

**ALUMNI JOB ANALYSIS & SUCCESS RATE PREDICTION
BY MACHINE LEARNING TECHNIQUES**

BY

K. M. Mostakim

ID: 172-15-10176

Anisur Rahman

ID: 172-15-10177

AND

Sohanur Rahaman

ID: 172-15-10199

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering.

Supervised By

Nishat Sultana

Lecturer

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

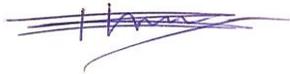
DHAKA, BANGLADESH

JUNE 2021

APPROVAL

This Project titled "**Alumni Job Analysis & Success Rate Prediction by Machine Learning Techniques**" by Machine Learning Techniques, presented by **K. M. Mostakim**, ID No: **172-15-10176**, **Anisur Rahman**, ID No: **172-15-10177**, and **Sohanur Rahaman**, ID No: **172-15-10199** to the Department of Computer Science and Engineering, Daffodil International University has been acknowledged as good for the fractional satisfaction of the necessities for the level of B.Sc. in Computer Science and Engineering and affirmed concerning its style and substance. The presentation has been hung on 01.06.2021.

BOARD OF EXAMINERS



Dr. Touhid Bhuiyan

Professor and Head

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Md. Tarek Habib

Assistant Professor

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Chairman

Internal Examiner



Saiful Islam

Senior Lecturer

Internal Examiner

Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Dr. Md Arshad Ali

External Examiner

Associate Professor

Department of Computer Science and Engineering
Hajee Mohammad Danesh Science and Technology University

DECLARATION

We thus announce that this project has been finished by us under the supervision of Nishat Sultana, Lecturer, Department of Computer Science and Engineering Daffodil International University. We additionally announce that neither this project nor any piece of this undertaking has been submitted somewhere else for the honor of any degree of recognition.

Supervised by:



Nishat Sultana

Lecturer

Department of Computer Science and Engineering

Daffodil International University

Submitted by:



K. M. Mostakim

ID: 172-15-10176

Department of Computer Science and Engineering

Daffodil International University



Anisur Rahman

ID: 172-15-10177

Department of Computer Science and Engineering

Daffodil International University



Sohanur Rahaman

ID: 172-15-10199

Department of Computer Science and Engineering
Daffodil International University

ACKNOWLEDGEMENT

To begin with, We specific our heartiest thank you and appreciation to Almighty Allah for His heavenly gift that makes us conceivable to finish last year's project/internship successfully.

We truly appreciate and wish our significant obligation to Nishat Sultana, Lecturer, Department of CSE, Daffodil International University, Dhaka. Profound Knowledge and unmistakable fascination of our supervisor in the field of "Machine Learning" to complete this task. Her unending persistence, academic direction, consistent consolation, steady and vigorous management, efficient analysis, critical counsel, perusing several substandard drafts, and revising them in any respect tiers have made it viable to complete this project.

We would possibly need to specific our heartiest thanks to Dr. Touhid Bhuiyan, Professor and Head of the Department of Computer Science and Engineering, for his thoughtful assistance to complete our project and furthermore to other faculties and the staff of the CSE branch of Daffodil International University.

We would possibly need to thank our entire path mate in Daffodil International University, who participated in this conversation while finishing the course work.

At long last, we should recognize with due regard the consistent help and patience of our parents.

ABSTRACT

Machine getting to know makes a specialty of the development of Computer packages which could get records and use it to discover on their own. This thesis centers on the job analysis of the Alumni. The job analysis helped out through a poll study follows. We want to build an “Alumni Job Analysis & Success Rate Prediction by Machine Learning Techniques”. Many people work at different farms or companies such as software firms, software companies. For the basis of Last Education, Programming Language, Monthly Salary and Company Rating. We can classify different types of classes such as Excellent, Very Good, Good, and Average. Our data is synthetic data. We collected our data from alumni. We have used Company rating from Google to make our dataset. At that point, machine learning classifications are applied to the dataset. Finally, a proficient model is created to predict a Success Rate. This model gives great classification measures with the dataset.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	I
Declaration	II
Acknowledgements	III

Abstract	IV
Table of Contents	V-VII
CHAPTER	PAGE
CHAPTER 1: INTRODUCTION	1-4
1.1 Introduction	1-2
1.2 Motivation	2
1.3 Rationale of the Study	2
1.4 Research Questions	2
1.5 Expected Outcome	3
1.6 Report Layout	3-4
CHAPTER 2: BACKGROUND STUDY	5-8
2.1 Introduction	5
2.2 Related Works	5-7
2.3 Comparative Analysis and Summary	7
2.4 Scope of the Problem	7
2.5 Challenges	7-8
CHAPTER 3: RESEARCH METHODOLOGY	9-13
3.1 Introduction	9
3.2 Research Subject and Instrumentation	9
3.3 Data Collection Procedure	9

3.4 Data Augmentation	10
3.5 Data Preprocessing	11-12
3.6 Training the Model	12
3.7 Implementation Requirements	12-13
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	14-18
4.1 Introduction	14
4.2 Performance Evaluation	14-15
4.3 Result Discussion	15-17
4.4 Confusion matrix	17-18
CHAPTER 5: Conclusion and Future Works	19-21
5.1 Summary of the study	19
5.2 Conclusions	20
5.3 Future Works	20-21
REFERENCES	22-25
APPENDIX	26
PLAGIARISM REPORT	27

CHAPTER 1

INTRODUCTION

1.1 Introduction

In this present era of hardship and struggle where the unemployment rate in Bangladesh is around 4.19 percent and 5.42 percent in the world, these not knowing the right path to be successful in life has become a major problem in our daily life. Software engineering is a dynamic and quickly developing territory that has a fundamental piece of the world that we live in today. In today's technology space every day new programming languages keep emerging continuously. Each of the programming languages intends to provide a better solution to complex problems. With having so many programming languages, sometimes the beginners(who want to learn about computer science) find it difficult to choose the programming language that they should learn to be successful in life. Every programming language has its own pros and cons and provides different kinds of applications. For example PHP one of the most used programming language for creating web applications while Python is most widely used in Machine Learning. But we cannot measure the success of any person only based on what language they learn. There are several criteria we use in order to measure what language would be best in modern days to successful income, experience, company they are working on right now. Numerous individuals may imagine that achievement relies upon how much cash you make however it's not 100 percent genuine money (Salary) is an important factor to pass judgment on individual achievement.

Another significant part is experience. an individual subsequent to finishing his BSc he finds a new line of work immediately at a decent compensation wherein a similar post someone else landed that position following 3,4 years of battle so their prosperity isn't the same. Pro software company stands firm on the main footing whereas devoting standard firm on the 50 positions. so the achievement of each company software engineer is not the same.

Hence it is very important to choose the right programming language to learn that benefits your career development. People do have questions like which is the most common

programming language used today? So to help all such aspiring people to know some of the top programming languages of the present IT world, here we are presenting this ' Alumni Job Analysis & Success Rate Prediction by Machine Learning Techniques' to help those who struggle to choose the right.

1.2 Motivation

The main reason for taking this project is to produce awareness among that kind of student who does not know what kind of skill he/she needs to learn in order to be a success in future. Many students fall under depression because they can't achieve their goals in their life for a lack of knowledge.

1.3 Rationale of the Study

As we want to solve that problem of unemployed, so try to find what other people do to succeed in their life after complete graduation. Undoubtedly there are so many reasons to an unsuccessful in their life. And we find the reason, why students can't achieve their goals. We know in recent times Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL) show significant results in solving many problems around many fields. In our study, we work with supervised learning which is a part of machine learning to improve these existing situations.

1.4 Research Questions

When we first time thinks to solve the problem of successful prediction, we can't understand what is our first work? And how we will solve this problem. It was so trying for us to finish this work. At that time, we thought some confusion questions.

- What algorithm we should use?
- Which programming language is perfect to implement our problems?
- What amount of data we need?
- Is it possible to get a good accuracy from our model?
- Is there enough time for completing our study?

1.5 Expected Outcome

We started our Thesis keeping in mind to solve one of the understandings of bachelor's students and restructure the educational experience around their new reality. Without Success rate prediction, It is not possible to understand new careers and our future students. Our goal is how we can provide a solution to find out our student's strengths from their successful career. After deciding our field of study as machine learning and deep learning specifically as Basic machine learning, we expected to use Basic machine learning in the classification of Success Rate. We want to build a Basic Machine Learning based predicted model to classify real life data which can be used as a part of Job Analysis.

1.6 Report Layout

Chapter 1: Introduction

In this chapter, introduction, motivation, rationale of the study, research questions and the expected outcome of our project will be discussed. To give a gentle introduction to our project, here we discuss those topics briefly.

Chapter 2: Background Study

In chapter 2, we will discuss on what already accomplished via way of means of the preceding worker/researchers on this place of job analysis, what the problem of their approaches is and why our method is better than their methods. We will also define some scope of the problem and highlight some challenges that we face during this project.

Chapter 3: Research Methodology

In chapter 3, our proposed model will be discussed theoretically. But before that we will mention some important procedures like data collection, data augmentation, data

preprocessing, etc. We will also provide some theoretical information about machine learning so that everybody can understand this concept. We will address some minimum requirements for this analysis at the end of chapter three.

Chapter 4: Experimental Results and Discussion

Chapter 4 discusses the performance comparison of our “Success rate” models with some popular pre-trained models such as Decision Tree, Random Forest, Gaussian Naive Bayes, Bernoulli Naive Bayes, Multinomial naive Bayes, K-Nearest Neighbor. After that we discuss complete result analysis of all models including our model.

Chapter 5: Conclusion and Future Works

In Chapter 5, we will make conclusion by summarizing our study. Some future works regarding our proposed model will be discuss to show its potentiality in the field of traffic management system.

CHAPTER 2

BACKGROUND STUDY

2.1 Introduction

In this section, we will dive down to describe why we are studying the field of machine learning to automate the process of identifying the type of job with the help of Computer Vision and Artificial Intelligence to be more specific with Machine Learning methods. Before moving towards our works, here we will discuss related works in the field of success rate prediction. Here we are able to talk about different studies paper, their method their strategies, and problems about their works. We will make a comparative analysis of what we learn from other projects and what problems they have. Then it will be clear that why we are doing this project.

2.2 Related Works

In 2017 Alaa Khalaf Hamoud, Aqeel Majeed Humadi, Wid Akeel Awadh, Ali Salah Hashim used Ali Bayes Algorithms for student success prediction. They used the Weka tool that provides built-in algorithms that help us to apply different classifiers and obtain results in an easy and flexible process. Bayes net and naïve Bayes these two algorithms used here.

In 2019 SOOHYUN NAM LIAO, DANIEL ZINGARO, KEVIN THAI, CHRISTINE ALVARADO, WILLIAM G. GRISWOLD, and LEO PORTER proposed A Robust Machine Learning Technique to Predict Low-performing Students .they used support vector machines (SVMs) with the radial basis function kernel to train one prediction model for each course.

In 2019 Xing Xua, Jianzhong Wanga, Hao Peng, Ruilin Wu presented a method to predict undergraduate academic performance from the internet uses behaviours by machine learning. Three common machine learning algorithms of the decision tree, neural network, and support vector machine were used to predict academic performance from these features.

Hassan Zeineddine, Udo Braendle, Assaad Farah proposed an Automated machine learning technique to enhance the accuracy of predicting student performance using data available prior to the start of the academic program. Here they used several machine learning

methods, logistic regression accuracy is around 70%. Decision tree accuracy is around 70% [2], and reaches 70% when using data gathered after joining the program. Artificial neural network accuracy is around 70%. Naive Bayes with a minimum accuracy of 50% and a maximum of 66%. K-nearest neighbors accuracy reached 73% when using data extracted from internal assessments, CGPA, and extra-curricular activities. Support vector machine accuracy is ` around 70%.

Huda Al-Shehri, Amani Al-Qarni, Leena Al-Saati, Arwa Batoaq, Haifa Badukhen, Saleh Alrashed, Jamal Alhiyafi, and Sunday O. Olatunji in 2017 presented a method of "Student Performance Prediction Using Support Vector Machine and K-Nearest Neighbor". here they applied both Support Vector Machine algorithm and K-Nearest Neighbor algorithm on the dataset to predict the student's grade and then compared their accuracy. The empirical study's outcome indicated that the Support Vector Machine achieved slightly better results with a correlation coefficient of 0.76, while the K-Nearest Neighbor achieved a correlation coefficient of 0.75.

In 2015 Fadhilah Ahmad, Nur Hafieza Ismail, and Azwa Abdul Aziz propose a method to "prediction of students' academic performance using Classification Data Mining Techniques". They contain the student demographics, previous academic records, and family background information Decision Tree, Naïve Bayes, and Rule-Based classification techniques are applied to the prediction model. Here decision tree accuracy value of 68.8% . Naïve Bayes accuracy value of f 67.0% and Rule-Based classification accuracy value is 71.3%.

In 2013 Dorina Kabakchieva proposed a method to predicting Student Performance by Using Data Mining Methods for Classification. Their study pointed toward dissecting the exhibition of different data mining classification algorithms on the provided dataset in order to evaluate their potential usefulness for the fulfilment of the project goal and objectives. Two rule learners, a decision tree classifier, two popular Bayes classifiers and a Nearest Neighbour classifier applied here.

Sujith Jayaprakash and Balamurugan attempt a method in which to predict students' academic performance in end-of-semester examinations by analysing student feedback and their performance in mid semester exams they used naïve Bayes algorithm. This study helps to identify the weaker students in advance and arrange necessary training before they take their final exams.

In 2011 R. R. Kabra and R. S. Bichkar proposed a method to predict the performance of Engineering Students using Decision Trees. The decision tree created from student .The accuracy of the model is 60.46 %. That is out of 346 instances 209 instances are correctly classified.

In 2013 Mrinal Pandey and Vivek Kumar Sharma attempt a method of Analysis and Prediction of student's performance using Decision Tree Algorithm. Four decision tree algorithms were applied here and J48 decision tree algorithm was selected for model construction. J48 accuracy 62.58 %, simple cart 61.46%, reptree is 61, 46 and NB tree is 59.77 %. This model helps to the management to identify weak students and can take appropriate decision to prevent them from failure.

2.3 Comparative Analysis and Summary

After analyzing those previous works, we observed that all models are complex to implement. They worked based on personality and job performance. That's why we focus on their educational experience. We want to build an "Alumni Job Analysis & Success Rate Prediction by Machine Learning Techniques" for bachelor's students and restructure the educational experience around their new reality. After finishing this work, we will find out our student's strengths from their successful career. Also, find the reason why he/she could not be successful in his/her career and the output helps our future students, especially in a country like- Bangladesh. And all those model that classifies students success based on our dataset. So that it can be useful for our future students.

2.4 Scope of the Problem

As unemployment has an effect on the different socio-economic aspects of a country. So, we try to solve this problem using new machine learning technology. Our study finds that no previous research like ours. So, we try to build a new dataset to build a new model to classify success rate prediction from alumni information.

2.5 Challenges

We confronted a few difficulties to finish our task. Our first challenge was to select a proper algorithm or method to make a model that can help in successful prediction classification. We finally find Supervised Learning that can help us in doing this job. Then our next challenge was a collection of data. As we probably are aware little information can't make a decent model, so attempt to track down an enormous measure of information.

But sadly, we found no dataset regarding this which can fulfill our requirement of success rate prediction. Then we decide to collect data manually from the internet and augment them to make an acceptable amount of data to train a model. Our next challenge was training our model on a good computational machine. So, we can say that our task was difficult, but we successfully complete it.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 Introduction

In this part 3, we will expand the work process of our novel way to classify success rate prediction from alumni information. In this exploration, we utilize a convolutional neural network for textual classification, because it is the most powerful and proven method of deep learning to classify textual data. We are going to use our own dataset to train our proposed model. Here we will discuss our proposed model.

3.2 Research Subject and Instrumentation

A simple definition of our study field is given by the research subject. We introduce and design our model in this section, gather perfect data, plan and train our model, discuss results, and then apply our model to function. To finish our errand we utilized the windows stage, python programming language with numerous bundles like Numpy, Pandas, scikit-learn, and so on are used to train and test our model. We prefer Python because of its simple syntax and a vast amount of use in complex algorithms readability for machine learning applications.

3.3 Data Collection Procedure

In this investigation, we utilize another public dataset to prepare our proposed model. We create the new dataset by collecting alumni information manually from the internet and publicly available information along with it. This dataset has four classes of success rate prediction as Average, Good, very good, excellent. This dataset has a total of 300 alumni information, 240 rows are used for training and 60 rows are used for testing purposes. In each class, we thought about 80% information for training and the excess 20% information for testing purposes.

3.4 Data Augmentation

Data is extracted from alumni (from various university) and company rating website. Note that, we extracted alumni information who was graduate. The extracted data from alumni contains the title of the name, last education, programming languages, company name, job experience, salary range. We are collect data by using Google Forms. Here, is some screenshot to collecting the data.

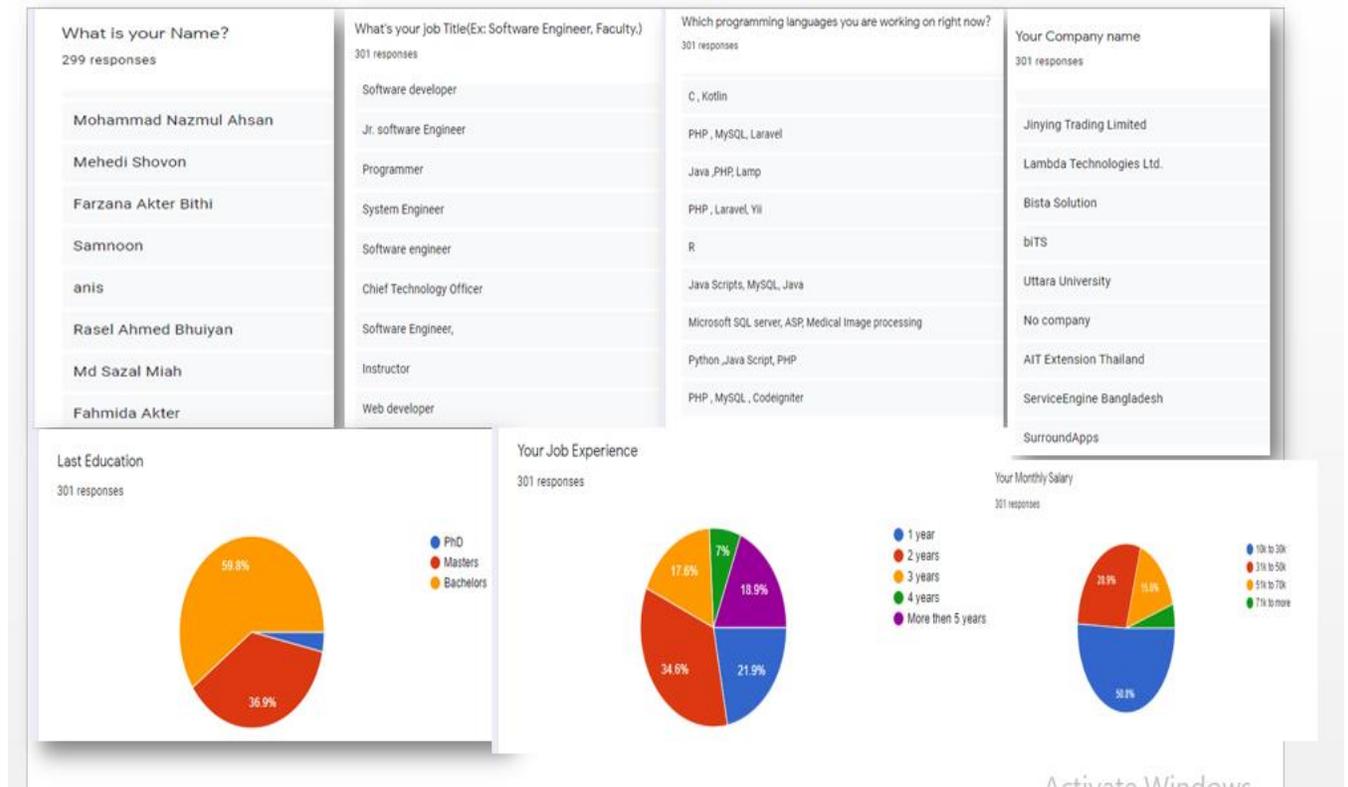


Fig: A small part of dataset

3.5 Data Preprocessing

In this pre-processed data we used last education, programming languages, job experience, company rating, salary as our main data but alumni name and company name are considered in our work so we remove them as our main data in this society people normally judge people success by their income that's why we are focus in salary is preferred in this work. we classified salary into four classes. if salary is 10k-30k its average successful, if salary is between 30k-50k its good if it's between 50k-70k its very good and more than this is excellent. Furthermore, the organization rating of each chose organization is removed from the organization rating site. Organization rating is turning out to be exceptionally helpful these days. That is the reason this stage is liked in this work. The organization rates each organization out of 5. Subsequently, evaluations are characterized into four classes average, good, very good, excellent. That is, assuming an organization rating is under 3.5, its class will be normal. On the off chance that the rating is between 3.5 to 4.0, the class will be acceptable. Along these lines, the class highlight of each organization in the dataset is either average or good or very good or excellent, based on its company rating.

Along these lines, the informational collection contains separated genuine information from both alumni and company rating websites.

Company Rating	Class
Less than 3.5	Average
3.5-4.0	Good
4.1-4.5	Very Good
4.6-5.0	Excellent

Salary Range	Class
10-30k	Average
31-50k	Good
51-70k	Very Good
71-more than	Excellent

3.6 Training the Model

To train our model we use our textual dataset. Our dataset has 300 people's information, among those, 80% of information is utilized for training and 20% of information is utilized for testing. All training procedure was done on jupyter notebook 6.0.3. For increasing accuracy and decreasing loss as possible, we update our model many times. After doing several types of hyper parameter tuning several times, we finally find the best performance of our model.

3.7 Implementation Requirements

Following a proper analysis of the important measurable or hypothetical ideas and techniques, a rundown of necessities for success rate prediction classification work has been made. The likely fundamental things are:

Hardware/Software Requirements

- Operating System (Windows 7 or above)
- Hard Disk (minimum 500 GB)
- Ram (Minimum 4 GB)
- GPU (Recommended)

Developing Tools

- Python Environment
- Jupyter Notebook

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 Introduction

In this section, we will discuss how we train our proposed Decision Tree model and what performance we get from it. We will also compare our model's performance with some pre-trained model like- Decision Tree, Random Forest, Gaussian Naive Bayes, Bernoulli Naive Bayes, Multinomial Naive Bayes, and K-Nearest Neighbor.

4.2 Performance Evaluation

Our Success rate model was a trained all alone dataset where 80% utilized as a training set and the other 20% in the testing set. We have done all the coding and training of our model in jupyter notebook 6.0.3. We use Python based library along with jupyter to build, train and test our model.

Table: Success rate Classification

Company + Salary	Class Name
Average + Average	Average
Average + Good	Good
Average + Very Good	Very Good
Average+ Excellent	Very Good
Good + Average	Average
Good + Good	Very Good
Good + Very Good	Very Good
Good + Excellent	Excellent
Very Good + Average	Good
Very Good + Good	Very Good

Very Good + Very Good	Excellent
Very Good + Excellent	Excellent
Excellent + Average	Good
Excellent + Good	Very Good
Excellent + very good	Excellent
Excellent + Excellent	Excellent

4.3 Result Discussion

This segment portrays the correctness of the classifiers at various phases of our work. Our system into one phase. The stage is to apply the classifiers to our dataset. As our informational index is adjusted, so its exactness is high. The correctness are represented in Table. Here, Bernoulli Naïve Bayes and K-NN, give a similar grouping exactness. And Random Forest is the most noteworthy among the six classifiers.

Consequently, the most noteworthy precision got by every classifier is recorded and appeared in Table. Among the Six classifiers, Random forest and decision tree gives the most noteworthy order exactness. Despite the fact that two classifiers give altogether great exactness here, that is over 80%. Decision Tree 90% and Random Forest 90%. Bernoulli Naive Bayes and K-Nearest Neighbor classifiers give the lowest accuracy in our model.

Thusly, the performance metrics of the six classifiers express the end that the decision tree and random forest classifier give the best results as far as exactness.

TABLE 4.1: CLASSIFICATION ACCURACY

Name of classifier	Accuracy
Random Forest	90%
Decision Tree	90%
Gaussian Naive Bayes	80%
Bernoulli Naive Bayes	45%
Multinomial Naive Bayes	65%
K-Nearest Neighbor	45%

4.3.1 Random Forest

Random forest area is supervised gaining knowledge of a set of rules this is used for each classification in addition to regression. Its balloting could be done for each anticipated result. Random forest gives 90% accuracy on our dataset. Random Forest performed very well for our dataset.

4.3.2 Decision Tree

The decision Tree goal is to create a model that predicts the value of a target variable by learning simple decision rules inferred from the data features. Decision Tree gives 90% accuracy on our dataset.

4.3.3 Gaussian Naive Bayes

For our textual dataset, Gaussian Naive Bayes gives 80% accuracy.

4.3.4 Bernoulli Naive Bayes

Bernoulli Naive Bayes classifier is a probabilistic classifier because of this that given an enter, it predicts the chance of the entry being labeled for all of the classes. For our textual dataset, Gaussian Naive Bayes gives 45% accuracy.

4.3.5 Multinomial Naive Bayes

The multinomial distribution normally requires integer feature counts. For our textual dataset, multinomial Naive Bayes gives 65% accuracy.

4.3.6 K-Nearest Neighbor

K-NN set of guidelines assumes the similarity of many of today's statistics and available statistics and locations the today's statistics into the magnificence that is most much like the available categories. K-NN set of rules offers 45% accuracy.

4.4 Confusion matrix

A vastly improved approach to assess the presence of a classifier is to take a gander at the confusion matrix. Each line in a confusion matrix addresses a genuine class, while every segment addresses a predicted class. The confusion matrix gives us a great deal of data, yet here and there we may lean toward a more concise metric.

Precision:

$\text{Precision} = (\text{True Positive}) / (\text{True Positive} + \text{False Positive}).$

Out of the multitude of positive classes we have anticipated accurately, the number of are really certain.

Recall:

$\text{Recall} = (\text{True Positive}) / (\text{True Positive} + \text{False Negative}).$

Out of the multitude of positive classes, the amount we anticipated effectively. It ought to be high as could really be expected.

F-measure:

$F1 \text{ Score} = 2 \times ((\text{precision} \times \text{recall}) / (\text{precision} + \text{recall}))$.

It is likewise called the F Score or the F Measure. Put another way, the F1 score passes on the harmony among precision and recall.

The F1 for the All No Recurrence model is $2 \times ((0 \times 0) / 0 + 0)$ or 0.

CHAPTER 5

CONCLUSION AND FUTURE WORKS

5.1 Summary of the study

It has presumably that there is a ton of examination work done in this field. Basically, we want to solve the problem of inept in the job sector. So, we use a Machine learning algorithm called Supervised Learning to give a solution to this problem. We proposed a Supervised Learning-based architecture to classify success rate. We made a dataset to train our proposed model. And finally, we find an optimal result of our model. We expect that this study will help our graduates to accretive their goals.

5.2 Conclusions

Alumni Job Analysis & Success Rate Prediction by Machine Learning Techniques gives the fundamental qualities of people who could fill the job successfully. The job analysis additionally characterizes a manner by which research is done to accumulate data expected to comprehend what the job will involve and what sort of skills are required.

As the Job area of the CSE department is getting tremendous step by step, rivalry here is additionally developing complex. Hence, anticipating class is developing complex too. Our model is created on a manufactured informational index and it is gathered from two stages, Alumni and Google. The model can likewise be utilized to foresee some other classification like Student Activity or Drug Addiction. Other than Employee Attrition, occupations Analysts, and so on can be anticipated by our model utilizing the highlights of our model. Depicts a few highlights having more impact on progress rate and some different highlights having less or no impact. As indicated by the compensation and friends rating is having a little sure impact however the name and friends name don't have any effect on the model. In this manner, our work should be possible by a weighted component arrangement to mirror these impacts. Alongside the graduated class understudy dataset, numerous college graduated class understudy datasets can be utilized to make the model more proficient. The information base can be enhanced by remembering the graduated class understudy for late years.

5.3 Future Works

With each business and association going computerized, the IT area today is a flourishing field when contrasted with different areas and there is a tremendous interest of Computer Science Engineers and developers.

In our dataset, all the data are synthetic data and this dataset only for Computer Science and Engineering department. But in the future, we will try to collect all departments of all university alumni information in Bangladesh in the dataset. We will also try to integrate more classes to accurately analyze the success rate.

In the future, this work can be a part of the mobile application. Where our model will work as a mobile app smartly. It can also give the result of his/her success easily. It will be different as a department. Then this work can be part of web. All synthetic information will be stored and we can update it every year.

REFERENCES

[1] Alumni: all alumni information is available at << <https://www.linkedin.com/home> >>.

[2] Company: Company rating data is accessible

At << <https://www.glassdoor.com/Reviews/Glassdoor-Reviews-E100431.htm>>>.

[3] Gurmeet Kaur, Williamjit Singh “Prediction of Student Performance Using Weka Tool”

‘2016’.

[4] Kolo David Koloa, Solomon A. Adepojub and John Kolo Alhassanb “A Decision Tree Approach for Predicting Students Academic Performance”. I.J. Education and Management Engineering, 2015, 5, 12-19.

[5] Ashkan Sharabiani, Member, IEEE, Fazle Karim, Anooshiravan Sharabiani, Mariya Atanasov, Houshang Darabi “An Enhanced Bayesian Network Model for Prediction of Students’ Academic Performance in Engineering Programs ”2014 IEEE Global Engineering Education Conference (EDUCON).

[6] Jason Brownlee on March 21, 2016 in over fitting and under fitting with Machine

Learning Algorithms

Available at << [https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-](https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learningalgorithms/#:~:text=Overfitting%3A%20Good%20performance%20on%20the,poor%20generalization%20to%20other%20data)

[learningalgorithms/#:~:text=Overfitting%3A%20Good%20performance%20on%20the,poor%20gener](https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learningalgorithms/#:~:text=Overfitting%3A%20Good%20performance%20on%20the,poor%20generalization%20to%20other%20data)

[alization%20to%20other%20data](https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learningalgorithms/#:~:text=Overfitting%3A%20Good%20performance%20on%20the,poor%20generalization%20to%20other%20data)>>.

[7] Rohith Gandhi Support Vector Machine — Introduction to Machine Learning Algorithms

Available at << <https://towardsdatascience.com/support-vector-machine-introduction-to-machinelearning-algorithms-934a444fca47> >>

[8] Onel Harrison Machine Learning Basics with the K-Nearest Neighbors Algorithm available at

<< <https://towardsdatascience.com/machine-learning-basics-with-the-k-nearest-neighborsalgorithm-6a6e71d01761> >>.

[9] R. R. Kabra and R. S. Bichkar “Performance Prediction of Engineering Students using Decision

Trees” International Journal of Computer Applications (0975 – 8887) Volume 36– No.11, December

2011.

[10] I. Witten, E. Frank, Data Mining: Practical Machine Learning Tools and Techniques, 2nd Edition,

Morgan Kaufmann, San Francisco, ISBN: 0-12-088407-0

[11] J. R. Quinlan, “Induction of decision trees”, Machine Learning, Volume 1, Morgan Kaufmann,

1986, 81-106.

[12] Alaa Khalaf Hamoud, Aqeel Majeed, Wid Akeel Awadh, Ali Salah Hashim “Students’ Success

Prediction Based on Bayes Algorithms”.

[13] Friedman, Nir, Dan Geiger, and Moises Goldszmidt. “Bayesian network classifiers.” Machine

Learning 29.2-3 (1997): 131-163.

[14] Nihat Cengiz , Arban Uka “ Prediction of Student Success Using Enrolment Data “

[15] Kovačić, Z.J. 2010, Early Prediction of Student Success: Mining Students Enrolment Data,

Proceedings of Informing Science & IT Education Conference (InSITE) 2010, Open Polytechnic,

Wellington, New Zealand.

[16] Anal Acharya, Devadatta Sinha “Early Prediction of Students Performance using Machine

Learning Techniques “International Journal of Computer Applications (0975 – 8887) Volume 107 –

No. 1, December 2014.

[17] Oladukun, V. O., Adebajo, A. T., Charles-Obawa, O. E., Predicting Students“ Academic

Performance using Artificial Neural Network: A Case Study of an Engineering Course

[18] Domingos, P, A Few Useful Things to Know about Machine Learning, <<<https://cacm.acm.org/magazines/2012/10/155531-a-few-useful-things-to-know-about-machine-learning/fulltext>>>

[19] Kotsiantis, S., Piarrekeas, C., Pintelas, P., 2007. Predicting Students“ performance in Distance. Learning using Machine Learning Techniques, Applied Artificial Intelligence, 18:411-426.

[20]. Domingos, P, A Few Useful Things to Know about Machine Learning,.

[21] . Kotsiantis, S., Piarrekeas, C., Pintelas, P.,2007. Predicting Students“ performance in Distance Learning using Machine Learning Techniques, Applied Artificial Intelligence, 18:411-426

[22].Rectified linear units (relu) in deep learning, available at <<<https://www.kaggle.com/dansbecker/rectified-linear-units-relu-in-deep-learning>>>,

[23].The most intuitive and easiest guide for convolutional neural network, available at <<<https://towardsdatascience.com/the-most-intuitive-and-easiest-guide-for-convolutional-neural-network-3607be47480>>>,

[24]Crash course in convolutional neural networks for machine learning, available at <<<https://machinelearningmastery.com/crash-course-convolutional-neural-networks/>>>,

[25]Convolutional neural networks, explained, available at <<<https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939>>>

APPENDIX

To complete our project, we have faced so many problems and the first was the selection of methodology. This project was so challenging because only some work was done before with different datasets and different ways. And nobody tries to solve the problem in the way that we have done it. So, we did not get that much information assistance from anyplace. Another big issue was that collection of data, it was a huge challenge for us. There was no dataset available on this kind of success rate prediction. But we overcome this challenge and successfully complete our work. To train our model, we collect data from various online sources and manipulate them. We have trained and compared several other models with our model and the best classification result was successfully obtained by our model. Our development and thesis work was difficult but it was very interesting too.

PLAGIARISM REPORT

Alumni Success Rate prediction

ORIGINALITY REPORT

17%	14%	9%	8%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3%
2	Hassan Zeineddine, Udo Braendle, Assaad Farah. "Enhancing prediction of student success: Automated machine learning approach", Computers & Electrical Engineering, 2021 Publication	2%
3	www.ijcaonline.org Internet Source	1%
4	ieeexplore.ieee.org Internet Source	1%
5	medium.com Internet Source	1%
6	www.afdc.doe.gov Internet Source	1%
7	machinelearningmastery.com Internet Source	1%
8	www.ripublication.com Internet Source	