IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Impact Factor: 6.252

Volume 2, Issue 1, November 2022

Predictive Analytics Accomplished by the Utilization of Social Big Data and Machine Learning

Mahadevi Somnath Namose[1] and Dr. Tryambak Hiwarkar[2]

Research Scholar, Department of Computer Science[1]

Professor, Department of Computer Science[2]

Sardar Patel University, Bhopal, MP, India

Abstract: *The ever-increasing quality and quantity of data generated from day-to-day business operations, in conjunction with the continuously imported related social data, has rendered the traditional statistical approaches inadequate to deal with such data floods. This inadequacy can be attributed to the fact that traditional statistical methods were developed before the advent of the internet. Because of this, academics have been compelled to design and develop advanced and complex analytics that may be incorporated to acquire useful insights that are beneficial to the commercial area. This chapter shines a light on fundamental characteristics that are the building blocks for social big data analytics and lays out those building blocks. In particular, the importance of predictive analytics within the scope of SBD is examined, and this analysis is bolstered by the presentation of a framework for SBD predictive analytics. After that, a number of different predictive analytical algorithms are discussed, along with their implementation in a number of essential applications, top-tier tools, and APIs. Experiments are presented alongside a case study that demonstrates how predictive analytics may be used to social data. This is done to demonstrate the significance and practicality of predictive analytics.*

Keywords: ML, Predictive Learning

I. INTRODUCTION

The prevalence of SBD provides businesses with the chance to maximise their usage of the vast amounts of information that is available to them in order to boost their revenues. As a result, there is an urgent need to capture, load, store, process, analyse, transform, interpret, and visualise a variety of social datasets in order to produce relevant insights that are specific to the domain in which an application is being used. In this context, organisations employ advanced social data analytics when building efficient marketing strategies and strive to utilise the interactive quality of online social services. This is done in order to compete in today's global marketplace. Therefore, in order to generate the necessary connection with their clients, businesses make use of a variety of contemporary modes of communication in order to entice customers and site visitors to their various online social platforms.

As a consequence of this, it is essential for businesses to conduct social media content analysis on their customers and categorise those customers into relevant groups based, for instance, on the subjects that they find interesting in order to convey the appropriate message to the appropriate group of customers. This final goal ought to make use of technology solutions that have the ability to infer the meaning of social content at both the user level and the post level.

An approach that is based on ontologies has been created in order to semantically analyses the social data at two levels, namely the entity level and the domain level. As a result, the semantics of textual data have been extracted, and the domain of data has been defined. This strategy has shown to be effective in bolstering the output of semantics analytical suppliers and providing support for their efforts (i.e. IBM Watson NLU). Nevertheless, there is a requirement to broaden the scope of this effort in order to develop a platform that can automatically categories and forecast the user's domain of interest. In particular, the current approaches to topic extraction, modelling, and classification rely on statistical bag-of-words techniques such as Latent Dirichlet Allocation (LDA) [6]. These methods have a number of drawbacks, some of which are as follows: I the number of designated topics is predetermined and must be known before the analysis [7]; (ii) the topics mined by these models do not take into account the temporal aspects [8]; (iii) these models are considered to be monolingual topic models, which means that they do not differentiate idioms of the same language [9]; and (iv) these models are unable to infer high.

IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 1, November 2022

Impact Factor: 6.252

II. AN ORGANIZATIONAL STRUCTURE FOR PREDICTIVE MODELLING OF SOCIAL BIG DATA

It is not a simple task to construct an effective framework that can analyse SBD and deduce its value. In point of fact, it is a unified process that calls for giving careful thought to each individual step in order to guarantee that the results of each step will be beneficial to the ones that come after them. When it comes to doing predictive modelling, particularly those that are related to SBD, there is a standard paradigm that is typically followed. This is the case despite the fact that the specific analysis could be different depending on the characteristics of each issue that is being tackled.

SBD is an essential platform for firms to acquire an added value that is beneficial to the business domain. This can be accomplished by conducting market research, listening to the Voice of the Customer (VoC), and conducting sentiment analysis in order to input data into applications for Business Intelligence [15]. For this reason, a methodology is required to infer the trustworthiness of unstructured data from various SBD sources such as social media networks, news agencies, and web logs, and to store a collection of trustworthy unstructured data utilising the solutions that are already available for data warehousing. When transferring data from unstructured sources like SBD to structured ones like data warehouses, another important duty is to semantically enrich the data by adding structure to it and then imposing that structure on the data. The discovery of the semantics of social data will improve the quality and accuracy of the data that is kept in data warehouses, which will have a significant impact on the process of decision-making as well as the quality of reports that are extracted. The implementation of semantic web technology provides a solution to the problem of ambiguous data and generates metadata, both of which make it easier to comprehend and correctly interpret data that is related to one another. Ontology, in the meantime, is utilised to define and collect concepts that are semantically connected, as well as relations between concepts [16]. This could be accomplished in particular through the utilisation of pre-existing ontologies (or the creation of new ones) that make the extraction of data semantics easier.

Before diving into the analysis, one of the most important phases in this paradigm is to clearly describe and articulate the nature of the problem at hand. According to John Dewey, who was both a psychologist and a philosopher, "a problem well-stated is half-solved" [11].

The inclusion of mathematical and advanced statistical methods into the construction of a model that can be used to forecast the future and identify predictive patterns from previous datasets is one definition of what is known as "predictive analytics." [17]. The nature of the desired goal or the embedded architecture of the predictive model can both be used as criteria for classifying different kinds of predictive models. To put it another way, predictive models can be classified as either regression models (which produce continuous output) or categorization models (nominal or binary output). The predictive models can also be categorised according to their technical implementations, and this categorization is possible based on the architecture of the embedding model.

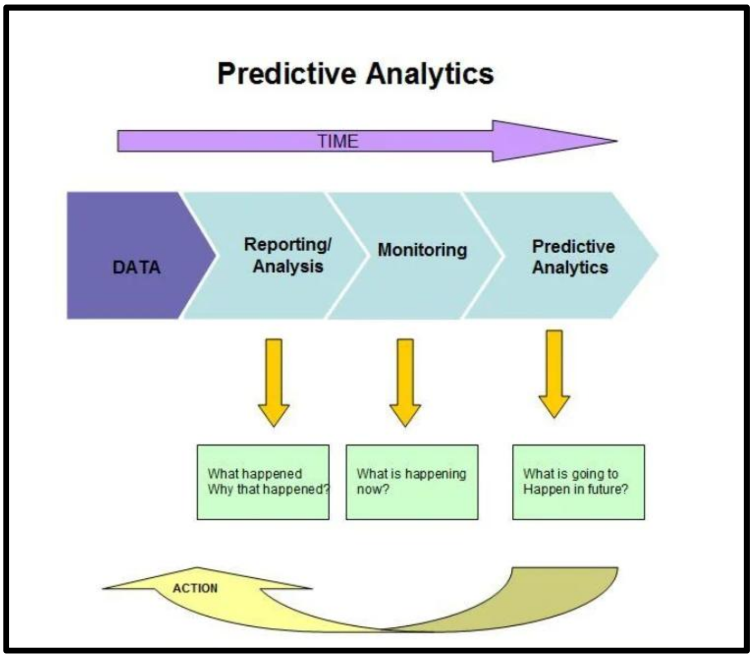

Figure 1: Predictive Analytics

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
Volume 2, Issue 1, November 2022
Impact Factor: 6.252

Linear models, Decision trees (often referred to as Classification and Regression Trees or CART), Neural networks, Support vector machines, Cluster models, and others are included in this category.

2.1 Definition of Extrapolative Analytics

Predictive analytics conversations often center around Big Data. In modern database management systems, Predictive Analytics is not a binary concept nor a discrete feature. Instead, it is a set of resources for analysing data and applying statistical principles. Predictive analytics is thus the result of the integration of Big Data with BI.

The goal of Predictive Analytics is to use past data and analysis to make predictions about the future. When numerous datasets are combined, it becomes possible to draw connections between previously unrelated areas, processes, and types of Big Data.

2.2 How Do Big Data and Predictive Analytics Interact?

The fundamental idea behind Predictive Analytics is that most phenomena can be "modelled." It is assumed here that some data parameters cause changes in other data parameters, and that as one data parameter changes (the cause), the other data parameters will also change (the effect) (effect).

The following is a high-level outline for implementing Predictive Analytics in organizations:

The process of collecting and compiling massive amounts of historical data.

Various statistical methods, including regression models, are employed to examine the information.

We then utilise these evaluations to speculate on what the future might hold.

These projections of the future can then be applied to aid in selecting choices, enhancing corporate procedures, cutting down on waste, and other similar activities.

2.3 Parallels between Big Data and Predictive Analysis

Big Data	Predictive Analytics
Big Data is concerned with the purification and interpretation of large amounts of data and can be applied to many business activities.	Predictive Analytics is a technique for predicting business and market events.
Big Data engines include built-in machine learning libraries, but integrating AI is still an R&D work for Data Engineers.	It deals with a platform based on mathematical calculations and probability.
The amount of data and the speed with which it is processed are enormous. It's not recommended to use Big Data platforms for small amounts of data because their performance is exponential.	The amount of data and the speed with which it is processed are both on the medium side. In terms of models and algorithms, very large and very small data sets can contribute to inaccurate predictions and discoveries.
Big Data comes with D3.js, Tableau, infogram, and other backend technology imports for Dashboards and Visualizations.	Predictive Analytics tools have built-in reporting integrations, such as Microsoft BI tools. So there's no need to get it from the source or a third-party seller.
The level of advancement for Big Data is high.	The level of advancement for Big Data is medium.
Big Data is a trendy topic right now. Everyone in the market wants to get into the Big Data business.	Predictive Analytics is popular, but it isn't the same as Big Data. It is dependent on the use cases and the type of organization that is putting it in place.
It's a tool for making data-driven decisions.	It is used in the assessment of risk and the forecasting of future results.
It's a best practice for handling large amounts of data.	It's a best practice for predicting the future with data

The Increasing Necessity of Employing Predictive Analytics on Massive Data Sets:

Large-scale strategic evaluations and assessments, as well as more nuanced, operational-level decisions, can all benefit from Big Data Predictive Analysis. Companies employ this method for a wide variety of purposes, including problem solving, opportunity discovery, and decision making.

In addition, organizations utilize these analytics to learn more about their customers, products, and partners, as well as to investigate connected data in order to spot risks and possibilities.

The following is a possible situation: A presentation to upper management on the results of your most recent data analysis study is in the works. New sales can be generated from your data, content for RFPs can be created, and interest in your marketing can be piqued. As the information is kept on the cloud, it can be quickly accessed and understood. Moreover, you have access to a dashboard with images demonstrating the amazing potential of the dataset. You're all set to make waves.

The question, "How will these data alter in the future?" comes a few minutes into the presentation from a high-ranking executive. One of the executives says, "How do we know this dashboard is telling us everything?" before you can even finish your thought.

You take a moment to collect yourself, stunned. You can confidently present this information to the executives because your QA staff has been deeply involved in testing for months. But is there any way to predict whether or not this information will shift in the future? The data in the dashboard and the dataset are, in reality, merely frozen moments. No one can see into the future and know what will happen.

Yet suppose you did succeed in approaching a suitable distance. Today's brands need more than simply real-time analytics. Companies in a wide variety of sectors are turning to Big Data Predictive Analytics to help them meet their three overarching objectives: lowering future risk, increasing sales and customer happiness, and streamlining operations.

It is crucial to be familiar with the current uses of Big Data and Predictive Analytics, as well as how they relate to the cloud and the science behind them.

Big Data Predictive Analytics has the ability to enhance decision support and operations across all industries, including those dealing with resource allocation and cost management.

2.4 The Five Best Models for Predictive Analytics

Numerous varieties of predictive analytics models can be utilized as a result of technological, data mining, and machine learning tool advancements. While developers use a variety of different predictive analytics models to accomplish their goals, the ones listed above are the ones most commonly employed. First, let's take a quick look at some of the most important predictive models:

1. Classification Model

Compared to other types of predictive analytics models, categorization models are the simplest and easiest to implement. Based on their analysis of the data, these models classify information into several buckets.

Classification models offer the "yes" and "no" answers necessary for a thorough investigation. For instance, with the use of such models, we may better understand:

In other words, is the user asking for the right thing?

Is there any information on whether or if the vaccination for a certain illness is commercially available.

Is there going to be a price increase for the company's shares on the market?

The greatest option for finding definitive answers is the categorization model of predictive modelling. The classification models' versatility and capacity to retrain with fresh data and provide in-depth research to address business concerns have led to their widespread use.

2. Clustering Model

Given that a dataset's components are likely to share characteristics, a clustering model may be used to organise the dataset's components into meaningful categories. For efficient advertising plans, this predictive analytics technique is your best bet for breaking the data down into subsets based on shared traits.

It's a huge hassle, for instance, for an eCommerce shop to sift through thousands of data records and create an efficient strategy for marketing campaigns. At the same time, the clustering algorithm may group comparable consumers based on shared traits and purchase history, making it easy to find the interested ones to contact.

Predictive clustering modelling may be further broken down into "hard" and "soft" clustering. The data point's membership in the cluster may be determined with the use of hard clustering. Soft clustering, on the other hand, aids in assigning the data likelihood of the data point when entering the group.

3. Forecast Model

Metric value forecasting is important to the predictive analytics forecast model, which is used to examine potential outcomes. With the aid of this predictive analytics methodology, firms may make educated guesses about the monetary worth of new data by analysing past data.

The fact that the forecast prediction model takes into account various input characteristics all at once is its greatest strength. As a result, the forecast model is widely adopted as a standard tool for predictive analytics. If a clothes manufacturer, for instance, needs to forecast manufacturing stock for the future month, the model will take into account all the aspects that might affect output, such as: Is there any upcoming festival? Can you tell me what the weather will be like next month?

You may use the prediction model everywhere that the corresponding numerical data exists in the past. A factory, for instance, can estimate how many units can be cranked out every hour. Concurrently, a monthly policy's potential customer base can be estimated by an insurance firm.

4. **Time Series Model**

When time is seen as an input parameter in predictive analytics, the time series model is the optimal option. This forecasting approach constructs the numerical metric using data points taken from the past, allowing for accurate forecasting of future trends.

The time series predictive model is the answer if a company needs to forecast how its operations or output will evolve over a given period of time. The process and interdependence of different business factors are discovered using the tried-and-true conventional technique in this model. Additionally, it takes into account the ad hoc elements and hazards that might have a significant long-term impact on the company's operations.

In terms of applications, this predictive analytics model may be used to estimate the volume of calls coming into a certain call centre over the course of the following week. It can also predict how many people will be admitted to the hospital the following week.

It is not true that development must follow a straight line or remain constant throughout time. Because of this, the time series model aids in achieving more consistent exponential growth and trend alignment.

III. CONCLUSION

The eras of Big Data and Predictive Analytics will continue on for some time. The most important advantage of Big Data Predictive Analytics is the value it brings to businesses in terms of dollars and cents. It also facilitates enhanced comprehension, deliberation, and automated procedure. There is also a "paradigm shift" in the way that analysis is conducted. Here, we see the transition from purely descriptive analytics to Predictive Analytics.

REFERENCES

[1]. Chan, K.Y., et al., Affective design using machine learning: a survey and its prospect of conjoining big data. International Journal of Computer Integrated Manufacturing, 2018: p. 1-25.

[2]. Abu-Salih, B., et al., CredSaT: Credibility ranking of users in big social data incorporating semantic analysis and temporal factor. Journal of Information Science, 2018. **45**(2): p. 259-280.

[3]. Abu-Salih, B., P. Wongthongtham, and K.Y. Chan, Twitter mining for ontology-based domain discovery incorporating machine learning. Journal of Knowledge Management, 2018. **22**(5): p. 949-981.

[4]. Wongthongtham, P. and B. Abu-Salih. Ontology and trust based data warehouse in new generation of business intelligence: State-of-the-art, challenges, and opportunities. inIndustrial Informatics (INDIN), 2015 IEEE 13th

DOI: 10.48175/IJARSCT-7570

IJARSCT

ISSN (Online) 2581-9429

IJARSCT

Impact Factor: 6.252

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

Volume 2, Issue 1, November 2022

International Conference on. 2015. IEEE.

[5]. Blei, D.M., A.Y. Ng, and M.I. Jordan, 10.1162/jmlr.2003.3.4-5.993. CrossRef Listing of Deleted DOIs, 2000. **1**(4-5): p. 993-1022.

[6]. Zhang, W., Y. Cui, and T. Yoshida, En-LDA: An Novel Approach to Automatic Bug Report Assignment with Entropy Optimized Latent Dirichlet Allocation. Entropy, 2017. **19**(5): p. 173.

[7]. Zoghbi, S., I. Vulić, and M.-F. Moens, Latent Dirichlet allocation for linking user-generated content and e-commerce data. Information Sciences, 2016. **367-368**: p. 573-599.

[8]. Li, C., et al. Topic Modeling for Short Texts with Auxiliary Word Embeddings. In Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval. 2016. ACM.