

# Enhancing Program Management with Predictive Analytics Algorithms (PAAs)

Bongs Lainjo

**Abstract**—The increase in the amount of big data and the emergence of analytics technologies has created the opportunity for applying algorithm development techniques using machine learning (ML) languages to predict future events. The use of predictive analysis algorithms provides a mechanism for the utilization of organizational databases, files in equipment, videos, images, and other types of data to predict future incidences or events. To conduct inclusive analyses of contemporary literature of existing relevant narratives with a focus on program management themes, including state-of-the-art methodologies on current plausible predictive analytics models. The methodology used is the review and applications of programming platforms available that can be applied in the analyses of data to predict future outcomes. Program management requires the utilization of the existing machine learning languages in understanding future events and enabling effective preparations among stakeholders to make strategic decisions that enable the achievement of their goals, objectives, and missions. The use of predictive analytics algorithms has gained thematic significance in automotive industries, energy sector, financial organizations, industrial operations, medical services, governments, and academic institutions. Predictive analytics algorithms are important in promoting the management of future events such as workflow or operational activities in a manner that organizations and institutions can schedule their activities and tasks in order to optimize performance. It also ensures that organizations use existing big data to predict future performance and mitigate risks. The improvements in information technology and data analytics procedures have resulted in the ability of businesses to make effective use of historical data regarding their performances to predict future events. This enables evidence-based planning, mitigating risks, and improvement of operational efficiency. PAA's models have enabled accurate prediction of performance of companies and planning for increased demand for the products and services they provide.

**Index Terms**—Models, predictive-analytics-algorithms, program-implications thematic significance.

## I. INTRODUCTION

This chapter examines the current knowledge and scholarly information about predictive analytics algorithms (PAAs) by focusing on the concept of working principles on which they are used to predict future events and the procedures followed in creating them. The PAAs have been used extensively in predicting future events in healthcare practice, manufacturing companies, businesses, education, sports, and agriculture.

The main programming languages used to create PAAs are Java, C, and Python amongst others. The forms of algorithms that are commonly used are brute force algorithm, simple recursive algorithm, backtracking algorithm, randomized algorithm, and dynamic programming algorithms.

## II. BACKGROUND

Over the years, the concept and principles of data management have remained mostly unchanged. What has changed, however, includes the introduction of a complex, state-of-the-art, sophisticated, and integrated technological ecosystem: big data, cloud computing, and analytics [1]. The dynamics of this system have moved the way data are managed to a higher level, and institutions (public, private, sports, healthcare, and more) have capitalized on this! They have maximized their respective productivity levels using these systems with no reservations. As expected, these innovative developments come with significant risks from reliability to privacy and security concerns. Data are only as good and useful as their level of validity and reliability. Analytics, mentioned earlier, is one of the major components of the ecosystem that is used in transforming data into information. It is a sub-system that is also as useful as the reliability of the data used in performing different analytical interventions. At the conceptual level, analytics is an algorithm-driven strategy [2]. It facilitates the transformation of complex (generally) historical data sets into meaningful outcomes used for predicting future events. Its effectiveness has transformed and refined different sets of intended results. Institutions have used its predictive capabilities to optimize resources, streamline activities and increase productivity—ultimately becoming more competitive. The key players involved in the management and utilization of these ecosystems are the service providers (SPs) and their clients (users) [3].

It has been difficult for equipment manufacturers to develop innovative products using hardware alone. Those involved in product development have been able to add capabilities by applying solutions that improve customer satisfaction and value creation. Predictive analytics programs and equipment have been effective in promoting the anticipation of failures and provide forecasts for energy requirements while reducing the cost of operations. Predictive analytic models are used by companies in developing forecasts and creating plans for better utilization of resources. Before PAAs are used, the developer must review the available data and create/test mathematical models that incorporate computational processes in predicting future outcomes. The models provide forecasts of future outcomes

based on a particular metric such as the associated parameter changes.

This paper looks at the scope, thematic applicability, challenges, and prognoses of predictive analytics with life case studies from different institutions. It also highlights limitations, implications, and potential vulnerabilities. In this study, a select number of key institutions are included. These serve as examples of classical life case studies meant to help readers resonate with their own different and unique challenges. The various organizations are reviewed and analyzed on multi-dimensional thematic platforms. These include problem statements, strategic approaches, relevant processes, algorithmic layouts, programming descriptions, pilot testing, process reviews, initial implementation, and challenges and lessons learned. The relevant contents of these themes are only limited by the inability to access reliable, valid, evidence-based, useful, and compelling sources of information. Every attempt is made to address these limitations, and at the same time, prioritize available sources based on their pragmatic perspectives, simplicity, and authenticity. The select institutions include business (e-commerce, banking, finance, marketing, and more), health, education, government, sports, agriculture, social media, and so on. One invaluable approach applied in developing this narrative is an extensive review of available and contemporary literature. While the topic remains new and evolving, available documentation does indicate an inclusive degree of participation by different stakeholders. Key limitations like technical inability to develop and implement the various models have not been a significant deterrent. Readers need to consider this paper as an evidence-based, knowledge-sharing cursory or soft-core and easy to understand demonstration of the strength, scope, and application of PAAs in optimizing program management challenges.

### III. QUALITY OF DATA (QOD)

My experience dealing with data of different types and categories spans over four decades. From attending a survey technician-training program after high school to studying in an engineering school, data management has played and continues to play a very significant role in my professional life. As well, the challenges encountered over this period of time continue to evolve exponentially! The most recent paradigm transformation in data management is in the proliferation of analytics — a domain that has enabled businesses, industry, academia, banks, etc. to exhale and address competing forces with might and vitality.

One adage that strongly and appropriately describes different forms of data is “garbage in garbage out” (GIGO). Interestingly, this adage is not just limited to conventional data as described in the previous paragraph—it also includes a human dimension. For example, healthy eating habits correlate positively with an improved quality of life and health.

The importance and significance of good data cannot be adequately emphasized in general, and more specifically and critically in data-intensive methodologies like analytics.

Here is a personal and professional life case study example. In 1992, I was recruited as a Senior Data Management Advisor by Columbia University (CU). My very first assignment was to recalculate the incidence rate of HIV/AIDS.

Four years earlier, CU had launched a project that was primarily managing an open HIV/AIDS cohort. That is a population of interest that recruited new members as the study progressed.

The project’s focus was to manage a cohort of over 13,000 participants and produce periodic reports (in this case every six months) on the dynamics of the epidemic. The milestones were morbidity rates — incidence and prevalence.

The week when my assignment began coincided with a scientific conference in Holland where Dr. Maria Wawer (my boss) and other colleagues were presenting papers on the project findings. During that first week of the conference, Dr. Wawer contacted me to inquire about what incidence rates I had come up with.

In the meantime, because of my limited knowledge of the data set, I recruited two experts who had been with the project as consultants during and since its interception. I identified what I believed were the most critical issues to be addressed before starting the computations and subsequent analysis.

The team was then assigned specific tasks. These included cleaning the relevant data set: generating frequency tables; identifying outliers; triangulating with both source data (original questionnaires), laboratory technicians (serology test results), and survey team members.

After completing this cleaning and validation process (including correcting the numerous inconsistencies), we proceeded to perform the calculations using the statistical package — Statistical Package for Social Sciences (SPSS).

This phase of the assignment went very well. After compiling the results, I then submitted the findings (as earlier agreed) to Dr. Wawer who was still at the conference in Holland. The recalculated rates this time were one infected case lower than what was being presented at the conference. And that, as it turned out, was a big deal! I received immediate feedback as anticipated, highlighting the fact that I was new to the project team with a limited understanding of the data sets.

During one of our weekly team meetings (post-conference), primarily to review what had gone wrong with our incidence rate, one of my colleagues was so embarrassed and distraught that he started shedding tears. Since no amount of consolation could calm him the meeting was immediately adjourned.

In the meantime, members of a similar and “competing” project were constantly and consistently asking us what the real incidence rate was. What should they quote in their papers? As the message continued to spread, our team agreed on a consensus response, which was that the team was still in the review and validation process after which the final and latest incidence rates would be disclosed. This resolution served very well in mitigating further concerns.

During this process, our team went back to the drawing board to confirm what the real rates were. After our earlier computations and part of the triangulation process, we had actually conducted a recount of the new infections. The numbers were consistent with our findings. This recounting exercise was again conducted in addition to further

calculations. And this time every degree of verification confirmed our results: there was one infected case too many!

And what is the message? PAAs and other quantitative methods are only as valid, reliable, and useful as the quality of data used.

### A. Objectives

The objectives of this paper are to examine:

- The current literature on PAAs with the focus on methods in which they are used to enable prediction of future events.
- Case studies of the use of PAAs in industrial applications
- The conceptual framework on which PAAs are used to develop a machine that enables prediction of future outcomes.

### B. Theoretical Frameworks

Descriptive highlights on which this framework's algorithm is based are as follows:

- A collection of literature materials explaining the concept of PAAs
- Relevant and applicable models used are reviewed
- And simultaneously analysing available literature material;
- An outcome report is compiled and
- Findings are presented to relevant parties
- The required theoretical framework is as illustrated in Fig. 1.

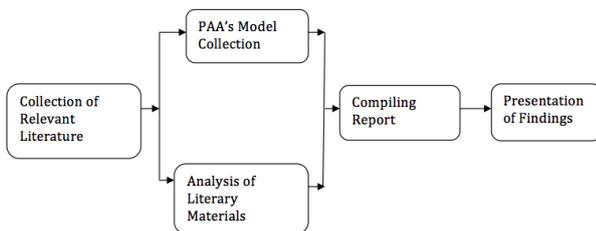


Fig. 1. Theoretical framework.

## IV. SCORECARD

### A. Description of the Conceptual Framework

A scorecard is a technique of measuring the performance of an organization in its entirety rather than focusing on a particular process or component of activities, tasks, and operations [4]. A balanced scorecard can be used to test the effectiveness of a program such as the ability of the program to be achieved at a reduced cost, increased efficiency, reduced efforts, and a high accuracy in producing the required outcomes. Previously, a balanced scorecard was designed to enable the assessment of the performance of companies and the extent to which its strategic decisions can be made to achieve the desired operational outcomes. It has been a relevant tool for companies in the assessment of the performance of internal processes and providing opportunities for learning and growth [5]. In spite of the perception that a balanced scorecard is used as a tool for measuring performance, it can be used in the measurement of other activities such as operational efficiency, effective time utilization, and the level of competitiveness of an organization in a particular industry.

### B. How It Works

A balanced scorecard (BSC) is used in deciding what a business is trying to achieve, to align resources in a manner that the regular activities of a business are achieved, and to create priorities for the provision of products and services to customers. It is composed of small boxes containing elements of mission, vision, core values of an organization, strategic areas of focus, and the activities in which a business will undertake to achieve continuous improvement [6].

BSC is primarily used by businesses, government agencies, and non-profit institutions. The working principle of a BSC is that an organization can be viewed from a number of perspectives, which can be used to create objectives, targets, and actions in relation to various points of views. The main perspectives of a BSC are listed below.

- **Financial performance:** The performance of an organization is viewed in terms of the effectiveness of its use of financial services.
- **Customers/stakeholder needs:** The BSC measures performance in terms of the ability to meet customer expectations.
- **Internal procedures:** The performance of an organization is viewed based on the quality and efficiency of production of a particular product, service, or major business processes.
- **Capacity of an organization:** From this perspective, an organizational performance is viewed based on its ability to utilize resources, technology, human capital, and other capabilities that create an environment for the achievement of a high performance.

### C. When It Is Used to Create PAAs

BSC can be used during the creation of PAAs by enabling the formulation of the performance features of the algorithms. The algorithms for analyzing an organization's performance can be analyzed using a BSC composed of capacity components such as the ability to be produced at low cost, ease of operation by the users, reduced likelihood of breakdown, and the ability to provide accurate forecast of an organization's performance [7].

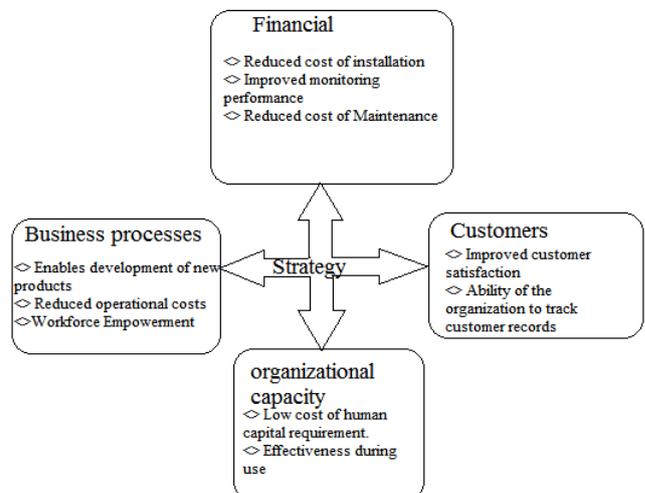


Fig. 2. Balanced scorecard for the implementation of PAAs [7].

### D. Strengths and Weaknesses of the Model

The strength of a balanced scorecard is that it provides the

opportunity for putting all the operations of a business into consideration. It also accounts for the impacts of different components on each other rather than examining the manner in which a particular component operates or achieves its intended goals [8]. When a BSC has been integrated into the functions of an organization, it can be used as a tool for monitoring the achievement of goals and objectives.

The disadvantage of a BSC is that it focuses on the impacts in general, which neglects the performance of an individual or a particular process within a set of processes. There is the possibility of perverting a scorecard by using it as a tool for monitoring employees rather than the performance of a business [9]. It also takes into account a large number of variables to constitute a practicable scorecard, making it challenging to manage.

In Louisiana University College (LCU) of Engineering, a ClearPoint Strategic balanced scorecard software is used to align the activities such as enrollment, assessment of students, and improvement of the infrastructure of the department according to its vision, mission, and goals. The outcomes of the balanced scorecard enabled members of the institution to understand their performances in relation to the target outcomes that need to be achieved [10]. Due to this strategic plan, there has been increased enrollment in the college and it is considered to be the fifth fastest growing college of engineering in the U.S.

## V. CURRENT MODELS OF DESIGNING PAAS

### A. Forecasting and PAAs

Forecasting and analytics algorithms are used to create a model of a future event. An example of a common future event forecasted in many businesses is sales volumes. PAAs are used by sales managers to compare the outputs of the algorithms with achieved results, and to discuss the variations with their representatives who examine them and make estimates [11]. Forecasting algorithms also provide salespeople with the opportunities to know when they need to communicate prospects based on changes in algorithms, which have an impact on the buying decisions of customers.

### B. Statistical Models

Time series algorithm is a common statistical model of PAAs and is categorized into frequency-based algorithms and time-domain algorithms. Frequency-domain algorithms consist of spectral and wavelength analyses, while time-domain methods include algorithms used during auto-correlation and cross-correlation analyses [12]. Another commonly used statistical algorithm is the market segmentation algorithm that is extensively used in customer profiling depending on particular characteristics or priorities of a business.

### C. Linear Regression Models

In simplistic terms, linear regression algorithms are used in modelling relationships between observed (dependent) and design (independent) variables. It is based on the least squares method that fits the best line and results into the minimal sum of squared errors between the expected and actual data points.

Linear regression algorithms are used to make decisions such as the most suitable marketing mix to achieve optimized sales when particular investment channels are used. An example of a case where linear regression is used is at Cable Company X in the United States, where a program is used to determine the effect of variables that predict truck rolls within seven days. The variables used are downstream power, upstream power, and downstream signal-to-noise ratio [13]. The results that are statistically significant provide an insight on the interventions that need to be made to prevent truck roll.

### D. Multiple Regression Models

Multiple regression analyses are used when product pricing is required across an industry such as real estate pricing and marketing organizations in order to establish the impact of a campaign. It is a broader category of regressions that incorporates both linear and nonlinear regressions and uses explanatory variables to perform an analysis [14]. The main application of multiple regression algorithms in practical situations is social science research, the analysis of the behaviour of a device, or in the insurance industry to estimate the worthiness of a claim. Multiple regression analysis was used to examine the factors that affected the outcome of a referendum in which the United Kingdom opted to leave the European Union. The research involved the application of multivariate regression analysis in which the Logistic (Logit) Model was combined with real data to determine the statistically significant factors that have an impact on the voting preference in a simultaneous manner, in addition to the odds ratio that supports Leave or Remain [15]. The results of the multiple regressions showed that the gender of voters, age, and level of education were statistically significant factors, while country of birth was a statistically insignificant factor.

### E. Multivariate Regression Model

In multivariate regression models, the value of a single variable is predicted using a number of independent variables. It is also used in the estimation of the relationship between predictors and responses. Predictors constitute continuous, categorical, or a combination of both. Multivariate analysis measures multivariate probability distributions in the context of their impacts on the observed data [10]. An example of such a model is multivariate analysis of covariance (MANOVA), which performs the analysis of variance that covers instances where more than one variable is analyzed simultaneously. Principal component analysis (PCA) is a multivariate analysis that enables the creation of a new set of orthogonal variables containing similar data as the original set. Multivariate regression analysis has been used by DHL, a global delivery company to predict future status of global trade, in its Global Trade Barometer program. A machine-learning language is used to input collected data related to different intermediate commodities that range from clothes, bumpers, or mobile devices [16]. The program leverages artificial intelligence and multivariate analysis PAAs to create a single data that enables understanding of the effects of a number of variables on a single variable. The output can be used by stakeholders to make decisions such as planning the capacity for future demands of their services and benchmarking on the forecasts to understand the industry's

competitiveness.

#### F. Decision Tree

Decision-tree algorithms are classified into supervised learning algorithms. They are used to create models for solving regression and classification problems. The goal of creating a decision tree is to generate values that can be used to predict the outcomes of a particular class or target variables by applying learning decision rules derived from past data [17]. The concept of tree representation of algorithms is used to solve a problem. Corresponding attributes are used in various internal nodes of the decision tree while class label is made at the leaf node. Pouch, a British plugin company developed an artificial intelligence (AI) chatbot, which informs customers of Black Friday discounts. The bot is available to users on Facebook Messenger and uses decision-tree logic to understand people's preferences [18]. The decision tree enables users to search the directory according to codes such as departments and their products, brands, and voucher codes of their preferences.

Milwaukee-based Aurora Health Care uses the technique of decision tree in the design of a "digital concierge," which operates on the principle of AI. The organization has cooperated with developers from Microsoft's arm of healthcare innovation in the development of a tool that simplifies decision-making in relation to patient care. The concept of decision tree is applied through a chatbot program, which can be accessed via a web browser [19]. This computer code enables mapping out symptoms and the common descriptions used by people to describe their health issues.

The input is provided through answers to a set of questions regarding the symptoms presented. The bot adapts to the answers and outputs possible causes and treatment plan suggestions. The algorithm enables the creation of a command for making a decision on whether the patient may need further clinical care by the patient clicking a section that reserves his or her place in a line at an Aurora urgent care center. The conceptual framework of the chatbot is illustrated in Fig. 3.

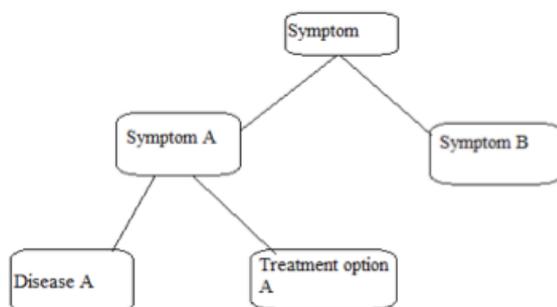


Fig. 3. Framework of decision tree used by Aurora Health Care [19].

#### G. Data Management

Testing data quality using predictive analytics algorithm takes place through the creation of a computer model for validity and reliability of data. The resulting computer model is usually a Pearson correlation that explains the relationship between response and design variables [20]. In measurement of reliability, the objective is to measure the extent to which the measured value is likely to change when the research is

replicated. Some computer algorithms measure reliability by performing random and systematic error analyses. Eco-Absorber is a panel acoustics commercializing company that uses reliability and validity algorithms to get accurate outcomes of its surveys [21]. The outcomes are used to determine the suitability of the survey findings to recommend a change in practice that addresses the 4Ps of marketing in an effective manner.

#### H. Program Management Implications of PAAs

A number of considerations must be made when applying PAAs in program management. Good prediction can be achieved only if there are good data such as past records, which can be used to predict future outcomes of a process or an activity. For instance, prediction of sales of an organization in the next six months is subject to the availability of historical data that, when analyzed, provide a better understanding of the trend of changes in sales [22]. Before data analysis is conducted, they must be organized to reduce redundancy and unnecessary fields must be discarded. In order to deploy the insights from predictive analysis into the systems, it is recommended that software applications should be used to integrate them into predicting performances of businesses [23]. Some of the software that can be used includes API calls, predictive markup language, and web services. The reliability of PAAs algorithms is subject to the use of original data that have been prepared effectively through calculation of aggregate fields, identifying missing data, and merging a number of sources. Each component of data analysis should be analyzed independently. In case of advanced requirements, more advanced algorithms may be required [24].

## VI. STAGES OF PAA DEVELOPMENT

This section explains a more streamlined and contextual version of cross industry standard process for data mining (CRISP-DM). It is a neutral framework that addresses data analytics from two perspectives: application and technical. It is commonly used in predictive data analytics. As we focus on these details, it needs to be pointed out here that conducting (PDA) should never be diploid simply for the sake of expressing curiosity or flaunting one's knowledge of an existing problem-solving strategy. PDA is meant to solve problems. And in order to solve these problems, significant efforts are required to justify its application. One important component of such an exercise is the identification of a relevant management challenge. Hard questions need to be asked. What specifically is the issue? What are some of the interventions that have been made? How have the intervention outcomes improved or addressed the problem? And how have these interventions contributed in mitigating these problems. A combination of these questions will help significantly in redirecting and focusing intervention strategies.

#### A. Problem Statement

In this stage, the business problem that needs to be addressed should be identified. The objective can be to perform a forecast of the future needs or to establish the likelihood of occurrence of a particular defect. The resulting

predictive algorithm should be one that promotes the attainment of the goals and objectives that have been identified [13]. Problem statement identification also involves the definition of performance metrics that a business needs to achieve. A plan should be devised that enables the measurement of the metrics when data are input into the algorithm.

#### *B. Intervention Strategies*

The intervention strategy involves making a decision about the right software or application to use in creating algorithms for resolving a particular operational procedure in a business. The intervention strategy may be to design an algorithm that enables understanding of the breakage of devices being manufactured, the likelihood of reduction in the number of purchases, or overall change in customer satisfaction.

#### *C. Processes*

The process of algorithm development will be determined by the goals to be achieved and the data to be analyzed. Algorithm development is achieved by the use of machine learning and data mining methods composed of relevant analytic platforms. The process of developing an algorithm can take different shapes according to the purpose to be achieved [25]. Some of the commonly used methods in creating algorithms are the creation of computer programs that enable processing of data input to perform a number of tasks such as regression analyses or estimation of variances. The relationships between an organization's data sets can be amassed by the use of unsupervised clustering algorithms. The processes to be followed during the design of algorithms can be illustrated using flow charts [26]. These are charts composed of activities to be performed, decisions to be made, the arrows which show the direction of a program, and conditions that must be satisfied before a program progresses to the next stage.

#### *D. Algorithm Design*

During algorithm design, the designer creates mathematical processes that can be used to solve problems. The concept used to develop algorithms is coding engineering. Algorithm design and implementation are achieved by the use of design patterns or template patterns and involve the use of data structures to create programs and subprograms that can be used to derive the mathematical output from a particular data input [27]. In order to develop an algorithm, mainframe programming languages that are recommended include ALGOL, FORTRAN, PL/I, and SNOBOL. The developer of an algorithm can create hand-written processes and a set of mechanical activities to be performed by hand before creating a corresponding algorithm using a computer program.

#### *E. Program Development*

During the program development stage, a code is written in the form of pseudocode and logic requirements to be followed in a particular programming language. Various coding language choices can be made in relation to a programming task depending on its characteristics and usability [18]. A relevant coding language is selected and syntax rules are followed with little deviation to improve the accuracy of the program.

#### *F. Pilot Testing*

In this stage, the written program undergoes a debugging stage in which the programmer identifies errors in the program. The identified errors can be syntactic or logic. In addition, the programmer explores other areas that are likely to make the program not run in a proper manner or not run completely [21]. The pilot testing stage is usually lengthy and tedious and often constitutes more than 50% of the program development process. However, when there is greater attention to program design and coding, it is possible to reduce the amount of time spent in the debugging stage. Syntax errors result in difficulty of executing a program and constitute simple errors such as misspelling or failure to comply with the syntax rules to be followed in a particular programming language [12].

#### *G. Pre-implementation Testing*

In this testing, test data is added to the program to determine its usability in providing the required outputs. Agile testing can also be performed by following the principle of testing from customers' perspectives [23]. This testing should be performed by the quality assurance (QA) team. User acceptance testing (UAT) is performed on the program to determine whether it is usable in the intended system when released. This is due to the fact that changes in software characteristics undergo changes as it is developed. The resulting changes can be misunderstood in a fashion that is not according to the objectives of users. When UAT is completed, if all requirements are met, the program is moved to production and made available to the users.

#### *H. Final Implementation*

The final implementation stage is where a program is used to conduct an analysis of a particular data to provide an output that can be used to predict future activities of an organization [14]. In the implementation stage, the data mined from an organization's database are input into the written computer program, processed (machine learning) and the resulting output is recorded and analyzed to enable prediction of a future characteristic of a program.

#### *I. Lessons Learned*

The programmer conducts an assessment of a written program to establish whether the expected output has been achieved. A program that results in a desired output such as the number of customers who purchase products in a particular time period and considered useful should be retained by the organization.

#### *J. Challenges*

A major challenge that is likely to be encountered during any programming activity is that some programmers may not use algorithms that produce the expected output. Some problems are difficult to solve because they do not have parallel codes that can be used to write their corresponding programs. Some parallel algorithms have complex features that make execution of programs difficult. Debugging is an important skill but most people do not have the ability to identify and correct errors due to the frustrations and difficulties encountered during this process. The design phase of a computer program can be challenging in terms of the need

to think about the program requirements that need to be put together in a manner that would facilitate future updates. When program design is not effective, the resulting program can be difficult to modify in the future.

## VII. LIFE CASE STUDIES OF THE USE OF PAAS IN INSTITUTIONS

In an attempt to simplify the conceptual complexities of PAAs, a select number of themes are included with life case studies. It is my hope that such an approach will enable readers to better internalize some of what has been accomplished and relate these accomplishments to their respective and unique themes.

### A. Health

Data analytics are used extensively to predict the resource requirements for hospitals. At Texas Hospital, predictive analytics algorithms have been used to enable reduction of its 30-day rate of readmission due to heart failure [25]. The data used to conduct the analysis are the admission number of patients who are readmitted and those having heart failure in the past months. The most commonly used method is a computer program that can be written using Java, JavaScript, or C in which an algorithm is created to establish a regression equation that can be used to predict future readmission rates. The independent variable is the number of patients with heart failures while the dependent variable is the number of readmissions in the past months. The resulting output provides a value of regression equation that can be used in combination with the current number of heart failure patients to predict future readmission rates.

At the Harris Methodist Hospital outside Dallas, predictive analytics algorithms are used to conduct scans on medical records to establish the most suitable care that can result in an improvement in patient outcomes. The algorithm accounts for a number of data characteristics such as blood pressure and the amount of glucose in blood to act as an identifier of patients who are at risk of experiencing heart failure [28]. The algorithm creates a 30-day risk score representing the likely heart failure incidence. This enables physicians to focus on patients who need to be provided with intensive care. The commonly used programming languages are Python and PHP. The risk score is determined by creating an algorithm that measures the p-value using a computer program. A particular level of significance is used to determine whether there is a likelihood of heart failure. The input variables are the amount of glucose in blood and blood pressure. The output of the analytic program is the level of significance, which may be 0.05 or any set value by the hospital. Patients whose values fall within the significance value are at risk of heart failure and effectiveness of treatment measures should be improved in promoting their health [29]. An algorithm is created that measures multiple regressions in which two independent variables are used; amount of glucose in blood and blood pressure. The resulting regression equation in a computer program contains the sections for input of the independent variables. The program is run and a regression value provided is used to predict the possibility of heart failure in a patient.

### B. Problem Statement

It has been necessary to determine methods of identifying patients who are at risk of heart failure with less human involvement. The existence of machine languages such as Java, Javascript, and Python has provided the opportunity for practitioners at Harris Methodist Hospital in Dallas to develop a machine-learning algorithm that enables distinction of patients at risk of heart failure in order to provide them with more intensive treatment.

### C. Intervention Strategy

The intervention includes the creation of a computer program based on machine learning languages in which the practitioners record patients' data and calculate the relationship between the values of blood glucose level and blood pressure to heart failure. This is where a notification is provided to the practitioners when blood pressure or blood-glucose levels reaches a particular value.

### D. Process

The process involved the installation of the machine learning languages into the systems at Harris Methodist Hospital, coding and testing of programs using sample patient values, training the employees to use the program, and its commission for use in identifying patients at risk of heart failure.

### E. Algorithm Design

The design of the algorithm was achieved by complying with the framework shown in Fig. 4.

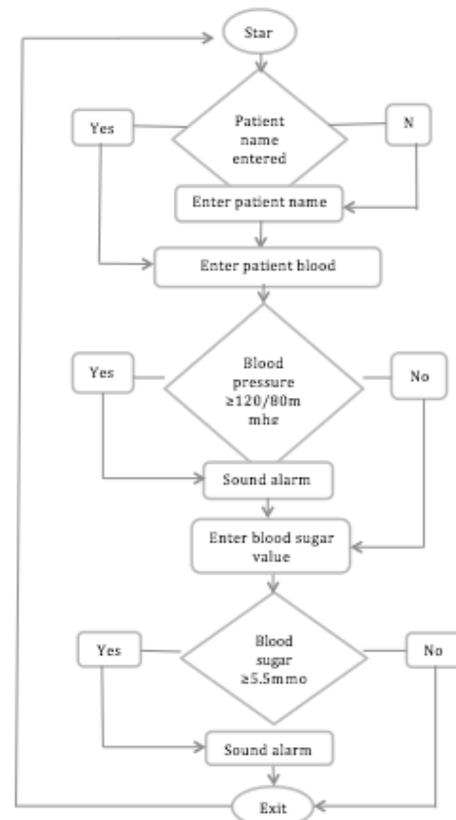


Fig. 4. Algorithm framework for testing patients at risk of heart failure [21].

### F. Pre-implementation Testing

Before the actual implementation of the algorithm, it is

tested by adding the value of blood pressures and blood glucose levels of patients to determine whether it is able to sound an alarm when the values are higher than the maximum amounts. The program is also debugged by removing syntax errors and misspelled words in order to improve its readability.

*G. Final Implementation*

The final implementation is the integration of the machine learning language in the diagnosis of patients who are at risk of heart failure. The implementation involves authorizing the use of the software in the organization as well as training the personnel involved in patient care to examine patients who are at risk of heart failure.

*H. Lessons Learned*

Machine learning algorithms can be created to enable healthcare professionals to make accurate decisions during the diagnosis of patients such as identifying those who are at risk of heart failure. The effectiveness of the program is determined by the nature of the machine language used, the competence of the personnel, and the dedication of the staff involved in monitoring blood sugar levels and blood pressure as determinants of heart failure.

*I. Challenges*

The major challenges that are likely to be encountered in the use of the program are the lack of staff motivation, difficulty in debugging due to failure to locate errors in coding, failure of organizations to allocate enough resources, and the practice of using machine learning language to diagnose patients for risks of heart failure.

*J. Education*

Many learning institutions have used predictive analytics to predict future performances by applying past performance scores of students in their institutions. At Southern Methodist University, an associate provost has contributed to student data management practices by applying predictive analytics algorithms that combine the grades attained by students in the past years to predict their performances in the future [11].

The analysis performed involves entering the raw data into the package and following the procedure of regression analysis. The preliminary result of the regression is a regression value that is related to the current performance of the student and is a factor that enables prediction of future performance. The final outcome is a standardized coefficient that acts as a predictor of the performance of a student in future tests based on the present performance.

*K. Problem Statement*

The need to achieve an accurate prediction of the future performance of students at the Southern Methodist University (based on their present performances) is unquestionable. The use of a machine learning (ML) program is regarded as the most suitable approach for achieving this objective.

*L. Intervention Strategy*

The intervention strategy that has been recommended is the use of an ML algorithm that calculates the regression value for the students' scores, which can be used to predict their

performances in the next academic periods. The recommended statistical package is GNU PSPP, which has features that enable calculation of statistical measures such as simple linear regression, multiple linear regression, cluster analysis, and reliability analysis [30].

*M. Process*

The process involved was the installation of the GNU PSP application into the computer system followed by the design of the machine codes that return particular values of performance when information is input. The computer program will be composed of input points and the points of making decisions regarding the required outputs.

*N. Algorithm Design*

The design of the algorithm will take place immediately after the installation of the GNU PSP computer application. The design involves the use of computer decision frameworks such as the flowchart shown in Fig. 5.

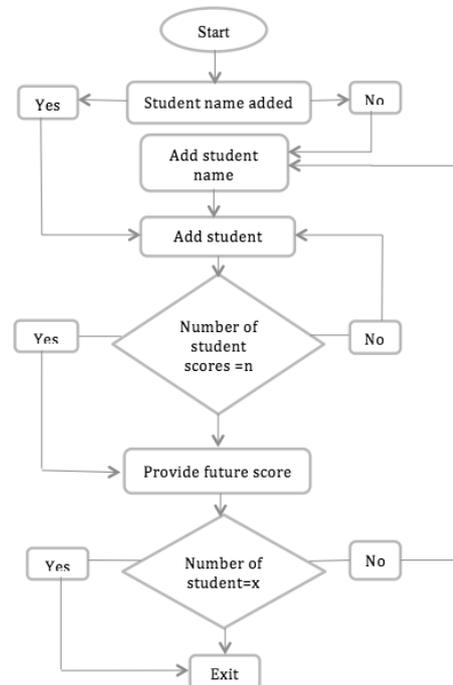


Fig. 5. Design of the algorithm for prediction of future student's performances [10].

*O. Pre-implementation Testing*

During the pre-implementation stage, the program is tested to determine whether there are any errors. Debugging is done to correct syntax errors and factors contributing to the failure of the program are examined. The ability of the program to be used in a particular system is tested.

*P. Final Implementation*

The program is authorized for use in predicting the future academic performance of students in an institution in which it is destined to be used [31]. The staff are trained to apply the program during the entry of students' previous performances. They also trained on the skills of interpreting the results of the program.

*Q. Lessons Learned*

The lessons learned from the program are that it is possible

to design an effective program if the desired outcome is established. The programmer also needs to have the relevant knowledge including the steps for writing a machine code containing greater details. When a program for predicting future performances is created, it provides an approximate future performance of a student so that potential low performances can be mitigated.

#### *R. Challenges*

The challenges that are likely to be encountered during the design of the computer program are the omission of particular procedures that enable analysis of the inputs to provide the accurate prediction of future outcomes. A challenge is also likely to occur in the debugging stage when the source of the error cannot be located.

#### *S. Agriculture*

AgDNA intends to solve the issue of excess nitrogen by implementing the PAAs concept, in which nitrogen requirements are optimally matched with site-specific conditions in the field, thus reducing the likelihood of the occurrence of high amounts of nitrogen in the atmosphere. The company has integrated next-generation cloud computing technology and techniques for big data analysis, soil characteristics analysis, and climate data as information that enables understanding the nature of a farming field and its suitability for crop production [32]. These inputs are then combined using the most recent precision nitrogen management (PNM) frameworks to provide a prediction of the required amounts of nitrogen. The methodology used is the creation of a computer program in which the characteristics of the soil are compared to the amount of nitrogen in order to determine whether there is significance in the relationship. The statistical measure used in the analysis is the p-value, which measures the level of significance of the relationship between various soil characteristics and the amount of nitrogen. The software used in the computation of the relationship is JavaScript, which is cloud computing software that enables the creation of programs for regression analyses. The analysis involves the input of the amount of nitrogen and the corresponding soil characteristics such as soil type, color, moisture content, and soil texture. The preliminary results are the p-values in which the values greater than the set significance value are soil types that have higher amounts of nitrogen that need to be regulated [33].

#### *T. Business: Marketing*

Business application of PAAs occurs at the New York Times (NYT) as a means of improving its business and operational model. Predictive analytics models have been created that enable subscription to the organization's social media sites and other business decisions. According to a report by Chris Wilgins in a Predictive Analytics conference, predictive analytics is used to influence customers [10]. The NYT uses natural language processing as a means of increasing reader engagement so that the most beneficial types of articles can be sold. The software used is C program, in which an algorithm is developed that enables recognition of words such as adjectives used by customers to demonstrate their satisfaction. The software also has a subprogram, which

enables the creation of a decision tree that matches the natural languages used by customers to make a particular decision. The preliminary result of the program is a tree diagram, which matches the natural language used by customers and the decisions that need to be taken to promote the sales of the NYT products.

#### *U. Business: Transportation*

Virgin Atlantic uses predictive analytics algorithms to determine the prices of tickets according to the likelihood of travel demands by customers [6]. The statistical packages used are either MATLAB or SPSS, which have features that enable the calculation of statistical measures such as regression analysis, multiple regression analyses, correlation analyses, and the T-test. The methodology used is the input of the raw data such as prices of tickets and the corresponding number of customers who board flights in a specified period such as a month or a year. The statistical measures conducted include regression analysis and significance analyses. The preliminary regression value is used as a measure of the relationship between independent variables (price) and the dependent variable (number of customers). A final prediction of future demand in ticket sales is established by the use of the regression coefficient to predict the likely number of customers.

#### *V. Sports*

A commonly used predictive analytic model in sports is Sports Performance Platform (SPP) from Microsoft, which incorporates an ML and AI in the creation of algorithms used to make decisions regarding the performance of athletes. This application provides solutions for the locker room, performance lab, and has an algorithm that enables prevention of injuries, making decisions pertaining to games, and changing training schemes to improve the performances of athletes [15]. An example of a sports club that uses PAAs is Midtjylland, a Danish club that was on the brink of bankruptcy but improved to nearly winning a championship title. The club made changes to the steering approach by implementing analytical procedures in which experts conducted an analysis of each player twice a month to obtain information that addressed the player's training needs. The experts also provided the coach with information such as when to change the game plan in accordance with the in-game statistics. Information from analytical models was used to recommend new players [34]. The programming software used for the analysis of the players was SPP. The methodology used was the creation of an algorithm that enabled input of player behaviours such as the number of passes, distances covered, number of touches of the ball, and the resulting team performance such as the number of wins, draws, and losses. The algorithm creation methodology also involved the creation of a code that measured the regression between the variables. The preliminary results were the raw player data in the computer program and the team's performance in the past matches. The final outcome was the regression value, which showed the relationship between each player's characteristics and the team's performance. This value is important in making decisions such as whether to

substitute a player in order to improve the performance of the club.

#### W. Social Media

Social networking companies such as Facebook have been using predictive analytics algorithms that enable updates regarding a brand to be available to the user after a product has been “liked”. Therefore, users are able to see posts, which improve their engagement rates with their individual networks such as posts that their friends have engaged with [16]. The programming language used is JavaScript due to its cloud computing feature and the ability to make changes to an already existing algorithm. The methodology used is the creation of an algorithm that enables the site to link a liked product to a user’s page. The process includes the statistical analysis of a decision tree in which the website is automatically coded to link a “liked” product to the user’s page. The final outcome is a user experience in which when a person likes a product, the updates regarding the product appear on their page in the future. This implies that Facebook will promote the ability of marketers to promote social engagement with customers.

#### X. Manufacturing

In manufacturing companies, machine-learning algorithms have been used to understand the machine problems that are likely to be encountered in order to apply preventive practices to keep the supply chain operational. At Georgia Institute of Technology, machine-learning algorithms provide the opportunity to promote forecasting the likelihood of machine failures, thus, enabling the technicians to perform maintenance practices [35]. The machine learning language used is a C program with capabilities for creating codes that enable calculation of statistical tests such as regression analyses, linear regression, and multiple regressions. The

methodology used is the creation of a computer algorithm in which past intervals of failures is added. The data are the failure times (the dependent variable) and the time interval (independent variable). A sub-program is created that enables the calculation of simple regression analysis, which establishes the relationship between machine failure times and the time interval. The preliminary results are the input values of failures of the machines against time interval. The outcome of the analysis is a regression coefficient, which can be multiplied by the current failure frequency to determine the next likelihood of the machine’s failure. This ML algorithm has been applied in the performance of regular maintenance tasks on lathes, grinders, saws, and gears.

#### Y. Government Institutions

In the United Kingdom (UK), the Ministry of Defense uses machine learning algorithms to explore and organize public documents. This is achieved by creating algorithms that enable the identification of documents depending on their subjects and conducts the analysis of information for the purpose of finding patterns and anomalies in data systems [25]. The algorithms are also implemented in the detection of fraudulent activities, transactions, or activities of any public official for personal gain. The algorithms have been effective in the detection of activities such as money laundering, the creation of counterfeit trade items or the duplication of business documents. The processes include the installation of machine learning languages into the systems of the organizations, the creation of computer programs, testing, and implementation [3]. The inputs are information regarding future activities such as the attempt to change the content of documents in order to achieve personal objectives or defraud the government. The program is capable of providing information about the perpetrators of the acts and insights on characteristics that can be used to trace them.

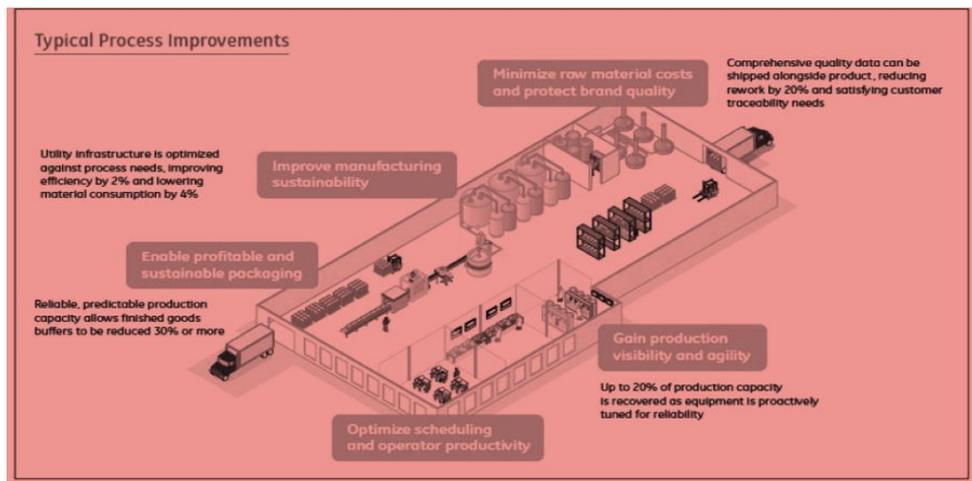


Fig. 6. Summary of the improvements made at general electric [35].

#### Z. Crowdsourcing

Bugcrowd Inc. uses crowdsourcing, in cooperation with Fortune 500 companies such as MasterCard Incorporation, to identify vulnerabilities that may be used by hackers to infringe on their infrastructure. This is achieved by the use of a machine learning language called a bug bounty program, which enables the engagement of the cybersecurity

community, providing them with monetary rewards for their contribution to the resolution of the vulnerabilities [2]. A major advantage associated with the company is the lack of a requirement for evaluation of claims of cyber threats using the crowd-sourced information to determine the areas of security where greater attention should be placed. Crowdsourcing also involves the use of application programming interfaces (APIs), a tool for software development that integrates the

sourced data into the current workflows or reports of business security analyses. The process involves the selection of a suitable programming language such as Python and installing it in the organization's system [5]. Professionals in machine code development develop machine codes that enable the recording of information from a number of sources. The output is a list of sources of information containing cybersecurity information that is usable for improving the organization's databases.

## VIII. INTERNATIONAL DEVELOPMENT PROGRAMS THAT USE PAAS

From a geopolitical perspective, I have also included case studies on themes that are universally applicable with specific emphasis on select themes that significantly contribute in making the world a better place and hence promoting a better quality of life.

### A. Natural Disaster Programs

The concept of predictive analytic algorithms has been implemented in the analysis of big data regarding past natural disasters and used to predict future incidences [23]. An example of an incident that provided data for fighting natural disasters is the earthquake that occurred in Haiti in 2010. Crowdsourcing has been used to obtain real-time images of disasters such as earthquakes while big data approaches in artificial intelligence (AI) have been used to determine meanings in messages such as SMS that were generated during the occurrence of natural disasters.

The processes involved the installation of machine learning language followed by the creation of an algorithm that enables the performance of mathematical analyses such as regression analysis and providing the output that can be interpreted to estimate the likelihood of occurrence of a similar incident such as another earthquake in the future [33]. The analytical procedures performed involve the input of information pertaining to disasters such as the magnitude of an earthquake, time of occurrence, and region into the machine language. The machine language performs an analysis of mathematical processes such as linear regression and multiple regressions to provide statistical coefficients that can be used to predict future disasters.

### B. Poverty Eradication Program

Predictive analytics have been used by the World Bank (WB) in poverty eradication initiatives such as the collection of information of affected areas, the analysis of the number of people who need relief services, and the relationship between their status with infectious diseases. This is in accordance with the WB objective of eradicating poverty by the year 2050. Countries conduct household surveys and provide WB with information used to classify the population according to the level of poverty [25].

The processes involve the creation of a machine language that enables input of raw data such as the economic statuses of families. Data from statistical offices in various countries are input into the machine learning language that has been designed in a customized fashion to enable the stratification

of families according to their gender, age, income levels, geographical location, race, or culture. The program has commands that enable the quick calculation of statistical measures such as linear regression or multiple regressions to provide coefficients that enable the prediction of poverty levels in the future [2]. The machine learning language has also been designed in a manner that enables the transfer of data from mobile phones to the program for analysis. This program has been implemented to measure the economic status of people in Togo, Tanzania, and Tajikistan to provide outputs that enable prediction of poverty status in the future. A similar program has been used by the WB in the measurement of the movements of nomadic people in Somalia to predict future migration patterns.

## IX. PROGRAMMING SOFTWARE

### A. Turn-Key Programming Model

A turnkey program (TKP) is one that is developed according to specifications because the owner has specified all the functional requirements. A TKP has the primary advantage of enabling the user to establish a program budget, inputs, and outputs in a scheduled manner. Turnkey programs do not provide easy flexibility in the management of changes and other features requested by the programmer.

### B. In House Programming Model

In-house programming, a program is developed by the IT department of the company rather than an outside company [32]. An example of in-house programming is Google's software development, which is done using its machines that are located in various parts of the computer network system.

### C. Outsourcing Programming Model

Outsourcing programming is the process in which a computer program is written by a third party and generally external institutions on a consulting basis. It is a more advantageous method of programming because an organization reduces the cost of undertaking a particular project. It is also a means of ensuring timesaving in the development of computer programs because it tends to be less time-consuming when a number of experts are assigned to complete program development. The risks and challenges involved in outsourcing are confidentiality, limited supervision, possible tardiness and service-provider loyalty.

## X. PROGRAMMING LANGUAGES, ARCHITECTURE DEVELOPMENT, PLATFORM, INTERFACES

### A. Java

Java is a major programming language used in building server-side programs for video games and apps in mobile phones. It is also popular in the creation of programs for operation on Android-based platforms. Java incorporates both compilation and interpretation techniques [35]. Java compiler is used to convert a source code into bytes. Java Virtual Machine (JVM) performs an interpretation of the

bytecode and the creation of a code that can be executed when the program is run. Java is highly recommended during the creation of web server programs, web commerce applications such as electronic trading systems, scientific applications, and enterprise databases.

**B. Python**

Python is an object-oriented programming language that is popular due to its simple and readable syntax. It is easy to learn and uses simple language for program coding. For instance, if the computer is required to write something, the command “print” is used. Python makes use of the concept of dynamic typing, reference counting, and detection of garbage in order to facilitate memory management [11]. It uses similar expressions to other programming languages such as C and Java.

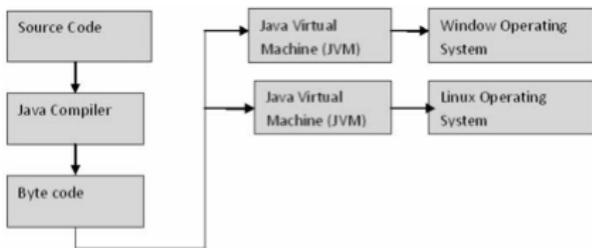


Fig. 7. A mapping of Java programming language architecture [35].

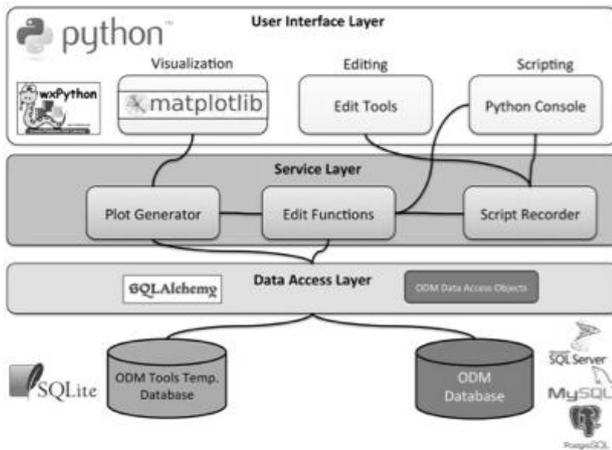


Fig. 8. Architecture of Python programming language [11].

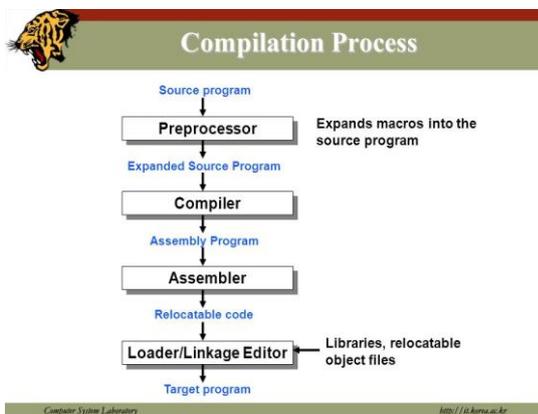


Fig. 9. Compiler architecture of a C program [2].

**C. C Language**

C is a compiler program that can be used to translate functions, declarations, and definitions into files that are executable. It has a simpler command procedure and performs

less programming tasks compared with other languages used in programming such as Python or Java. Executable files are created by the compiler translating source code into executable codes independently. It does not remember the defined variables while performing file processing [2]. This implies that a variable cannot be used if it has undergone previous declaration in the same file. C is similar to Java in functions such as loops and conditionals, but the former is simpler in other aspects, such as the structure of data definitions.

**XI. ALGORITHM DEVELOPMENT: EXAMPLES OF ALGORITHMS**

**A. Brute Force Algorithms**

Brute force algorithms enable enumeration of all integers from 1 to n and establish whether each number is divisible by n to obtain a whole number. With this type of algorithm, direct computation is performed based on a problem statement to be resolved and the corresponding concepts [7]. The search phase for the text can be done randomly. It is an algorithm that is commonly used in the solution of problems such as sorting, searching, and binomial expansion.

**B. Simple Recursive Algorithm**

A recursive (self-executing) algorithm is one that uses smaller input values and applies simple operations to them in order to obtain the result. It applies the principle of solving a problem by dividing it into smaller versions, which can then be solved by the use of recursive algorithms. If a function is represented recursively, the corresponding recursive algorithm for the computation of its members is a mirror of the definition.

**C. Backtracking Algorithms**

A backtracking algorithm is an algorithm that is used to find solutions to computational problems such as conditional problems. The process of programming starts with a particular move out of a number of alternatives [36]. If it is possible to reach a solution using the selected move, the solution is printed; otherwise, the program backtracks and selects another move to try.

**D. Randomized Algorithms**

Randomized algorithms use the concept of randomness to determine the task to be performed anywhere in the algorithm. Their preferred use is for the analysis of expectation of worst cases, in which all likely values of the random variables are considered and the corresponding time by a possible value is evaluated.

**E. Dynamic Programming Algorithms**

Dynamic programming is the process where algorithms are created for breaking down a problem into a number of sub-programs. These problems are solved just once and the result is stored so that when a similar problem occurs in the future, a solution is looked up amongst the stored solutions [35]. This basically involves creating a program that memorizes the results of a particular state and using it to solve a sub-problem.

### F. Highlights

This paper has reviewed and analyzed contemporary documentation pertaining to the use of PAAs, the processes involved in their development, their application in the computation of mathematical procedures, such as linear regression and multiple regression, and prediction of future outcomes. The stages in which PAAs undergo until the outcome is achieved include problem statement, intervention strategy formulation, processes, algorithm design, program development, pilot testing, pre-implementation testing, the analysis of lessons learned, and examination of the challenges encountered.

The concept of PAAs has been used in most machine-learning languages to develop computer programs that provide an output, which enables understanding future events in healthcare, education, manufacturing, governance, and natural calamities such as earthquakes or poverty levels. In healthcare practice, it has been possible to develop a PAA that uses blood sugar levels and blood pressure to predict the patients who are at risk of heart failure so that intervention measures can be implemented. In educational institutions, PAAs have been developed that enable the input of the student's performance in the present period to predict future performances in various fields of specialization. In agriculture, big data PAAs have been used to formulate soil characteristics in the future based on the current characteristics such as soil moisture content, the amount of nitrogen in the soil, and the amount of salts. The output has been used, for example, as a guide on the measures that can be taken to reduce the amount of nitrogen in the soil. Other areas where PAAs have been used are player performance prediction in sports, sales predictions in businesses, predictions of unauthorized acts in government departments, and crowdsourcing to promote organizational cybersecurity.

### G. Discussion

The euphoria created by the advent and exponential evolution of predictive analytics seems to have left many stakeholders in awe. From every level of business to different institutional categories, the best and optimal performance seems to be in sight with no establishment being left behind. While the positive outcomes achieved so far continue to escalate, institutions at large need to take one step backwards to do some stocktaking. This process involves asking critical and provocative questions, including: Are we doing the right thing? How evidence-based are our strategies? Are they sustainable? How reliable are our data sets? Is client data adequately protected from potential cybercriminals? Have all the ethical concerns been adequately addressed? What is the gold standard?

If PAAs' dynamics are any indication, the learning curve is bound to be long, steep, and daunting. One major reason for this possibility is the growing complexities of managing data and the institutions involved in processing them. There is also the challenge of establishing a diverse team of experts involved in developing problem solutions. Members of such a complementary group serve as an invaluable backbone to any potential success. The problems are complex, ranging from good quality data to the nuances that accompany risks and assumptions of selecting and applying the appropriate

algorithms.

As already indicated elsewhere in this paper good quality data is *sine qua non* to any successful analysis (quantitative and qualitative). Mark Twain's characterization of lies, "lies, damned lies and statistics," should always serve as a compelling reminder that the information generated from data through the machine learning (ML) process is only as useful as the quality of data used. Having and using the appropriate and reliable piece of information is a catalyst for making informed decisions. PAAs are no exception! ML processes continue to gauge significant amounts of data. This data is transformed through the ML process to predictive outcomes (information) used in making informed decisions. ML's propensities to process big data sets have made cloud computing an inevitable requirement. The arrival of quantum computers (QC) has made the transformation process faster, reliable, and more efficient. These QCs, which have miniaturized the binary digit (bit), have moved computing to a higher level. According to an IBM definition, "Quantum computers, on the other hand, are based on qubits, which operate according to two key principles of quantum physics: superposition and entanglement. Superposition means that each qubit can represent both a 1 and a 0 at the same time." Access to good quality data is one way of optimizing the utilization of these QCs.

In one of my series of lectures given to graduate students at the University of the West Indies in Kingston, Jamaica, a student wanted to know why program managers firmly believe that in any strategic framework — "logframe" for example — outputs (and their indicators) always contribute to outcomes, especially given the potential for misleading and unreliable results reported at the output level.

In my response, I agreed with the student while elaborating on the data collection and reporting vulnerabilities, especially in environments where very little appreciation is given to data that are subsequently converted to information. I explained the trade-offs that managers and other stakeholders are faced with. I described what it takes to address issues like these, including conducting a control study. I further shared an anecdote with the group; an experience I had conducting a program evaluation for a UN agency. In this case, the agency had spent 4.5 million dollars over a three-year period on national capacity strengthening. The participants, who were medical health workers, were trained both nationally and internationally. This was identified as one of the output indicators that contributed to a corresponding relevant indicator — improved quality of health services — at the outcome result level. During the evaluation assignment, I followed up (something that was never done after training), and as it turned out, most of those who benefitted from the training had moved on; some changed ministries, others had left the country, and some had even changed professions! Obviously, any planning decisions made using that training report would undoubtedly be erroneous, misleading, and deceptive at best.

It is quite conceivable that the evolving, inclusive, and streamlining dynamic of PAAs will continue to have positive and unquestionable consequences on how programs are managed. The myriad implications are unfathomable with synergies that collectively yield both intended and unintended

outcomes. If current thematic applications are any indications, introducing analytics in any intervention will continue to be a win-win initiative.

While different institutions condition their interventions towards their respective strategies, the ultimate outcome is improved productivity and optimization of resource (human, financial, and material) utilization. There is also the human (quality of life) dimension that can revolutionize, reverse, and mitigate certain nuances that affect our wellbeing. For example, academic institutions now apply some models for improving student performance. By using historical data these institutions are able to identify vulnerable students, counsel them on an individual basis, and enable them to set more achievable objectives based on their academic performance with respect to the average group standing. The ultimate outcomes demonstrate counterfactuals that are obvious. And the results have been quite impressive. Some students in some cases have even encouraged themselves to become their own agents of change.

There is also gradually and increasingly, an inclusive element of analytics that continues to encourage and involve members of different community populations: crowdsourcing. This strategy has mushroomed and generated an astounding dynamic amongst communities. It remains to be seen to what extent the strategy will contribute to improving people's quality of life.

In general, business institutions are ahead of the curve with marketing as one of the trailblazers. The competition remains extensive and brutal.

#### CONFLICT OF INTEREST

The author, Bongs Lainjo declares no conflict of interest.

#### AUTHOR CONTRIBUTIONS

The first and only author, Bongs Lainjo, conducted the research design, analyses and wrote the manuscript for the paper.

#### REFERENCES

- [1] N. Khan, I. Yaqoob *et al.*, "Big data: Survey, technologies, opportunities, and challenges," *The Scientific World Journal*, 2014.
- [2] B. M. Purcell, "Big data using cloud computing," *Journal of Technology Research*, vol. 5, no. 1, 2014.
- [3] M. D. Assunção, R. N. Calheiros, S. Bianchi, M. A. Netto, and R. Buyya, "Big data computing and clouds: Trends and future directions," *Journal of Parallel and Distributed Computing*, vol. 79, pp. 3-15, 2015.
- [4] O. Zizlavsky, "The balanced scorecard: Innovative performance measurement and management control system," *Journal of technology Management & Innovation*, vol. 9, no. 3, pp. 210-222, 2014.
- [5] M. A. El Deen and M. M. Solayman, "Maximizing strategic performance results: Adopting balanced scorecards and BI tools," *International Journal of Computer Applications*, vol. 117, no. 10, 2015.
- [6] K. Chitra and B. Subashini, "Data mining techniques and its applications in banking sector," *International Journal of Emerging Technology and Advanced Engineering*, vol. 3, no. 8, pp. 219-226, 2013.
- [7] M. Gera and S. Goel, "Data mining-techniques, methods and algorithms: A review on tools and their validity," *International Journal of Computer Applications*, vol. 113, no. 18, 2015.
- [8] E. A. Awadallah and A. Allam, "A critique of the balanced scorecard as a performance measurement tool," *International Journal of Business and Social Science*, vol. 6, no. 7, pp. 91-99, 2015.
- [9] L. Stefanovska and T. Soklevski, "Benefits of using balanced scorecard in strategic and operational planning," *Universal Journal of Management*, vol. 2, no. 4, pp. 165-171, 2014.
- [10] S. B. Soumya and N. Deepika, "Data mining with predictive analytics for financial applications," *International Journal of Scientific Engineering and Applied Science*, vol. 1, pp. 310-317, 2016.
- [11] N. Mishra and S. Silakari, "Predictive analytics: a survey, trends, applications, opportunities & challenges," *International Journal of Computer Science and Information Technologies*, vol. 3, no. 3, pp. 4434-4438, 2012.
- [12] N. Padhy and R. Panigrahi, "Data mining: A prediction Technique for the workers in the PR Department of Orissa (Block and Panchayat)," *arXiv preprint arXiv:1211.5724*, 2012.
- [13] S. K. Mandal, "Performance analysis of data mining algorithms for breast cancer cell detection using Naïve Bayes, logistic regression and decision tree," *International Journal of Engineering and Computer Science*, vol. 6, no. 2, 2017.
- [14] H. Hassani and E. S. Silva, "Forecasting with big data: A review," *Annals of Data Science*, vol. 2, no. 1, pp. 5-19, 2015.
- [15] N. Razali, A. Mustapha, F. A. Yatim, and R. A. Aziz, "Predicting player position for talent identification in association football," in *Proc. IOP Conference Series: Materials Science and Engineering*, vol. 226, no. 1, p. 01208, 2017.
- [16] S. Pippal, L. Batra, A. Krishna, H. Gupta, and K. Arora, "Data mining in social networking sites: A social media mining approach to generate effective business strategies," *International Journal of Innovations & Advancement in Computer Science*, vol. 3, no. 1, 2014.
- [17] E. H. Thomas and N. Galambos, "What satisfies students? Mining student-opinion data with regression and decision tree analysis," *Research in Higher Education*, vol. 45, no. 3, pp. 251-269, 2004.
- [18] A. Semenov, P. Romov, S. Korolev, D. Yashkov, and K. Neklyudov, "Performance of machine learning algorithms in predicting game outcome from drafts in dota 2," in *Proc. International Conference on Analysis of Images, Social Networks and Texts*, pp. 26-37, Springer, Cham., 2016.
- [19] Anderson, *Framework of Decision Tree Used by Aurora Health Care*, 2018.
- [20] N. Bhargava, G. Sharma, R. Bhargava, and M. Mathuria, "Decision tree analysis on j48 algorithm for data mining," *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, vol. 3, no. 6, 2013.
- [21] Z. Obermeyer and E. J. Emanuel, "Predicting the future — big data, machine learning, and clinical medicine," *The New England Journal of Medicine*, vol. 375, no. 13, p. 1216, 2016.
- [22] C. Perry, *Machine Learning and Conflict Prediction: A Use Case*, 2013.
- [23] K. Danjuma and A. O. Osofisan, 2015, "Evaluation of predictive data mining algorithms in Erythematous-Squamous disease diagnosis," *arXiv preprint arXiv:1501.00607*.
- [24] S. Ganas, "Data mining and predictive modeling with excel 2007," in *Proc. Casualty Actuarial Society E-Forum*, Spring 2010.
- [25] V. Vijayan and A. Ravikumar, "Study of data mining algorithms for prediction and diagnosis of diabetes mellitus," *International Journal of Computer Applications*, vol. 95, no. 17, 2014.
- [26] S. K. Yadav and S. Pal, "Data mining: A prediction for performance improvement of engineering students using classification," *arXiv preprint arXiv:1203.3832*, 2012.
- [27] N. Kinkade, L. Jolla, and K. Lim, "Dota 2 win prediction," Technical Report. Tech. Rep., University of California San Diego, 2015.
- [28] M. Alkhatib, A. Talaei-Khoei, and A. Ghapanchi, 2016, "Analysis of research in healthcare data analytics," *arXiv preprint arXiv:1606.01354*.
- [29] B. Milovic and M. Milovic, "Prediction and decision making in health care using data mining," *Arabian Journal of Business and Management Review*, vol. 1, no. 12, p. 126, 2012.
- [30] P. Thaka, "Performance analysis and prediction in educational data mining: A research travelogue," *arXiv preprint arXiv:1509.05176 r*, 2015.
- [31] S. P. Karkhanis and S. S. Dumbre, "A study of application of data mining and analytics in education domain," *International Journal of Computer Applications*, vol. 120, no. 22, 2015.
- [32] M. Kaur, H. Gulati, and H. Kundra, "Data mining in agriculture on crop price prediction: Techniques and applications," *International Journal of Computer Applications*, vol. 99, no. 12, pp. 1-2, 2014.
- [33] Arumugam, A. (2017), "A predictive modeling approach for improving paddy crop productivity using data mining techniques," *Turkish Journal of Electrical Engineering & Computer Sciences*, vol. 25, no. 6, pp. 4777-4787.

- [34] A. Maszczyk, A. Golaś *et al.*, “Application of neural and regression models in sports results prediction,” *Procedia-Social and Behavioral Sciences*, vol. 117, pp. 482-487, 2014.
- [35] P. Bastos, I. Lopes, and L. C. M. Pires, “Application of data mining in a maintenance system for failure prediction,” in *Proc. 22nd European Safety, Reliability and Risk Analysis: Beyond the Horizon*, 2014, vol. 1, pp. 933-940.
- [36] I. M. A. D. Suarjaya, “A new algorithm for data compression optimization,” arXiv preprint arXiv:1209.1045, 2012.



**Bongs Lainjo** is an RBM systems consultant with expertise in program management and evaluation. Bongs is also a former UN senior advisor of Program Management, Reproductive Health Commodity Security (RHCS) and Evaluation. Prior to that, he worked for USAID as a Logistics and Management Information Systems (LISA) advisor. He also served as a COP/senior data management advisor for Columbia University after teaching as a professor at a number of Canadian academic institutions. His many papers have been presented at international conferences.