# A TRAJECTORY OF SCIENCE TECHNOLOGY ENGINEERING AND MATHEMATICS (STEM) EDUCATION IN PAPUA NEW GUINEA

**A. K. Bilson**
*Ural Federal University named after the First President of Russia B. N. Yeltsin*
*Ekaterinburg, Russia*
*andrew.k.bilson@gmail.com*

***Abstract.*** For young students choosing the right career path is of fundamental importance. One has to be practical, logical and rational in deciding a career pathway. A correct choice may lead to prosperity while a wrong choice may lead to failure and lifelong frustration. Each individual has unique skills and talents, and they also have a unique set of triggers that help them acquire knowledge. Unfortunately, schools train, assess, and evaluate students in a very generic fashion, and therefore it is often difficult to pinpoint each individual's interests and capabilities to give them a clear road map towards possible STEM career pathways. In this paper we will describe a solution to predict STEM career pathways concerning individual academic performance in the different subject areas studied by STEM students in Papua New Guinea (PNG).

***Keywords:*** *Science Technology Engineering Mathematics (STEM), STEM Career, Holistic Psychometric Assessment, Machine Learning, Career Development*

*Introduction*

Over the last decade, we have witnessed a dramatic shift in the education space with an increase in the number of career options and opportunities. These shifts have been direct results of advancement in science and technology which has opened up new streams of career pathways in many different fields. With a variety of career pathway options at their disposal, students are often faced with a dilemma in making these choices. Oftentimes, students' select career pathways that do not reflect their true capabilities. There are many factors that determine their selections including unawareness, pressure from parents, or simply misguided by advice from academic guidance officers' etcetera. Those are challenges that they encounter while making these lifelong decisions.

It is wise to consider all factors before choosing a career path simply because these are lifelong decisions. The world is changing every day. In this age of innovation life is very complex and demands specialization. Therefore, preparing young adults for success in the future requires a different educational experience than it did a generation ago.

This paper will delve into STEM secondary education in Papua New Guinea, career pathways for STEM students, specific assessment techniques and strategies to identify student capabilities and talents, and student career prediction models in education utilizing machine learning algorithms.

The distinctiveness of this study is that STEM education was introduced in PNG for the first time in 2021. PNG is a developing country in the Pacific Region and has its unique educational challenges very much influenced by the cultural and ethnic diversity of the country. This STEM program enrolls only the top 5 % of students from all over the country and it is pioneered by six National School of Excellence (NSOE) specifically chosen by the PNG government for implementation.

It is hoped that students coming through the STEM education will not only gain productive employment in the knowledge-based economy but become creators, innovators and inventors of knowledge and ultimately commercialize their creativity through development of their intellectual property rights.

Guiding students towards the right career pathway is vital not only for them but for the future success of the STEM program as a whole in PNG.

*A. Existing system*

Most career prediction systems use student grades input taken from general assessment tasks to compile datasets and train their machine learning models. This category of grading system does not indicate or measure an accurate benchmark of a student's ability in the different academic strands they are studying. According to Predictive Index editorial by Thad Patterson, student grades provide only 1 % predictive ability and analysis also reveals that GPA is not a good predictor [1]. For instance, a student scores an overall grade of 95 % in Programming during semester one mostly from theory assessment tasks. This grading cannot be used in the dataset unless it is validated that it represents more than one aspect of a student general cognitive learning ability. Therefore, it is highly unlikely that this data gives us an indication that the student is capable to take Software Engineering as a career pathway.

Student grade or scores given as input into the datasets should be taken from a grading system that measures more than one aspect of the student ability. Student grades or scores should come from Holistic Psychometric Assessment which has multidimensional gauging points of students learning ability and performance. These aspirations are in line with the United Nations' Sustainable Development Goal (SDG) number 4 — Quality Education — which can be attained by integrating data-driven decision-making techniques into the educational sphere [2].

The success and accuracy of a good career prediction system depends entirely on the validity of a reliable dataset. Liu R, Tan A. (2020) in their STEM career prediction model revealed that the key to good prediction is proper feature enrichment in the beginning stage of data analysis [3]. Ajay Kumar Pal (2013) also emphasized that data quality is very crucial for making informed decisions [4, 5].

Most career prediction systems use overall grading scores of subjects studied by students to train a data set for their machine learning models. The validity of this data is good to an extent but not always reliable. This is because most fields today demand specialization, therefore it is also important to grade students under the specific topics studied to maximize data validity towards specialized areas in different STEM career fields.

For example, under ICT strand of study there are topics such as Programming, Cyber Security, Networking, Database, etc.

These are specialist field areas in the IT industry and therefore students must be graded under the specific topics to fully understand their performance abilities in each module. Consequently, it will create a clear road map towards career options choices in those specialist areas. For instance, a student gets the following scores through assessments as shown in the breakup below.

- Programming — 20 %
- Networking — 95 %
- Cyber security — 85 % • Database — 49 %.

All these topics come under Computer Science & ICT. On a face value analysis of this data we can confirm that the student will thrive in the field of Network Engineering or as a Cyber Security Analyst given the 95 and 85 percentage scores respectively. According to onlinedegrees.unr.edu [6], surveys show that it is through the transformation and compilation of quality data collection enables management or businesses to make informed decisions. This process is no different from student career prediction systems.

*B. Implementation*

Papua New Guinea STEM syllabus is designed to capture all aspects of student learning ability guided by the Holistic Psychometric Assessment procedures. Prior to data collection, data integrity checks on student grades and scores must be validated in these two areas to ensure due process of assessment delivered as instructed in the syllabus.

- Teachers (Assessment Implementers) — Ensure that they implementing Holistic Psychometric student assessment on lesson topics and grades they record is an accurate result of this type of holistic assessments approach.
- Academic Head of Departments — Responsible for quality assurance. Their task is to validate every assessment piece before it is given to STEM students by the implementers.

By this point of the project, the identification of all possible STEM career fields in line with the subject/module topics taught would have been completed and the final dataset features for data collection would be confirmed and finalized.

Papua New Guinea STEM program is currently implemented in six National School of Excellence secondary schools. This project identifies them as Site1, Site2, Site3, Site4, Site5 and Site6. The pilot project will be conducted on Site1 which has a total of 40 current STEM students.

Data Selection — In this project we will select a subset of the data set to build the prototype. The subset will come from CS&ICT Department. Even though using more data from all the STEM departments is better for high accuracy, the advantage of using a small data set takes less time to process, easy to manage and uses less computer memory. The primary focus here is on quality and not quantity.

Having the subset data for prototyping on hand, it will be pre-processed to a format that the algorithm will understand. This will involve:

- Data Formatting — source data format has to be unified with the required format to fit inside the selected machine learning algorithm because sometimes data come in different formats.
- Data Cleaning — At this step any unwanted data is removed and instances of missing data are fixed. It is confirmation that there are no missing records and features in the data set.
- Data Sampling — This project will use a small data subset from CS&ICT Department to prototype. This step is essential as it will save time, computing power and will be faster solution for exploring prototypes.

It is extremely important that data must be kept in an organized format to train the model effectively.

Data Transformation — this step will involve scaling to ensure data received has similar properties and no odds, decomposition to pick the category that suits our ML Algorithm model and aggregation where the raw data set is aggregated with the purpose of extracting and normalizing data where seen fit if required.

*C. Machine Learning Algorithms*

To train the model, this project will use classification method supervised machine learning algorithm. This is because the input (student grades) and the output (career pathways) are already labeled with the aim is to predict a label or class. The machine will be provided with a label set of input and output data in the training phase itself. Basically, we will feed the output of the algorithm into the system before the machine starts working on it.

In this approach of supervised learning, all the algorithm has to do is map the student grades input to the career pathways fields which is the output (Fig 1).

*D. Result Analysis*

In supervised learning there is a direct feedback mechanism between the input and the output since the machine is trained with labeled input and output. According to the data set of this project, student grades and career pathway are both known. Results will be analyzed to see if the algorithm is working according to the objectives of this project.

As illustrated in Fig. 2 and Fig. 3, the goal of reinfusing output back into the model as training data in order to prevent model drift (i. e. the gradual degradation of a machine learning model's performance over time) [7, 2]. This will in turn ensure optimal performance of the model.

*E. Conclusion*

In general, students face a dilemma when it comes to choosing the right career pathway. In this age where the world demands specialization on every field, students need to be guided well when making lifelong career choices. Career Prediction System developed using machine learning is a very effective in predicting

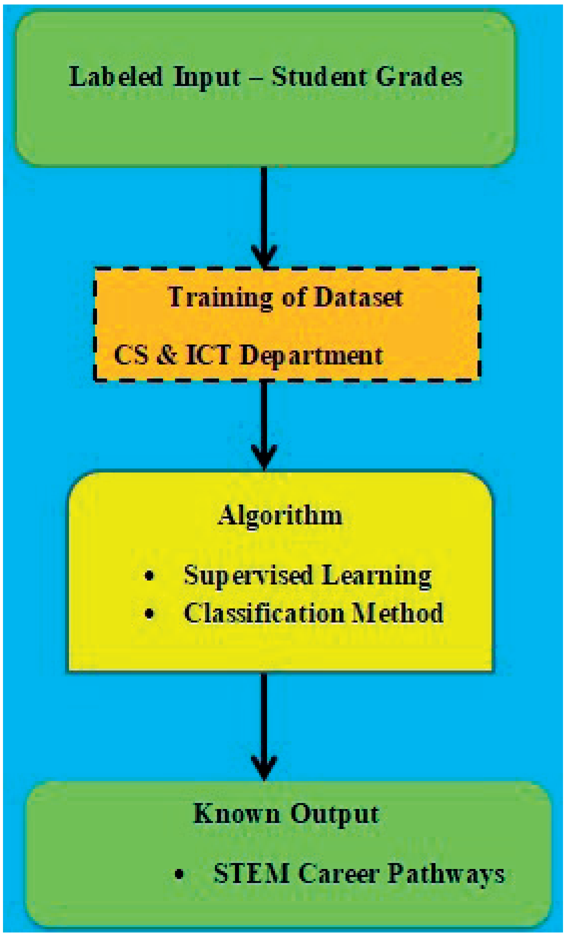


Fig. 1. Proposed Model Workflow

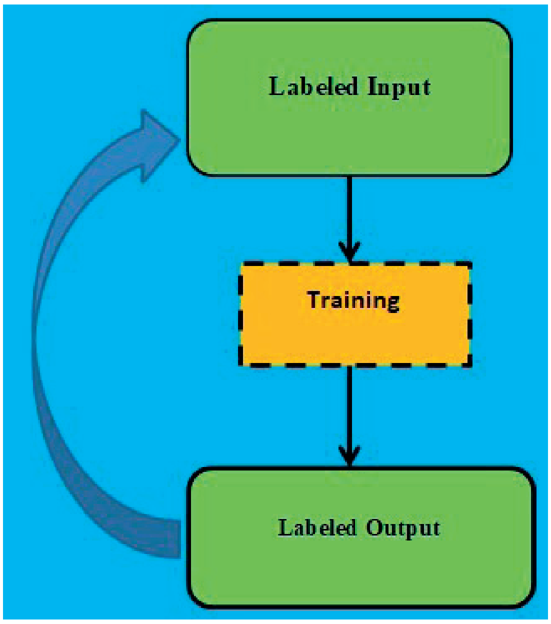


Fig. 2. Output Feedback

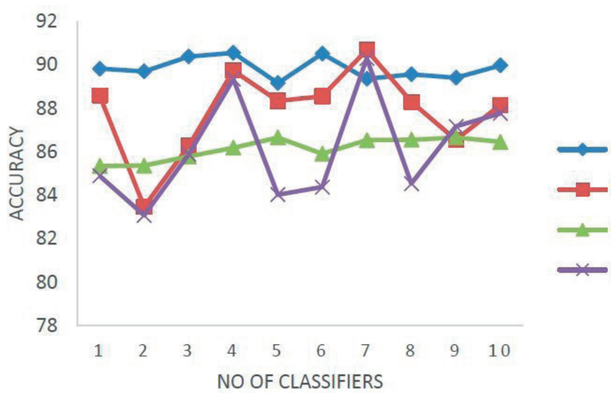


Fig. 3. Proposed Output Results

suitable career pathways based on the student cognitive ability. Students' career prediction or recommendation systems have their own advantages and disadvantages. Designing and building an effective system without a doubt will play a very important role in decision making on student career choices.

**References**

1. Abraham, A., Gressel, G. and Achuthan, K., 2019, December. Temporal Resilience of Phishing Detection Models in Machine Learning. In Proceedings of the 5th International Conference on Cyber Security & Privacy in Communication Networks (ICCS).

2. Agbozo, E., 2018. The role of data-driven e-government in realizing the sustainable development goals in developing economies. Journal of Information Systems & Operations Management, 12(1), pp.70–77.

3. Liu, R, Tan, AF.: Article title. Towards Interpretable Automated Machine Learning for STEM Career 12(2), 19–32 (2020).

4. Ajay Kumar Pal 2013 Classification Model of Prediction for Placement of Students, I. J. Modern Education and Computer Science

5. Madhan Mohan Reddy, V.: Career Prediction System, Department of Computer Science and Engineering Madanapalle Institute of Technology and Science, Andhra Pradesh, India,(2021).

6. Page Blog, URL: https://www.onlinedegrees.unr.edu/blog/

7. Nelson, K., Corbin, G., Anania, M., Kovacs, M., Tobias, J. and Blowers, M., 2015, May. Evaluating model drift in machine learning algorithms. In 2015 IEEE Symposium on Computational Intelligence for Security and Defense Applications (CISDA) (pp. 1–8). IEEE.

8. Roshani, A, Deshmukh, P.R S., Efficient Knowledge Transformation System Using Pair of Classifiers for Prediction of Students Career, Department of Computer Science and Engineering, Sipna College of Engineering and Technology, SGBAU, Amravati, 444607, India (2014).

9. S. Sathish, C. R. Rene Robin "An Optimized Result Analysis System for Institutions in India having Credit System" Vol 03, Issue 01; January-April 2012 ICICES-2012-SAEC