Predictive and Prescriptive Analytics in Big Data Era

Prachi Deshpande

Abstract The notion of data analytics and its real-time application is important in the big data era owing to the voluminous data generation. Predictive and prescriptive analytics provides the future trends from the available data effectively. This will help to decide the usability of the data and thereby its retention for future applications. The paper reports the predictive and prescriptive analytics notion in big data regime, various platforms for its analysis, and the future research directions.

Keywords Analytics · Big data · Data science

1 Introduction and Driving Force

The present era is of information explosion with Internet of Things (IoT) and very soon the Internet of Everything (IoE) will ensure the connectivity of millions of gadgets and devices to the Internet [1]. According to [2], today 2.5 Exabyte (10^{18}) of data is created every day and 90% of the data is created in the last 2 years. Such an enormous data, generated from various user-driven applications, paved the way for many new avenues across different fields starting from household to business applications. Academicians and industry referred such a huge data with a notion of big data.

The term *big data* was first announced by *Cox and Ellsworth* in 1997 [3] in the context of data handling and processing capacity of available computers. Soon the term *big data* became the catchphrase for the academicians and industry. Big data can be characterized by five Vs as shown in Fig. 1.

Increased social media applications have increased the volume as well as the variety of data generated. It is expected to have 5 million mobile phone users in the world. The biggest challenge in front of big data is its processing and veracity. Only real-time processing of such a voluminous data will help to capitalize the usability

P. Deshpande (✉)
Nanded, India
e-mail: deprachi3@gmail.com

of the data in the decision-making process. Value of the generated data is another
important concern of big data. To increase the maximum gain out of the generated
data, its analytics in different forms are required.

1.1 Classification of Data in Big Data Era

The data generated from different sources may differ in terms of its scale, distribution,
domain, representation, and its density. Figure 2 depicts the classification of data in
big data era.

Data science technology had defined, categorized, and processed such a data in
efficient manner. Data science is an interdisciplinary field, which extracts meaningful
information from huge data by using scientific techniques, methods, and processes.
With the advent of data science technology, it becomes easy to predict the behavior
of users. This aspect is known as data analytics. The increased use of social media
has given rise to the big data in terms of online social network data.

Mobile and IoT data is another source of big data. With the emergence of 5G
technology, the analysis of such data becomes an important concern for data scien-
tist/analytics. With the advent of the concept of smart cities, the online social media
data will very soon consist of geographical data in real-time mode. This will shift
the research paradigm toward building of 3D geographic database of the locations.
This advancement will shift the data science to a new direction of spatiotemporal
data, the data with many features for selection and analysis. The spatial big data may
consist of vector, raster, and networked data. The advancement in the spatiotemporal
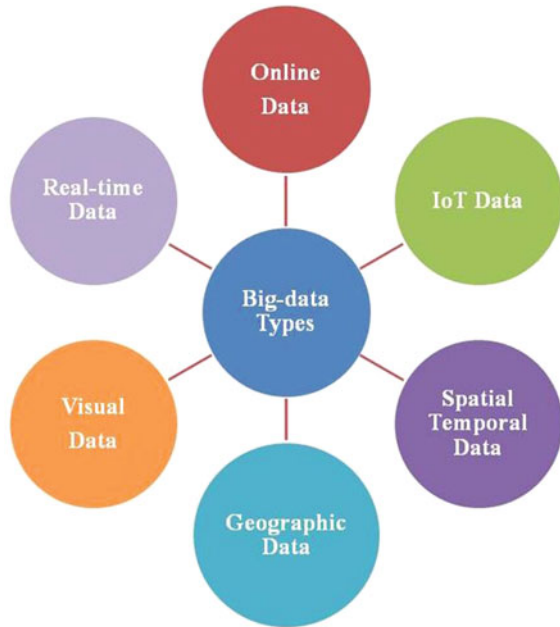
Fig. 2 Data types in big data
Era

Table 1 Normal data and big data

Parameter	Normal data	Big data
Volume	In Gigabytes (GB)	In Tera (TB) or Peta (PB) bytes
Data management	By RDBMS	Hadoop, MapReduce, NoSQL, and high-performance computing
Growth interval	Measured in hours or daily basis	No such fixed scale for measurement
Data format	Structured	Structured, semi-structured, or unstructured
Data access	GUI-based interactive	Batch processing or real-time system

data will give rise to the notion of real-time streaming data due to the real-time IoT-based day-to-day applications. Further, increased use of multimedia data in social networking, especially image data arises as a new source of big data. The storage, analysis, and retrieval of such data will be a challenge in the coming era (Table 1).
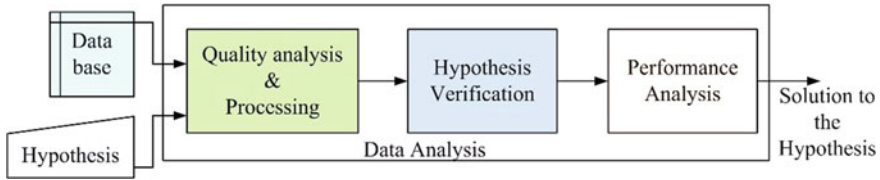
Fig. 3 Data analysis in big data Era

1.2 Data Processing in Big Data Regime

The problem of voluminous data processing has been solved by the data science in the last decade. The data science used various tools for the analytics of such a huge data. Figure 3 depicts the block diagram of a typical data processing architecture.

However, *value* of the gathered data will play a big role in next-generation communication technology. The principle issue in big data is its storage and processing. The data storage aspect is resolved by the advent of cloud-based distributive computing. The availability of the voluminous data from various sources poses a challenge of its category-wise processing and further analysis to decide its usability (value). To decide the value of the gathered data and to ripe maximum output from it, the notion of predictive and prescriptive analysis will be useful.

Section 2 describes the fundamentals of prescriptive and predictive analytics. The upcoming research challenges and the open issues in the prescriptive and predictive analytics are described in Sect. 3, and the paper is concluded in Sect. 4.

2 The Predictive and Prescriptive Analytics in Big Data

2.1 The Predictive Analytics

Data analytics has a wide range of methodologies and techniques to analyze the available data. They are broadly classified as descriptive, diagnostic, predictive, and prescriptive analytics [4]. Figure 3 shows the evolution of the predictive and prescriptive analytics.

The descriptive analytics is generally performed at the initial stage of data accumulation. It helps out to get a reasonable insight into the nature and pattern of the data. The descriptive analytics mainly concentrated on "*what*" with the help of classification, clustering, and segmentation of the data.

Once it is understood "*what*", the next stage is to concentrate on "why". This aspect of analytics is known as diagnostic analytics, which is carried out with the help of machine learning algorithms. Both the descriptive analytics and diagnostic analytics are largely dependent on the incidences, which have already occurred in the past. However, these two approaches have limitation to predict the futuristic

behavior of the user and the cause-action strategy to deal with it. Here comes the need of prediction-based forecast and the redial action (prescription of the pattern) for the user behavior based on available data. Predictive analytics uses machine learning algorithms and statistical analysis techniques to analyze current and historical data to make predictions about future trends, behavior, and activity.

2.2 The Prescriptive Analytics

The business opportunities are very stringent, and hence the entrepreneurs always wish to capitalize the available opportunities. However, having the insight on previous behavioral patterns and futuristic trend forecast will not be sufficient to take advantage of business opportunities. There is a trade-off between the insights generated and optimal operational courses of actions [5]. Hence, it is the need of the hour to utilize the analytics to transform information into valuable data (insights) so as to act upon them effectively to meet their objectives [6–8]. Prescriptive analytics helps to overcome this gap and is considered as the next frontier in the business analytics [9].

Prescriptive analytics is related to both descriptive and predictive analytics. While descriptive analytics aims to provide insight into what has happened and predictive analytics helps to model and forecast what might happen, prescriptive analytics seeks to determine the best solution or outcome among various choices, given the known parameters.

It provides organizations with adaptive, automated, and time-reliant sequences of operational actions. It answers the questions like "What", "Why", and "when" and very importantly "how" should be done [10]. However, in the preview of big data, analytics engine must be dynamic enough to provide the best possible cause–effect relationship for the decision-makers in the enterprise. This requirement can be met by the virtue of an adaptive mechanism, which will best utilize the predictive and prescriptive analytics at a given time of instance. Figure 4 shows the conceptual block diagram of an adaptive prescriptive and predictive analytics mechanism (Table 2).

Table 2 Predictive versus prescriptive analytics

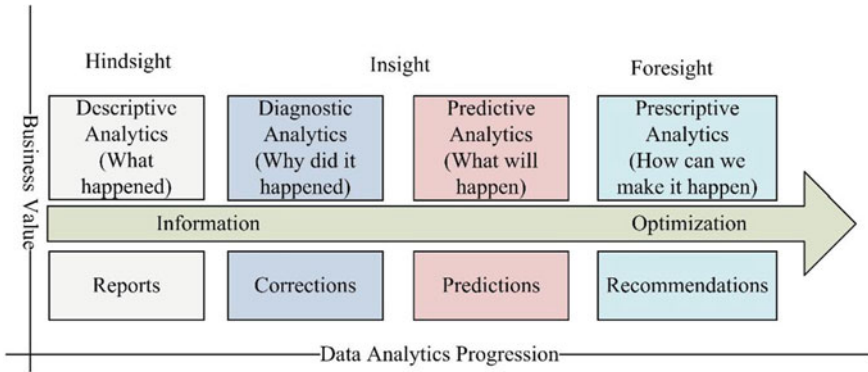| Parameter	Predictive analytics	Prescriptive analytics
Nature of outcome	Provides only a prediction of outcome	Provides deterministic outcome
Base data requirement	Historical and present data	Historical and present data
Type of insights	Projections of various outcomes	Ascertainment of a set of outcomes
Processing of big data	Machine learning algorithms	Artificial intelligence
Decision-making process based on	Descriptive analytics	Both descriptive and predictive analytics

Fig. 4 The predictive and prescriptive analytics

2.3 The Prescriptive Analytics Platforms

There are several commercial and open-source platforms available for the predictive and prescriptive analytics. This section briefly reviews the various available tools and their features. A beginner may use one of these tools for the research work (Tables 3 and 4).

Table 3 Predictive analytics tools

Name	Technology	Key features	Availability
RapidMiner studio	Machine learning	Visual workflow design and unified platform	Proprietary
KNIME platform	Modular data pipelining	Churn analysis, credit scoring, and social media sentiment analysis	Open source
IBM predictive analytics	Statistical analysis	Hands-on predictive analytics and direct deployment into business processes	Proprietary
Minitab	Statistical analysis	Smart data import and automatic graph updating	Proprietary
TIBCO spotfire	Data mining	Location analysis, big data analysis, and data wrangling	Proprietary
DataRobot	Machine learning	Drag and drop dataset and machine learning automation	Proprietary

Table 4 Prescriptive analytics tools

Name	Technology	Key features	Availability
Ayata	Machine learning and operation research	Adapted and automated, integrated prediction and prescription, and secure SaaS delivery model	Proprietary
AIMMS	Web-based solutions	Customized saluting, on-site or cloud-based support, and virtual solution environment	Proprietary
NGData	Personalization and engine software	Adaptive and just in time operation	Proprietary
LIONoso	Machine learning	Healthcare solutions, R&D facility, learning from data, and optimization	Proprietary
Profitect	Data warehousing	Intelligent data transform, multiple business technology support, and real-time customer feedback	Proprietary
IBM prescriptive analytics	Data mining and machine learning	Automate complex decisions and trade-offs, risk analysis, increase agility, and efficient resource management	Proprietary

3   The Open Research Issues of Predictive and Prescriptive Analytics

Big data is available in almost all aspects of life as data generation takes place in every form of human behavior. It benefits both research and industrial fields such as health care, financial services, and commercial recommendations. Slowly but steadily, business and enterprise houses are adopting the prescriptive analytics as the best alternative to drive valuable insights from the user-driven big data. This paves the way to explore many new avenues for research and development and subsequent employability aspect too. The trusted area for the futuristic research will be as follows:

A. *Smart City*

The notion of smart cities has paved the ways for great changes in fundamental facilities of transportation. Possessing certain features such as a large scale, diversified foreseeability, timeliness, city traffic data, sanitation, and security represent the scope of big data. To ensure hassle-free life, accurate decisions need to be initiated

in these areas, which in turn depend on the valuable insights available from the big data generated from the users.

B. *Healthcare Applications*

The notion of automated health assistance is now becoming prevalent all over the globe. With the advancement of IoT technologies, tiny body area sensors are deployed with the users, which continuously monitor the human vital signs. Mechanisms are under test, which generates automated alerts for the human health to the self as well as the medical service providers. In this process, a huge data is generated in terms of basic signals and images too. The storage, processing, and retrieval of such a huge data will be the next-generation research challenge. With the help of predictive and prescriptive analytics, pinpoint decisions may be initiated to provide better services to the users.

C. Content Recommendation

Increased use of multimedia content on social network has now acquired a day-to-day activity of human life. Users are now more cautious about expressing themselves. For this purpose, users seek the help of various online services for multimedia applications. With the help of predictive and prescriptive analytics, service providers may put forward the best alternatives for the users.

D. User Behavior Prediction

Many of the network big data predictions are based on data from online social networks. Big data is used for predictions based on ranked data such as elections, car performance, and other areas in business and politics. Using the predictive and prescriptive analytics, policy-makers may frame deceive policies regarding the various sociopolitical, economical, and educational strata.

E. Security and privacy of user data

With the global business digitalization driving the corporate world toward the risk of cyberattacks more than ever before. The predictive and prescriptive analytics of big data has the potential to offer protection against the cyberattacks. Based on the huge data, the security service providers can easily predict the behavior of eavesdroppers and prescribe the corrective measure against them. It is believed that the security analytics will emerge next generation.

F. Application of Artificial Intelligence

Along with big data, the notion of artificial intelligence (AI) and machine learning is storming the world these days. These are the set of technologies that empower connected machines and computers to learn, evolve, and improve upon their own learning by reiterating and consistently updating the data bank through recursive experiments and human intervention. These characteristics empower the machines to predict and thereon prescript the user trends and behavior (Fig. 5).

In the future, it is expected that the IoE and big data boundaries will vanish and it will become a synonym. In view of this, the scope of research in the area of
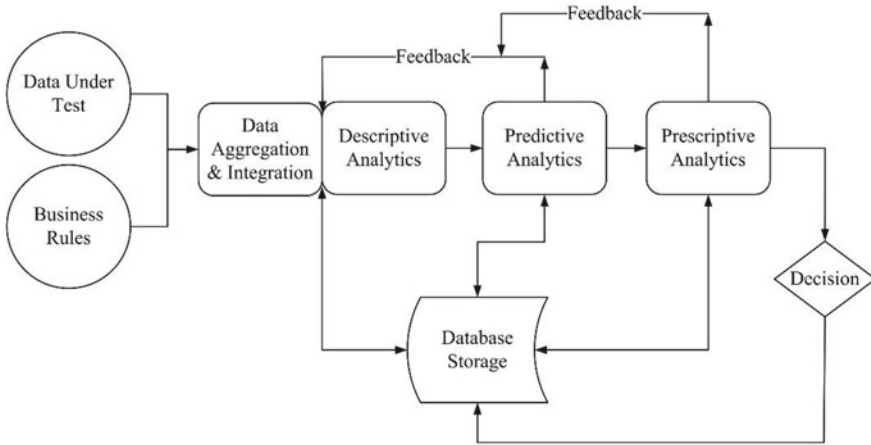
Fig. 5 Conceptual block diagram of a prescriptive analytics engine
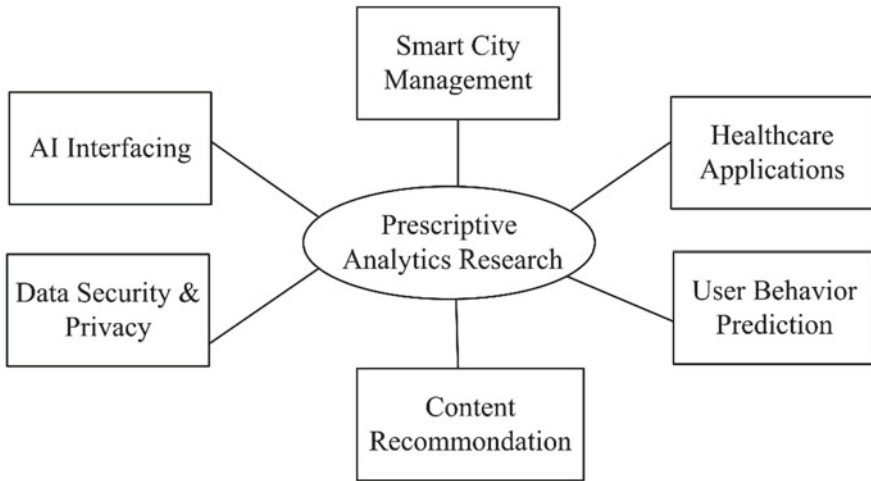

Fig. 6 Research directions in the prescriptive analytics

prescriptive analytics will accommodate all areas of engineering and technology, science, and humanities (Fig. 6).

4 Conclusions

The paper reports the predictive and prescriptive analytics in big data. It is believed that the next-generation data processing technology will be largely dominated by

the predictive and prescriptive analytics along with big data and IoT. The concept of IoE will be totally dependent on the predictive and prescriptive analytics. In future, research may be undertaken on the AI-based prescriptive and predictive analysis of big data.

References

1. Deshpande, P., Iyer, B.: Research directions in the internet of every things (IoET). In: International Conference on Computing, Communication and Automation (ICCCA), Noida-India, pp. 1353–1357, May 2017
2. Number of mobile phone users worldwide 2013–2019. https://www.statista.com/statistics/274 774/forecast-of-mobile-phone-users-worldwide
3. Cox, M., Ellsworth, D.: Application Controlled Demand Paging for Out-of-Core Visualization. NASA Ames Research Centre Report, Mountain View, CA (1997)
4. Hardoon, D., Shmueli, G.: Getting Started with Business Analytics: Insightful Decision Marking. CRC Press, Boca Raton, FL (2013)
5. Barga, R., Fontama, V., Tok, W.H.: Predictive Analytics with Microsoft Azure Machine Learning: Build and Deploy Actionable Solutions in Minutes. Apress (2014). ISBN-13 (pbk): 978-1-4842-1201-1 and ISBN-13 (electronic): 978-1-4842-1200-4
6. Baker, P., Gourley, B.: Data Divination: Big Data Strategies. Delmar Learning (2014). ISBN: 1305115082 9781305115088
7. Chen, H., Chiang, R.H., Storey, V.C.: Business intelligence and analytics: from big data to big impact. MIS Q. **36**(4), 1165–1188 (2012)
8. Kaisler, S.H., Espinosa, J.A., Armour, F., Money, W.H.: Advanced analytics-issues and challenges in a global environment. In: 2014 47th Hawaii International Conference on System Sciences (HICSS), pp. 729–738. IEEE (2014)
9. Evans, J.R., Lindner, C.H.: Business analytics: the next frontier for decision sciences. Decis. Line **43**(2), 4–6 (2012)
10. Delen, D., Demirkan, H.: Data, information and analytics as services. Decisi. Support Syst. **55**(1), 359–363 (2013)