**Predicting CSE Undergraduate Students’ Careers Using Machine Learning Algorithms in the Context of Bangladesh.**

**Abstract/Summary:**

This thesis presents a predictive analysis of students' career paths for Computer Science and Engineering (CSE) students in Bangladesh using machine learning techniques. The rapid evolution of the CSE field and the diverse career opportunities it offers make understanding students' career trajectories crucial for educational institutions, policymakers, and career counselors. However, predicting career paths accurately is challenging due to the multifaceted nature of factors influencing career decisions.

**Introduction:**

Computer science and Engineering(CSE) mainly focuses on computers and their associated aspects including hardware, software, algorithms, computational theory and  also chose non CSE job sector in Bangladesh. Understanding and predicting students’ career paths can provide valuable insights for educational institutions, policymakers, and career counselors to better support students to achieve their professional aspirations. This thesis proposal outlines a research study aimed at utilizing machine learning techniques to predict the career path of CSE students in Bangladesh.

**Related Literature Review:**

Previous research has explored various aspects of **Career Path Prediction Using Machine Learning Classification Techniques globally**.However, there is a dearth of studies specifically focusing on the context of Bangladesh, particularly in the CSE domain.This thesis will review existing literature on career development theories, factors influencing career decisions, and machine learning applications in educational research.

**Research Methodology:**

Gather data from CSE students in Bangladesh through surveys, questionnaires etc. The dataset will encompass a wide range of variables, including demographics, academic performance metrics, various skills, extracurricular activities, internship experiences, and their preferred job.

Some factors are list down:

General question:

|  |  |  |
| --- | --- | --- |
| 1. Name | 2.University | 3. Studying year |
| 4. Communication skill | 5.Adaptability | 6.Continuous Learning |
| 7.Teamwork | 8.Problem solving skill | 9.Management skill |

Academic Skill:

|  |  |  |
| --- | --- | --- |
| 10.Programming skill | 11.Software Engineering Principal | 12.Data structure and Algorithm |
| 13.Database Management | 14.Data Analysis | 15.Web developing skill |
| 16.Understanding of computer architecture and system | 17.Understanding of OS | 18.Networking Concept |
| 19.Cyber Security | 20.Research skill | 21.current cgpa |

Influencing factor:

|  |
| --- |
| 1. Social acceptance 2. Job security 3. Job environment 4. Selary |

Prefer job

|  |
| --- |
| 1. Software Development Company 2. Bank job 3. BCS 4. Business 5. Teaching 6. Researcher 7. Govt job 8. Abroad 9. other |

**Expected Results:**

Present the findings of the predictive analysis, including insights into significant factors affecting students’ career paths in the CSE field. Discuss the implications of the results for educational institutions, policymakers, and career counselors. Compare the performance of different machine learning algorithms and highlight the strengths and limitations of the predictive models.

**Significance and Implications of the Study:**

The significance of this thesis proposal lies in its ability to leverage machine learning to forecast career paths for Computer Science and Engineering (CSE) students from Bangladesh. By providing insights into students' career trajectories, this research can guide educational institutions, policymakers, career counselors, and students themselves in making informed decisions about education, workforce development, and individual career aspirations.

**References:**

<https://ijsrcseit.com/paper/CSEIT21832.pdf>

<https://www.internationaljournalssrg.org/IJCSE/2022/Volume9-Issue3/IJCSE-V9I3P102.pdf>

<https://www.researchgate.net/publication/370984871_Prediction_of_the_Computer_Science_Department's_Educational_Performance_through_Machine_Learning_Model_by_Analyzing_Students'_Academic_Statements>

<https://www.webology.org/abstract.php?id=2388>

# Introduction:

Traditionally, student performance can be estimated using questionnaires. However, this method takes a lot of time. Computers now play an important role in many fields. Machine learning is one of the new technology trends. In this digital world, machine learning, imaging, classification, analysis, regression, etc. It is increasingly used in many fields and industries.

It is possible to create and research automation without special instructions. There are three types of machine learning: supervised machine learning, unsupervised machine learning, and reinforcement machine learning algorithms. In short, machine learning is the science of learning and acting like a human. Assessing the student's ability is very important and should be well received. In this study, the concept of machine learning was used to determine the next learning level of students. This estimate can be used for any type of school, recruiter, etc. is important for Based on the results of the correct forecast, schools can identify weak employees and provide the necessary training to improve their performance. Employers also spend a lot of money choosing the right candidate

face difficulties in choosing careers that satisfy them as well as their skills and interests. Choosing the right career path can have a huge impact on a student's future professional and personal success. To make an informed decision, students must consider many factors, including their strengths, passions, values, and career prospects. They will also need to research a variety of job and career opportunities, conduct interviews and seek guidance from a mentor or mentor. In such situations, predictive models and tools can provide valuable information to help students evaluate career options and make more informed decisions.The model uses a data-driven approach to identify patterns and trends in the market and match skills and personal preferences to career opportunities. we will explore some of the key factors affecting student per formance and discuss the benefits and limitations of predicti ve models in this context. We will also provide some tips an d ideas to help students make better career decisions. The results of the prediction model are also used to find out whether students are willing to study or want to study their higher education. This research project is mainly focused on estimating the work of undergraduate students. Machine learning algorithms such as SVM, DT, RF and Adaboost classifiers are used to build the model. Among the above products, RF performs better. These classifiers are used with the help of the python programming language because most of the time problems are easily implemented by this programming language. The next section presents the theories and methods used by various authors in forecasting research.

Literature Review 1. Iqbal et al discuss various machine learning techniques for predicting student performance on a variety of subjects. Methods such as matrix factorization, classification and regression were used in the analysis of data collected by ITU Pakistan. They evaluated performance using machine learning methods and found that RPM is the best among many machine learning methods (Iqbal, 2017). 2. Vaidu et al. used machine learning based on student performance to predict job skills. They used KNN and Naive Bayes model to divide students into different groups. The KNN algorithm predicts the student function with an accuracy of 95.33%, which is the purpose of Naive Bayes67. 3. Predicting our future performance after students, Byung-Hak et al. using a deep learning-based Grit Net algorithm. According to this research paper, Grit Net gives more accurate results based on logistic regression. They received information from Udacity Nanodegree Programs (Schaar, 2017). 4. Jie et al. also proposed a machine learning method to predict student performance in higher education programs. In this study, students' past and present performances are evaluated. It uses a two-tier model with multiple levels of prediction and a data-driven approach for bottom-up prediction based on quality results. Research data show that the proposed method gives better results than the basic method. 5. Pojon Murat reviews machine learning algorithms used to predict student performance. Pojon Murat used three different algorithms: Linear Regression, Naive Bayesian Classification and Decision Trees on two different datasets, of which Roberson et al. is a prominent engineering version. According to the results, Naive Bayes is the best method since it gives 98% accuracy for the first data, while Decision Tree is the best method because it gives 78% accuracy for the second data.

6. Singh et al. They used some machine learning techniques to predict students' performance in engineering courses.By examining the grades of the courses in the previous semester, they predict the grades of the students in the next courses. For this, decision tree classification and Naive Bayes method were used and it was seen that the decision tree gave more accurate results than Naive Bayes (Bendengnu Ksung, 2018). 7. Using machine learning like support vector machines, random forests, gradient boosting and Naive Bayes, Pushpa S and others. It can predict whether a student will fail or pass a grade in the previous semester. As for the prediction, Random Forest has an accuracy of 89.06%, which is more than other algorithms (Pushpa, 2017). 8. Bendengnu Ksung et al. Using DNN Models, eg. Deep Neural Networks for Predicting Student Webology (ISSN: 1735-188X) Vol. 18, no. 6, 2021 3509 http://www.webology.org reports. A research paper by Bendengnu Ksung et al. compared DNNs to machine learning algorithms such as Naive Bayes, Neural Networks, and Decision Trees. Accordingly, DNN achieved an accuracy of 84.3%, which is the performance of machine learning (Gerritsen, 2017). 9. To predict student performance, Gerritsen L. et al. using data from learning management about data learned from the use of neural networks. For this article, it is accepted that the Moodle log dataset contains the data of 4601 students. In this paper, the performance of the neural network is compared with six known as K-Nearest Neighbors, Naive Bayes, Decision Trees, Support Vector Machines, Logistic Regression and Random Forests.According to the data, the neural network gives more accurate results than the other six groups (Hernandez, 2018). 10. To estimate a school's exit list, Martin S. et al. It uses four machine learning methods to measure performance. The four methods include random forest, support vector machine, logistic regression, and neural networks.For this research data, data from students from the Costa Rican Institute of Technology (ITCR) between 2011 and 2016 were used. Among the four algorithms, the random forest algorithm is the best choice for college education cost estimations.

11. K. Sripath Roy (2018) Using machine learning algorithms such as SVM decision tree and XG boost to build a student work prediction model.Among these algorithms, SVM reaches 90.3% and gives more accurate results. 12. Mubarak Al Baraka Umar Jesus uses neural network technology to predict student learning. In this study, a neural network model was developed to predict a student's GPA using the student's personal information, opposition, and academic records.According to this model, the prediction accuracy is 73.68% 13. Pardos, Baker, San Pedro, Gowda, and Gowda (2014) also studied Service, but focused on discussing students' thinking and collaborative behavior. Betting math tests can help students get bored, focused, confused, etc. They developed an intervention and collaborative behavior system to measure outcomes using eight technologies to learn to predict outcomes. They also developed a model to predict students' math test scores and showed that the developed tool helped the model achieve high-level predictions. 14. Baker, Berning, Gowda, Zhang, and Hawn (2019) present a case study that uses data on student discipline, attendance, academics, and grades to identify students at high risk of failing in high school. The regression model not only helps authors select student at-risk

One of the main challenges of schools and universities today is providing programs that are aligned with the policy of commission on higher education (CHED) and delivering consistent outcomes that can be accepted not only in the Philippines but also in other countries. The higher educational institutions shift their paradigm from just simply educating the students to developing life-long learning skills. The CHED in the Philippines initiated reforms for education sectors through the conversion of the traditional way of teaching known as teacher-centered to an outcomes-based education (OBE) curricula that are commonly known as student-centered, as per CHED-CMO No. 46 s. 2012 [1]. The higher education institutions’ programs offering and syllabus were based on the principles of OBE by which accentuate the type of delivery of its services to the students. To strengthen and assist the country by producing graduates with critical thinking, behavioral and life-long learning skills and competencies aligned with institutional learning outcomes, industry desired values, and international standards [2]. Graduate readiness study has also been conducted by other scholars, where the aim of their paper is,(1) to investigate student’s experience when it comes to essential skills acquired in the university ; (2), student perception when it comes to potential job role when they graduate [3]. According to a recent survey by the Malaysian Ministry of Education assessing “(youth)” unemployment and graduates, In particular, just 53 percent of the 273,373 graduates find jobs within six months of graduation in 2015, 24 percent of graduates have no job after graduation and 18 percent were engaged in continuing education. That is why, as mentioned, only 53% of students were employed because of “the discrepancy between the education provided at the universities and the skills needed by the industry. According to the researchers in Malaysia and China, most of the university curriculum where they conducted their studies reflect the current skill requirements of the industry [4]. In a Alsore researchers have studied graduate employability. Research utilizing data mining and modeling methods has been carried out that highlights particular computing and data management challenges. [5]. Using the data analytics, the results were evaluated by (1) monitoring the job status of graduates by giving them prompts and invitations; (2) encouraging them to maintain track of the position they wanted; and (3) determining the jobs approaches function well, especially in the sector-specific region. [6]. Unfortunately, the complexity of the workplace and the advent of modern technologies have shifted, and the varied demands. of consumers demand the definition of “globally competitive” for employable graduates. It also challenges the capacity of universities to satisfy the need for graduates who are suitable for the job in the industry [7]. These concepts of career development and employability lead college education to evaluate their program offerings and to test their effectiveness and congruence with the needs of the sectors to deliver qualified students who can quickly be consumed by industries. The assessment of being suitable for the job often follows the theory of human capital, in which individuals’ personal and technical growth are called assets in human capital that acts as factors for their degree of employment and personal earnings. Graduates will then make substantial improvements in their intellectual resources to allow their potential employers more marketable [8]. However, It is increasingly important for individuals to retrain or develop new knowledge and abilities to address the requirements of the rapidly growing workforce and the multifaceted entrepreneurial environment. They will develop their professional abilities, values, and work experience and adapt to the changing labour market requirements. Most published researches and studies used data mining techniques to predict employability. Some of the techniques were Tree of Decision, Naïve Bayes, and Vector Machine Support [9]. Often used in data mining techniques are the Logistic Regression, K-Nearest Neighbor, Random Woodland, SVM (Linearsvc), Quadratic Discriminant Analysis (QDA), and Multi-class Ada Boosted [10]. The application of Machine learning when it comes to forecasting employability is in the infancy period. They compared numerous algorithms in the analysis carried out by Ohio University where the datasets used were from business education. The aforementioned research will not find the datasets regarding the mismatch [11]. This paper seeks to establish a machine learning method to forecast the employability of the applicant and to examine the signs of their skill set. This paper is in the production stage of a model focused on machine learning to forecast the employability of students. The researchers were inspired to perform the study in the light of emerging areas such as operational intelligence or instructional analytics to strengthen and encourage certain ability sets found that will lead to the enhanced jobs of engineering students.

Engineering is one of the best career streams apart from medicine, which most students are opting for, some due to interest and some due to parental pressure, as it is the most defined career option in the world. Many engineering students come out of college every year. Many students choose their stream after their graduation. Opting for the right carrier has become a complex science nowadays, as there are multiple career options and job competitions in the market. Researchers have even suggested machine learning classification technology to explore the right career option. Students face problems choosing the right career path without proper guidance from professional services. They often mismatch their career path regarding their personality, skills and interest. Students are even forced to opt for a career stream in engineering as pressure from family and the greed for high pay. The students in the past who have passed engineering and started working for MNC but still lack interest and skills make them unhappy. As a result, the upcoming generation has now started to opt for the streams that interest them. Machine learning technique for career guidance has been developed for engineering graduates who have completed their graduation or in the last semester and are still confused about which part of the field they should opt for. It’s a big challenge for those students to make the correct decision regarding the career they choose, as their complete future depends upon this. Thus, we have considered other aspects that will help us choose the right career path based on the academic score and personality, which is important for making their decision. This research aims to develop a classification model for predicting career options after engineering undergraduate students. This study aims to govern whether the student’s academic performance is determined by aptitude or personality and develop a model to analyse students’ performance. Machine learning is the best technique to automatically identify and analyse the data and use them to perform the predictions automatically. Processes for many data are analysed to discover patterns and rules after data. A computer can easily process and define these generated rules and patterns to characterise the new data. It is an automatic process that helps to improve the updated data. Therefore, as a result, it helps the students improve their learning activities and helps them analyse their career path for the future. Classification techniques like KNN, SVM, SGD, Logistics Regression, Decision Tree, and Neural Network are applied to calculate educational performance at the global level after students. The prediction model in terms of student-related variables is assessed. In this paper, the future career options for students are predicted based on their skills, interest, hobbies, links, etc. the rules learned are portrayed in the context of the decision tree. In this work, the SVM algorithm is preferred to improve prediction accuracy. The application of Machine Learning algorithms involves data analysis, visualisation of data, performance prediction, providing feedback and recommendations, and grouping of students. For this, student performance data is pre-processed to extract features and select them. It also involves data cleaning, tokenisation, sentimental analysis, and removal of words. At the end of the preceding processes, the student’s final performance is obtained, which will help them analyse the best career options they should opt for after graduation in engineering. The rest of the paper is summarised as follows:- Section 2 addresses the literature review. Section 3 represents the proposed work. Section 4 discusses the research design and methodology. Section 5 describes the results and discussions of our research work. Section 6 shows the conclusions of the entire research. Finally, Section 7 lists the references. 2. Literature Review In this section, we have reviewed some papers in the related area. Iqbal et al. have discussed various machine learning techniques to predict grades after students in various courses. Models such as matrix factorisation, classification, and regression are used to analyse the collected data from ITU, Pakistan. They have evaluated performance using machine learning techniques, and it has been found that RPM is the best among various machine learning techniques. (Iqbal,2017) Vaidu et al. have implemented machine learning techniques based on student performance to predict their employability skills. They have used KNN and Naïve Bayes models to classify the students into numerous groups. The prediction of the students’ employability from the KNN algorithm is 95.33% accurate, which is all for the Naïve Bayes is 67.67% accurate. (vadiu, 2017) To predict our future performance after students, Byung Hak et al. have used a GritNet algorithm based on deep learning. As per the logistic regression, GritNet gives more accurate results, according to this research paper. They have taken data from the Udacity Nanodegree Program. (B.H,2018) Jie et al. also proposed a machine learning approach to predict student performance in degree programs. In this investigation, the past, as well as present the performance of the students is evaluated. It uses a bi-layered structure that compromises multiple phase predictors and a data-driven approach based on efficient factors to base prediction. This research paper has shown that the proposed method gives a more accurate result than the benchmark approaches. (Jie,2017) The machine learning algorithms examined by Pojon Murat et al. are used to predict student performance. Pojon Murat et al. have used three different algorithms, Linear regression, Naïve Bayes classification, and decision tree, on two separate data sets, Roberson and another one featuring an engineering version. As per the result, Naïve Bayes is the best technique used for the first data set as it gives an accuracy of 98%, while the Decision Tree is the best technique for the second database as it gives an accuracy of 78%. Singh, M. et al. have used some machine learning techniques to predict the academic performances of the students’ subjects wise in their engineering field. To analyse the subject’s scores based on the previous semester, they predict the success scores of the students in the ongoing courses. For this purpose, decision tree classifier and Naive Bayesian techniques have been used, and it has been shown that the decision tree gives a more accurate result than Naïve Bayes. (Singh, M., 2013) Using machine learning techniques like Support Vector Machine, Random Forest, Gradient Boosting, and Naïve Bayes, Bendangnuksung et al. predict the student’s performance, whether they will fail or get a pass in the previous semester. As per the prediction, the accuracy rate of Random Forest is higher than other algorithms, that is, 89.06% (Bendangnuksung, 2018). Pushpa et al. have used the DNN model, Deep Neural Network, to predict student performances. The research paper by Pushpa et al. compares the DNN with machine learning algorithms like Naïve Bayes, ANN, and Decision Trees. According to this, DNN achieves 84.3% accuracy, which is better than Machine Learning Techniques. (Pushpa,2017) Gerritsen L. et al. used data from Learning Management System about educational data using Neural Networks to predict student performance. For this paper, a Moodle Log data set is considered that has a file that contains 4601 students’ information. In this paper, the performance of Neural Networks is compared with six classifiers named K- Nearest Neighbor, Naive Bayes, Decision Tree, Support Vector Machine, Logistic Regression, and Random Forest. According to this paper, Neural Network is more accurate than the other six classifiers ( Gerritsen, L., 2017). To predict the University’s dropout student list, Hernandez et al. have used four techniques of Machine Learning to analyse the performance. These four techniques include Random Forest, Support Vector Machines, Logistics Regression, and Neural Networks. For this research paper, the dataset of institute Technological de Costa Rica (ITCR) students is used who enrolled between 2011 and 2016. Among these four algorithms, the Random Forest algorithm is preferred to be best to predict University dropouts. (Hernandez,2018) García-Peñalvo et al. have used the new artificial Neural Networks to predict the students’ career paths based on the dataset. He proposed a data-driven system to collect the data to predict all the future career paths available. (García-Peñalvo, 2018) The machine learning algorithms such as the SVM decision tree and X.G. boost are used by K.Sripath Roy et al. to create a model of student career predictions. Among these algorithms, the Support Vector Machine gives the most accurate result, that is 90.3%.(k. Sripath Roy,2019) Mostafa et al. proposed artificial neural network technology to predict students’ academic performance. In this study, the model of neural networks is created, which predictor students’ GPA by using their personal information, place of resistance, and academic information. According to this model, the accuracy of prediction is 73.68%. (Mostafa,2021) 3. Proposed Work Intense websites and web applications help students know their suitable career paths. Still, the drawback of this system is that they only use personality traits to predict the career, which might not give a consistent result. Similarly, numerous websites suggest students opt for a career per their interests. These systems cannot understand whether the student can survive in that particular field or not. Beth Dietz-Uhler & Janet E. Hurn’s paper suggests a need for learning analytics to predict and improve student performance to enlighten the importance of students’ interests, trends, abilities etc. (UD Beth, 2013). Lokesh Katore, Jayant Umale, and Bhakti Ratnaparkhi’s paper predicted that different classifiers accurately predict a student’s career (K.S. Lokesh, 2015).10 Let’s look at these machine learning algorithms used to develop the classifier. It will help students to predict careers after their graduation in engineering. 3.1. K Nearest Neighbor The KNN algorithm is used for boat regression and classification problems. This algorithm categorises all the cases into new ones based on key neighbours’ majority votes. It is most suitable for classification rather than regression. Here are the following steps that are involved in the process of K&N: Select some neighbours, for example, K =10; • Now calculate the Euclidean distance, that is, the distance between two different data points; • Now let’s categorise the neighbours based on distance; for example, the four nearest neighbours are in category A; after that three nearest neighbours to be category B, and the remaining three be in category C; • Now, this is the new data that has been prepared by using KNN. 3.2. Support Vector Machine Supervised learning at Corrigan is used for classification and regression problems. The goal of the SVM algorithm is to mark the decision boundaries that are also called a hyperplane. This algorithm divides different training data sets into various classes based on the hyperplane. In this algorithm, the n-dimensional space is plotted where it belongs to the number of features with the value of a particular coordinate. It helps to maximise the classifiers’ margin. 3.3. Stochastic Gradient Descent The Stochastic Gradient Descent classifier is the most efficient technique for machine learning classification problems under the Linux support vector machine and logistic regression. The Stochastic Gradient Descent classifier is merely an exaggeration technique that has not corresponded to the specific dataset of machine learning classification techniques. It is the most efficient algorithm and easy to implement. Here, the linear support vector machine has been used as a classifier so that the Stochastic Gradient Descent algorithm can be applied to optimise the accurate result. 3.4. Logistic Regression It is a regression technique of machine learning used to perform classification problems. Logistic regression is used to calculate the probability of the given class-specific data. If there is more than 50% accuracy, the value is assigned to that class, and if the result is less than 50%, it will assign the value to the other class. Thus, it is stated as a binary classifier. 3.5. Decision Tree This technique is used to conclude some conditions and flag this as a node and a leaf node. The first decision mode tells us about the attribute to be selected, while the leaf node tells us about a class. It is a primary node, also called root node pictures selected based on two different methods: • Gini index Method • information Gain Method