ISSN 2278-3091

Volume 5, No.3, May - June 2016

International Journal of Advanced Trends in Computer Science and Engineering

Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse02532016.pdf

Predictive Analytics: Cloud Computing

Satish Kumar
Bahra University
Shimla Hills
chandelsatish672@gmail.com

Parul Sharma
Bahra University
Shimla Hills
parulgazta@gmail.com

Madhusudan
Himachal Pradesh University
Summerhill Shimla
msudan_07@outlook.com

***Abstract*: The increasing use of computing machine in our day to day lives, and our dependence on them has created a demand for better, reliable and efficient machines which can help in making our decisions better. Predictive analysis one of such field of computer science and data engineering which generally predicts some occurrence or probability of some event based upon its previous data. Predictive analytics uses data-mining techniques in order to make predictions about future events, and make recommendations based on these predictions. The process involves an analysis of historic data and based on that analysis to predict the future occurrences or events. There are some models designed for the same which involve rigorous data analysis, and are widely used in business for segmentation and decision making. The predictive models make it possible to make more right decisions, more quickly, and with less expense. They can provide support for human decisions, making them more efficient and effective, or in some cases, they can be used to automate an entire decision-making process. The Big Data phenomenon, ever-improving tools for data analysis, and a steady stream of demonstrated successes in new applications. In this study, we analyze data related to cloud computing, big data and application of machine learning techniques in data mining to predict the demand of cloud recourses by the users in future.**

Keywords: **Predictive Analytics, Big Data, Cloud Computing, Machine Learning**

1. INTRODUCTION

Predictive Analytics is the combination of two words predicts and analysis, but it works in the reverse order first analyze the raw data and then make prediction on the basis of that data [1]. It is the human nature to enquire and predict what the future hold. Predictive analytics uses data-mining techniques in order to make such predictions about future events, and make recommendations based on these predictions [2]. Predictive analytics encompasses a variety of statistical techniques from predictive modeling, machine learning, and data mining that analyze current and historical facts to make predictions about future. Predictive analytics is an area of data mining that deals with extracting information from data and using it to predict trends and behavior [18]. It establishes strong relationship between the data and identifies the interesting patterns to make the prediction. This process uses Machine learning technique and data mining tools supported by neural network which play an important role to identify the required patterns for forecasting about the future [19].Predictive analytics is used to determine the probable future outcome of an event or the likelihood of a situation occurring. It is the branch of data mining concerned with the prediction of future probabilities and trends. Predictive analytics is used to automatically analyze large amounts of data with different variables; it includes clustering, decision trees, market basket analysis, regression modeling, neural nets, genetic algorithms, text mining, hypothesis testing, decision analytics, and more [16]. The core element of predictive analytics is the ‘predictor’, a variable that can be measured for an individual or entity to predict its future behavior. For example, a credit card company may consider age, income, and credit history as predictors to determine the risk factor in issuing a credit card to an applicant. Predictive analytics combines business knowledge and statistical analytical techniques which, when applied to business data, produce insights [15]. These insights help organizations understand how people behave as customers, buyers, sellers, distributors, and so on. Predictive analytics play greater role to utilize the cloud recourses in a proper way that helps the cloud services provider to predict the demand of cloud recourses in advance.

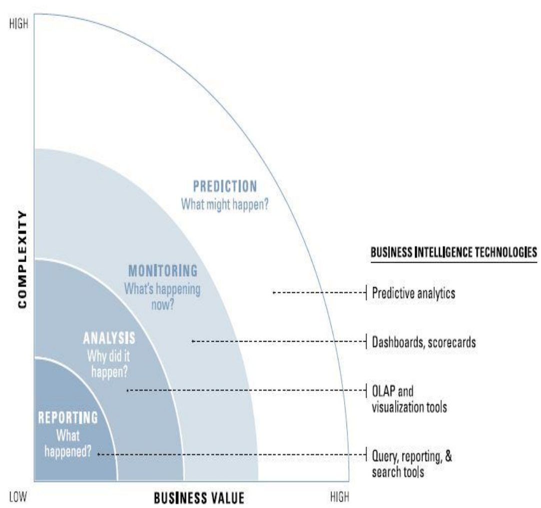

Figure 1. Predictive analytics

1.1 Reflections on Predictive Analytics

Predictive analytics is used to determine the probable future outcome of an event or the likelihood of a situation occurring. It is the branch of data mining concerned with the prediction of future probabilities and trends. Predictive analytics is used to automatically analyze large amounts of data with different variables; it includes clustering, decision trees, market basket analysis, regression modeling, neural nets, genetic algorithms, text mining, hypothesis testing, decision analytics, and more [1] [17]. Recent academic discourse on information systems speaks of a paradigm shift toward data-intensive computing and "big data"-based studies [22]. A new level of connectedness among peers creates a huge database by providing new ways for the dissemination and consumption of data and ever-easier means of collecting vast amounts of public data from various on- and offline resources, including posts, tweets, Web documents, and news feeds. Evolving data mining technologies and the increasing processing power of today's computers support the desire to appropriately analyze at least parts of today's growing public

data deluge in real time, and thus tackle the core question of what meaningful information can be derived through algorithmic analyses and what predictive value can be inferred in automated fashion from public data [21] [20].

1.2 Application of Predictive Analytics

- **Direct marketing and sales**: **-**Leads coming in from a company's website can be scored to determine the probability of a sale and to set the proper follow-up priority. Campaigns can be targeted to the candidates most likely to respond.
- **Customer relationships:** Customer characteristics and behavior are strongly predictive of attrition (e.g., mobile phone contracts and credit cards). Attrition or "churn" models help companies set strategies to reduce churn rates via communications and special offers.
- **Pricing optimization:** With sufficient data, the relationship between demand and price can be modeled for any product and then used to determine the best pricing strategy. Analytical pricing and revenue management are used extensively in the air travel,
- **Health outcomes**: Models connecting symptoms and treatments to outcomes are seeing wider use by providers. For example, a model can predict the likelihood that a patient presenting a certain set of symptoms is actually suffering a heart attack, helping ER staff determine treatment and urgency.
- **Insurance fraud**: Many types of fraud have predictable patterns and can be identified using statistical models for the purpose of prevention or for after-the-fact investigation and recovery.
- **Improper public benefits payments and fraud**: Health, welfare, unemployment, housing and other benefits are sometimes paid when they should not be, wasting taxpayers' money and making benefits less available to those who deserve them. Models similar to those used in insurance fraud help prevent and recover these losses.
- **Tax collections**: Likely cases of additional tax owed (due to non-filers, underreporting and inflated refunds) can be identified. The IRS and many state governments use revenue collection models and are continually improving them.
- **Predicting and preventing street crime, domestic abuse and terrorism**: In addition to link analysis techniques for investigating crimes, predictive models help determine high-risk situations and hotspots for preventive action [2] [15] [17] [18].

1.3 Models of Predictive Analytics

Some models have been proposed for predictive analysis based on statistics, Bayesian theory, and machine learning modules. Three models have been discussed in this study which are widely used in this field.

1. **Predictive model:** Predictive models are models of the relation between the specific performance of a unit in a sample and one or more known attributes and features of the unit. The objective of the model is to assess the likelihood that a similar unit in a different sample will exhibit the specific performance. This category encompasses models in many areas, such as marketing, where they seek out subtle data patterns to answer questions about customer performance, or fraud detection models. Predictive models often perform calculations during live transactions, for example, to evaluate the risk or opportunity of a given customer or transaction, in order to guide a decision. With advancements in computing speed, individual agent modeling systems have become capable of simulating human behavior or reactions to given stimuli or scenarios. The available sample units with known attributes and known performances are referred to as the "training sample." The units in other samples, with known attributes but unknown performances, are referred to as "out of [training] sample" units. The out of sample bear no chronological relation to the training sample units. For example, the training sample may consists of literary attributes of writings by Victorian authors, with known attribution, and the out-of sample unit may be newly found writing with unknown authorship; a predictive model may aid in attributing a work to a known author. Another example is given by analysis of blood splatter in simulated crime scenes in which the out of sample unit is the actual blood splatter pattern from a crime scene. The out of sample unit may be from the same time as the training units, from a previous time, or from a future time [1] [18].
2. **Descriptive Models:** Descriptive models quantify relationships in data in a way that is often used to classify customers or prospects into groups. Unlike predictive models that focus on predicting a single customer behavior (such as credit risk), descriptive models identify many different relationships between customers or products. Descriptive models do not rank-order customers by their likelihood of taking a particular action the way predictive models do. Instead, descriptive models can be used, for example, to categorize customers by their product preferences and life stage. Descriptive modeling tools can be utilized to develop further models that can simulate large number of individualized agents and make predictions [19].
3. **Decision Models:** Decision models describe the relationship between all the elements of a decision — the known data (including results of predictive models), the decision, and the forecast results of the decision in order to predict the results of decisions involving many variables. These models can be used in optimization, maximizing certain outcomes while minimizing others. Decision models are generally used to develop decision logic or a set of business rules that will produce the desired action for every customer or circumstance [18] [19].

2. PREDICTIVE ANALYTICS AND CLOUD COMPUTING

Cloud computing is the recent technology that enables cloud customers to enjoy the on demand high quality applications and services from a shared pool of configurable computing resources through storing their data remotely in the cloud [7]. It refers to both the applications delivered as a service over the Internet and the hardware and software in the data centers that provide those services [27]. Cloud computing is a network-based environment that focuses on sharing computations and resources. The amount of resources needed is rarely static, varying as a result of changes in overall workload, the workload mix, and internal application phases and changes [23]. Under-provisioning resources will cause service level objective (SLO) violations, which are often associated with significant financial penalties. Over-provisioning wastes resources that could be put to other uses [28]. To avoid both problems, the amount of resources allocated to applications should be adjusted dynamically, which brings two main challenges: (1) deciding how much resource to allocate is non-trivial since application resource needs often change with time and characterizing runtime application behavior is difficult; (2) application resource needs must be predicted in advance so that the management system can adjust resource allocations ahead of the needs. Furthermore, resource-management systems should not require prior knowledge about applications, such as application behavior profiles, and running the resource management system itself (including its prediction algorithms) should not be costly [3].So using predictive analytics we can aware in advance about the demand of cloud recourses and invest money in better way.

2.1 Cloud Computing: Trends and Directions

Cloud computing is a way to increase the capacity or add capabilities dynamically without investing in new infrastructure, training new personnel, or licensing new software. It extends Information Technology's (IT) existing capabilities. In the last few years, cloud computing has grown from being a promising business concept to one of the fast growing segments of the IT industry [4].Cloud computing providers have setup several data centers at different geographical locations over the Internet in order to optimally serve needs of their customers around the world. However, existing systems do not support mechanisms and policies for dynamically coordinating load distribution among different Cloud-based data centers in order to determine optimal location for hosting application services to achieve reasonable Quality of Service (QoS) levels [5] [7].Cloud computing is said to be an emerging new computing paradigm for delivering computing services. This computing approach relies on a number of existing technologies, e.g., the Internet, virtualization, grid computing, Web services, etc. The provision of this service in a pay-as-you-go way through (largely) the popular medium of the Internet gives this service a new distinctiveness [6].A consumer with an instantaneous need at a particular timeslot can avail computing resources (such as CPU time, network storage, software use, and so forth) in an automatic (i.e. convenient, self-serve) fashion without resorting to human interactions with providers of these resources [7].

2.2 Cloud Service Model

Cloud computing is a delivery of computing where massively scalable IT-related capabilities are provided ―as a service across the internet to numerous external clients. This term effectively reflects the different facets of the Cloud Computing paradigm which can be found at different infrastructure levels. Cloud Computing is broadly classified into three services: ―IaaS", "PaaS" and "SaaS" [20].

- **IaaS (Infrastructure as a service) model**: The main concept behind this model is virtualization where user have virtual desktop and consumes the resources like network, storage, virtualized servers, routers and so on, supplied by cloud service provider. . Usage fees are calculated per CPU hour, data GB stored per hour, network bandwidth consumed, network infrastructure used per hour, value added services used, e.g., monitoring, auto-scaling etc. Examples: Storage services provided by AmazonS3, Amazon EBS. Computation services: AmazonEC2, Layered tech and so on [7].
- **PaaS (Platform as a service) model**: It refers to the environment that provides the runtime environment, software deployment framework and component on pay to enable the direct deployment of application level assets or web applications. PaaS is a platform where software can be developed, tested and deployed. It means the entire life cycle of software can be operated on a PaaS. This service model is dedicated to application developers, testers, deplorers and administrators. Examples: Google App Engine (GAE), Microsoft Azure, IBM Smart Cloud, Amazon EC2, salesforce.com and jelastic.com and so on [4].
- **SaaS (Software as a service):** Through this service delivery model end users consume the software application services directly over network according to on-demand basis. For example, Gmail is a SaaS where Google is the provider and we are consumers. Other well-known examples of PaaS include billing services provided by Arial system, op source. Financial services: Concur, workday, Backup and recovery services and so on [23].

2.3 Types of Cloud Model

There are four primary cloud computing deployment models which are available to service consumer as shown in fig-1.2.1.

- **Public cloud/external cloud**: This model allows cloud environment as openly or publically accessible. Public cloud is off premise in which various enterprises can be used to deliver the services to users by taking it from third party.
- **Private cloud/internal cloud**: This model referred to on-premise cloud which is managed or owned by an organization to provide the high level control over cloud services and infrastructure. In other words private cloud is built specifically to provide the services within an organization for maintaining the security and privacy.
- **Hybrid cloud/virtual private cloud model**: This model compromised both private and public cloud models where cloud computing environment is hosted and managed by third party (off-premise) but some dedicated resources are privately used only by an organization.

- **Community model**: It allows the cloud computing environment which is shared or managed by number of related organizations.

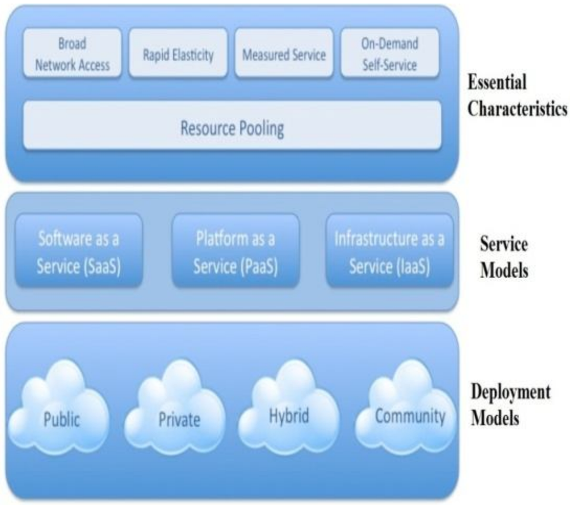

Figure 2. Cloud Services and Types

3. MACHINE LEARNING IN PREDICTIVE ANALYTICS

Machine learning is a subfield of computer science that evolved from the study of pattern recognition and computational learning theory in artificial intelligence [24]. Machine learning explores the study and construction of algorithms that can learn from and make predictions on data. The Machine Learning field evolved from the broad field of Artificial Intelligence, which aims to mimic intelligent abilities of humans by machines. In the field of Machine Learning one considers the important question of how to make machines able to “learn”. Learning in this context is understood as inductive inference, where one observes examples that represent incomplete information about some “statistical phenomenon” [13]. Unsupervised learning one typically tries to uncover hidden regularities (e.g. clusters) or to detect anomalies in the data (for instance some unusual machine function or a network intrusion). In supervised learning, there is a label associated with each example. It is supposed to be the answer to a question about the example. If the label is discrete, then the task is called classification problem – otherwise, for real valued labels we speak of a regression problem [10]. Based on these examples (including the labels), one is particularly interested to predict the answer for other cases before they are explicitly observed. Hence, learning is not only a question of remembering but also of generalization to unseen cases.

3.1 Reflections on Machine Learning

The Machine Learning field evolved from the broad field of Artificial Intelligence, which aims to mimic intelligent abilities of humans by machines. In the field of Machine Learning one considers the important question of how to make machines able to “learn”. Learning in this context is understood as inductive inference, where one observes examples that represent incomplete information about some “statistical phenomenon” [8][30]. Machine learning and data mining are research areas of computer science whose quick development is due to the advances in data analysis research, growth in the database industry and the resulting market needs for methods that are capable of extracting valuable knowledge from large data stores [9]. Machine learning algorithms can figure out how to perform important tasks by generalizing from examples. This is often feasible and cost-effective where manual programming is not [10].

3.2 Types of learning

The learning types are described below with suitable examples.

1. **Supervised Learning:** An important task in Machine Learning is classification, also referred to as pattern recognition. In supervised learning computer is presented with example inputs and their desired outputs, given by a "teacher", and the goal is to learn a general rule that maps inputs to outputs [29]. Training data includes both the input and the desired results. It discover patterns in the data that relate data attributes with a target (class) attribute and these patterns are then utilized to predict the values of the target attribute in future data instances. A pattern is described by its features. These are the characteristics of the examples for a given problem [12]. Face recognition, voice recognition, signature recognition and customers discovery are the example of supervised learning For instance, in a face recognition task some features could be the color of the eyes or the distance between the eyes. Pattern classification tasks are often divided into several sub-tasks:
   a. Data collection and representation.
   b. Feature selection and/or feature reduction.
   c. Classification.

2. **Unsupervised learning:** In unsupervised learning no labels are given to the learning algorithm, the model is not provided with the correct results during the training, leaving it on its own to find structure in its input.

Unsupervised learning can be a goal in itself to discovering hidden patterns in data. Clustering is the example of unsupervised learning [8] [11].

4. CONCLUSION

Both predictive analytics and cloud computing are service-oriented and fulfill the demands of user whether they are about infrastructure, resources or some other kind of data/information. Since to make correct and judicious decisions, right knowledge and data is required, but due to ever increasing use of computing machines and internet technologies, data produced over the servers has grown many folds and as such there is need for analysis and maintenance of this data which is covered under big data. Big data is analyzed with the help of special types of databases which can implement machine learning and supplies with correct and precise data to the user. Cloud computing has decreased the load on the start-ups and small organizations by supplying the resources on rent basis. This led to implementation of cloud in such a way so as to fulfill the needs of the user in a much efficient and reliable manner with the help of elastic cloud. To help adjust the size of the cloud so as to provide the resources in profit-making manner, the recent trends of demand over a cloud can be analyzed with predictive analytics. The future of Data Mining lies in Predictive Analytics. Predictive Analytics can also helpful in Cloud Computing. If we know the resources provided by the cloud server along with the data containing its demand, we predict the utilization and performance of these resources and help in managing them so as to make the cloud elastic and self-managed. It will help the cloud service provider to know about the demand of cloud recourses in advance so that he can take proper action on time. Predictive Analytics make him to invest on those recourses that is highly demanded by the cloud's users. Predictive Analytics is an area of interest to almost all communities and organizations. Predictive analytics is using business intelligence data for forecasting and modeling. Proper data mining algorithms and predictive modeling can refine search for targeted customers. Predictive Analytics can aid in choosing marketing methods, and marketing more efficiently.

REFERENCES

[1] Nishchol Mishra, Dr.Sanjay Silakari, "Predictive Analytics: A Survey, Trends, Applications,Oppurtunities & Challenges", School of IT, Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India, Vol. 3 (3) , 2012.

[2] Stefan, Krauss, Schoder," PREDICTIVE ANALYTICS ON PUBLIC DATA – THE CASE OF STOCK MARKETS".

[3] Zhenhuan Gong, Xiaohui Gu,"PRESS:PRedictive Elastic ReSource Scalingfor cloud systems", North Carolina State University.

[4] S. Subashini, V. Kavitha," A survey on security issues in service delivery models of cloud computing", Anna University Tirunelveli, Tirunelveli, TN 627007, India, July 2010.

[5] Rajkumar Buyya, Rajiv Ranjan, Rodrigo N. Calheiros," InterCloud: Utility-Oriented Federation of Cloud Computing Environments for Scaling of Application Services".

[6] Nabil Sultan," Cloud computing for education: A new dawn?", Liverpool Hope University, Hope Park, Liverpool, L16 9JD, UK,2010.

[7] Tharam Dillon, Chen Wu and Elizabeth Chang " Cloud Computing: Issues and Challenges" Curtin University of Technology Perth, Australia, 2010 24th IEEE International Conference on Advanced Information Networking and Applications.

[8] Gunnar R¨atsch," A Brief Introduction into Machine Learning", Friedrich Miescher Laboratory of the Max Planck Society, Spemannstraße 37, 72076 T¨ubingen, Germany.

[9] J. F¨urnkranz et al., Foundations of Rule Learning, Cognitive Technologies, DOI 10.1007/978-3-540-75197-7 1, © Springer-Verlag Berlin Heidelberg 2012

[10] Pedro Domingos," A Few Useful Things to Know about Machine Learning", University of Washington.

[11] Mike Schuster and Kuldip K. Paliwal ," Bidirectional Recurrent Neural Networks", Member, IEEE, VOL. 45, NO. 11, NOVEMBER 1997

[12] Akshay Sharma,Babita Sandooja ,Deepika Yadav," Extreme Machine Learning: Feed Forward Networks" Dronacharya College of Engg.(CSE), Volume 3, Issue 8, August 2013

[13] XIN YAO," Evolving Artificial Neural Networks", PROCEEDINGS OF THE IEEE, VOL. 87, NO. 9, SEPTEMBER 1999

[14] Ms. Sonali. B. Maind, Ms. Priyanka Wankar" Research Paper on Basic of Artificial Neural Network", IJRITCC January 2014,

[15] R. Banjade and S. Maharjan, "Product Recommendations using Linear Predictive Modeling," 2011.

[16] Debahuti Mishra et al., “Predictive Data Mining: Promising Future and Applications”, Int. J. of Computer and Communication Technology, Vol. 2, No. 1, pp. 20-28, 2010.
[17] Sanfilippo et al., “Technosocial Predictive Analytics in Support of Naturalistic Decision Making,” Symposium A Quarterly Journal In Modern Foreign Literatures, no. June, pp. 144-151, 2009.
[18] http://en.wikipedia.org/wiki/Predictive_analytics .
[19] http://analyticsweb.com/predictive-analytics.
[20] Puneet Singh Duggal, Sanchita Paul,” Big Data Analysis: Challenges and Solutions” International Conference on Cloud, Big Data and Trust 2013, Nov 13-15, RGPV
[21] Jafar Raza Alam, Asma Sajid, Ramzan Talib, Muneeb Niaz,” A Review on the Role of Big Data in Business”, IJCSMC, Vol. 3, Issue. 4, April 2014, pg.446 – 453
[22] Munesh Kataria, Ms.Pooja Mittal,” BIG DATA: A Review”, M. D. University, Rohtak, IJCSMC, Vol. 3, Issue. 7, July 2014, pg.106 – 110
[23] Farzad Sabahi,” Secure Virtualization for Cloud Environment Using Hypervisor-based Technology”, Member, IEEE, Vol. 2, No. 1, February 2012
[24] Hai Jin, Li Deng, Song Wu, Xuanhua Shi, Xiaodong Pan,” Live Virtual Machine Migration with Adaptive
[25] Memory Compression”, Huazhong University of Science and Technology, Wuhan, 430074, China, December 1, 2009
[26] Ashima Agarwal, Shangruff Raina,” Live Migration of Virtual Machine in Cloud”, Department of Computer, MIT College of Engineering Pune, India, Volume 2, Issue 6, June 2012
[27] Nikolas Roman Herbst, Samuel Kounev, Ralf Reussner,” Elasticity in Cloud Computing: What It Is, and What It Is Not”, Karlsruhe Institute of Technology, Germany
[28] Guilherme Galante and Luis Carlos E. de Bona,” A Survey on Cloud Computing Elasticity”, Department of Informatics Federal University of Parana, 2012 IEEE/ACM Fifth International Conference on Utility and Cloud Computing
[29] Ahmed Ali-Eldin, Johan Tordsson and Erik Elmroth,” An Adaptive Hybrid Elasticity Controller for Cloud Infrastructures”, Department of Computing Science, Ume˚a University, 2012 IEEE
[30] T. M. Mitchell. Machine Learning. McGraw-Hill, New York, NY, 1997.