International Journal of Business and Management Invention (IJBMI)
ISSN (Online): 2319 – 8028, ISSN (Print): 2319 – 801X
www.ijbmi.org || Volume 13 Issue 4 || April, 2024 || PP: 149-152

Machine Learning andPredictive Analytics – A Paradigm Shift in Decision Making

Philomine Roseline. T
Research Scholar, Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirappalli.

Dr. J. G. R. Sathiaseelan
Associate Professor&Head, Department of Computer Science, Bishop Heber College (Autonomous), Tiruchirappalli.

ABSTRACT:*Machine learning (ML), a branch of Artificial Intelligence (AI), is revolutionizing the way the decision making process was done before. A myriad of algorithms- for varied situations, inputs and constraints, being developed and successfully implemented stands as a proof of their efficacy and versatility.*
Predictive analytics is an amalgamation of machine learning algorithms, statistical techniques and data analytics which reveals the hidden pattern amongst the data and enables the algorithms to predict or forecast the outcome of a given situation based on the past or historical data.
The objective of this paper is to elucidate all the different types of predictive analytics models under the different types of machine learnings, their concepts as well as areas of applications. It is impossible to study predictive analytics without prior knowledge of machine learning. This paper would serve as an introduction to the vast subjects.
KEY WORD*Artificial Intelligence, Machine learning, Predictive Analytics models, big data and data analytics.*

I. INTRODUCTION TO MACHINE LEARNING

Cognitive skills are those that distinguishes human beings from others. Learning from the patterns, past mistakes as well as success helps us to evolve. If we can successfully program these cognitive skills into a computer then most of our mundane task can be taken care by futuristic machines.

Machine learning (ML) is a term coined by Arthur Samuel in 1959. As the name suggests, it aims to infuse the learning capabilities of human in the form of an algorithm, model, architecture and/or techniques clubbed with data analytics, into a machine- computer. It could also be an equation or tools and techniques that is programed, which when fed with data, learns by getting trained repetitively and is finally able to make predictions or forecast the output as a human would in the given situation.

II. TYPES OF MACHINE LEARNING

Machine learning can be broadly classified as supervised learning, unsupervised learning and reinforcement learning. Between the supervised and unsupervised lies the semi-supervised learning. The data- usually big data, used to feed the algorithms, is ideally partitioned into 70% and 30% and used for training and testing purposes respectively. Data being labelled or unlabelled determines the type of learning. Pre-processing of the data to handle missing values, noisy data and outliers is a crucial step that determines the accuracy level of the algorithms. Various tools and techniques are available for handling data pre-processing, the most popular being Python and it's rich library functions.

Supervised learning- also known as predictive learning, uses labelled data. Here the algorithm maps the labelled input with the corresponding output. The algorithm learns to tag the input with the correct output and is trained with the data until it makes the predictions correctly. It is similar to student learning with the supervision of teacher. These algorithms are best suited for simple task.

Unsupervised learning- also known as descriptive learning, deals with unlabelled data. The algorithm studies the underlying pattern by identifying the attributes and their association between them and groups them together. Since the data are not labelled, this learning model must analyse the input to find out the similarities and classify them accordingly. These types of algorithms are used for complex task.

Semi-supervised learning- deals with both labelled and unlabelled data. This model of learning algorithm tries to harness the power of both supervised and unsupervised leaning models. The accuracy of this model depends heavily upon the quality of the unlabelled data. If they contain noisy, missing, wrong data then the data pre-processing step is crucial to predict the output more accurately.

Reinforcement learning- uses feedback mechanism to learn whether the predictions made are right or wrong. If the output predicted is good, then it gets a positive feedback and learns to perform the same way and tries to enhance it. If the output is not good, then a negative feedback is given and the algorithm learns to avoid that in future. This is similar to a child's learning process. It is the most complex type of learning using unlabelled data.

III. INTRODUCTION TO PREDICTIVE ANALYTICS

Predictive analytics is a technique that uses the past data (also known as historical data) to analyse and identify the underlying pattern to predict the future trends or events. It uses statistical tools, machine learning algorithms and data analytics to study and uncover the hidden facts amongst the data to answer the question "What would happen in future?"

The systematic flow of steps in predictive analytics are data pre-processing, feature engineering, model-selection, training and testing, evaluation of metrics and finally deployment. **Data pre-processing** makes the data set complete without any missing values, redundant data, erroneous inputs and outliers. There are various statistical methods to handle outliers, missing data and programming tools to transform the data ready to be processed.**Feature engineering** consists of feature creation, transformation, extraction and selection.Relevant attributes must be selected that will contribute to the prediction. Irrelevant attributes can also impact the accuracy of the model. For example, the correlation between the various attributes can be evaluated using the heat map in Python. Principal Component Analysis (PCA) is the most popular method in which the already existing features are combined to create a new meaningful feature that will aid in the decision making process. Once the data set is ready the model to be used is identified. Accordingly the algorithm is developed based on whether it is supervised, unsupervised, semi-supervised or reinforcement learning. **Training and testing** is an iterative process and continues till the desired prediction is achieved with highest accuracy level. The entire data set is divided ideally into 70% and 30% for training and testing respectively. **Various metrics** are used to calculate and substantiate that the model developed yields the desired accuracy. The choice of metrics used depends upon the model developed. There is no single standard metric for evaluation. Each formula in metrics depicts various aspects of the algorithm. **Deployment** or implementation of the prediction algorithm is done once we have proved that the model performance is far better than its predecessor. Each of the above given steps need various machine learning algorithms and tools (like Python, R, Tableau and Power BI) for implementation.

Data visualization helps in the interpretation of the results and therefore forecasting the future. Risk factors can also be identified and handling them in timely manner helps the organization to prevent or contain the damage that is likely to dampen the success of the institution. Thus by virtue of predictive analytics model, tools and techniques the organizational objectives can be attained by maximizingthe profit and minimizing the loss by predicting the likelihood of the event to occur.

IV. TYPES OF PREDICTIVE ANALYTICS MODELS

Predictive analytics models harness the power of machine learning and statistical tools and techniques to predict the future. Based on the type of data- categorical, cluster, numerical, anomalous and time stamped, they fall under the following categories as listed in the table:

CLASSIFICATION	CLUSTERING	FORECASTING	OUTLIERS	TIMESERIES
1. Logistic Regression	1. K-Means clustering	1. Straight line	1. EVA– Extreme Value Analysis	1. Auto Regressive (AR)
2. Naïve Bayes	2. Mean-Shift	2. Moving Average	2. Probabilistic & Statistical models	2. Moving Average (MA)
3. Stochastic Gradient Descent	3. DBSCAN (Density based spatial clustering of applications with noise)	3. Simple linear regression	3. Linear models	3. Auto Regressive Moving Average (ARMA)
4. K- Nearest Neighbor (KNN)	4. EM (Expectation maximization) using Gaussian mixture model (GMM)	4. Multiple linear regression	4. Proximity based models	4. Auto Regressive Integrated Moving Average (ARIMA)
5. Decision Tree	5. Agglomerative hierarchical clustering		5. Information Theoretic models	5. Seasonal Auto Regressive Integrated Moving Average (SARIMA)
6. Random Forest			6. High dimensional outlier detection	6. Vector Auto Regression (VAR)
7. Support vector Machine (SVM)				7. Vector Auto Regression Moving Average with Exogenous Regressors (VARMAX)

V. AREAS OF APPLICATIONS OF PREDICTIVE ANALYTICS

Depending on the type of data set, model of the machine learning algorithm applied as well as the required accuracy there are various avenues wherein predictive analytics can be implemented. Few of the domains and the respective applications are as mentioned below:

Medical and Health care: Detection of cancer, thyroid, diabetes and their classifications; segmentation of tumours, bone fracture classification, DNA sequence classification, genetics research,drugs and symptoms classification, drug discovery, prediction of heart failure, robot assisted surgery, personalised health care etc.

Social Media: Identifying fake news,spam filter, sentiment analysis and blogs mining for opinion, intent analysis, chatbots, smart search engine optimization, social media influencer content generation, movie and media recommendations, fake image and video recognition, automatic language translation etc.

Human Resources: Talent Management System (TMS) - talent acquisition, performance appraisal, on boarding, employee engagement, retention strategies, succession planning and attrition prediction; staffing, human resources analytics, resume screening, payroll management, virtual personal assistant,etc.

Corporate Business: customer relationship management, sales relationship management, stock comparison, dynamic pricing, digital marketing, search engine optimization, product recommendations,demand forecasting, risk identification and mitigation, strategic planning, decision making support, online fraud detection etc.

Engineering Sector:Autonomous vehicles, pedestrian detection in automotive car driving, simulation and modelling of intelligent systems, quality control predictions and automations, resource automation, flaw detection in 3D printing, air traffic control and predictions, object tracking, motion detection, robotics etc.

Agricultureand Marine Sciences: Soil analytics, soil classification,intelligent harvesting, crop yield prediction, water management, livestock management, weed and disease detection, optimization of irrigation schedules, fertilizer and pesticides management, marine wildlife predictions, fisheries management etc.,

Computer Technologies: Detection and classifying network traffic, wireless sensor network based applications, , image processing and segmentation, speech and handwriting recognition, malware detection and filtering, e-commerce applications, pattern matching and recognition, generative adversarial networks etc.

Apart from the above ML and predictive algorithms are used in banking and insurance sectors, automobile industry, sports and entertainment, weather predictions, pandemic predictions, logistics and supply chain management, retail management, departmental stores, public safety and security, space explorations, geological analysis and predictions etc.

VI. CONCLUSION

The above is a glimpse into the dynamic world of decision making complemented with machine learning and predictive analytics algorithms. Though they rely heavily upon historical data, require specialized skills and may pose a security concern dealing with sensitive data, ongoing research and the plethora of the applications in varied domains are path breaking in the way they make lives better. The fact that many leading companies are using these algorithms to improvise their functionalities, enhance strategic planning and effective allocation of resources and customer satisfaction oriented deliverables undeniably states that this technology is here to stay and has already brought about a paradigm shift.

BIBLIOGRAPHY

[1]. Ameet V Joshi (2020), "Machine Learning and Artificial Intelligence", Springer.

[2]. SaikatDutt, Subramanian Chandramouli, Amit Kumar Das (2019), "Machine Learning", Pearson India Education Services Pvt. Ltd.

[3]. Richard V. McCarthy, Mary M. McCarthy, Wendy Ceccucci, Leila Halawi(October 2021), "Applying Predictive Analytics: Finding Value in Data", Springer International Publishing.

[4]. Steven Finlay (July 2014), "Predictive Analytics, Data Mining and Big Data - Myths, Misconceptions and Methods", Palgrave Macmillan UK.

[5]. Eric Siegel (2016), "Predictive analytics: the power to predict who will click, buy, lie, or die", Wiley.

[6]. MitanshiRustagi, Neha Goel (2022), "Predictive Analytics: A study of its Advantages and Applications", International Association of Research Scholars, International Research Journal, vol. 12.

[7]. Fatimetou Zahra Mohamed Mahmoud (May 2017), "The Application Of Predictive Analytics: Benefits, Challenges And How It Can Be Improved", International Journal of Scientific and Research Publications, Volume 7, Issue 5.

[8]. Nishchol Mishra, Dr.SanjaySilakari (2012), "Predictive Analytics: A Survey, Trends, Applications, Oppurtunities& Challenges", International Journal of Computer Science and Information Technologies, Vol. 3 (3).

[9]. Vaibhav Kumar, M. L. Garg (July 2018), "Predictive Analytics: A Review of Trends and Techniques", International Journal of Computer Applications (0975 – 8887) Volume 182 – No.1.

[10]. Folashade Gloria Olaniyi, OluwaseunOladejiOlaniyi, Chinasa Susan Adigwe, Anthony Abalaka, NishantHemantkumar Shah

(Nov., 2023), "Harnessing Predictive Analytics for Strategic Foresight: A Comprehensive Review of Techniques and Applications in Transforming Raw Data to Actionable Insights", Asian Journal of Economics, Business and Accounting, Vol 23, Issue 22, Page 441-459.

[11]. https://www.sap.com/india/products/technology-platform/cloud-analytics/what-is-predictive-analytics.html

[12]. https://www.ibm.com/topics/predictive-analytics