



Sparkling Light Publisher

Sparklinglight Transactions on Artificial Intelligence and Quantum Computing (STAIQC)



Website: <https://sparklinglightpublisher.com/> ISSN (Online):2583-0732

Use of Supervised Learning Algorithms in Predictive Analytics

Geetha Poornima K^{a,*}, Vinayachandra^b, Rajeshwari M^c, Bishwas Mishra^d

^{a,b,c}Assistant Professor, Dept. of Computer Science, St Philomena College, Puttur, D.K. Karnataka, India

^dMaharishi International University, Fairfield, IA, USA

Abstract

Human life has been made easier and more comfortable thanks to technological advancements. Predictive analytics is a revolutionary technique that utilizes a significant amount of historical data to create predictions about the future. Its goal is to analyze specific data in order to forecast the future and identify the risks connected with a certain decision. Using data-driven predictive models, decisions that were the product of several mathematical computations can be made more quickly and accurately. Banking, education, healthcare, entertainment, and other industries employ technologies to create difficult decisions and forecast future trends. The goal of predictive analytics is to make accurate and cost-effective predictions. The data required for the analysis comes from a variety of sources and will be in a structured, semi-structured, or unstructured format. The classification of a large volume of data during the data analytics process is a tough challenge. The purpose of classification is to turn accessible data into knowledge that will be useful in future research. It is possible to learn from the training data set using machine learning, and the knowledge gathered this way can be applied to effective decision-making. Classification algorithms examine at the training data and use that knowledge to categorize the test data. To maximize their profitability, organizations acquire experts in critical decision-making. Using human intelligence to make key decisions is costly, dangerous, and time-consuming. As a result, predictive analytics is getting lots of attention these days. It makes the most out of available data in order to make better and more informed decisions. It can be used to discover different patterns and relationships in data in order to forecast future events. Data analysis delivers useful insights and reliably identifies potential hazards. The predictive model and attributes chosen for analysis determine the accuracy of the prediction. The use of an incorrect model and erroneous data can be catastrophic for an organization. Artificial intelligence, cloud computing, machine learning, and other emergent technologies are used to collect, store, and analyze data effectively. The quality of the data acquired and the models employed for analysis are both important factors in forecasting. To analyze the data and make predictions, many supervised learning approaches can be applied. The authors of this paper attempt to provide a thorough overview of the many supervised learning approaches prevalent in machine learning. They also attempt to investigate several application areas in which these strategies are employed to aid decision-making.

© 2021STAIQC. All rights reserved.

Keywords: Classification, Machine Learning, Predictive Analytics, Supervised Learning

E-mail address of authors: ^{a,*}poornima.sanjay@gmail.com, ^bveeciashu@gmail.com, ^crajimuraleedhar@gmail.com, ^dbmishra@miu.edu

© 2021 STAIQC. All rights reserved.

Please cite this article as: Geetha Poornima K, Vinayachandra, Rajeshwari M, & Bishwas Mishra (2021). An Analysis into the use of Supervised Learning Algorithms in Predictive Analytics. *Sparklinglight Transactions on Artificial Intelligence and Quantum Computing (STAIQC)*, 1(2), 1-20. ISSN (Online):2583-0732

1. Introduction

A vast volume of data is generated in this world which is fully driven by information technology. If presented properly, this information could be useful to someone. Machine learning (ML) has progressed as a result of the abundance of data. Many businesses utilize data analytics to make better and more accurate decisions, allowing them to make more money from their operations. Because the data, which is the most important component of analysis, is plentiful, it may be analyzed to gain better insights. Data analysis enables one to understand the relationship between previous and contemporary data, which may be used to make appropriate decisions [1]. Data-driven decision-making is more supportive of an organization's progress. It is a cost-effective method of making precise decisions. Analyzing data will allow a company to improve the quality of its products or services, helping the company to increase revenue. When data is transformed into knowledge, indisputable proof is created that may be used to make a better decision or reach a good conclusion. Institutions can foresee potential dangers and obstacles using historical data before they become a full-fledged catastrophe [2].

Organizations can measure the effectiveness of a strategy by analyzing data in a systematic way. When a new strategy is implemented to address current challenges, a study of the findings will allow policymakers to assess whether or not modifications in decision are required. Organizations can uncover the underlying cause of problems via effective data analysis. Data analysis can be used to illustrate the relationships between different events that occur in different areas. Administrators will be able to construct multiple theories and determine the most effective solutions to a problem by viewing data points concurrently [3]. Data analysis is an important part in demonstrating for a specific decision. It serves as evidence for authorities to present a convincing argument for or against a particular decision. The authorities will be able to facilitate the desired changes in the system if they can demonstrate employing historical data as evidence. The authorities will be able to explain the outcomes of decisions to various stakeholders in an organization thanks to data analysis. It will also allow for prediction of the outcome based on concrete evidence rather than speculation. Data analysis will also enable authorities to focus on specific sectors in order to better manage resources and people, identify areas for improvement, and recognize skilful workforce, among other things. Companies will be able to use data analysis to define targets, standards, and performance metrics in order to keep progressing [4].

Predictive analytics (PA) is concerned with extracting necessary features from a massive quantity of data. A predictive model is developed from the available data, which is then utilized to forecast future events. By evaluating valuable patterns in massive amounts of data, analysts can find trends and behaviors in an organization. To transform data obtained from numerous sources into knowledge, PA employs a variety of techniques such as machine learning (ML), statistical modeling, and others. Based on historical data, PA makes estimates about future outcomes. It predicts future occurrences using machine learning and other statistical approaches. PA is a revolutionary innovation that can be used to predict the future with a high degree of accuracy. With the help of PA tools and models, any organization may leverage its extensive historical data as well as current data to forecast future trends [5]. PA will estimate the likelihood of an event occurring. It automatically analyzes large amounts of data with a variety of factors. The most crucial aspect of PA is a variable or parameter whose value, when measured, may be used to forecast future trends. Multiple variables can be combined to form a predictive model that can be used to forecast future events with accuracy. PA is a combination of business expertise and statistical approaches that produce relevant and actionable insights when applied to business data [6].

The most sophisticated statistical analysis technology is machine learning (ML). ML can fine-tune parameters so that they are consistent with the models that are being used. These jobs may take a long time to complete manually due to their complexities. Furthermore, machine learning can reject data points and change parameters to fit the model. To accomplish difficult mathematical calculations, ML employs a variety of methods and processing resources. It modifies the weights of parameters in order to provide the most accurate forecast of future events or trends. Learning is a method by which computers utilize a variety of ways to analyze large amounts of data and make recommendations based on the results. Facebook's "People you may know" feature, Amazon's product recommendations, Netflix's movie suggestions, etc. rely on ML [7]. ML is a ground-breaking approach that allows machines to learn without having to be explicitly programmed. The algorithms used in ML can learn from previous data. When a machine learns, it simply feeds data into an ML algorithm, which develops logic and delivers results based on the supplied data. The accuracy of the ML system improves when it does the same task with multiple

inputs. The ML algorithm uses training data to create a model. When given new test data, the model produces predictions. The accuracy of the results is determined once they have been obtained. If the results do not meet the acceptable accuracy level, the algorithm is retrained. This technique is repeated until the desired level of precision is reached. The ML algorithm learns to produce the best feasible response on its own in this technique. Once the ML model's output matches the desired accuracy level, it is integrated in the user's organization[8]. The authors make a reasonable attempt in this paper to discuss different forms of learning, commonly used supervised learning algorithms, and issues and challenges in PA employing ML.

2. Related Study

The paper cites several scholarly articles published between 2010 and 2021 that make recommendations on a variety of supervised learning algorithms utilized in PA. Mokhaet al. 2016 [9] used different statistical techniques to predict the possibility of injuries among athletes. When the data of necessary parameters is stored in a centralized repository it is possible to retrieve the required data, apply statistical techniques to predict the possibility of injuries Abdulla et al. (2016) [10].

Tang, F., &Ishwaran [11]used RF algorithm d to explain how to manage missing data and various forms of data. When constructing the forest, it was also explained how the RF method is effective in handling premature, impute, and pre-impute data.Manogaran, et al. (2019) [12] performed survey of different ML algorithms to categorize clinical data associated with the various diseases and trained the same using various machine learning algorithms. The selection of attributes for classification is entirely dependent on the weight of the attribute. AI and ML models are used to accurately forecast injuries, and the models are tested using injury data from various athletes and shown to be accurate (Naghlaet al. (2018)) [13]

.Apostolou&Tjortjis (2019) [14] Illustrated the use of ML algorithms to pinpoint the cause of a deterioration in an athlete's performance.Patients' safety was assessed using tree-based machine learning methods. The investigation uncovered a previously undiscovered. Simsekleret al. (2020) [15] used RF algorithm to evaluate the safety of the COVID-19 patients so that the errors in the healthcare assessment and risks are reduced. Oytun et al. (2020)[16] developed ML models to find the non-linear relationship between different parameters of athletes and to boost the possibility of winning.

Iwendiet al. (2020) [17] demonstrated the use of RF algorithm to predict the patient health. Used boosting technique to increase the accuracy.Also verified the accuracy of results. Krishna Prasad et al. (2021) [18] developed a statistical model that calculates the probable COVID-19 infestations and used an online tool that uses artificial intelligence technique to predict the trends based on the stringency index imposed by the government.

3. Objectives

- To review different supervised learning algorithms
- To understand issues and challenges associated with different supervised learning algorithms
- To elucidate different application areas of PA that use ML algorithms
- To understand the usefulness boosting and bagging in supervised learning algorithm

4. Methodology

This paper is developed to analyse different SL algorithms that are extensively being used in PA. The secondary data from journal articles, conference proceedings and some official websites is used as reference for this study. Applications of different SL algorithms in PA are analysed by referring a good number of papers published in peer reviewed journals.

5. Overview

Data analytics (DA) is a science of transforming raw data into knowledge to arrive at conclusions about the data collected. It reveals the trends and hidden relationships among the data. The main purpose of DA is to improve the

business process to increase the efficiency of an organization. It is essential because it helps the organization to improve overall performance. DA can be of four types as shown in Figure -1. They are in their increasing level of business value generated as well as their complexity levels.

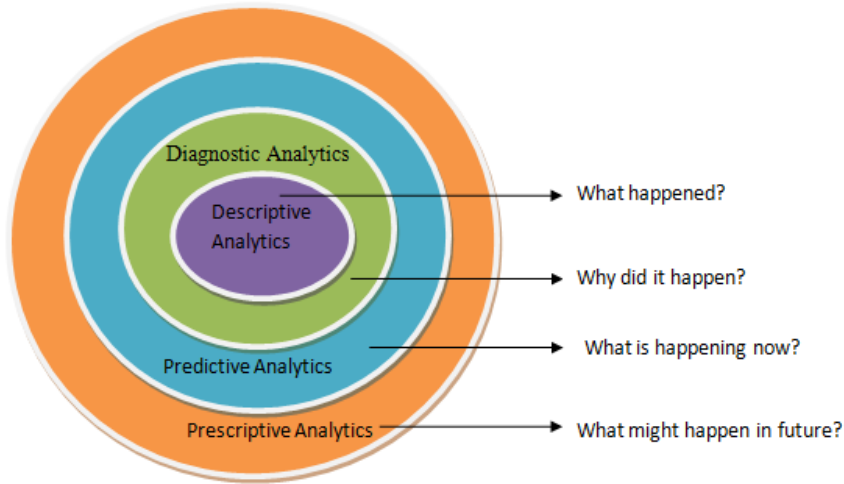


Fig. 1: Types of Data Analytics

5.1 Descriptive analytics

It is used to describe what happens over some time. It produces explanatory information. Since it operates on the live data, it is comprehensive and precise. It organizes the raw data collected from various sources to generate valuable insights into the past. The findings will describe what went wrong without finding the reason.

5.2 Diagnostic analytics

It aims at finding the reason or root cause for a particular problem. Based on the historical data it performs hypotheses. It gives in-depth insights into a particular problem. To do so, detailed information about the problem must be available.

5.3 Predictive analytics

It concentrates on what is likely to happen shortly. Historic data is used to predict the possible outcomes using algorithms. When compared to the first two, this is an advanced type of DA that uses sophisticated models and technologies to predict the future. Forecasting requires vigilant action and escalation [19].

5.4 Prescriptive analytics

It suggests an action for a particular situation. Depending on the current circumstances different strategies are suggested using advanced analytics techniques. It concentrates mainly on what action to take to eliminate the problems that occur in the future. Among the four this is the most type of DA.

In the current era, data dominates everyday life. Volumes of data get generated every fraction of a second. When the generated data is analyzed properly several hidden relationships among the data can be uncovered. In the past, organizations used to hire experts to perform research activities. The use of human beings for decision-making has its limitations such as the possibility of biasing, inaccurate decisions. As computers are accurate and cost-effective, they are used to perform PA. To provide valuable insights for an organization PA is extensively used. When the software, technology, and business infrastructure are continuously evolving there is a need to adopt a systematic procedure for effective decision-making. This is done by using PA which not only identifies the likely outcomes or

potential predictions but also explores new avenues and trends based on the historical data. The term "predictive analytics" may sound like intricate expertise which is used by large scale organizations. In reality, it is not so. PA can be applied to any process related to any organization irrespective of the size and type [20].

PA is considered as an innovative process that predicts the future trends accurately using AI and ML techniques. PA and forecasting are completely different in terms of operational outcomes. Forecasting is a method that analyzes the current data to elucidate future trends. Future trends and their impact on the organization are pre-determined using forecasting. Some important differences between PA and forecasting are listed in Table-1.

Table 1: Difference between PA and forecasting

| Forecasting | Predictive Analytics |
|---|---|
| Operates on time series data to predict the future | Requires different types of historic data to perform predictions |
| Does not require a vast amount of data | Requires volumes of historic data |
| Forecasting is very specific such as weather, sales, production, etc. | It covers many industries |
| Does not require complex software and models | Requires sophisticated software and models |
| Based on several probabilistic features | Entirely based on past performance |
| Major risks and uncertainties are not identified or rather hidden | Major risks and uncertainties in the system will be uncovered |
| Mostly based on single trend | Based on multiple trends and patterns |
| It has a short-term perspective, where what will happen in the near future will be identified | It has long-term perspective rather than predating what is going to happen in the near future |
| Tries to analyse the possible relationship between related data | Identifies hidden relationship among different unrelated data |
| Example: Predicting the oil price of tomorrow | Example: Predicting the impact of change in oil price |

PA begins with the statement of business objectives. It aims to make the best use of available data to minimize waste, save time, and cut down the expenses. The entire process makes the maximum use of massive heterogeneous data into models to generate actionable insights [21].

6. PA Models

A model is a logical or mathematical representation of a thing or a process. Models are developed to simulate real-world phenomena for investigation. Models make the concept clear, simple, and understandable. Whenever a specific problem is to be solved, a model is created. To develop a model that represents an event or an object, historical data related to the same needs to be gathered. When an adequate quantity of high-quality data is gathered, prediction becomes easy and accurate. Predictive models are designed to build predictions about future trends based on historical data.

Steps in creating a predictive model

Predictive modeling demands a team approach. It needs a group of people to solve the problem, perform analysis of the data, and generate the desired output. To improve performance, the model is to be refined several times. Developing a predictive model for a given problem is an iterative process. It requires both historical and real-time data. Historical data is used to train the model and real-time data is used to evaluate the accuracy of prediction. The Figure-n shows the block diagram of various steps involved in creating a predictive model.

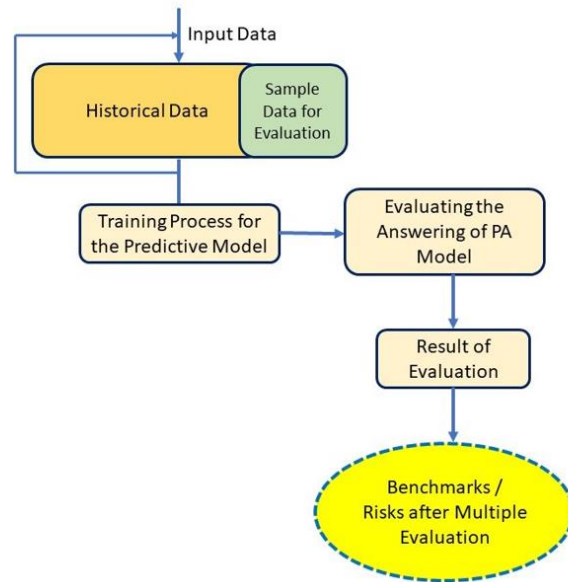


Fig. 2: Steps in developing a Predictive Model

Data scientists and information technology specialists are responsible for developing a PA model for an organization. They must create a PA model that is most suitable to the organization's business processes. In the case of SL, the learning algorithm generates the model by analyzing various samples. The model's most valuable input is the historical data. Fig. 2 depicts the various phases involved in constructing a PA model using supervised learning. The model is trained using 75 percent of the historical or training information. A model is constructed based on this data to forecast the future outcomes. The model must be constructed in such a way that it can make accurate predictions with the least amount of loss or error. In this scenario, the loss is due to a poor prediction. The model is trained repeatedly until it produces a result with the desired accuracy. If the forecast is correct, the loss value will be zero; otherwise, it will be a number greater than zero. After the model has been trained with the training data, it is tested with test data and the correctness of the outcome is assessed. A risk baseline is determined based on an assessment. During testing, the outcome of the PA model is evaluated to check if it makes the required contributions to the overall growth of the business organization. Training the model is repeated to improve the performance of the PA model [22].

Predictive models are data-driven as they require a lot of data. Different applications demand different types of models. Different types of predictive models are given in Fig. 3.

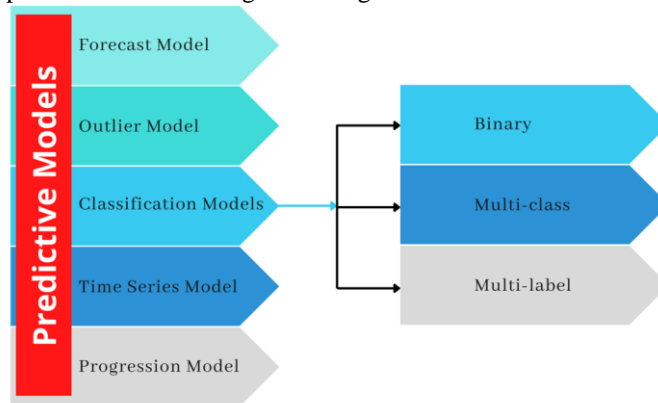


Fig. 3. Types of predictive models

6.1 Forecast Model

This is the most extensively used predictive model. It is associated with the prediction of metric value to estimate a new numeric value based on the currently available historical data. This model requires many parameters to be taken care of. If a hotel owner needs to estimate the number of customers during the week, the model has to take not only the historic data but also to consider the event that is going to happen in the locality.

Examples:

- A shop can determine the number of certain items to be stocked to meet the market demand during a specific period. This will help to avoid the out of stock or excess stock situations.
- A company can estimate the number of customers who turn out during a specific season.
- Hospital authorities can predict the number of patients with a specific illness come to avail treatment.
- When used for weather forecasting, it can predict the natural calamity such as a flood.

This model requires many parameters to be taken care of. If a hotel owner needs to estimate the number of customers during the week, the model has to take not only the historic data but also to consider the event that is going to happen in the locality.

6.2 Classification Models

It is the simplest of all PA models. It puts the data into different categories based on knowledge gathered from the historical data. It places the data into several categories based on the knowledge that is gathered from the historical data. In the case of the classification model, one class of data is assigned to the data set. There are mainly three types of classification models. They are

- Binary Classification Model
- Multi-Class Classification Model
- Multi-Label Classification Model

6.2.1. Binary Classification Model

It categorizes the data into two categories. It is used in situations where the model has to produce 'yes' or 'no' as an answer. Since there are only two class labels, it is called binary classification. This model can be applied to many different industries.

Examples:

It can be used to predict the behavior of the customer such as whether the customer is going to buy or not.

For the loan provider, it can be used to check whether there will be default or in time payback.

In the education sector, it can be used to predict whether a particular student will clear an exam or not.

In healthcare, it can be used to predict whether a patient is to be admitted to ICU or not

6.2.2. Multi-Class Classification Model

It refers to the classification task that has more than two class labels. Here the given dataset is classified into one among a range of known classes. Based on the problem the number of classes may vary. The sample is assigned to only one class label.

Examples:

Can be used to predict which class label a particular fruit belongs to. Fruit can be either apple or orange or pear, not all simultaneously. In this case, the number of class labels is limited.

Can be used in word prediction application that will predict the next word based on the one which is currently typed. Depending on the word, there may be several predictions.

Can be used in face recognition problem where a face needs to be compared with thousands of faces that are stored. If face recognition is used in an office, the number of comparisons required is equal to the number of employees.

Can be used in optical character recognition which will compare the image of the handwritten character to the one that is stored.

6.2.3. Multi-Label Classification

In this type of classification, a set of target labels are assigned to each sample. This model is capable of producing multiple outputs.

Examples:

In a picture containing multiple known objects, the model predict the objects such as ‘bicycle’, ‘tree’, ‘flower’, ‘people’, etc.

When a video clip is given, the model can classify it based on the contents as ‘political’, ‘game’, ‘art and culture’, ‘educational’, etc.

Based on the symptoms and the clinical parameters obtained the model can map the record into one of the diseases.

Based on the story of the movie, it can be classified as ‘comedy’, ‘drama’, ‘action’, ‘suspense thriller’, etc. by using the PA model.

6.3 Outlier Model

The forecasting and classification models are used to analyze the historical data where as outlier model works with anomalous data. The data that deviates from the normal range is called anomalous one. It is nothing but the unusual patterns in the data or the isolation of data in relation. This model is used in the financial industry for effectively detecting fraudulent transactions. The outlier model will detect fraud before it occurs thereby saving money. Therefore this model is exceptionally valued in the financial sector.

Examples:

When there is an unusual spike in the support calls indicates product failure.

Based on unusual transactions such as withdrawal or deposit of a huge amount of money a fraud can be detected.

Based on the unusual behavior of a person, an attempt to suicide or a crime can be detected.

When there is a deviation in the transactions, false insurance claims can be detected.

6.4 Time series model

This model focuses on the data where time is also an important parameter. It works on the data points that are collected at different time intervals and develops predictions. It is used by organizations when they want to know how a variable changes over time. It is also called the temporal model.

Examples:

Predicting the number of ACs to be stocked during the summer season.

Predicting the number of customers likely to visit during a specific time

Predicting the hotel bookings during an event in the city

Predicting the number of patients who are likely to visit during a specific season

6.5 Regression model

This model works on numeric data. It predicts the data values using statistical techniques based on the training data set. It maps several classes of events to either 0 or 1.

Examples:

It is used to predict an event as pass or fail

It is used to predict the winning probability in a game based on the required parameters.

It can be used to assess whether a person is healthy or sick based on the data provided.

When there is an image containing several objects such as lion, dog, tree, etc. it will map to 1 if an object is detected [23-24].

7. PA Process

7.1 Elements of PA Process

Fig. 4 shows different elements of the PA process. PA collaborative work of three different entities namely management, models, tools, and technologies. The three components are in the form of a triangle which means that these three elements collaborate to ensure the proper functioning of the PA process.

Management of the organization plays a key role in the PA process. The PA process has to be monitored continuously to ensure the right issue is addressed efficiently. Every project requires strong leadership or ownership for monitoring and control. Hence management is considered as the bedrock of the PA process.

The model used in the PA process analyzes the data to find out hidden relationships. PA model is considered as the heart of the PA process. It should ensure a cost-effective and efficient solution is provided to the problem under consideration.

PA process analyzes a huge amount of data. PA tools and technologies have to ensure that the right data is stored for analysis. Issues such as scalability, security, fault tolerance, etc. ensured by using appropriate tools and technologies.

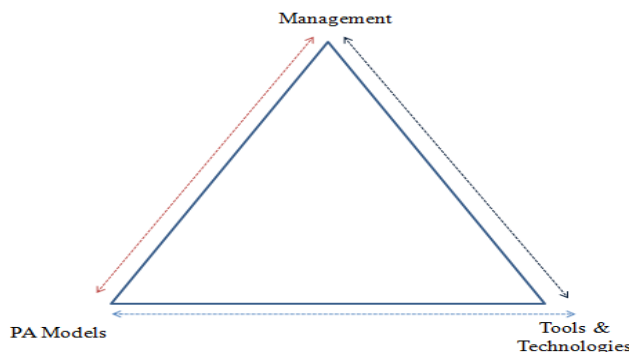


Fig. 4: Elements of PA process

6.2 Steps in PA Process

PA is a continuous process that tries to predict the future by identifying patterns in the available data. Various algorithms and techniques are used and the outcomes are evaluated. To obtain maximum success from PA steps shown in Fig. 5 are to be followed.

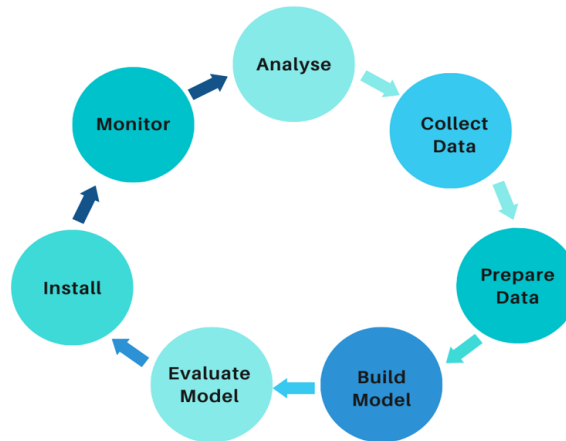


Fig. 5: Steps in PA

1. Problem analysis and identification of objectives: The problem must be defined properly, and the goals must be specified for PA to be effective. PA should be used in such a way that it contributes favorably to the existing system's performance.

2. Collection of required data: PA necessitates large amounts of data from different sources. Data can be gathered from both internal and external sources. Internal sources include numerous procedures relating to the organization, whereas external sources include social media, government, and a variety of other things.

3. Data preparation: This is the most critical aspect of the PA procedure. Because the data obtained from diverse sources will be in different forms, it cannot be used as is. 12-01-2019, for example, might mean both the 12th of January and the 1st of December 2019. This ambiguity must be clarified. The data may also contain noise. Bad data can't be used to make good decisions. As a result, data cleansing is critical. A training set is created by storing the cleaned data in the data storage.

4. Development of predictive model: Building a model is a common feature that everyone does. This concept is extensively used by architects, fashion designers, and engineers. Data analysts analyze the problem thoroughly and use one or more predictive models for PA. Several algorithms and tools are applied to construct a better PA model.

5. Evaluation of the model: The outcome of all the PA models is probabilistic. One model has to be chosen based on the accuracy of the result produced by the same. The output of models is evaluated for accuracy. If the result is found to be satisfied that the PA model is selected. If none of the PA models produce results with the desired accuracy that means data chosen is not good enough for PA.

6. Installation: This is the last step of the PA process where the model is deployed in the business environment and is executed based on real-time data.

7. Monitoring: It is very much essential to assess the effectiveness of the model as the results generated by it is probabilistic. As time progresses business scenarios and policies are going to change. New business rules will have to be added. To obtain favorable results from PA the entire process is to be repeated. This will ensure the sustainability of the business process in a competitive environment [25].

8. Analytical Techniques

Learning is very much essential to extract knowledge from the growing volumes and wide varieties of data sets. It is a powerful and affordable mechanism that analyzes complex data and produces accurate results. PA uses different types of learning techniques as shown in Figure-5.

1. **Supervised Learning:** In this technique, there will be a training data set that is already tagged with a proper label. When the new data set is given, the learning algorithm analyses it and produces proper outcomes based on the training data set. There are two types of supervised learning techniques:

- a. **Classification:** It is a technique that is used to fit the categorical data into a known structure called class labels. Common classification algorithms include decision trees, NB classifiers, neural networks, etc.

- b. **Regression:** It is a technique that works on statistical data. It is mainly used to understand the relationship between two different variables. It is applied to analyze how the value of a dependent variable changes when the value of the independent variable is changed. There are two types of regressions-linear regression and logistic regression.
2. **Unsupervised Learning:** In this technique, the data is not labeled or classified. The system tries to learn by itself based on the similarities without any guidance or supervision. Common unsupervised learning techniques include clustering, association rule mining, etc.
3. **Semi-Supervised Learning:** Semi-supervised learning is a ML technique that involves training using a tiny portion of labeled data and a large quantity of unlabeled data. It falls between unsupervised learning and supervised learning. Applications such as text document analysis, speech analysis, internet content classification, etc. use this technique.
4. **Reinforced Learning:** The goal of reinforcement learning is to make judgments in a sequential manner. In simple terms, the output is determined by the current input's state, and the next input is determined by the previous input's output. The initial input or the input delivered at the start is decided in supervised learning. Because decisions in reinforcement learning are dependent, we assign labels to sequences of dependent decisions. Gaming, robotic process automation, navigation etc. are the different types of applications that use this technique [26-27].

9. Adaptive Technologies

9.1 Internet of Things (IoT)

IoT is a revolutionary technology that allows "people" and "things" to connect "anytime," "anywhere," with "anything," and "anyone." The new technology will make an individual's day-to-day existence easier, simpler, and better. The healthcare industry is one of the most popular and widespread uses of IoT. Its goal is to offer people with 'smart' healthcare. It is a technology that connects several heterogeneous devices. It is much used in the automated data collection process. Sensors are deployed on devices that generate data so that they send the acquired data to the PA system automatically. This enables the PA process to collect the required data quickly. IoT is extensively used in healthcare PA where wearable sensing devices are given to patients. These devices measure the required parameters such as blood pressure, heart rate, etc, and send the same to the PA system. Manufacturing industry and agriculture are two other sectors that extensively use sensing devices in PA [28]

9.2 Cloud Technology

The huge amount of heterogeneous data from various sources that are needed for the PA is stored efficiently. This requires a central repository that is flexible as well as scalable. DBMS may not be the best solution to store volumes of data because of its inherent limitations. Cloud computing provides resources for storing large amounts of data. It also provides computational power. Cloud offers exceptional flexibility and unlimited storage. Data, files, reports, and softwares can be stored on the cloud platform and can be accessed remotely using the Internet. The cloud platform is reliable and affordable. When storing sensitive data on the cloud, security, and privacy issues need to be taken care of [29].

9.3 Artificial Intelligence (AI)

AI can be treated as the driving force behind the success of PA. AI-based decision support systems are increasingly used for performing predictions in several fields such as manufacturing, marketing, healthcare, and many more. It is through AI the institutions analyze missing or incomplete data. The much needed computational power of AI algorithms is extensively used in building PA models. The techniques of AI such as natural language processing (NLP) and rule-based expert systems are extensively used in PA. It is through AI self-driving cars, robots and recommendation systems are realized [30].

9.4 Machine Learning (ML)

ML employs statistical and probabilistic techniques to enable computers to learn from previous instances and uncover difficult-to-find patterns in large amounts of noisy or complex data. In predictive analytics, several machine learning methods are widely employed. They are categorized into supervised, unsupervised, semi-supervised, and reinforcement learning strategies depending on the nature of the learning. Learning from examples is supervised learning. The supervised learning approach trains a model to provide realistic predictions for the response to new data using a known set of input data and known responses to the data (output). There is no training set in the case of unsupervised learning. The new test data provides the necessary insight. Both labeled and unlabeled training data are used in semi-supervised learning. It is useful in cases where data labeling is more expensive. Reinforced learning is a method of training a computer to make accurate decisions without the need for human intervention. It's usually employed in online games, robots, and navigation.[31].

10. PA Industry Applications

10.1 Banking

In any financial institution, there will be several risks that creep up stealthily. When small losses are summed up it may lead to misfortune for an individual or an organization. Hence banks and other financial organizations use PA to predict the defaulters of mortgage loans. They also make use of PA to identify fraudulent financial transactions. In banking, PA is used

- To provide the best services to the customers based on their preferences
- To perform resource and workforce optimization
- To increase the efficiency and security of transactions
- To assign credit scores to the customers based on how they pay back the loans
- To prevent identity theft.

10.2 Retail

In the retail industry, PA is used to understand the buying habits of the customers. 80% of the customers who buy eggs are likely to buy bread and onions. Such facts and hidden relationships exist between several items. These facts will be used in prize optimization. Customer behavior is analyzed to gain profit. Market trends are analyzed by using PA to avoid excess stock as well as out of stock situations. To stay ahead in the competitive market, online retailers extensively use PA

- To suggest different products based on customers' choice
- To suggest personalized offers to specific customers
- To maintain optimum inventory level
- To analyze customer behavior and trends.

10.3 Healthcare

The healthcare sector extensively uses PA for providing rapid, accurate, and cost-effective solutions. Healthcare data comes from various sources such as sensors, wearable devices, clinical records, clinical reports, electronic health records, and so on. The analysis of such data is a challenging one. The clinical data is used for effective decision-making. Prediction of diseases before the actual symptoms are observed will have greater benefits. It is extensively used in the clinical decision support system to determine which patients are at the risk of developing certain chronic diseases like diabetes, asthma, heart disease, and other prolonged ailments. PA in healthcare aims at providing better solutions to improve the quality of life of patients at considerably reduced healthcare costs. PA can be used in healthcare

- To analyze the exact condition of the disease and to predict its future progress
- To provide continuous care for the patients
- To effectively manage the resources and services available in hospitals
- To respond efficiently during emergencies.

10.4 Education

An increase in the dropout ratio has become a problem for colleges and universities. The educational institutions are increasingly using the demographic data of students to keep track of their courses and provide the required support to ensure that the student does not fall behind. Analyzing historic data related to students and predict what the student is going to do has helped the institutions a lot in terms of their advertising and enrollment policies. Colleges make use of PA in the following situations:

- To generate alerts based on the academic and demographic data of the student. If a student is irregular and does not show interest in academic activities, the PA system can analyze the student data and suggest solutions to the problem.
- To use adaptive technologies for making the student better understand the subject.
- To facilitate a customized learning environment by identifying the skills, knowledge level, and understanding ability of students.
- To provide support such as counseling, career guidance, medical, financial, transport, etc based on the necessity.

10.5 Manufacturing

PA helps the manufacturers to forecast the demand and manufacture the items based on the demand. PA provides insights by identifying unknown relationships among different variables used for manufacturing. It is used for the following purposes:

- To identify the demands and trends
- To forecast workforce efficiency
- To perform quality improvement
- To carry out resource management
- To comprehend preventive maintenance
- To improve the efficiency of the manufacturing process.

10.6 Supply Chain Management (SCM)

SCM uses PA to forecast the customer requirements accurately for maintaining the stock levels in the inventory. The use of PA has made the SCM process more efficient and transparent. Various fields of SCM that use PA include product pricing, inventory management, demand forecasting, logistics, and maintenance. SCM uses PA for

- Managing volatility of demand
- Supply chain optimization
- Making the supply chain more transparent
- Improving inventory planning in the case of large retailers.

10.7 Smart Services

This domain is extensively using PA for developing ‘smart’ devices and applications. Smart homes, smartwatches, smart energy management, smart appliances, smart traffic, etc. make the best use of PA for distributing real-time data. Business organizations combine PA and IoT to provide better services to their customers. Logistic industry is the best example of this. It uses e-logs, sensors, and other devices as well as PA to resolve the challenges in transportation. Smart cold storages, smart traffic management, smart homes, etc. can save a lot of scarce resources such as electricity and fuel. [32].

10.8 Agriculture

PA is extensively used for quick decision-making in the agriculture sector. It is used to increase efficiency in crop production. When used in agriculture, PA will help farmers to produce higher yields thereby increasing profitability. It can be used

- To analyze the spatial data and suggest the best crop suitable
- To forecast weather to avoid loss of crop
- To suggest measures to be taken for a crop such as spraying of pesticides
- To suggest an adequate quantity of pesticides and fertilizers .

10.9 Competitive Sports and Games

Sports analytics has gained a lot of focus these days. sports data is extensively being used by different teams to increase the competitive edge. PA enables athletes and coaches to gain more insights on the data and enhance the performance. Sports use PA to

- Predict possible injuries because of inappropriate moves
- Predict fatigue and mental stress of athletes
- Predict the performance of athletes based on their fitness
- Identify winning strategy
- Provide customized training to players
- Pre-match and post-match evaluation, etc.

10.10 Government

These days governments are more focusing on preventing the problems rather than reacting to them when they occur. The advancement in PA has enabled governments to focus towards preempting the problems. From spotting frauds to combatting with the pandemic, a small preventive measure by the government will be of great help to the citizens. Governments extensively use PA to

- Identify hotspots of criminal activities and take necessary measures
- Identify suspicious activities such as human or child trafficking, drug peddling, etc
- Predict cyberattacks
- Prevent accidents
- Prepare for natural disasters such as floods [33-35]

11. Supervised Learning Algorithms

SL uses labelled data needed for the analysis. This data is collected from several resources and will be there in different forms such as text, numbers, video, images, etc. and a learning process is developed by using training data as input. Some predominantly used SL algorithms are explained in this section.

11.1 Linear Regression

It is a well-known ML algorithm used to identify the hidden relationship between single input attribute X and its outcome Y. It is used in problems where the data is continuous. It attempts to find the hidden relationship between two attributes. It works effectively on continuous numeric data. The method of least square is used to calculate the accuracy of prediction. Here the value of output variable Y can be calculated in terms of input variable X using the equation. The equation is $y = a_0 + a_1x + \epsilon$ where a_0 and a_1 are coefficients and ϵ is the error term. It is extensively used in problems such as sales forecasting, product pricing, predict the risk of investment, etc.

Strengths

- It is simple
- Works well for the continuous data
- Best when there is only one dependent variable and one independent variable

Weaknesses

- Not suitable when there are several dependent variables.
- Not recommended for practical applications as the analyser tries to build a linear relationship between the variables

- The outcome is much dependent on outliers.
- It assumes that the attributes selected for analysis are independent.

11.2 Logistic Regression

This algorithm works on both numeric and categorical data. It uses statistical functions to perform future predictions. It is used to calculate the probability of the occurrence of an event Y based on the value of a parameter X. It is most suitable for binary classification. The logistic function used to calculate the probability of occurrence of event Y based on a variable X is also called as sigmoid function as it takes a 'S' shaped curve whose values lie between 0 and 1. It works well on categorical dependent variable using one independent variable. It is based on maximum likelihood estimation is used to measure accuracy. The equation is $\log \left[\frac{y}{1-y} \right] = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$ where Y is the dependent variable and x_i are the values of independent variables and $b_0 \dots b_n$ the coefficients of variable x_i . This algorithm can be used for spam detection, credit card fraud detection, cancer prediction, etc.

Strengths

- It is simple and easy to implement, fast and less prone to overfitting
- Performs a probabilistic interpretation
- Used when the dependent variable is binary (True/ False, Yes/No, 0/1)
- Gives the relationship between positive and negative association of variables

Weaknesses

- Not flexible for complex relationships
- Non-linear problems cannot be solved using this technique
- It is sensitive to outlier data.
- Requires large datasets and sufficient training examples for all categories

11.3 Support Vector Machine (SVM)

It is a classification algorithm that is used to categorize the given data into different classes. Popularly it is a binary classification algorithm that assumes given data is partitioned into two classes. It is a simple kernel-based algorithm that classifies given data into two classes. It is extensively used to handle problems associated with image analysis and healthcare predictive analytics. In this algorithm, the data is plotted in a n-dimensional space for analysis where n is the total number of features. Application areas include Face detection, facial expression classification, handwriting recognition, image classification.

Strengths

- Simple and accurate
- Capable of reducing redundant information
- Stable because the output is not affected by a small change of data

Weaknesses

- Not suitable when the size of dataset is huge
- Shows poor performance when there are overlapping classes
- Does not perform well if there is noise in the dataset
- There is no probabilistic explanation for classification

11.4 Decision Trees

It is a SL algorithm that is used to classify the training data based on decisions. In this technique the training data is divided based on certain parameters. It contains two components namely 'nodes' and 'leaves'. Every node in a decision tree represents a decision and leaves outcomes of a decision. It can be used to represent data graphically without knowledge of statistical functions. It gives good visual representation of the decision-making process. It works on both numerical and categorical data and is suitable for both classification and regression problems. For

problems with a Yes/No answer, a decision tree is used to predict the outcome. It can be used to make decisions in almost any field, including healthcare, education, banking, and manufacturing. A decision tree's building is a process in which the tree grows exponentially in response to the data. The larger the decision tree, the more complex it becomes, but it is always better for analysis. Starting at the root, the tree is created upside down. There will be multiple splits spanning a massive tree when the problem under investigation is big.

Strengths

- Decision Trees usually mimic human thinking ability while making a decision, so it is easy to understand.
- It can be very useful for solving decision-related problems.
- It helps to think about all the possible outcomes for a problem.
- Applicable where the given data results in two outcomes
- There is less requirement of data cleaning compared to other algorithms.

Weaknesses

- There is a possibility of overfitting issue
- If there are more class labels its complexity will increase
- Looks complicated when the splitting criterion yields to more than two outcomes

11.5 Naïve Bayes Algorithm

It is a simple classification technique where the class labels are represented in the form of vectors. It can be used to classify the given input effectively. It uses the maximum likelihood approach of probability when categorizing the given data. It is based on the Bayes theorem for classification. It is called naïve as it assumes the probability of occurrences of a feature is independent of probability of occurrence of some other features. It classifies vast amount of training data quickly and efficiently. It predicts the outcome on the basis of probability theory by using the formula $P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)}$ where $P(A/B)$ is the probability of hypothesis A on the observed event B, $P(A/B)$ is the probability of evidence given that the probability of the evidence given that the hypothesis is true. $P(A)$ is the probability of hypothesis before observing the evidence and $P(B)$ is the probability of evidence. This algorithm is extensively used in problems such as spam filtering, sentiment prediction, text classification, etc.

Strengths

- It is simple and hence is easy to implement
- Large amount of training data is not necessary to train the predictive model
- Suitable for multi-class training data
- Works efficiently on categorical data rather than numeric

Weaknesses

- Assumes all predictor attributes are independent of each other. In real-life it is not so. There will be some relationship between different attributes
- There is more chance for improper predictions as it assigns value 0 when it is not able to make predictions [36-37].

11.6 Random Forest Algorithm

RF is a supervised learning algorithm involving "forests," which are collections of numerous decision-trees. The decision-trees generated by using either the "boosting" or "bagging" techniques. It is widely used for accurate decision-making in the healthcare, business, and other sectors. It can be used to solve problems involving classification and regression. To produce predictions, random forests use predicted values from a set of trees. Each tree forecasts the result as a function of the predictor variables' values. The RF algorithm's decision-making process begins with the root node X and then divides into many nodes. This step is repeated until the leaf node is reached. There will be a question in each node, and the branches will reflect several options that the node's inquiry could lead to. RF is made up of multiple trees, each of which has the identical nodes but uses different data to get to the leaf node. Finally, the trees are merged, and the result is the average of all decision trees.

Strengths

- Suitable for both numerical and categorical data
- Applicable for classification and regression type of problems
- Suitable for handling complex tasks involving multiple variables
- Accuracy of prediction is high
- It is robust to outliers and detects the outliers automatically
- It automatically checks for missing data

Weaknesses

- Creates numerous decision trees and hence requires more computational resources
- Requires more time to train the data [38]

12. Issues and Challenges**Benefits**

- **Easy to predict:** With the help of PA it is easy to predict future performance as it performs ‘what-if analysis’. The PA can reduce the number of risks by predicting them in advance.
- **Optimum allocation of resources:** Resource allocation plays a major role in the success of any business. PA ensures the resources in the organization are utilized optimally.
- **Fast, efficient, and cost-effective:** Use of advanced technology in the analysis process makes it fast, efficient without causing additional financial burden on the organization.
- **Decision-making:** It is a good tool that provides better decision-making capability which is very much essential for the success of a business organization.
- **Identify suspicious trends:** It is easy to identify suspicious trends in the transactions and avoid future loss.
- **Identify customer needs:** PA can easily identify the needs of the customers and forecast future trends. This helps in product optimization which in turn increases the profitability.

Challenges

- **Overwhelming with information:** PA requires a wide range of data several processes of an organization. In addition to this, it may also require data from external sources. Many organizations find it difficult to leverage the information obtained from their everyday transactions. A large amount of data gets collected too much of the collected data is not used in the analysis.
- **Risk analysis is risky:** The important task of PA is to uncover the risks related to the problem under consideration. It requires the right data to be given to the right person for the analysis. If one thing goes wrong major risks remain uncovered.
- **Biases in selecting algorithms, tools, and models:** The outcome of PA depends heavily on the algorithm, tool, and predictive models. Since the models, techniques, and tools are selected by humans, there is a chance of bias when choosing them.
- **Storage and backup for the data:** when there is a large amount of data, selecting the right storage medium is a very important factor because the proper storage medium has a high impact on the time required for retrieval.
- **Data quality:** Good quality data leads to more accurate results. Acquiring high-quality data in a consistent format is a difficult task as data will be there in different formats. Added to this, there will be missing as well as invalid data. Hence acquiring good-quality data is a real challenge.
- **Multi-criteria based decision:** In a real-life business environment there will be multiple processes with conflicting criteria that are to be evaluated for decision-making. Prediction based on the multiple conflicting criteria is a challenging one [39-40].

13. Discussion

All predictive models that are used in SL are data-driven. Hence it is very much essential to consider all attributes that are crucial for analysis. There is a possibility that the inclusion of a new variable can change the performance. Hence it is very much essential to examine the accuracy of prediction after adding or removing an attribute. With

regards to data which is the key ingredient for prediction there are several issues and challenges such as data validity, storage and management. If these issues are addressed properly then only the prediction can be done easily. The accuracy of predictions depends upon the predictive model used to implement the issues related to an organization. Unfortunately, the performance of predictive models is not accurate as expected. Hence a flexible, robust, reliable and sustainable predictive model which is most suitable for the given business application is the need of the hour.

Predictive technology is growing at a rapid rate and it is becoming more and more sophisticated each passing day. But the level of knowledge and expertise within the organization is not advancing at the same pace. There is a huge scarcity of analysts with right skillsets. Missing essential skillsets acts as a hinderance to the digital transformation. Organizations are looking for someone who have a good understanding of PA and how it impacts the company. In reality, there is a disconnect between PA and business processes as analysts are well-versed in PA but not in the subtleties of the enterprise. The data analysts must be capable of analyzing the business application, evaluating the essential data, forming the hypothesis and performing the investigations. It will be advantageous if a single person understands both PA and its implications on the business.

Undoubtedly PA has become a powerful tool which can be adopted by any organization for making faster, more accurate and smarter decisions. But in reality, when going for digital transformation balancing people, data and technology is really tough task. This ideological shift requires not just huge investments but training the staff from the ground up in data analysis.

The same SL algorithm can generate different results for the same data across various study settings. When data is given to multiple SL algorithms, the outcome may be of different accuracy. Hence selecting the right algorithm for a given problem is very much essential. Also, in this study the authors considered the SL algorithms that are extensively used in making predictions. Their variants and sub-classifications are not considered.

14. Conclusion

In a highly volatile global economy, the only way to survive is to apply the knowledge or insights gained through analysis. PA has become a significant guidance in the decision-making process as data quality has improved and modern computational technologies have been used. Companies, the manufacturing industry, banking, e-commerce, hospitality, healthcare, government, charity, enforcement agencies, and other fields have all benefited from PA. All industries have grown intrinsically tied to technology. The algorithm used to construct the predictive model, as well as the precision of the data, determine prediction accuracy. It also depends on the attributes that were considered to conduct the analysis. The quality of data acquired has a significant impact on data analysis. Collecting high-quality data necessitates extreme efficiency and use of a large number of scarce resources. Effective data analysis will enable the company to be pro - active, and will help to ensure the long-term viability of quality standards. According to the "no free meal hypothesis" of machine learning, there is no single algorithm that performs well for every application. As a consequence, developing an efficient algorithm for the given problem is challenging. Because the prediction system's outcome is so crucial, analysts must select the optimum solution. The complexity of the problem, the availability of training data, and processing power all play a role in the success of the machine learning technique employed to address a specific problem. Furthermore, the predictive model will be developed in accordance with the organization's requirements. Developing a predictive model is entirely dependent on the data analyst's perception. In a nutshell, an effective prediction system will be an immeasurable asset to an organization.

References

- [1] Geetha Poornima K. & Krishna Prasad, K. (2020). Data Analytics Solutions for Transforming Healthcare Information to Quantifiable Knowledge – an Industry Study with Specific Reference to ScienceSoft. *International Journal of Case Studies in Business, IT, and Education (IJCSBE)*, 4(1), 51-63.
- [2] Geetha Poornima K. & Krishna Prasad, K. (2020). Integrated Prediction System for Chronic Disease Diagnosis to Ensure Better Healthcare. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 4(1), 25-39.
- [3] Krishna Prasad, K., Aithal, P. S., Bappalige, Navin N., & Soumya, S., (2021). An Integration of Cardiovascular Event Data and Machine Learning Models for Cardiac Arrest Predictions. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 5(1), 55-54
- [4] K., Geetha Poornima & Krishna Prasad, K. (2020). Application of IoT in Predictive Health Analysis—A Review of Literature. *International Journal of Management, Technology, and Social Sciences (IJMTS)*, 5(1), 185-214.
- [5] Nithya, B., & Ilango, V. (2017, June). Predictive analytics in health care using machine learning tools and techniques. In *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)* (pp. 492-499). IEEE.

- [6] Zhu, P., & Sun, F. (2020). Sports Athletes' Performance Prediction Model Based on Machine Learning Algorithm. In *Advances in Intelligent Systems and Computing* (Vol. 1017). Springer International Publishing.
- [7] Krishna Prasad K., Aithal P. S., Geetha Poornima K., & Vinayachandra, (2021). Tracking and Monitoring Fitness of Athletes Using IoT Enabled Wearables for Activity Recognition and Random Forest Algorithm for Performance Prediction. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 5(1), 72-86
- [8] Asri, H., Mousannif, H., Al Moatassime, H., & Noel, T. (2016). Using machine learning algorithms for breast cancer risk prediction and diagnosis. *Procedia Computer Science*, 83(1), 1064-1069.
- [9] Mokha, M., Sprague, P. A., & Gatens, D. R. (2016). Predicting musculoskeletal injury in national collegiate athletic association division II athletes from asymmetries and individual-Test versus composite functional movement screen scores. *Journal of Athletic Training*, 51(4), 276-282.
- [10] Abdullah, M. R., Maliki, A. B. H. M., Musa, R. M., Kosni, N. A., & Juahir, H. (2016). Intelligent prediction of soccer technical skill on youth soccer player's relative performance using multivariate analysis and artificial neural network techniques. *International Journal on Advanced Science, Engineering and Information Technology*, 6(5), 668-674.
- [11] Tang, Fei; Ishwaran, Hemant (2017). Random forest missing data algorithms. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 2(1), 1-15.
- [12] Manogaran, G., & Lopez, D. (2017). A survey of big data architectures and machine learning algorithms in healthcare. *International Journal of Biomedical Engineering and Technology*, 25(2-4), 182-211.
- [13] Naglah, A., Khalifa, F., Mahmoud, A., Ghazal, M., Jones, P., Murray, T., ...& El-Baz, A. (2018). Athlete-customized injury prediction using training load statistical records and machine learning. In 2018 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT) (pp. 459-464). IEEE
- [14] Apostolou, K., & Tjortjijis, C. (2019). Sports Analytics algorithms for performance prediction. 2019 10th International Conference on Information, Intelligence, Systems and Applications (IISA), 1-4
- [15] Simsekler, M. C. E., Qazi, A., Alalami, M. A., Ellahham, S., & Ozonoff, A. (2020). Evaluation of patient safety culture using a random forest algorithm. *Reliability Engineering & System Safety*, 204(1), 1-9.
- [16] Oytun, M., Tinazci, C., Sekeroglu, B., Acikada, C., & Yavuz, H. U. (2020). Performance Prediction and Evaluation in Female Handball Players Using Machine Learning Models. *IEEE Access*, 8, 116321- 116335.
- [17] Iwendi, C., Bashir, A. K., Peshkar, A., Sujatha, R., Chatterjee, J. M., Pasupuleti, S., Mishra, R., Pillai, S., & Jo, O. (2020). COVID-19 patient health prediction using boosted random forest algorithm. *Frontiers in Public Health*, 8(2), 1-9.
- [18] Krishna Prasad, K., Aithal, P. S., Geetha Poornima, K., & Vinayachandra, (2021). An AI-based Analysis of the effect of COVID-19 Stringency Index on Infection rates: A case of India. *International Journal of Health Sciences and Pharmacy (IJHSP)*, 5(1), 87-102.
- [19] Khalifa, M. (2018). Health Analytics Types, Functions and Levels: A Review of Literature. *ICIMTH*, 137-140.
- [20] Doyle, A., Katz, G., Summers, K., Ackermann, C., Zavorin, I., & Lim, Z. (2014). The EMBERS Architecture for Streaming Predictive Analytics. 11-13.
- [21] Husák, M., Komárková, J., Bou-Harb, E., & Čeleda, P. (2018). Survey of attack projection, prediction, and forecasting in cyber security. *IEEE Communications Surveys & Tutorials*, 21(1), 640-660.
- [22] Crisci, C., Ghattas, B., & Perera, G. (2012). A review of supervised machine learning algorithms and their applications to ecological data. *Ecological Modelling*, 240, 113-122.
- [23] Alanazi, H. O., Abdullah, A. H., & Qureshi, K. N. (2017). A critical review for developing accurate and dynamic predictive models using machine learning methods in medicine and health care. *Journal of medical systems*, 41(4), 69.
- [24] Candanedo, I. S., Nieves, E. H., González, S. R., Martín, M. T. S., & Briones, A. G. (2018, August). Machine learning predictive model for industry 4.0. In *International Conference on Knowledge Management in Organizations* (pp. 501-510). Springer, Cham.
- [25] Ogunleye, J. (2014). The concepts of predictive analytics. *International Journal of Developments in Big Data and Analytics*, 1(1), 86-94.
- [26] Vashisht, P., & Gupta, V. (2015, October). Big data analytics techniques: A survey. In 2015 International Conference on Green Computing and Internet of Things (ICGCIoT) (pp. 264-269). IEEE.
- [27] Praveena, M., & Jaiganesh, V. (2017). A Literature Review on Supervised Machine Learning Algorithms and Boosting Process. *International Journal of Computer Applications*, 169(8), 32-35
- [28] Gubbi, J., Buyya, R., Marusic, S., & Palaniswami, M. (2013). Internet of Things (IoT): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7), 1645-1660.
- [29] Garcia Lopez, P., Montresor, A., Epema, D., Datta, A., Higashino, T., Iamnitchi, A., Barcellos, M., Felber, P., & Riviere, E. (2015). Edge-centric Computing. *ACM SIGCOMM Computer Communication Review*, 45(5), 37-42.
- [30] Davenport, T., & Kalakota, R. (2019). DIGITAL TECHNOLOGY The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, 6(2), 94-102.
- [31] Marcelino, P., de LurdesAntunes, M., Fortunato, E., & Gomes, M. C. (2021). Machine learning approach for pavement performance prediction. *International Journal of Pavement Engineering*, 22(3), 341-354.
- [32] Mishra, N., & Silakari, S. (2012). Predictive analytics: A survey, trends, applications, opportunities & challenges. *International Journal of Computer Science and Information Technologies*, 3(3), 4434-4438.
- [33] Shin, S. J., Woo, J., & Rachuri, S. (2014). Predictive analytics model for power consumption in manufacturing. *Procedia CIRP*, 15, 153-158.
- [34] Schoenherr, T., & Speier-Pero, C. (2015). Data Science, Predictive Analytics, and Big Data in Supply Chain Management: Current State and Future Potential. *Journal of Business Logistics*, 36(1), 120-132.
- [35] Harris, S. L., May, J. H., & Vargas, L. G. (2016). Predictive analytics model for healthcare planning and scheduling. *European Journal of Operational Research*, 253(1), 121-131.
- [36] Gianey, H. K., & Choudhary, R. (2018). Comprehensive Review On Supervised Machine Learning Algorithms. *Proceedings - 2017 International Conference on Machine Learning and Data Science, MLDS 2017, 2018-January*, 38-43.
- [37] Decision Trees Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39(4), 261-283.

- [38] (Random Forest) Naghibi, S. A., & Ahmadi, K. (2017). *Application of Support Vector Machine , Random Forest , and Genetic Algorithm Optimized Random Forest Models in Groundwater Potential Mapping*.
- [39] Bravo, C., Saputelli, L., Rivas, F., Pérez, A. G., Nikolaou, M., Zangl, G., ...& Nunez, G. (2014). State of the art of artificial intelligence and predictive analytics in the E&P industry: a technology survey. *Spe Journal*, 19(04), 547-563.
- [40] Thakkar, A., & Lohiya, R. (2021). A review on machine learning and deep learning perspectives of IDS for IoT: recent updates, security issues, and challenges. *Archives of Computational Methods in Engineering*, 28(4), 3211-3243.
