ISSN No:-2456-2165

Subject Stream Prediction: A Machine learning Approach to Select the Suitable Subject Stream for Senior Secondary Students in Sri Lanka

K. G. Kaushalya Abeywardhane , Anjalie Gamage
Faculty of Graduate Studies and Research Sri Lanka
Institute of Information Technology Colombo, Sri Lanka

Abstract:- Education is an important factor that measures the nation's wealth and education directly affects the country's future development. All children at all levels in Sri Lanka are entitled to free education up to the university level. During that period, the students have to face two essential examinations to complete their senior secondary education. According to the Sri Lankan education schema, students happen to select one subject stream to start their senior secondary education key stage 2. Most of the students select that subject stream without thinking deeper, thereby that decision will cause them to create a good future as well as not create. A prediction system which is called 'Subject Stream Prediction' predicts the appropriate subject stream to begin senior secondary education based on students' previous examination results as well as their skills, and preferred working area for their target career. If some student does not satisfy with one predicted answer, the model proposes ten appropriate subject streams with relevant jobs and educational and technical qualifications that need for those careers based on the above features. I have done a performance analysis between four machine learning algorithms to select the best-suited algorithm to predict the suitable subject stream by accessing their accuracy levels. That analysis demonstrates that the 'Random Forest Classifier' algorithm gives high accuracy (72).

Keywords:- *Machine Learning Algorithm, Subject Stream, Prediction System.*

I. INTRODUCTION

Education is a very important part of any society. Education makes people who can live in society well. When the child learns to read and write they feel more confident in their abili- ties. Education is also a human right in all nations. Therefore, it is the responsibility of every nation to solve the problems in the education systems and find the appropriate solutions. I plan to develop this system according to the academic circulars (2016/13) which were published by the educational ministry of Sri Lanka. In my research, I found a machine learning solution to solve problems that are occurring in senior secondary education in Sri Lanka. Machine learning is the newest tech- nology that enables the machine to learn from historical data automatically. Wide range of machine learning techniques: Supervised, Unsupervised, and Reinforcement is used to build various kinds of mathematical models for making predictions.

The proposed system predicts the suitable subject stream based on previous General Certificate of Education Ordinary Level (G.C.E.(O/L) examination results as well as students' inborn talents and main extra curriculum activities that they had done during school time. The ultimate goal of this system is to provide an appropriate subject stream to individuals based on his/her exam results, innate abilities, and preferred target career field. The student can see the other possible suggestions to reach his/her career with relevant educational and technical qualifications. To build the SubjectStreamPredict system, the relevant data were collected from 1000 employees who are doing various occupations in different working sectors. The gathered data have been used to train and test the four types of supervised machine learning algorithms: Decision Tree, Ran- dom Forest, K-Nearest Neighbor -KNN, and Support Vector Machine algorithm – SVM to predict the appropriate subject streams: Art, Maths, Biology, and Commerce. The results of the experiments demonstrated that the Random Forest algo- rithm outperforms the other algorithms when comparing the prediction accuracy score and error calculations. The Random Forest algorithm gives a high score (72) with the lowest error rate. I have added an extra function to the Random Forest algorithm to improve the prediction output. That function was implemented by giving scores to the user inputs and obtaining the results in descending order from the dataset.

II. BACKGROUND

According to the current education schemas in Sri Lanka, secondary education lasts eight years. It comprises two major parts. Junior secondary education is considered from grade 6 to grade 9. Senior secondary education is considered from grade 10 to grade 13. The students who obtain the government free education should complete two types of major examinations conducted by the education ministry of Sri Lanka. The General Certificate of Education (Ordinary Level-G.C.E.(O/L)) exam is happened to face at the end of grade 11. The General Certificate of Education (Advanced Level-G.C.E.(A/L)) exam is holding the end of grade 13. The student happens to select one major subject area to begin his/her senior secondary education part two and he/she happens to face the Advanced Level examination from that stream. [1] There is a huge competition among students to pass the A/L examination and enter a university because the career path of most students depends on the results of the A/L examination. That means a person's career journey is determined by the subject stream that person chooses to pursue in his/ her senior secondary

ISSN No:-2456-2165

education key stage 2. If a student chooses the right curriculum to begin senior education, it will be a big advantage to pass the A/L exam. According to current rules and regulations in the Sri Lankan education system, students used to select the A/L subject stream by depending only on the O/L examination results. Although a person's talents are somewhat determined by past examination results, it is not enough to recognize the hidden skills of that person. Therefore, most people miss the opportunity to decide their career path according to their skills. The results of this matter are that many students drop out of their senior secondary education. After facing trouble they started a new learning path but it wasted time and also money. Another thing is, that although the Sri Lankan government provides free education to students, the advantage that is gained from the government is less.

III. RELATED WORKS

According to the 'Recommender System Survey' conducted by J.Bobadilla and others, They have used collaborative filter- ing methods and techniques such as content-based knowledge and techniques to implement the system. They mainly used two collaborative methods: used-based and item-based. The used-based collaborative methods are important to find the similar features among two customers by analyzing their purchased items and services earlier . [3] M.C.B.Natividad and two others have implemented the system to recommend careers to help students who are learning in high schools to guide them to decide on their careers. The feature selection technique has been used to select the best attributes and prepare the data as crisp inputs and used an engine Fuzzy-based to predict a suit- able recommender. This system was implemented according to the education system in the Philippines. That means this system is working only for students in the Philippines. [11] According to the research, which was conducted by Dassari, Flaws, Abu, and three others, they have implemented the 'CareerRec' system. They used a machine-learning algorithm to suggest the career path of IT graduates. According to the implemented system, the career path is recommended based on some basic details related to IT graduates such as soft skills that are related to the IT field. They implemented that system by collecting data from job holders who are doing their jobs in IT field in Saudi Arabia. Then this system can be used only for persons that are relevant to the IT field and suitable for Saudi Arabic people only. They conducted a performance compression test of five machine learning algorithms: K-Nearest Neighbors (KNN), Decision Tree (DT), Bagging Meta-Estimator, Gradient Boosting, and XGBoost. According to their experimental results, the recommended XGBoost algorithm outperforms other algorithms and gives high accuracy of a percentage of 70.47 [1] Authors in [8] have implemented a system for making professional recommenda- tions and skill recommendations for a specific profession. This system has been focused on the educational background, work experiences, and skills of people who already doing the jobs. They analyzed these training data to recommend professions to job seekers by matching their profiles like their educational background and skills. They have used data mining techniques such as rule mining to develop this system. Kumar and Dr. P. Thambidurai also proposed a web recommendation system using the method of Fuzzy association rule mining. According to this system, relevant web pages are recommended support to the users. [9] Akshay Nagpal and Supriya P. Panda have implemented a career path suggestion system for graduates to reach their target careers according to their current education qualifications. String matching methods and the decision tree concept were used to develop the algorithm. They have imple- mented a system by considering only one most recent educa- tion qualification to predict the appropriate career. According to the current job market, most careers require more than one qualification, skill, experiences from job seekers. [10] Saqib Hakak, Amirrudin Kamsen, and five others have surveyed to propose a new classification and related possible challenges in the area of string-matching algorithms by focusing on the present exact string-matching algorithms. String matching can be used for matching the text, natural language processing, speech processing pattern recognition, etc. [5] Min Nie and four others have implemented a system which is called 'The approach Cluster Centers Based on XGBOOST' to predict students' career choices. They have constructed four types of features: Student's interesting reading, Mastery of professional skills, Behavior regularity, and Family economic status for the career choice prediction process. According to my point of view, some of the prediction categories that they used to implement to develop the algorithm are unfair for all kinds of students to predict the student's feature career. [12] Authors in [13] have implemented a recommender system for selecting training that related to followed degrees in 2019 by students. The data were collected via an online survey of 200 participants. The C4.5 algorithm was used to classify the data. The model that was implemented by the Decision Tree, obtained a high accuracy (percentage of 78.84). The model was developed with a web app that can be entered preferences of students to see company recommendations. Bassam Zafar and two others have implemented a Learning Management System (LMS) to investigate the grades of online activities. The data was collected from 241 undergraduates who flow six different courses at King Abdulaziz University. The five algorithms are used for their study: Random Forest, Decision Tree, sequential minimum optimization, multilayer perceptron, and Logistic Regression. According to their analysis results, the random forest algorithm was recommended to predict students' academic performance. [2] Ramraj .S. and three others have done a quantitative comparison of the accuracy and speed of the XGBoost algorithm in the multi-threaded single system and Gradient Boosting with different data sets.

Their research results have revealed that the accuracy of the XGBoost was not always higher than the Gradient Boosting algorithm but the speed of the execution of XGBoost has always been seen to be superior. [14]

ISSN No:-2456-2165

IV. RESEARCH METHODOLOGY

The Subject Stream Predict system was developed under five phrases. The main five pharses are in figure 1.

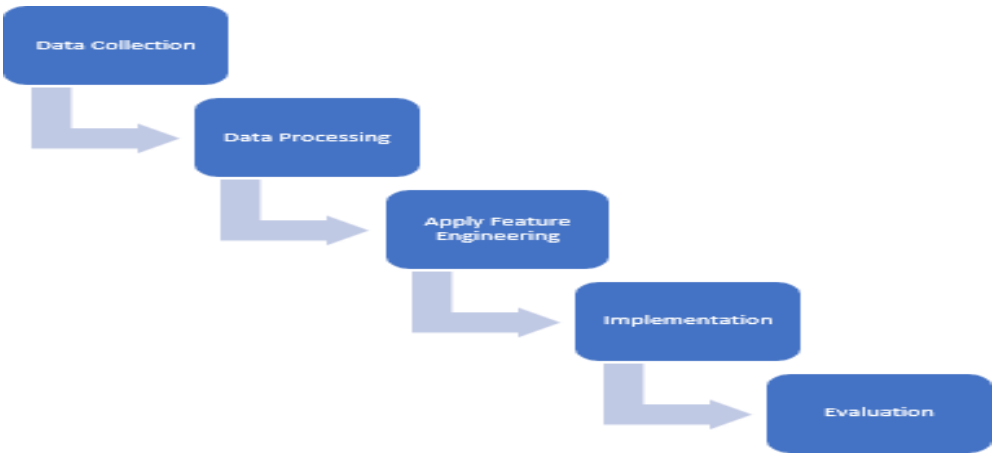

Fig. 1: Flow Diagram of Proposed Model

A. Data Gathering

I have collected data under 30 features: examination results, the subject stream that faced A/L, inborn talents, extra curriculum activities in school time, working sector, job role, education, and technical qualification that needed to follow that job, etc. from professionals, academicians and various kinds of job holders. Here I have removed the data of respon- dents who faced A/L in the technological subject stream. The technological subject stream was introduced in 2013 to the Sri Lankan education system thereby it was difficult to find job holders who did A/L in that stream. My target population size was 1000 job holders who are doing their job satisfactorily. To cover my population size, I collected data via an online survey by issuing Google forms. In my questionnaire, I asked to fill in the O/L results of main subjects, inborn talents, and extra curriculum activities in school time: the first category for initializing clubs or societies that students bear some post or membership, the second category for initializing students' behavior by confirming whether the student has done cadets, scouting, etc., the third category for collecting the details that student has attended and the last category for getting the achievements and victories of students. I used the above extra curriculum activities to measure the skills and preferences of the students. I asked to fill the inborn talents of respondents in levels wise as high, medium, and low. The job role, working sector, educational and technical qualifications details which they needed to reach their current job position were also asked to fill out. Finally, my question was about satisfaction with the job and if not, I have mentioned the reason for not satisfaction also. Some basic characteristics of collected data and its expansion is shown in figure 2. I gathered unique extra curriculum activities, talents, and skills for particular four main subject streams. The extra curriculum activities, talents, and skills that belong to the Art subject stream are Computer Ability, Singing, Public Speaking, Painting, Mem- orizing, Creativity, Acting, Writing Skills, Communication, Cartooning, Carpentering, Photographing and etc. Computer Ability, Singing, Public Speaking, Writing Skills, Communi- cation, Mathematical Ability and etc. kinds of skills, inborn talents, and extra curriculum activities are belonging to Com- merce subject stream. Critical Thinking, Analytical Ability, Strategic Thinking, Mathematical Ability, Decision Making, Problem Solving, Computer Ability and etc. kinds of skills, inborn talents, and extra curriculum activities are belonging to Science and Mathematical stream. [15], [6], [4]

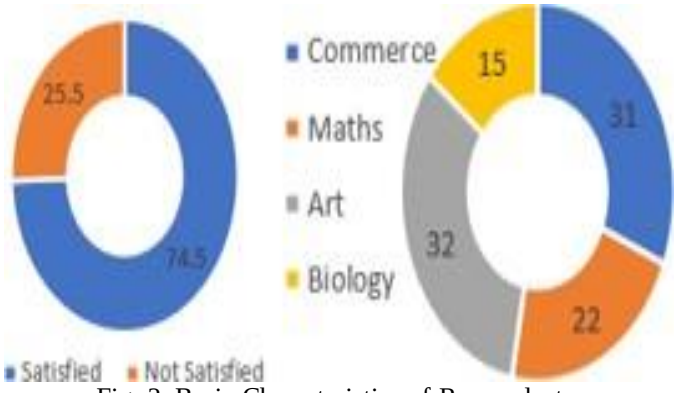

Fig. 2: Basic Characteristics of Respondents

ISSN No:-2456-2165

B. Data Pre-Processing

By using data preprocessing techniques, I rearranged the collected data to make it more suitable to build the model. I directly contacted job holders via their emails and mobiles and filled in the missing values and made corrections to fix the inconsistencies in the records. The participants used different kinds of names for the same features. It is harder to implement a machine learning model thereby improving the accuracy of the prediction, I have categorized those independent features as shown in table 1.

C. Feature Engineering

I have applied feature engineering methods to remove irrele- vant data such as contact details, gender data, and job location data. Label encoding was applied for other all categorical variables. The target variable: 'ALSubjectStream' was encoded with a value between 0 and 4.

D. Building the Model

Four machine learning algorithms were used to build the model.

- **K-Nearest Neighbors (KNN):** It is a supervised machine learning algorithm that is used to solve classification and regression problems. The techniques used there are that store all available data and classify new cases by comparing the similarity.

Independent Features	Categorized Responses
Inborn Tal- ents	Critical Thinking, Analytical Ability, Strategic Thinking, Mathematical Ability, Decision Mak- ing, Problem Solving, Computer Ability, Singing, Public Speaking, Painting, Computer Ability, Memorizing, Creating New, Innovations, Creativ- ity, Acting, Writing Skills, Communication
Extra curriculum Activity 1	Media Unit, Science Club, IT Club, Literacy Association, Math Club, English Unit, Physics Club, Robotics Club, Art Club, Music Club, Web Team, Student Council, Commerce Club, Sports Club, Other
Extra curriculum Activity 2	Cadets, Website Creating, Prefect Board, Blog- ging, Scouting, Eastern Band, Reading, Announc- ing, Drama, Western Band
Extra curriculum Activity 3	Cricket, Swimming, Athletics, Carrom, Basket- ball, Chess, Badminton, Volleyball, Netball, Table Tennis, Hokey, Other, Planting, Organizing Func- tions, Other
Extra curriculum Activity 4	School Level Art, School Level Math, School Level Science, School Level Commerce, District Level Art, District Level Math, District Level Commerce, Provincial level Art, Provincial level Math, Provincial level Science, Provincial level Commerce, School Level Technology, District Level Technology

Table 1: Independent Features Categorization

- **Decision Tree (DT):** DT also uses supervised machine learning techniques and mostly used classification prob- lems. Decision nodes are used to take the decision and leaf nodes are used to show the output of those decisions.
- **Random Forest Classifier:** Random Forest algorithms also can be used to solve classification problems and regression problems. It uses the concept of ensemble learning by combining multiple classifiers to solve com- plex problems.
- **Support Vector Machine Algorithm (SVM):** SVM also can be used to solve regression and classification problems and is mainly used for solving classification problems. Choosing an extreme point creates a decision boundary which is called a hyperplane that caused to put new data in the correct category easily.

V. EXPERIMENTS AND RESULTS

In this phase, I have analyzed the collected data and chosen appropriate features to train the above-mentioned algorithms.

A. Analyzing the Dataset

In this section, the analysis was done to find out which subject stream the respondents belonged to. I have removed the 25 percent of responses from data collection in which respondents mentioned that they do not satisfy with their job and are not like to go further from their career positions. By using respondents' data, I calculated the total count of unique responses for each independent feature in the subject stream- wise and calculated the percentage of those responses and got the total count. Figure 3 shows how those calculations were made for the 'Art Subject Stream'. Next, I calculated the total count of each unique feature of categorical variables related to extra curriculum activities and inborn talents. And after I calculated the total count of each feature in the subject wised. Next, I calculated the total count of each feature in the subject wise for three kinds of extra curriculum activities and inborn talents. Figure 3 shows the part of the total responses for each category related to the Art Subject Stream and figure 4 shows the total count of responses for each category in the subject stream wised related to Extra Curriculum Activity 1. Like wised I calculated the above values for other independent features: Extra Curriculum Activity 2,3, 4, and Inborn Talent.

ISSN No:-2456-2165

Subject Stream	Extra curriculum (Category 1)		Extra curriculum (Category 2)		Extra curriculum (Category 4)		Inborn Talents	
Art	Media Unit	48	Cadets	27	School Level Art	82	Critical Thinking	48
	Science Club	9	Website Creating	18	School Level Physical Subject Stream (Maths)	6	Analytical Ability	9
	IT Club	26	Prefect Board	38	School Level Science	15	Strategic Thinking	26
	Literacy Association	42	Blogging	24	School Commerce	21	Mathematical Ability	42
	Physical Subject Stream (Maths) Club	5	Scouting	34	District Level Art	69	Decision Making	5
	English Unit	30	Eastern Band	37	District Level Physical Subject Stream (Maths)	0	Problem Solving	30
	Physics Club	10	Reading	45	District Level Commerce	14	Computer Ability	10
	Robotics Club	10	Announcing	43	District Level Science	1	Singing	10
	Art Club	49	Drama	0	Provincial level Art	89	Decision Making	49
	Music Club	53	Western Band	35	Provincial level Physical Subject Stream (Maths)	10	Public Speaking	53

Fig. 3: Percentage and Total Count for Art Subject Stream

I could draw the following conclusions based on the above participants' responses.

- When considering extra curriculum activity category 1, more than 30 percent were obtained for Media Unit (48), Literacy Association (42), Art Club (49), Music Club (53), and Student Council (35) in Art Stream. More than 30 percent were obtained for IT Club (41), Physical Subject Stream (Maths) Club (30), and Commerce Club
- (75) in Commerce subject stream. The highest values were gained for Science Club (55), Physics Club (20), and Robotics Club (25) in Biology subject stream. The highest values were gained for Science Club (24), IT Club (30), Physical Subject Stream (Maths) Club (42), Physics Club (55), and Robotics Club (40) in Physical Subject Stream (Maths) subject stream.
- When considering extra curriculum activity category 2, the highest values gained for Prefect board (38), Cadets (27), Blogging (24), Scouting (34), Eastern band (37),
- Reading (45), Announcing (43), Drama (30) in Art sub-ject stream. More than 30 percent were gained for website creating (45), Blogging (40), and Website creating (47) in Commerce subject stream. The highest values were gained for Blogging (32), and Reading (28) in Physical Subject Stream (Maths) subject stream.
- When considering extra curriculum activity category 4, the highest values got for achievements that related to selected subject streams by respondents.
- According to the analysis, it seems that the respondents have paid considerable attention to their external talents while choosing the A/L level subject stream.
- When considering the inborn talents of the respondents, the features that got the highest values related to their selected subject stream. That means those responses have paid their attention to inborn talents when selecting a subject stream for the A/L examination.

From the responses given under the independent variable (Extra curriculum activity category 3), it was unable to take the significant opinion about determining the A/L subject stream. This independent variable was not getting for predicting the A/L subject stream. Based on the above analytical data, I could draw the following graph which is shown in figure 4. figure 4 shows the total count of responses for each category in the subject stream wised related to Extra Curriculum Activity 1. Like wised I calculated the above values for other independent features: Extra Curriculum Activity 2,3, 4, and Inborn Talent.

ISSN No:-2456-2165

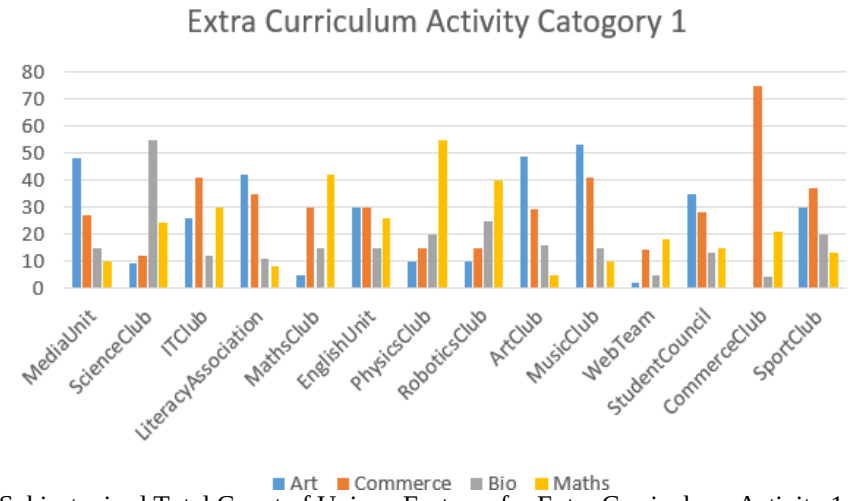

Fig. 4: Subject wised Total Count of Unique Features for Extra Curriculum Activity 1

The results of the respondents obtained in the O/L exami- nation were classified as shown in table 2 and calculated each grade percentage. Here I considered 'high capability' those who obtained distinction and very good passes for subjects, 'medium capability' those who obtained credit passes for subjects, 'low capability' those who obtained ordinary passes, and 'very low capability' those who obtained failure.

Main Sub-jects	High	Medium	Low	Very Low
Mathematics	50	36	13	
Science	44	38	14	33
Language	59	35	7	
English	33	37	21	9
Religion	49	37	12	2
History	45	37	15	3

The percentage of Mathematics high (50) and the percentage of science high (44) are higher than the percentage that did Maths Stream for A/L. It implies that if respondents obtained the highest results for Mathematics and Science, they faced different subject streams other than Maths and Biology. One possible reason is that if they have passed those subjects with high marks, their inborn talents vary. According to their inborn talents and skills, they have selected suitable subject streams for A/L, passed the A/L, and did their jobs satisfactorily. Other subjects like Languages, English, and Region were passed by most of the respondents and obtained distinction and very good passes by the respondents who selected the Art stream for A/L. The matter is the G.C.E(O/L) is not the proper way for measuring the skills and inborn talents of a student. This reality is not understood by most students. SubjectStreamPrediction tries to catch all inborn talents, skills, and students' preferences to predict the AL subject stream.

B. Predictive Analysis

All four algorithms were trained by using 745 data. I used the confusion matrix to determine the performance of the above 4 ML algorithms. The obtained results of the confusion matrix for those ML algorithms are shown in table 3.

Feature	DT	Random Forest	KNN	SVA
Accuracy Score	0.50	0.72	0.28	0.33
Micro Pre-cision	0.50	0.72	0.28	0.33
Micro F1-Score	0.50	0.72	0.28	0.33
Macro Pre-cision	0.39	0.60	0.09	0.20
Macro F1-Score	0.39	0.60	0.09	0.20
Weighted Precision	0.48	0.73	0.14	0.26

Table 3: Confusion Matrix Results

The accuracy of the 4 ML models shows in figure 7 graph- wise. According to the graph results Random Forest algorithm achieved the highest accuracy score (72.2) and with closer micro and macro precision.

ISSN No:-2456-2165

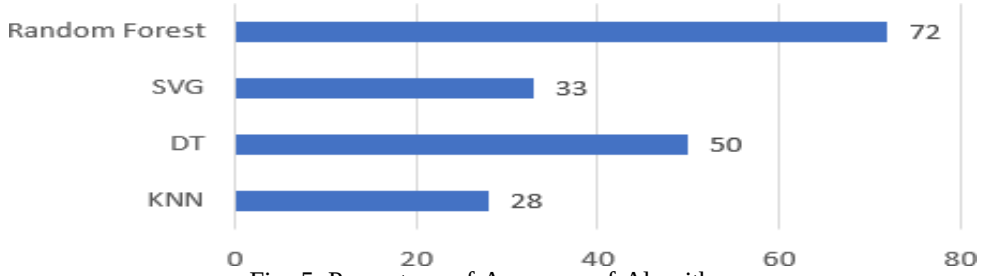

Fig. 5: Percentage of Accuracy of Algorithms

C. Improve the Predictive Output

The Highest Score Algorithm: Random Forest Classifier was selected to develop my ML model: SubjectStreamPrediction system. Streamlit open-source framework was used to develop the model as a web application. Python language was mainly used to build the code. I added an extra feature to improve the performance of the predicted output which was gained from the Random Forest Algorithm. Many researchers developed Random Forest Regression models with multi-output to predict more than one feature in the output. [7] The researchers in [16] have developed their 'Disease Prediction System' by using the hybrid method which combined Random Forest and multivariate adaptive regression splines. But Random Forest Classier models did not use that multivariate output function. According to my ML model just giving one suggestion with one feature to users is not reasonable because the predictive answer affects to decide the whole future of the user. I tried to give at least 10 suitable suggestions with four features to the user as predictive output from this model. For that, I have developed a scoring system to predict more than one output with more than one feature. According to the user input values newly added function calculates the score by using values that the system assigns earlier, and the system gives ten outputs that have high scores gained in descending order. While Random Forest Classifier predicts the best subject stream, this function helps the user to know what are jobs the user can do according to their preferable working fields, educational paths to reach those job positions, and technical and professional qualifications that needed to do those suggested jobs. This feature will help the user if the user does not satisfy the one answer which has fewer features. The main interfaces of the web application are shown in figure 7.

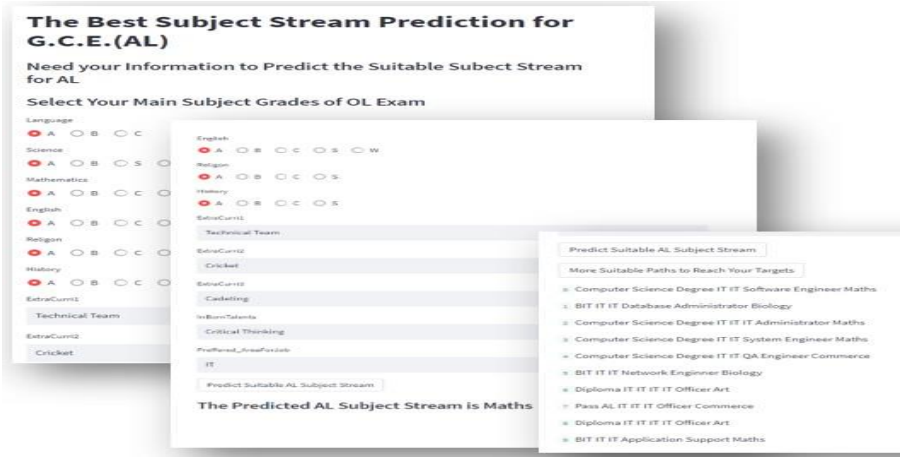

Fig. 6: Main Interfaces

I use the technique of cross-validation (CV): K-Fold CV to evaluate the implemented model. First, I checked the parameters currently used by the Random Forest and got the results as n-estimators:10, random-state:10, etc. Next, I created the parameter grid to find the best parameters suited to the model. According to the grid search results, I narrowed the range of hyperparameters. For evaluation purposes, I compared the best model with the base model. Then the average error of the best model (0.4737) was less than the average error of the base model (0.5763). For evaluation of the model with users, I randomly selected the 20 job holders to check whether the model predict the suitable output for the users. According to the calculated results, the percentage of the system predicting the correct subject stream was 75 and the percentage of the system predicting the correct job role was 65.

VI. CONCLUSION

From this research, I have proposed Subject Stream predic- tion system to predict the best subject stream by depending on the O/L results inborn talents and skills of users. To implement the model four machine learning algorithms were used to build the system. All four algorithms were trained and tested using 745 of user responses. By analyzing the accuracy levels of four machine learning algorithms, Random Forest Algorithm outperforms the other three algorithms in predicting the best subject stream. To improve that output, more data should be used to train the model.

ISSN No:-2456-2165

REFERENCES

[1.] Hmood Al-Dossari, FA Nughaymish, Z Al-Qahtani, Mohammed Alkahlifah, and Asma Alqahtani. A machine learning approach to career path choice for information technology graduates. *Engineering, Technology & Applied Science Research*, 10(6):6589–6596, 2020.

[2.] Amal Alhassan, Bassam Zafar, and Ahmed Mueen. Predict students' academic performance based on their assessment grades and online activity data. *International Journal of Advanced Computer Science and Applications*, 11(4), 2020.

[3.] Jesu´s Bobadilla, Fernando Ortega, Antonio Hernando, and Abraham Gutie´rrez. Recommender systems survey. *Knowledge-based systems*, 46:109–132, 2013.

[4.] Anja Brezovnik. The benefits of fine art integration into mathematics in primary school. *CEPS journal*, 5(3):11–32, 2015.

[5.] Saqib Iqbal Hakak, Amirrudin Kamsin, Palaiahnakote Shivakumara, Gulshan Amin Gilkar, Wazir Zada Khan, and Muhammad Imran. Exact string matching algorithms: Survey, issues, and future research directions. *IEEE access*, 7:69614–69637, 2019.

[6.] Tim N Ho¨ffler, Christine Ko¨hler, and Ilka Parchmann. Scientists of the future: an analysis of talented students' interests. *International Journal of STEM Education*, 6(1):1–8, 2019.

[7.] Kevin Hoffman, Jae Yoon Sung, and Andre´ Zazzera. Multi-output random forest regression to emulate the earliest stages of planet forma- tion. In *2021 Systems and Information Engineering Design Symposium (SIEDS)*, pages 1–6. IEEE, 2021.

[8.] Maria Kiran, Hira Asim, and Malik Tahir Hassan. Career and skills recommendations using data mining technique: Matching right people for right profession, in pakistani context. 2019.

[9.] A Kumar and P Thambidurai. Collaborative web recommendation systems based on an effective fuzzy association rule mining algorithm (farm). *Indian Journal of Computer Science and Engineering, Engg Journals Publications*, 1(3):184–191, 2010.

[10.] Akshay Nagpal and Supriya P Panda. Career path suggestion using string matching and decision trees. *arXiv preprint arXiv:1505.06306*, 2015.

[11.] MCB Natividad, BD Gerardo, and RP Medina. A fuzzy-based career recommender system for senior high school students in k to 12 edu- cation. In *IOP Conference Series: Materials Science and Engineering*, volume 482, page 012025. IOP Publishing, 2019.

[12.] Min Nie, Zhaohui Xiong, Ruiyang Zhong, Wei Deng, and Guowu Yang. Career choice prediction based on campus big data—mining the potential behavior of college students. *Applied Sciences*, 10(8):2841, 2020.

[13.] Adewale Opeoluwa Ogunde and Joy Oseremen Idialu. A recommender system for selecting potential industrial training organizations. *Engi- neering Reports*, 1(3):e12046, 2019.

[14.] S Ramraj, Nishant Uzir, R Sunil, and Shatadeep Banerjee. Experi- menting xgboost algorithm for prediction and classification of different datasets. *International Journal of Control Theory and Applications*, 9(40), 2016.

[15.] Robert Root-Bernstein and Michele Root-Bernstein. Artistic scientists and scientific artists: The link between polymathy and creativity. 2004.

[16.] Dengju Yao, Jing Yang, and Xiaojuan Zhan. A novel method for disease prediction: Hybrid of random forest and multivariate adaptive regression splines. *J. Comput.*, 8(1):170–177, 2013.