decision-making process. The B-DAD (big data analytics and decision) framework was created to map the big data tools, its architecture, and analytics for the several decision-making steps by the adoption of methodology based on design science. The ideal goal and offerings of the framework are adopting big data analytics in order to intensify & aid decision making for the organization using an integration of big data analytics into the corresponding decision-making process. Thus, the experiment was carried out in the retail domain to test the framework. As an end result, the results showcased the value-added if big data analytics is integrated with corresponding decision-making activity.

Keywords: Big Data, Big Data Analytics, Business Analysis, Decision-making process, Software Testing, Business Analytics, Cost of doing business.

1. Introduction:

The modern era of digital technology evolution has changed the way how organizations operate on a daily basis based on the business-driven approach. The extensive amount of data is easily available since the storage type has increased with the data collection. With each second passing by more and more data is being generated from several sources. [11] Such data is required to have some mechanisms such that it can be stored for analysis to draw the value. Organizations should push for capturing the maximum value out of the huge data repository. In addition to that, organizations and corresponding stakeholders hold the technology and devices allowing them to create and store data based on different grouping or buckets. [12] Each user these days can have access to personal devices i.e. laptops, smartphones, and such devices consist of larger data volumes which can be important to the organizations. [3] Such type of data is referred to big data where data varies based on volume, variety, and velocity which becomes difficult to maintain and manage with the existing set of tools.

Big data can consist of various types i.e., sentiments, click stream, video, audio, website session tracking for

user activity along with location-based data. [1] Therefore, it requires several methods of big data analytics depending on size, variety, and consistent changes along with storage and analysis methodologies. In addition to the above, such big data should be analyzed carefully to draw out valuable and related information. Organizations are looking for a solution and guidelines for big data management due to the significant request for employing the big data in order to make the best of the available opportunities. [4] The question in our research article is *how to employ big data analytics using the integration with the decision-making process?* The goal of the research is to construct and test the framework using the integration of big data techniques and tools with the decision-making process. [15] The adoption of the framework will allow the decision-makers to intensify the standard of the decision-making process and potentially increase the decision quality as a byproduct. [13] The framework comprises several supreme characteristics of big data analytics i.e. The big data analytics life cycle, required architecture and infrastructure along with necessary tools for their mapping of the different decision-making processes. [6]

2. The Background.

Big Data is applicable to a set of data that grows exponentially such that it becomes hard to work with using the existing traditional database management system. [14] However, the dimensions of big data have lengthened far off the capability of generally used software tools and storage mechanisms to store and manage the process of data in a certain timeframe. [5] The V's are the main characteristics of big data i.e. volume, velocity, and variety. The Volume can be considered as data in size, velocity can be considered as a dynamic form of data that changes at a certain rate or based on how it is created over time. Lastly, variety is considered as a different type of data form with formats. [6]

Big Data analytics is adopted where advanced analytics techniques can be put in on big data sets. [19] The Analytics found in a larger data set can support, disclose and strengthen the business alternatives. The larger set of data throws more challenges when it comes to managing it. [8] In order to ameliorate the decision making, risk depletion, and discovering hidden insights using the data, knowledgeable analytics can come in handy else the valuable intel inside the data will remain hidden forever. [7] Fewer times the decisions are not always required to be automated instead, they should be cross-questioned using the analysis based on data, big data techniques, and technologies. This will help the individuals to understand better in terms of how extracted information is valuable. [10]

Besides, the topics related to managerial decision-making processes are much important and they've been covered in so much research so far. The decision-making consists of four-phase activities: intellect, Plan, alternative, and execution. There are several ways to the big data analysis conveyor but they come with their own hurdles and require decision making. [16] Such a decision can include what data should be acquired and how to represent the data post-extraction, clean up and its integration with additional origins in order to reach the decision based on results using analysis. [17] All of these hurdles and decisions should be planned successfully for the big data analysis to generate the true value. [18]

The decision-makers should be able to recognize and make the best use of big data to furthermore intensify

the conventional decision-making process as they are always in awe to execute the informed decisions whenever the opportunity is available to them.[2] Therefore, the research should be able to abstract how tools and methods can be integrated with the decision-making process using big data for intensifying decision making in order to generate the important insights for decision-makers. [9]

3. B-DAD Framework -

Our Research applies the design science technology and the hence six-phase design science activity is equipped for creating and assessing the framework. As per the first two-stage the problem was identified and the objective is defined for a solution using the investigative research. Thus, the corresponding knowledge was acquired from the knowledge base along with the business requirements for the environment to transform the B-DAD Framework. This is accomplished by going through research and testing the key techniques related to big data analytics such that they can be inherited in the framework. Based on the above research diligence along with relevance were accomplished.

At a later stage after the development of the B-DAD framework, it was examined and showcased using the adoption of big data analytics for supporting the decision-making process. The illustrated idea is static as per the experiments on real data and Real-time business use cases for supporting the applicable context for accessing. The framework evaluation was completed based on the observation of to what extent the framework was a success when applied the big data analytics all over the decision-making process to support the decision based on insights backing it up. Furthermore, the ease of the process and relevance of the framework to several scenarios were perceived. Based on the results, we navigated back to the framework design and creation phase for incorporating the modifications which later resulted in our final B-DAD framework. The detailed process is listed below.

3.1. Creation of the Framework

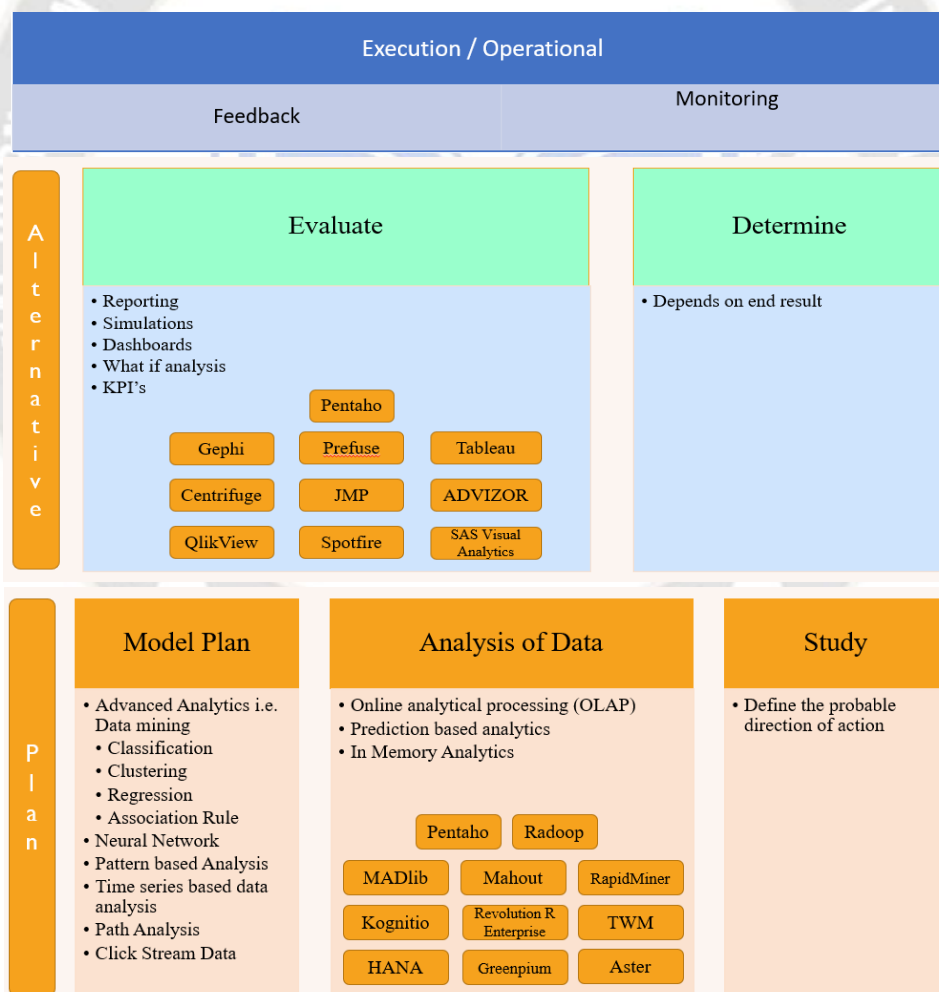
The Big Data Analytics and decision (B-DAD) framework was created to plot the big data tools, analytics, and architecture into several decision-making

process steps. The Big refers to the three aspects which are not only limited to the data but plot several aspects into one. The depth analysis of related literature and the technologies in big data analytics lead to the development of a designed framework for us.

It is impossible to summarize all the big data technology, tools, and analytics as this is a conceptual form using which the support can be added to the decision-making process with big data analytics. The precondition for the framework is that the domain of decision is known and doesn't require any exploration with regards to the extraction of the problem for which we are looking for a resolution. The framework is listed in **Figure 1**.

The First step in the decision-making process is the intellect phase, in which the data is utilized to recognize

the opportunity and issues which is furthermore gathered using the external and internal data origins. The identification of big data sources is a must and data should be captured using several sources, cleaned, saved, and moved to the end-user. In other words, the first step in the framework is about identifying which big data should be utilized for the analysis. The key difference lies in the diversity of data types identified from the various sources. Other than the traditional or transactional data, data related to social media, images & video exist. In addition to the above, data is generated by machines and devices such as log files, data captured from sensors, or location-based data. Besides location-based data is extremely valuable when combined with internet data, XML, or clickstream files.



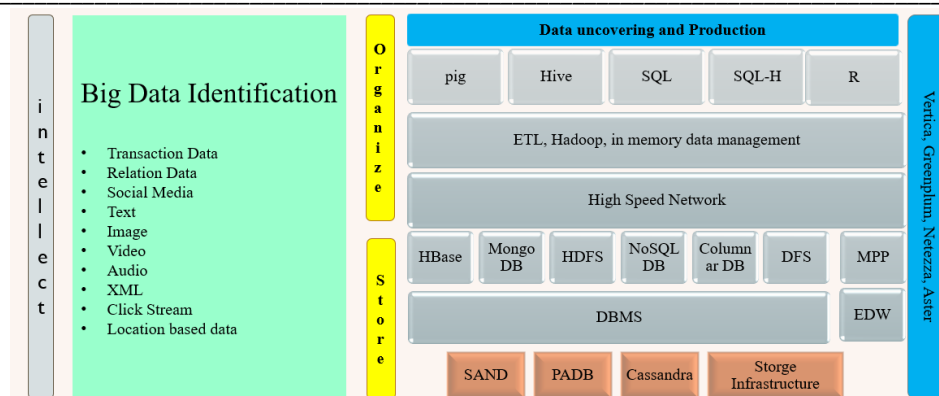


Figure 1: B-DAD Framework

In order to define the analysis source and type of data are required here as later on the selected data is obtained and stored. The Selected data can be stored or saved in any big data storage unit or management tool. Such tools can consist of conventional Database management systems. i.e. My SQL (open source) or MPP DBMS like Cassandra, SAND, PADB. Adding to this the distributed file systems HDFS is utilized for big data storage along with MongoDB (transformed on top of HDFS).

In the data preparation phase, Once the big data is stored and obtained, it is categorized, prepared, and processed for the data analytics lifecycle. Once the data is processed, it enters the organized phase inside the integrated information architecture. This is accomplished using the high-speed network along with ETL or a big data processing tool. For the data processing Hadoop, MapReduce & in memory management is utilized. The data can be obtained using query, computation and the processing can be executed using the several languages starting from Pig, Hive, and R (preferred for statistical analysis) to SQL-H & SQL used for accessing the data stored in Hadoop. The group of such tools can contribute to the big data findings and create the required analysis.

Few vendors offer a pool of tools and platforms for supporting big data storage and management for organizations. This facilitates extensive solutions based on big data along with more features in a single package. i.e., Greenplum, Vertica.

The next step in the decision-making process is the Plan Phase where the possible direction of actions is defined, created, and analyzed using the core fundamentals or based on a model of the problem. The framework consists of the following steps: Model Plan, Analysis of

data, and Study. As per the Model Plan step, a corresponding model for data analytics is opted and planned. The corresponding applicable model and algorithms are selected depending on the available data type and analysis. A set of few models and analyses chosen here are displayed in the framework.

The set of conventional data mining and analytics techniques i.e., classification, a regression can opt for machine learning and AI techniques i.e., decision tree, and pattern analytics. To analyze the data point sequence, time series analysis can be used which will reflect the values at consecutive time intervals. If the big data is in the form of text or social media data, the text and social media analysis can be performed for the above. To represent the complex path analysis or network, graph analysis can be used which can elaborate on the direct dependency between the variables. For the cluster of dense data or location-based data, spatial or density analysis can be performed. For Web data and analysis of clicks on a web page, the click stream analysis can opt.

In the analysis of the data step, the opted model is applied. In order to predict better for the future, predictive analytics can opt for historical and current data and it can also comprise Online analytical Processing (OLAP). To intensify the speed and access for scoring in the analytics model, in-memory processing and analytics can be combined with big data. Diverse tech and analytics tools are used in this phase i.e., Kognitio, and HANA built on top of R Language, TWM, Mahout, and MADlib. Radoop or RapidMiner is an extension used to integrate the offerings of data analytics into Hive. Mahout serves solutions for Hadoop based on data analytics which can execute

OLAP, prediction analytics, or capable to integrate NoSQL with Hadoop & analytics DB.

Inside the Study phase, the end result of past steps along with analytics results goes for analysis. Later on, the probable direction of action is defined. The action is later selected based on an alternative form per the next phase.

The alternative phase is the next step in the decision-making process where each method is utilized to identify the significant influence of the suggested solution, and direction of action based on the plan phase. The framework consists of two steps: evaluate & determine. Inside the evaluation step, evaluation is carried out for the evaluation steps, suggested direction of action along with the consequences are checked with prioritization. This is accomplished using Reporting,

Simulations, dashboards, KPI & what-if analysis. The data visualization tools come in handy i.e., Gephi (popular graph-based visibility), Tableau, and SAS Visual analytics.

The next step in the Alternative phase is to determine the best direction of action. The actual decisions are executed based on the results of the evaluation for the probable best direction of action.

At last, the final phase in the decision-making process is the execution or operational phase in which the suggested solution is implemented based on the outcome of previous results in the phases listed above. Thus, in order to monitor the decision results, big data tools and technology can be utilized which provide real-time data and feedback for the end results of execution.

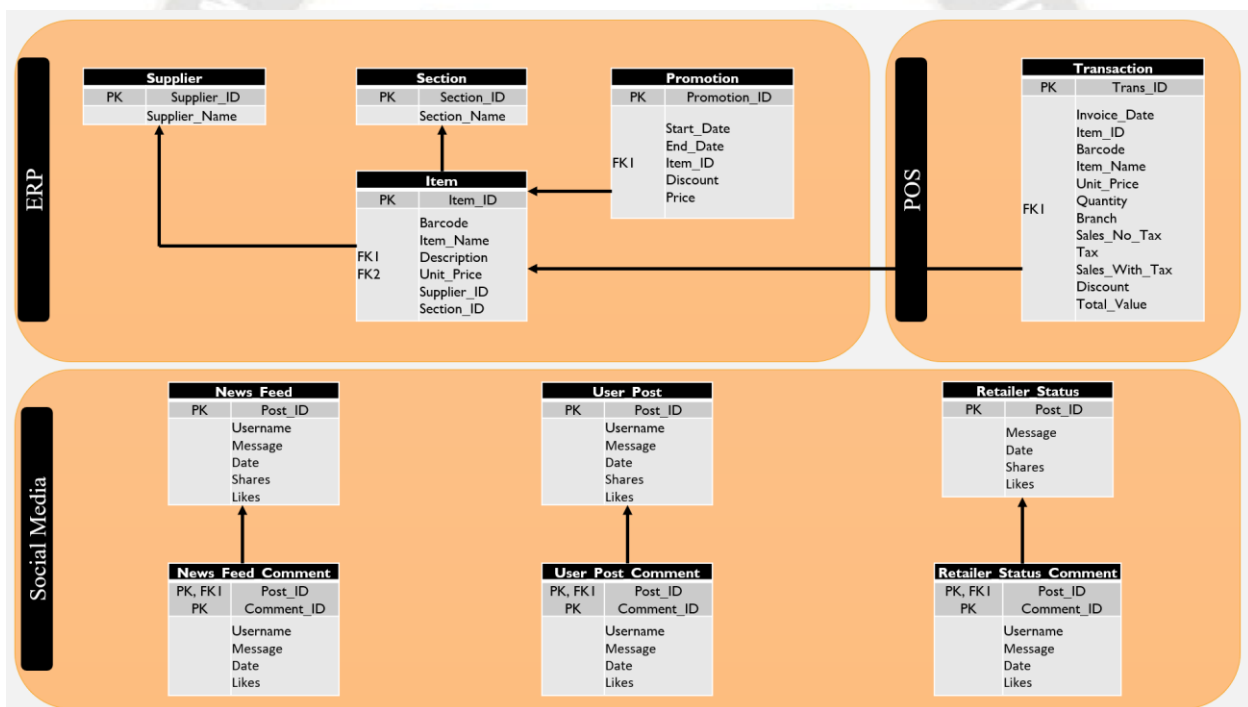


Figure 2 – Data Model

3.2. Retail Domain: framework evaluation

Post the competition of IT infrastructure, the framework should be evaluated furthermore. An experimental method was selected to test the B-DAD framework. Therefore, the illustration of the framework was generated with real-world data. Though the accessible solutions i.e., TWM, Aster, and several other DBMS were tested per lab findings. The flow and integration in the middle of several tools were captured and studied.

The findings were executed in the retail domain for evaluating the B-DAD framework. The decision testing was executed to identify the effectiveness, sentiment impact, & social media impact on Sales. Thus, the decision lies in which product the promotions should be offered, what is the instance on which the promotion should be offered, does the marketing based on social media are effective or should be focused on or not. To support the decisions following set of knowledge should

be available: analysis related to purchase data, feedback, and customer revert to the posts on social media. All the different phases in the framework are listed in the later section. Thus, the implementation phase was not tested since when it comes to decisions they have to be executed and tracked over time and that was not feasible with the given scope of the findings.

1. Intellect Phase:

In this phase of the framework, the big data user is required to be captured. In our experiment, a combination of social media, text, and relational data is used. Our data model is listed in Figure 2. We selected the supermarket data using the POS (Point of sale) and ERP (enterprise resource planning) to capture the relational data related to the retail purchases. The supermarket listed more than 80,000 consumer items with the two largest branches segmented into 30 sections. The daily user visit counter to the outlet can hit up to 50,000. For our usage, we captured the samples of data from POS and ERP. The data was captured using Teradata tools using Teradata DBMS.

The Social Media data was captured using the customer posts and comments on the supermarket's Facebook fan page. We were unable to find the tweets related to supermarkets on Twitter. Thus, we utilized the open-source APIs to capture the post and related comments in a given timeframe based on sales data available for us. We also captured the post and comments related to fan pages for the given timeframe.

2. Plan Phase

The second tested phase in the framework is the Plan phase. Here the actual model planning takes place and the relationship is required to be recognized and explained. To discover the relation between several retail aspects we focused on purchase and discount. Thus, this generated several models in the planning phase. The model and analysis are explained below in detail –

- **Visibility analysis** – Firstly, it is required to understand how the distribution of purchased quantity and the sum of item purchase work across sections or branches, and hence we begin using Tableau for adding visibility to relationships.

Using the visibility, we were able to capture the supermarket branches and departments that are

generating maximum discount, profit, or sales. In order to add visibility to the relationship between total sales, discounts offered, and quantity of sales, a time series analysis was executed for a certain time duration. Some hidden facts were discovered related to the product and peak time impacting the sales, relation with the promotion, and discount.

- **Regression & Association Analysis** – An association analysis was carried out on TWM to discover the relationship among variables. We discovered that the higher association to the discount were sales and quantity values when generating the association matrix. Logistics regression was executed on TWM to identify the connection between if there is a discount or not and the left independent variable. This, the projection of a scenario with a discount or not was discovered on the basis of 9 independent variables. Such variables were used in our model.

- **Cluster Analysis** – The cluster analysis was carried out to group the items on the basis of regular occurrence as per discount. We begin to use cluster analysis in TVM where K means the planned algorithm and the two clusters were used as we focus on the scenario where the discount should be offered to have Boolean values 0 and 1. 0 represents no discount and 1 represents the discount. The resultant metric displayed 78% of the items in “Cluster 1” and “Cluster 2” had 22% of the items.

At a later stage, the clustering was executed for Weka with the help of the k-means algorithm where k is equal to 2 clusters based on discount categories. Discount is a class where clusters are recorded using evaluation class to cluster. Once the cluster model was developed based on training data, it is tested by mapping each class value to the cluster. To measure the validations of clustering the rule was verified. i.e., the discount number of points matches the class in which it was clustered. The difference was found in the results when compared the cluster 0 and 1 with TWM. “Cluster 0” was recorded with 43% of the instance & “Cluster 1” was

recorded with 57% of the instance. The discrepancy can occur since Mahalanobis distance is used to calculate the distance among

the points & mean for a cluster in TWM. On the other hand, Euclidean distance was used in the case of Weka to calculate the distance.

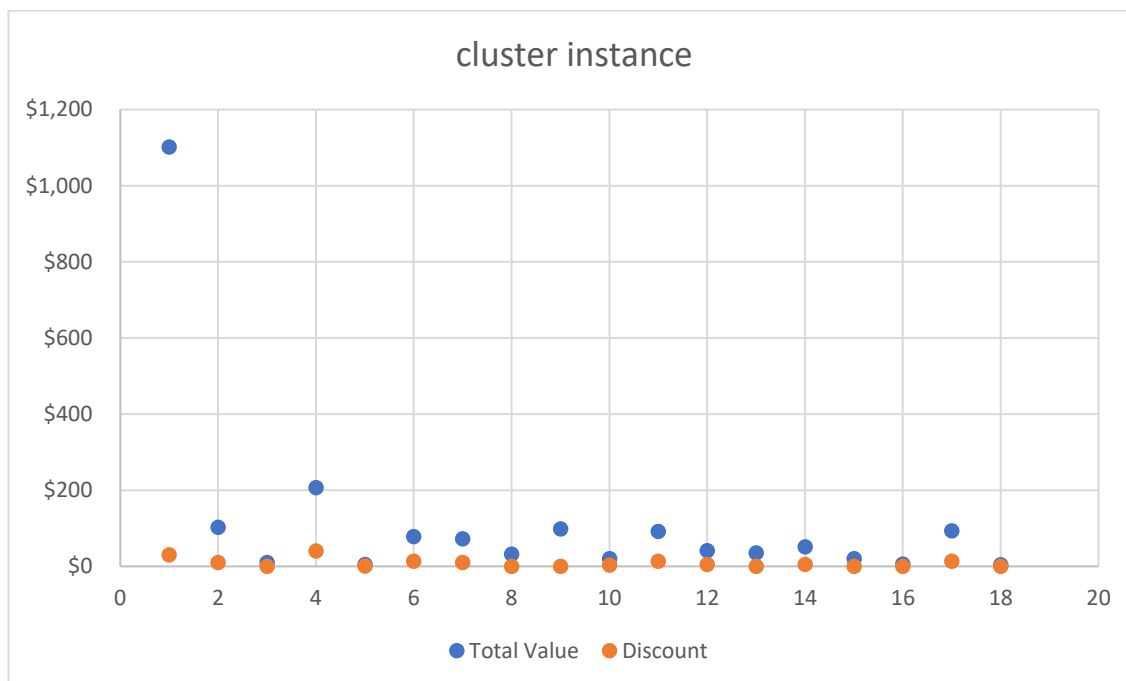


Figure 3 – K-means clustering visualization

Figure 3 adds the visibility of the cluster instance. “cluster 0” can be identified by a orange point where the discount is not available & “Cluster 1” can be identified by blue points where the discount is available. Based on the classes to cluster findings, the overlapping represents the incorrect clustered instance, the orange circle refers to items that have a discount but were mapped to the cluster which holds no discount. The Blue circle refers to an item that doesn’t have the discount but was mapped to the discount cluster. Now the promotions can be mapped to the blue circle. The reason for the error or rise in the orange circle could be because of the fact that 80% of the promotion-based items are supermarket

discounts & discounts were identified using the promotion plan with the help of a supplier instead of the supermarket.

The different end results can be generated based on changes in the clustering algorithm. We can utilize EM (Expectation-Maximization) rather than k-means as we have only clusters to represent. The Expectation maximization clustering runs through each instance and a probability distribution is mapped to each instance to measure the instance probability of belonging to the corresponding cluster. Using the cross-verification the no of clusters was decided. The algorithm bisects the data into 10 clusters and the % of cluster placement is stabilized.



Figure 4 - Expectation-Maximization cluster visualization

Figure 4 adds the visibility on clustered instances founded on the Expectation-Maximization algorithm. The X-axis notation refers to item code and the Y-axis notation refers to the total value where the colors refer to 10 clusters. The higher item code was discovered in blue “Cluster 0”, the “cluster 5” instance aka magenta was found to have the lowest total value & the highest total value was discovered in orange “Cluster 7”.

- **Association Analysis** – The association analysis consists of association rule mining to realize the undiscovered relations and to notice the most usual combination among the variable of data. In our research, we transformed the attribute of numeric code into normal attributes and later segregated the rest of the numerical variables with the help of equal frequency binning such that they can be converted into category-based variables for our analysis. At a later stage, we executed association rule mining by utilizing Apriori & Weka algorithm where minimum support was set to 0.1 with minimum confidence as 0.25. No rules of value will result If the support was lifted up.

Thus, various interesting rules were discovered. i.e., There is no discount applicable if the purchased quantity is low having the confidence ratio as 84% and the second branch listed the confidence as 45%. The “Chocolate & candy” section doesn’t list any discount on 80% of the products. In the case of the first branch, no discount was observed on the 74% of the products in a given time but in the case of the second branch, no discount was observed on the 72% of the products in a given time. Close to 60% of the purchases were discovered in the “Cosmetics” section in a given time and then later on the same was made available for the first branch. There are many more association rules which can be extracted based on the results. We were unable to capture the rules which rely on the purchase of some items together (pack item) since the data doesn’t consist of the individual cart instead it listed daily aggregated invoice.

- **Decision tree** – The decision tree was developed as a part of classification analysis. We have used TWM to construct the tree model with the split strategy on the gain ratio. The

selected dependent variable was the Boolean discount variable whereas the liberated variables are i.e., purchased item quantity, branch, item code, section, supplier code, and sales. The decision tree metric indicated the higher accuracy of the model along with the % of true classification as 77.21 % whereas the incorrect classification percentage was observed to be 22.79%. The end result of the decision tree i.e., the resulting rule indicated the factors impacting the decision if the item should have a discount or not.

- **Text mining and social media analysis** – We executed the text mining on the data captured from social media to realize or identify people's opinions about the supermarket and the listed products along with the reverts to the supermarket post and marketing. Next, we were required to discover the commonly used words on such posts and the association within those words. This helped us to learn about sentiments, products or what people think of supermarket outlets. For example, we discovered that the people were not happy with a certain product or they were happy with some promotions, such information can be added to the model to enhance the decision.

we used RapidMiner to execute the text mining. Reading the data from the database, choosing the expected attribute for analysis, 7 appending the final set of data into an illustration set are the first step to start with the process. Next, refine the data into a document form where each text has some tokens assigned and to filter the English stop words, the operator is used to filter them and remove them from the document. The tokens are cleaned based on length to eliminate the very long or short tokens such that n-grams are produced. Once the data processing is finished, Nominal and numeric attributes should be transformed into binomial for executing the association rule mining. In order to capture the frequency of itemset, the FP-Growth algorithm is employed with the help of an operator which builds the association rule.

The process was never able to go beyond the FP-growth operator post several days of process execution. The largest execution time was for close to 2 days but no result was generated. This could be due to the fact that the language was in Unicode format and it became very hard for processing. Due to the grammatical error or wrong sentence formation, the FP-Growth operator might not be able to perform in such a scenario.

Results generated by document processing went for depth analysis using the TF-IDF (term frequency-inverse document frequency) algorithm which gives the importance to relative words in the document when differentiated from the words in the rest of the document. We were able to locate the number of reoccurring words which are frequently used in social media posts and comments. i.e., “quantity”, “offer”, “promotion”, “orange”.

- **Sentiment Analysis** – To explore the negative and positive opinions shared by the user on the supermarket social media page, we executed the sentiment analysis on RapidMiner. We begin by building the label data set using the samples for 100 posts which were stored in documents based on folders with different prioritization or if they represent the negative or positive posts. The document was, later on, went for processing, and tokens are assigned to each of them. The word vector was trimmed to eliminate the regular and reoccurring terms, such ratio was limited to less than 3% or close to 95% based on time limitations. The vector wordlist was constructed from the data and the same was stored as the “naive bayes classification model” with the help of cross-validation. The data was segmented into training and test sets and the model construction begin with the help of the training set once done, it was applied to the test set based on the accuracy of the performance was calculated.

One of the documents out of 64 was forecasted to be negative whereas 40 of them were classified as negative correctly. This led to a precision ratio of close to 62.5%, and a recall

ratio as 100% as the negative document was marked as negative after classification was completed. Speaking of 60 documents, into which 36 of them were classified successfully based on correctness resulting in 60% recall, and 36 of the documents were forecasted to be positive results with 100% accuracy. Thus, the model accuracy was 76%.

The created model is implied on the posts without any labels to have them filled into the negative or positive category. This was not a practical approach for our case knowing it to be a traditional and widely used sentiment analysis form. This needs each of the posts to be stored in a different document and later manually labeling was required for the corresponding positive or negative training set. Later, after successfully implying the model and marking the documents to be negative or positive it needed to go back into the document and verify how the corresponding post was classified by opening each of them.

Repustate is social media and sentiment analysis website. It consists of APIs which count for sentiment analysis in the corresponding languages. The sentiment analysis was executed on the posts using the trial version of Repustate on the supermarket's social media page. We were able to locate the negative posts based on time to identify when the negative comments were posted and at which time throughout the day. We were able to identify the set of customers frustrated with the product or service experience and how they ended up posting negative feedback or opinion. i.e., this helped us to identify that the quality of delivering services was concerning for the second branch.

3. Alternative Phase – in this phase, we are required to adopt the proposed solution based on the defined results in the previous phase. For our case what item, when refers to the promotions created for customers. We don't seem to have complete background knowledge on how supermarket operates or function or utilizes the KPIs for evaluating them.

Thus, we opted to go for visualization as the way of measuring step in the experiment as it turns out to be adding more value in our case. The analysis in the last phase generated the visibility which can be utilized in the alternative phase. New visualizations were created in this phase to foresee the relationship between variables and how they tend to differ over time based on time series. We opted for the specific instance as an example to realize the impact caused by social media adverts on customer purchasing. For instance, to visualize the best method for the promotion of specific juices Tableau was used since the customer feedback using posts and opinions about the purchase is equally important.

To capture the social media activity of the users, the user engagement with social media posts was visualized. This enabled us to pivot on the specific days which had higher user activity in terms of published posts. We were able to capture the likes and shares for each post to be able to identify which post had the higher user engagement ratio along with comments and feedback.

The higher likes on the corresponding promotion post reflected that the user is more engaged and showcased a positive opinion and vice versa.

Thus, using the analysis for item sales & discounts on the items over time and including the user feedback enabled us to identify the accuracy of the online promotions and sentiments based on purchase patterns. This leads to another analysis where we can forecast which date and time are best for offering the promotions.

3.3. Findings in the experiment –

As per our experiment, we have showcased the decision on which corresponding items should have promotions offered and how social media can be used for marketing, resulting in what extent this can influence the customer purchase. This helps to capture the customer feedback for each purchase. Our framework is accompanied by opting for mapped big data analytics tools and methods build in the corresponding phase of the decision-making process. Through the process, data was stored in big data and later went for processing of the same. Later the analytics was executed and the results were analyzed to enable the decision-making

with enhanced visibility. Using big data analytics, the decisions were supported with the additional set of knowledge. Steps in the framework were easy and added value with visibility. The required set of changes is explained below.

We discovered a few limitations and disadvantages not specific to the framework but related to the experiment. They are related to the Turkish language as it is difficult to work with. Unlike another language, it is argued if the said word or phrase reflects the positive or negative sentiment and later it comes down to the context since it is the important aspect here. Turkish is more difficult than the rest of the languages when trying to identify the root word to opt for, tense, and grammar. Consequently, the same word can have different meanings or interpretations. For instance, the word “helwa” can refer to sweet or candy and the adjective can refer to taste, something which tastes like sugar or nice. In our case, we can never know if the intention to use the word refer to an item or taste, or probably someone like the item, or its an expression.

Though there was complexity, we have successfully extracted the valuable insight with the help of analysis based on social media. We merged the results obtained from several analyses such that we can gain out-of-the-ordinary visibility on insight based on which decisions can be made. The social media data including the post and comments helped us to understand the relationship between attributes by executing the analytics on relational data. We might have not realized the impact of social media on customer engagement with purchase patterns. This helped to identify the reason for the spike in sales overtime after the online promotion. At last, we never are able to have the user perspective & sentiment as a basis for the decision and try to understand how the user thinks or feels about service, promotion, or items.

4. Result –

Based on the outcome of our experiment, we have realized observations and improvements related to the framework. Firstly, as defined in the framework development previously, the framework enables the use of conceptual approaches for executing big data and analytics to brace the decision-making process. It was not all about big data analytics and tools. Some of them

already exist and the corresponding offering is increasing over time which makes it impossible to have the list of feasible solutions. Various solutions are not intended to be released as big data tools could be utilized for other usages. As per our experiment, we have used “Weka” which is not a big data analytics tool instead this is an open-source machine learning tool that brings ease to learning, is useful and best fit for analysis. Thus, it was not comprised of a big data tool. It was moreover used in the experiment to generate further knowledge and visibility.

We discovered that the visibility was the intention to be used in the intellect phase in the process of discovery for data and in the alternative phase at the time of assessing the possible direction of action, was found to be serving as analysis on their own for the planning phase based on which the important understanding can be captured. Therefore, we can foresee the valuable relationship well in advance which can be included in the added analysis or the results of the analysis can be used for a comprehensive study. This was required to be included in the plan phase of the framework. A similar goes for the statistics which was initially planned for the intellect phase at the time of data findings. Nevertheless, it was identified that the statistical analysis could be utilized in the Plan phase in the form of analysis or can be used with other analyses using integration. Consequently, in the Analysis of data step, Prediction & descriptive analytics can be utilized for value and to add insights.

It was later discovered that there is no need for two steps in two different phases in order to analyze and evaluate the end results related to big data analytics instead they can merge in one step for refinement and better understanding. The evaluation and analysis must be carried out on the analytics which relate to the decision domain instead of the probable direction of action. This is because of the fact once analytics are completed, the generated outcomes are analyzed to obtain the insight which can enable more knowledge and helps at the time of decision making. It's not always possible to have the best scenario where the probable direction of action is defined first and later start evaluating each of them to retrieve the best selection for the decision.

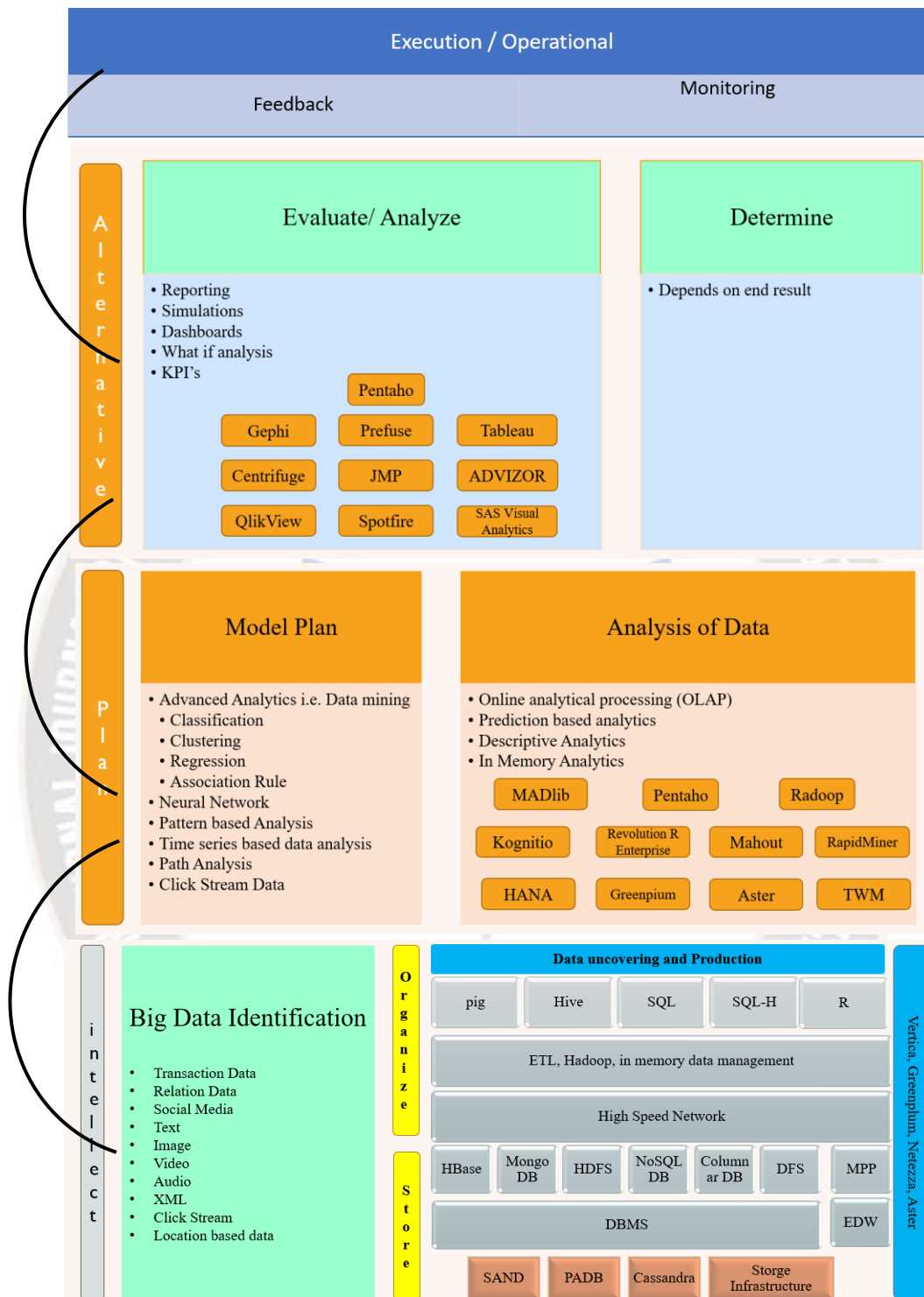


Figure 5 – Updated B-DAD Framework

Big data analytics has diverged from that of traditional ones in terms of being more structured, ability to have a method in order to detect the business problem, capturing the data, and analyzing it instead of opting for an unstructured method where firstly data is gathered

followed by analysis i.e. capture the max information out of the data and later trying figure out how to support business decision using the extraordinary analysis. We don't have to know about the decisions which are about to happen in advance and such a case is required to be

supported by the framework. Consequently, in a single step analyzing and evaluating must be merged using which one or both can be applied to the end result of the analysis.

The framework is not a tool for generating adequate informed and structured decisions instead it's moreover conceptual when it comes to the mapping tools with the decision-making process. Besides the framework was more of a flexible process instead of being sequential. We must be able to navigate back forth within the steps of the phases. As we stated previously that big data analytics is not fully predictable so one can opt for an unstructured method where no one is aware of the decisions in advance and in such a case it will be more meaningful if one wants to move back to the prior stage. For example – while performing the analysis for the plan phase, we might find ourselves in a situation where more diverse data increases the knowledge of our model. Therefore, a framework must enable the ability to move back and forth between the steps.

The obtained findings of the experiment were inherited in the updated framework. The newly checked B-DAD framework is displayed in figure 5. No changes are made to the intellect phase. Though the plan phase only consists of a model plan and analysis of data steps. Descriptive analysis was a new addition to the analysis of data step. The analysis and evaluation are combined into one step located in the alternative phase. The step doesn't reflect analyzing or evaluating the probable direction of action instead analyzing the previous analysis and assesses the impact on the decisions. The bilateral arrow was included in the middle of phases to reflect the ability to move back and forth all around the framework.

5. Conclusion –

We have inspected the innovational aspect of big data in this research which has captured lots of attention due to its offering and opportunities. The current era of information technology consists of high-velocity data which is produced on daily basis and they hold the hidden information inside of it and it is required to have it extracted. Thus, utilizing big data analytics business can be leveraged resulting in enhanced business decision making with the adoption of advanced analytics method on big data and identifying the hidden and valuable knowledge or insights.

Enforcing analytics on big data can help to extract valuable information and it can be utilized to intensify decision making and help the informed decision. We have implied the data science methodology to discover the solution to the problem of how to integrate big data analytics into the decision-making process? This research contributed to the development and testing of the B-DAD framework enabling us to know the decision-making process helped by big data analytics and several tools or methods.

The research-based on data science must contribute towards the knowledge base along with the application in corresponding environments. Our research adds the contribution to the knowledge base by capturing various theories, frameworks, and methods along with opted data analysis methods from research in the past to construct the reliable B-DAD framework. Hence, it creates a cluster of so many aspects & integrates them to transform the one framework. The testing of such a framework in real-life scenarios adds more diligence and reinforces the framework evaluation.

The research adds contributions to the environment. The B-DAD framework can be utilized both in research and in the industry or organization. The prime goal of the decision-makers or stakeholders in the organization is to intensify decision-making and capture the hidden knowledge with facts and insights. This research has produced the B-DAD framework for organizations by showcasing how it can be integrated with the big data analytics for each phase in the decision-making process to generate enhanced and impactful decisions.

It's not easy to intensify decision-making and capturing of hidden insights or knowledge using big data analytics. One of the key challenges we faced in our case was access to big data.

We trust in the fact that in the time of overflow of information big data analytics will be a huge help to foresee the hidden insights and by offering the advantages to the decision-makers in so many aspects. Once it is explored and utilized efficiently, big data analytics have a lot to offer at the scientific and technological levels.

References –

- [1]. Harrison, E. Frank. "A process perspective on strategic decision making." *Management decision* (1996).

-
- [2]. Nutt, Paul C. "Investigating the success of decision making processes." *Journal of Management Studies* 45.2 (2008): 425-455.
- [3]. Sagioglu, Seref, and Duygu Sinanc. "Big data: A review." *2013 international conference on collaboration technologies and systems (CTS)*. IEEE, 2013.
- [4]. Hurwitz, Judith, et al. "Big Data." *New York* (2013).
- [5]. Labrinidis, Alexandros, and Hosagrahar V. Jagadish. "Challenges and opportunities with big data." *Proceedings of the VLDB Endowment* 5.12 (2012): 2032-2033.
- [6]. Reichert, Ramón. *Big data*. Bielefeld: transcript Verlag, 2014.
- [7]. George, Gerard, Martine R. Haas, and Alex Pentland. "Big data and management." *Academy of management Journal* 57.2 (2014): 321-326.
- [8]. Madden, Sam. "From databases to big data." *IEEE Internet Computing* 16.3 (2012): 4-6.
- [9]. Davenport, Thomas H., and Jill Dyché. "Big data in big companies." *International Institute for Analytics* 3.1-31 (2013).
- [10]. Che, Dunren, Mejdil Safran, and Zhiyong Peng. "From big data to big data mining: challenges, issues, and opportunities." *International conference on database systems for advanced applications*. Springer, Berlin, Heidelberg, 2013.
- [11]. Russom, Philip. "Big data analytics." *TDWI best practices report, fourth quarter* 19.4 (2011): 1-34.
- [12]. Kambatla, Karthik, et al. "Trends in big data analytics." *Journal of parallel and distributed computing* 74.7 (2014): 2561-2573.
- [13]. Fisher, Danyel, et al. "Interactions with big data analytics." *interactions* 19.3 (2012): 50-59.
- [14]. Silva, Bhagya Nathali, Muhammad Diyan, and Kijun Han. "Big data analytics." *Deep Learning: Convergence to Big Data Analytics*. Springer, Singapore, 2019. 13-30.
- [15]. Duan, Lian, and Ye Xiong. "Big data analytics and business analytics." *Journal of Management Analytics* 2.1 (2015): 1-21.
- [16]. Vassakis, Konstantinos, Emmanuel Petrakis, and Ioannis Kopanakis. "Big data analytics: applications, prospects and challenges." *Mobile big data*. Springer, Cham, 2018. 3-20.
- [17]. Bumblauskas, Daniel, et al. "Big data analytics: transforming data to action." *Business Process Management Journal* (2017).
- [18]. Santos, Maribel Yasmina, et al. "A big data analytics architecture for industry 4.0." *World Conference on Information Systems and Technologies*. Springer, Cham, 2017.
- [19]. Shi, Yong. *Advances in big data analytics: theory, algorithms and practices*. Springer Nature, 2022.