

Entrepreneurs' prediction using machine learning algorithms

BY

Farid Uz Zaman

ID: 221-25-113

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

Supervised By

Dr. Sumit Kumar Banshal

Assistant Professor

Department of CSE

Daffodil International University



DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JANUARY 2023

APPROVAL

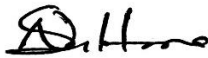
This Project/Thesis titled “**Entrepreneurs’ prediction using machine learning algorithms**”, submitted by **Farid Uz Zaman**, ID No: **221-25-113** to the Department of Computer Science and Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of M.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 17th January, 2023

BOARD OF EXAMINERS



Chairman

Dr. S M Aminul Haque, PhD
Associate Professor & Associate Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



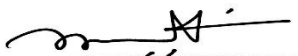
Internal Examiner

Ms. Naznin Sultana
Associate Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



Internal Examiner

Mr. Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University



External Examiner

Dr. Mohammad Shorif Uddin, PhD
Professor
Department of Computer Science and Engineering
Jahangirnagar University

DECLARATION

I hereby declare that, this project has been done by me under the supervision of **Dr. Sumit Kumar Banshal, Assistant Professor, Department of CSE** Daffodil International University. I also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



Dr. Sumit Kumar Banshal
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:



Farid Uz Zaman
ID: 221-25-113
Department of CSE
Daffodil International University

ACKNOWLEDGEMENT

First, I express my heartiest thanks and gratefulness to almighty God for His divine blessing makes me possible to complete the final year project/internship successfully.

I really grateful and wish my profound my indebtedness to **Dr. Sumit Kumar Banshal, Assistant Professor**, Department of CSE Daffodil International University, Dhaka. Deep Knowledge & keen interest of my supervisor in the field of “*Altermetrics, Scientometrics and Text Data Mining*” to carry out this project. His endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

I would like to express my heartiest gratitude to Head, Department of CSE, for his kind help to finish my project and also to other faculty member and the staff of CSE department of Daffodil International University.

Finally, I must acknowledge with due respect the constant support and patients of my parents.

ABSTRACT

From the beginning of the world, entrepreneurs are deeply involved in every sector of our life. The important component is that entrepreneurs are the primary riding pressure of the worldwide economy. This research focuses on the personality traits and entrepreneurship skills of students. With the help of machine learning algorithms implemented in WEKA, I tried to match up personality traits and entrepreneurship skills. I tried to find out the level of actual entrepreneurs' skills. I have collected data from 408 volunteers who agreed to give data on their concerns. I have collected these data in two ways and there are two types of data. One is 'The Big Five Personality Test' made by the IPIP organization which is to check personality traits. Another one is the Entrepreneurial Self-Assessment survey and it is based on some personal information. The data are taken from various regions of Bangladesh.

TABLE OF CONTENTS

CONTENTS	PAGE
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
 CHAPTER	
 CHAPTER 1: INTRODUCTION	 01-04
1.1 Introduction	01
1.2 Motivation	01
1.3 Rationale of the Study	02
1.4 Research Queries	02
1.5 Expected Outcome	03
1.6 Thesis Duration	03
1.7 Layout of The Report	03

CHAPTER 2: BACKGROUND STUDIES	05-08
2.1 Terminologies	05
2.2 Related Works	06
2.3 Analysis and Summary	08
2.4 Challenges and Barriers	08
CHAPTER 3: RESEARCH METHODOLOGY	09-14
3.1 Research Subject and Research Instrumentation	09
3.2 Data Collection Procedure/Dataset Utilized	09
3.3 Applied Methodology	10
3.4 System Requirements	14
CHAPTER 4: EXPERIMENTAL RESULTS AND DISCUSSION	15-24
4.1 Experimental Setup	15
4.2 Questionaries	15
4.3 Experimental Results & Analysis	19
4.4 Discussion	23
CHAPTER 5: IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY	25-27
5.1 Impact on Community	25
5.2 Impact on Economy	25
5.3 Ethical Perspective	26
5.4 Sustainability Plan	27

CHAPTER 6: SUMMARY, CONCLUSION, RECOMMENDATION AND IMPLICATION FOR FUTURE RESEARCH	28-29
6.1 Summary of the Study	28
6.2 Conclusions	28
6.3 Implication for Further Study	29
REFERENCES	30-33

LIST OF FIGURES

FIGURES	PAGE NO
Figure 3.1: Classification of Exponent	10
Figure 4.1: Performance Graph of Different Classifiers	20
Figure 4.2: Graph For ROC, PRC, and 10-Fold Cross Validation	21
Figure 4.3: Precision-Recall-F Measure Graph	21
Figure 4.4: Matthews Correlation Coefficient Values of Different Classifiers	22
Figure 4.5: Kappa statistics, MAE and RMSE classifier	23

LIST OF TABLES

TABLES	PAGE NO
Table 1.1: Project Summary	03
Table 3.1: Details of Dataset	09
Table 3.2: Confusion Metrics of Random Forest	11
Table 3.3: Confusion Metrics of Random Tree	11
Table 3.4: Confusion Metrics of Hoeffding Tree	12
Table 3.5: Confusion Metrics of Logistic Model Tree (LMT)	12
Table 3.6: Confusion Metrics of Decision Stump	13
Table 3.7: Confusion Metrics of Bayes Net	13
Table 3.8: Confusion Metrics of Naive Bayes	14
Table 4.1: Performance Metrics of Different Algorithms	19
Table 4.2: Kappa Statistics, MAE and RMSE of Different Classifiers	22

CHAPTER 1

Introduction

1.1 Introduction

Human are always hungry for inventing new things from the beginning of the world. The constant madness of these innovations made this modern world ambitious. For this continued consistency, the earth moved to the modern age from the primitive age. In this challenging arena, it is so hard for entrepreneurs to be a successful entrepreneur than just survive. For that reason, an entrepreneur might have some personality traits with some qualitative needy characteristics of self-assessment which could have an effect on or expect his organizational conduct.

Without any proper justification or any kind of examination, these personality traits cannot be declared. Following to the 'Big Five Personality Test', there are 5 traits to justify any people's overall personality known as OCEAN stands for Openness to experience, Extraversion, Agreeableness, Neuroticism, and Conscientiousness. If I can predict the traits individually, then it is possible to predict overall personality perfectly. With the help of IPIP questionnaire, the perceptions of each trait could be recognized. On the other hand, the Entrepreneurial Self-Assessment calculation is depended on some common features of an entrepreneur. The different entrepreneur ability levels graded outstanding, satisfactory, inappropriate, and avoid entrepreneurship are determined using the survey. Then, I collaborate personality traits data with entrepreneurial data and with the help of Waikato Environment for Knowledge Analysis (WEKA), I implemented some machine-learning algorithm to get the predicting value of my dataset. I used only supervised machine-learning algorithm for this research.

1.2 Motivation

Entrepreneurship is the artwork of beginning a commercial enterprise and it is an interest full of creativity. An entrepreneur is an inventor or a writer who presentss something unique to the company or financial system. To be a hit entrepreneur, it is very critical to have managerial abilities and sturdy group-constructing skills. Leadership attributes are a

signal of hit marketers. The motivation of this file is to locate the correlation between human character tendencies and Entrepreneurial skills. It has to be stated that tertiary degree students in Bangladesh typically look for jobs after commencement. But compared to the graduating students in Bangladesh, the activity is insufficient. As an end result, many knowledgeable youths continue to be doing nothing or do now something but not get proper employment based on their qualifications.

Entrepreneurship is crucial for our use. It has the ability to enhance our requirements for dwelling. It has the capability to dispose of poverty and motive marketers to create a task area. When an entrepreneur invents a new product and services, it creates a probability for new job region.

1.3 RATIONALE OF THE STUDY

My reason is to construct a computational version in order that I will expect the entrepreneurship abilities of tertiary-level college students. After predicting entrepreneurship capabilities of our country's unemployment problem increases day by day. The task quarter isn't developing consistently with our populace. At present I see that once finishing graduation students follow for jobs however there have no sufficient jobs. Following the 2020 report of the international labor organization (ILO), a specialized agency from the UN, the global rate of youth employment is 13.6 percent. Relation between entrepreneurship and unemployment has been quite a captivating topic for researchers. Globally interesting and across the world comparable a look made every year by way of GEM (Global Entrepreneurship Monitor) notes that an increasing number of human beings are keen to establish themselves as an entrepreneur.

1.4 RESEARCH QUERRIES

Is it possible to find or predict skills of entrepreneurship using machine learning algorithm?

I have already collected data on personality and entrepreneurship skills. By mapping those data, I will be able to predict the entrepreneurship skills of humans.

1.5 EXPECTED OUTCOME

From studies, I count on that we are capable of locating the destiny of entrepreneurs. In survey, I ask two sorts of questions. In the first class, questions use for gathering information on personality than any other organization of questions uses for accumulating information on their entrepreneur self-evaluation. So, I expect after mapping the one's records we will locate our destiny marketers.

1.6 THESIS DURATION

To complete this thesis, I spent my time on this around 5 months.

Table 1.1: Project Summary

Activities	Timeline
Planning for the domain	1-month
Knowledge Gathering	1-month
Data Collection	1.5 Months
Data Preprocessing	15 Days
Implementation	15 Days
Report Generation	15 Days
Total	5 Months

1.7 LAYOUT OF THE REPORT

Total 6 chapters in this report. Explanations of chapters are the following:

Chapter 1: Chapter 1 is all about introductory things like Introduction, Motivation, Objectives, Expected Outcome, and Layout of the Report.

Chapter 2: Chapter 2 is about Terminologies, Related Works, Comparative analysis and summary, Scope of the problem, and challenges.

Chapter 3: Research Subject, Statistical analysis, proposed methodology, data collection procedure, implementation requirements are described in chapter 3

Chapter 4: Chapter 4 describes about experimental setup, experimental result and analysis and discussion.

Chapter 5: Chapter 5 is about the impact on society, ethical aspects, impact on the environment, and sustainability plan.

Chapter 6: Lastly, chapter 6 is all about the summary of the thesis, recommendation, conclusion and implication for further study.

CHAPTER 2

BACKGROUND STUDIES

2.1 TERMINOLOGIES

WEKA:

“WEKA” stands for Waikato Environment for Knowledge Analysis. It is a data mining tool developed at the University of Waikato in New Zealand. It is a set of machines getting to know algorithms used for records mining responsibilities. Weka contains tools for data pre-processing, association rules, clustering, classification, regression, and visualization. It is likewise well-ideal for growing new system-studying schemes. Weka is an open-source machine learning software under the general public license. It is used for research, thesis, education, and application. The tool gathers a comprehensive set of data pre-processing tools, learning algorithms and evaluation methods, graphical user interfaces, and an environment for comparing learning algorithms.

MACHINE-LEARNING:

Machine learning in short ML is the study of computer algorithms that improve amicably through time. Machine learning collaborates computer science with math, statistics, and so on. Machine learning is a subdivision of Artificial Intelligence and an area of the current computing world that analyze records, identify patterns and generate choices with minimum human intervention. It is the technique of facts evaluation that automates analytical version building wherein a model is trained from the records to find patterns and capabilities in big quantities of statistics to generate decisions and predictions based totally on new information. If we provide more data to a machine-learning algorithm then we can get more accurate performance from that.

SUPERVISED ALGORITHMS:

Supervised machine learning algorithms are works as like student and teacher. We have to train the machine with a well labelled dataset. It means some data of this particular dataset

are already labelled with correct answer. After the training session we can ask from it to show us result and it can produce the correct answer from tagged data.

2.2 RELATED WORKS

Entrepreneurship is one of the maximum essential research topics for all international locations. Cause an entrepreneur can improve the monetary increase and decrease unemployment. Therefore, various kinds of research have been done to determine the effect of Personality Traits on business skills.

Using China Family Panel Studies information, the relationship between persona development and the entrepreneurial probability of individuals in self-employment, agriculture, and non-agricultural employment is explored. Shows outcomes that persona traits affect entrepreneurial possibility but don't have any huge results on non-agricultural employment. Personality traits: conscientiousness and agreeableness are definitely related to the entrepreneurial opportunity of farmers in rural regions.

"Risk propensity variations between managers and marketers and among low- and high-increase marketers: A respond in a greater conservative vein" this paper affords four-character kinds capable of accomplishing entrepreneurial achievement.

Another research right here, Verify the relationship between locus of control, need for achievement, and entrepreneurial intention of adolescents. And how entrepreneurial training consolidates entrepreneurial abilities. This study targets to higher understand the main drivers of entrepreneurial motivation amongst college students and to determine whether or not entrepreneurship education has a moderating impact on enhancing the effect of expertise base and entrepreneurship talents on entrepreneurial motivation.

Interesting research was about to predict entrepreneurship skills by measuring the surroundings of a people. The main key factor is social, psychological, economical, environment and personal greediness. The paper pursuits to speak about the application of those models in distinct settings and now not indicate which the nice is. The paper is beneficial for the further research on the entrepreneurial behavior & aim escalation. It will

be also useful in designing educational and education packages for entrepreneurial development.

The elements that have an effect on the selection to internationalization and the overall performance of internationalization for SMEs had been seriously considered and studied in worldwide business literature. For instance, the Traditional Process Theory of Internationalization claims that firm-degree features, firm resources, and the company's expert level in its nearby market may provide possibilities for SMEs to take thriving actions while penetrating international markets. Nonetheless, the research built on this theory does not keep in mind or consist of person psycho-cognitive developments of the managers.

This article offers an issue for a way genetic factors can also have an effect on the tendency of humans to interact in entrepreneurial hobby and describes four mechanisms via which genetic factors ought to function. It also explores approaches those researchers can use quantitative and molecular genetics to examine entrepreneurship, and discusses the capability implications of a genetic angle for management studies on entrepreneurship.

Universities offer entrepreneurship-unique schooling (ESE) to equip students with the talents required to pursue entrepreneurial careers in new companies and revolutionary non-public and public sector agencies. Building upon insights from entrepreneurial event idea, cultural values principle and human capital concept, the cause of this paper is to explore the linkage between perceived nearby cultural surroundings, ESE investment, and the depth of entrepreneurial purpose with reference to becoming an entrepreneur.

The higher training industry within the GCC is growing wherein; international locations like Bahrain are making investments thousands and thousands in this regard. In order to attract and keep college students, it's far crucial that higher training institutes are able to fulfill them. Keeping the necessity of this detail in view, the existing examine tried to take a look at how the classroom surroundings can possibly make an impact on college students' delight. Therein, the prevailing has a look at examined lecture room environment, classroom format, and net centers on college students' delight among the students of a private university in Bahrain. A total of 325 questionnaires using a self-administered

approach had been allotted amongst undergraduate students in a personal college in Bahrain. Results from the structural equation modeling located a sizable courting between study room environment and pupil satisfaction and among internet facilities and scholar pleasure. However, the study did not locate any giant relationship between study room environment and student pleasure. The examiner forwards theoretical and sensible implications observed with the aid of the scope for similarly studies.

2.3 ANALYSIS AND SUMMARY

There have been several forms of research on the impact or impact of character tendencies on entrepreneurial chance. This paper examined the relationship of youngsters' entrepreneurial intention with character traits such as locus of manipulation and need for success. And how a lot of entrepreneurial schooling impacts entrepreneurial skills. I discovered no studies for predicting the Entrepreneurship abilities based totally on Big Five persona developments.

2.4 CHALLENGES AND BARRIERS

The purpose of my research is to predict the entrepreneurship skills of human beings. I ask some questions in my survey for collecting data. It is so challenging for me to collect data because I asked 40 questions to the volunteers. The questionnaires contained 20 personality traits questions and 20 entrepreneurial questions respectively. All questions were ranked with 1 to 5 number. Moreover, most people were not willing to share their personal information. That's why it became difficult for me to collect information.

CHAPTER 3

RESEARCH METHODOLOGY

3.1 RESEARCH SUBJECT AND RESEARCH INSTRUMENT

Most of the students in our country start searching for jobs after finishing their studies. There are only a few college students who need to be marketers or self-reliant after their studies. There are many college students who, in spite of having abilities, no longer want to be marketers due to fear of failure or aren't aware of their abilities. In my paper, I am that specialize in predicting entrepreneur talents primarily based on character tendencies and entrepreneurial self-evaluation. Here we've accrued forms of records. One is Personality traits and the opposite is entrepreneurial self-assessment.

3.2 Dataset Utilized

There have types of information series approaches for this paper. I have gathered statistics bodily from college students via fill-up the bureaucracy with the aid of traveling the school room. I also gathered records thru Google Forms. And most of the information was accumulated thru Google Forms. The members had been asked to top off the google shape and printable hard reproduction. Every personality trait and entrepreneurial self-assessment question was labeled between 1 to 5 score, where 5= strongly agree, 4=agree, 3=neutral, 2=disagree, and 1=strongly disagree. I have collected data from 408 participants, among them 95 were female and 313 were male. All the volunteers were Bangladeshi. All statistics from members are kept anonymous and they duly signal the consent form for records privacy and security concerns.

Table 3.1: Details of Dataset

Properties	Value
Number of Participants	408
Total Exponent	408
Number of Outstanding class Exponent	36

Number of Satisfactory class Exponent	219
Number of Inappropriate class Exponent	153
Number of male Exponent	313
Number of female Exponent	95

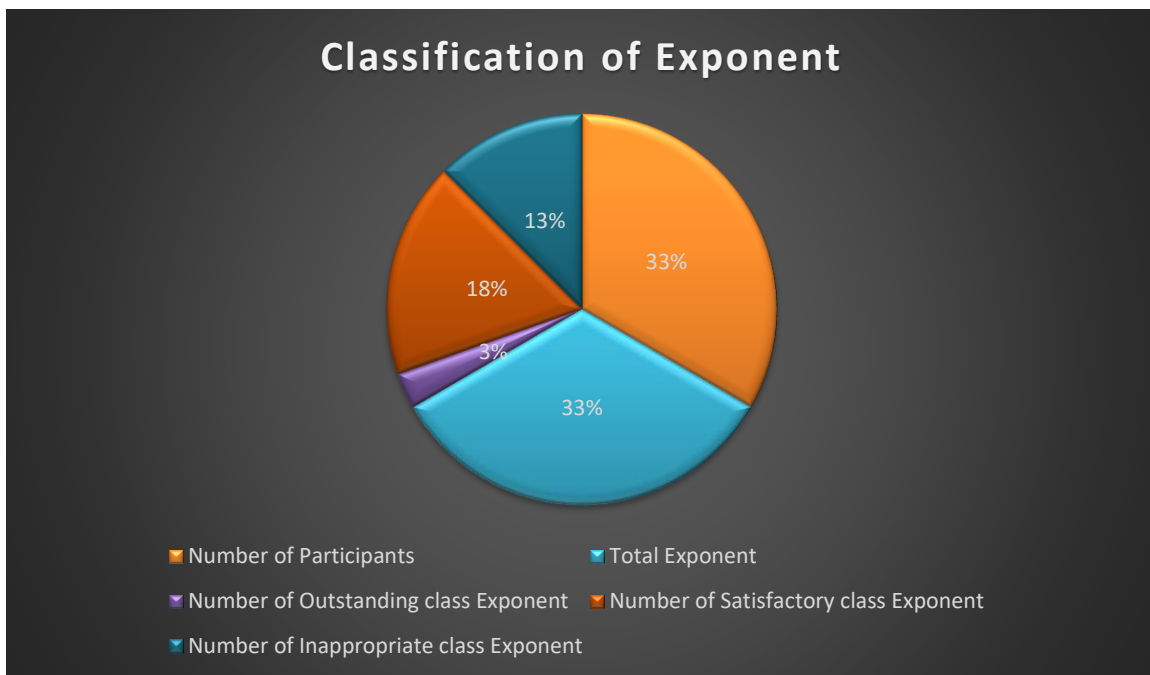


Figure 3.1: Classification of Exponent

3.3 Applied Method

In this thesis total of 7 types of machine learning algorithms were implemented here. By using those algorithms, I have trained and tested my dataset. And then I got different results and accuracy for different algorithms. The following algorithms I have implemented were:

- Random Forest
- Random Tree
- Hoeffding Tree
- Logistic Model Tree (LMT)

- Decision Stump
- Bayes Net
- Naive Bayes

Random Forest [1]

Random forest is my first implemented algorithm. It is used for classification and regression. It is mainly used for classification problems. Random forest is a decision tree-based algorithm which is building a big wide variety of trees at the records set whilst training and assembling them. At the final stage is selects the best output that is the mode of the classes and means regression of every tree.

Table 3.2: Confusion Metrics of Random Forest

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	219	0	0
Outstanding	19	17	0
Inappropriate	1	0	152

Random Tree [2]

Random tree is my second implemented algorithm. It is a supervised classifier algorithm that works with both classification and regression problems. The random tree is a combination of two machine-learning algorithms.

Table 3.3: Confusion Metrics of Random Tree

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	160	18	41
Outstanding	19	13	4

Inappropriate	56	4	93
---------------	----	---	----

Hoefding tree [3]

My third machine-learning algorithm is the Hoefding tree. This is an incremental and selection tree-based totally set of rules that uses a decision tree induction approach for mastering. It is so capable of learning from a large data stream.

Table 3.4: Confusion Metrics of Hoefding Tree

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	201	10	8
Outstanding	5	31	0
Inappropriate	1	0	152

Logistic Model Tree (LMT) [4]

It is a popular method perfect for class problems and works high-quality for the prediction of nominal and numerical capabilities. The logistic Model Tree is a mixture of logistic regression models and tree induction. Where logistic regression capabilities continue to be as leaves with a preferred choice tree.

Table 3.5: Confusion Metrics of Logistic Model Tree (LMT)

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	219	0	0
Outstanding	2	34	0
Inappropriate	0	0	153

Decision Stump [5]

Decision stump is the fifth algorithm that I have implemented on the dataset. Decision Stump is a classifier machine learning algorithm. It makes decisions by making one level decision tree. Decision stump has an internal node known as root and a terminal node known as leaves and the immediately connected each other. It's decision is known as one-rule because this algorithm makes decision just using a single input feature.

Table 3.6: Confusion Metrics of Decision Stump

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	219	0	0
Outstanding	36	0	0
Inappropriate	0	0	153

Bayes Net [6]

A Bayes net is a probabilistic graphical model for representing know-how about an unsure domain where each node corresponds to a random variable and every edge represents the conditional possibility for the corresponding random variables. It works with linear regression, ridge regression, logistic regression and all are supervised learning algorithms.

Table 3.7: Confusion Metrics of Bayes Net

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	212	7	0
Outstanding	0	36	0
Inappropriate	0	0	153

Naive Bayes [7]

My last but not least implemented algorithm is Naive Bayes machine learning algorithm. This algorithm is so fast that it can build a model and quickly predict the output. Naive Bayes is a probabilistic classifier. It is used for recommendation system, spam filtering, sentiment analysis. It is fast and easy method for predicting a statement.

Table 3.8: Confusion Metrics of Naive Bayes

Prediction \ Observed	Satisfactory	Outstanding	Inappropriate
Satisfactory	182	15	22
Outstanding	1	35	0
Inappropriate	5	0	148

3.4 SYSTEM REQUIREMENTS

- Windows operating system (Windows 8 or above)
- Waikato environment for knowledge analysis software (WEKA)

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1 EXPERIMENTAL SETUP

The main purpose of my research is to make prediction about entrepreneurial skills of human being based on their entrepreneurial self-assessment and personality traits. For data collection, I gathered data through physical and online Google Forms. Volunteers were asked to fill-up the any of online or offline form. Every personality trait and entrepreneurial self-assessment question were labeled between 1 to 5 score, where 5= strongly agree, 4=agree, 3=neutral, 2=disagree, and 1=strongly disagree.

For using this information in machine learning, I even collected these primary data and transformed them right into a dataset. This dataset contains the five personality traits values (Openness-to-experience, Extraversion, Agreeableness, Neuroticism and Conscientiousness) and entrepreneurial class too. Personality traits values and entrepreneurial magnificence are determined the use of pre-set formulation. For every participant, that information is accrued in each row considering a feature set.

For the algorithm implementation, I used the CSV file format in Waikato Environment for Knowledge Analysis (WEKA) developed by a New Zealandia university named university of Waikato. Added that it is free open-source software that provides tools for data preprocessing, implementation of several Machine Learning algorithms, and visualization tools with many classification algorithms.

4.2 Questionaries

I asked a total of 20 questions to participants based on Entrepreneurial Self-Assessment Survey.

1. I am willing to work 50 hours or more per week regularly.
2. My family will support my going into business.
3. I am willing to accept both financial and career risks when necessary.
4. I don't need all the fringe benefits provided by conventional employment.

5. I would like to take full responsibility for the successes and failures of my business.
6. I would experience more financial success by operating my own business.
7. I feel a great deal of pride when I complete a project successfully.
8. I have a high energy level that can be maintained over a long time.
9. I enjoy controlling my own work assignments & making all decisions affecting my work.
10. I believe that I am primarily responsible for my own successes and failures.
11. I have a strong desire to achieve positive results even when it requires a great deal of additional effort.
12. I have a good understanding of how to manage a business.
13. I can function in ambiguous situations.
14. One or both of my parents were entrepreneurs.
15. I believe that my abilities and skills are greater than those of most of my coworkers.
16. People trust me and consider me honest and reliable.
17. I always try to complete every project I start, regardless of obstacles and difficulties.
18. I am willing to do something even when other people laugh or belittle me for doing it.
19. I can make decisions quickly.
20. I have a good network of friends, professionals, and business acquaintances.

These 20 questions were translated in Bengali for our native people.

1. আমি নিয়মিত প্রতি সপ্তাহে ৫০ ঘন্টা বা তার বেশি কাজ করতে ইচ্ছুক।
2. আমার পরিবার আমার ব্যবসায় যেতে (আর্থিক/মানসিক) সহায়তা করবে।
3. আমার যদি প্রয়োজন হয় আর্থিক এবং ক্যারিয়ার উভয় ঝুঁকি গ্রহণ করতে ইচ্ছুক।
4. আমার প্রচলিত কর্মসংস্থান দ্বারা সরবরাহকৃত সব ধরনের উপকারের প্রয়োজন নেই।
5. আমি আমার ব্যবসায়ের সাফল্য এবং ব্যর্থতার পুরো দায়িত্ব নিতে চাই।

6. আমি আমার নিজের ব্যবসা পরিচালনা করে আরও আর্থিক সাফল্য অর্জন করব।
7. আমি যখন একটি প্রকল্প সফলভাবে সম্পন্ন করি তখন গর্ববোধ করি।
8. আমার উচ্চ কর্মশক্তি আছে যা দীর্ঘ সময় ধরে বজায় রাখা যায়।
9. আমি আমার নিজের কার্যকারিতা নিয়ন্ত্রণ করা এবং আমার কাজকে প্রভাবিত করে সমস্ত সিদ্ধান্ত গ্রহণ উপভোগ করি।
10. আমি বিশ্বাস করি যে আমি আমার নিজের সাফল্য এবং ব্যর্থতার জন্য প্রাথমিক ভাবে দায়বদ্ধ।
11. আমি ইতিবাচক ফলাফলগুলি অর্জন করার ইচ্ছা যখন থাকে, যদিও তা অর্জনের জন্য অতিরিক্ত প্রচেষ্টার দরকার হয়।
12. কীভাবে কোনও ব্যবসা পরিচালনা করতে হবে তা সম্পর্কে আমার ভাল ধারণা আছে।
13. আমি বিভ্রান্তিকর পরিস্থিতিতে কাজ করতে পারি।
14. আমার বাবা / মা / দুজনেই উদ্যোক্তা ছিলেন।
15. আমি বিশ্বাস করি যে আমার দক্ষতা আমার বেশিরভাগ সহকর্মীর চেয়ে বেশি।
16. লোকেরা আমাকে বিশ্বাস করে এবং আমাকে সৎ ও নির্ভরযোগ্য মনে করে।
17. আমি সবসময় বাধা এবং অসুবিধা বিবেচনা না করেই শুরু করি এবং প্রতিটি প্রকল্প শেষ করার চেষ্টা করি।
18. আমি একটি কাজ সম্পন্ন করি যদিও অন্য লোকেরা যখন আমাকে এটি করার জন্য হেয় করে।
19. আমি দ্রুত সিদ্ধান্ত নিতে পারি
20. আমার বন্ধু, পেশাদার এবং ব্যবসায়িক পরিচিতদের একটি ভাল নেটওয়ার্ক রয়েছে।

After these 20 questions I asked participants another 20 questions which are on Personality traits.

1. I am the middle of social events.
2. Empathize with the feelings of others.
3. I can finish housework easily.

4. Frequently mood swings.
5. Have a vivid imagination.
6. Don't talk too much.
7. Not interested in other people's problems.
8. Often forgets to put things in their proper place.
9. Feel relaxed most of the time.
10. Not interested in unrealistic ideas.
11. Talk to different people at social events.
12. I feel the feelings of others.
13. I like to obey orders.
14. I get upset easily.
15. Difficulty understanding new/unrealistic concepts.
16. Hide yourself at social events.
17. I am not really open to others.
18. Keep things untidy.
19. Sometimes I feel disappointed.
20. No good imagination.

These 20 questions were translated in Bengali for our native people.

1. আমি সামাজিক অনুষ্ঠানের মধ্যমনি
2. অন্যের অনুভূতির সাথে সহানুভূতি প্রকাশ করি
3. ঘরের কাজ সহজেই শেষ করতে পারি
4. ঘন ঘন মেজাজ বদলায়
5. প্রাণবন্ত কল্পনা শক্তি আছে
6. বেশি কথা বলি না
7. অন্যান্য লোকের সমস্যায় আগ্রহী নই
8. প্রায়শই জিনিসগুলোকে তাদের যথাযথ স্থানে রেখে দিতে ভুলে যায়
9. বেশিরভাগ সময় রিল্যাক্স বোধ করি
10. অবাস্তব ধারণাগুলিতে আগ্রহী নই
11. সামাজিক অনুষ্ঠানে বিভিন্ন ব্যক্তির সাথে কথা বলি

12. অন্যের অনুভূতি অনুভব করি
13. আদেশ মানতে পছন্দ করি
14. সহজেই মর্মান্বিত হয়ে যাই
15. নতুন/ অবাঞ্ছিত ধারণা বুঝতে অসুবিধা হয়
16. সামাজিক অনুষ্ঠানে নিজেকে লুকিয়ে রাখি
17. আমি অন্যের প্রতি সত্যই অগ্রহী নই
18. জিনিসপত্র অগোছালো রাখি
19. মাঝে মাঝে হতাশ বোধ করি
20. ভালো কল্পনা শক্তি নেই

4.3 EXPERIMENTAL RESULTS AND ANALYSIS

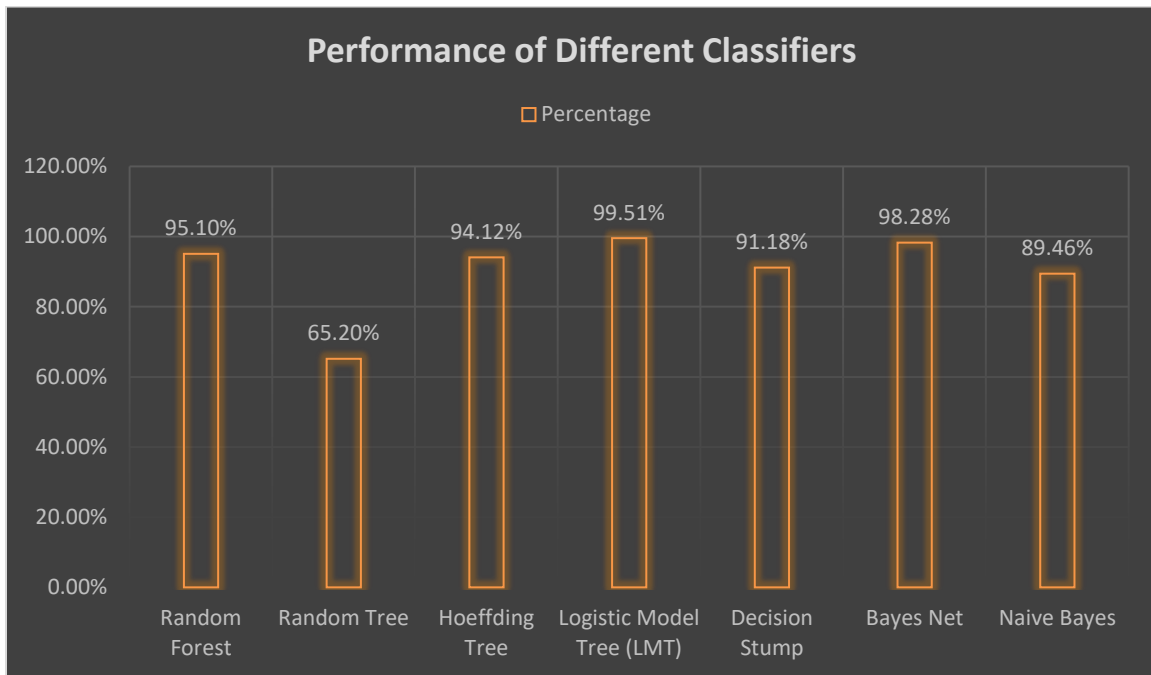
In this research paper, I used machine-learning algorithms are Random Forest, Random Tree, Hoeffding Tree, Logistic Model Tree (LMT), Reduced Error Pruning Tree (REP), Decision Tree (J48), Decision Stump, Bayes Net, and Naive Bayes. I have used 10-fold cross validation for each algorithm. 10-fold cross validation means use the part of dataset 9 times for training and the final tenth is for testing purpose. Algorithms are built into the WEKA and we can use the software to apply the dataset directly to the collected data.

Table 4.1: Performance Metrics of Different Algorithms

Classifier Name	Precision	Recall	F-Measure	Accuracy	MCC	ROC Value	PRC Value
Random Forest	0.955	0.951	0.944	95.098%	0.918	0.999	0.999
Random Tree	0.651	0.652	0.650	65.1961%	0.373	0.689	0.574
Hoeffding Tree	0.944	0.941	0.942	94.1176%	0.902	0.964	0.958
Logistic Model Tree (LMT)	0.995	0.995	0.995	99.5098%	0.992	0.970	0.950
Decision Stump	0.9295	0.912	0.962	91.1765%	0.917	0.923	0.843

Bayes Net	0.986	0.983	0.983	98.2843%	0.974	0.971	0.956
Naive Bayes	0.908	0.895	0.896	89.4608%	0.825	0.962	0.964

Figure 4.1: Performance Graph of Different Classifiers



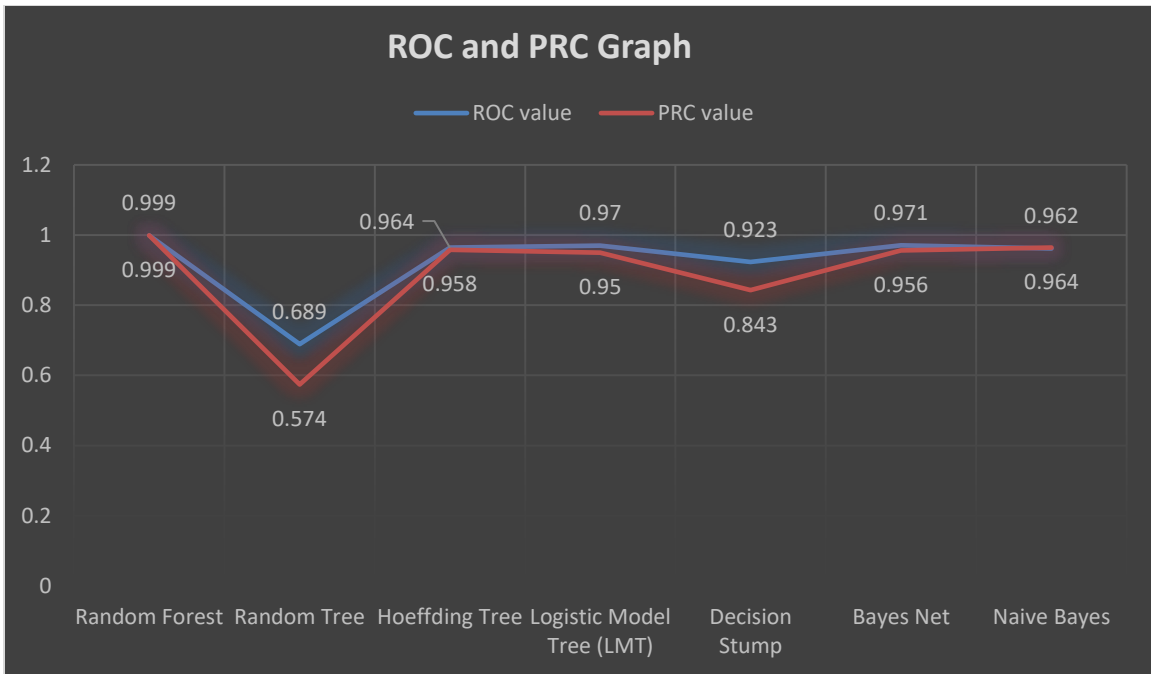


Figure 4.2: ROC and RRC graph for 10-Fold Cross Validation

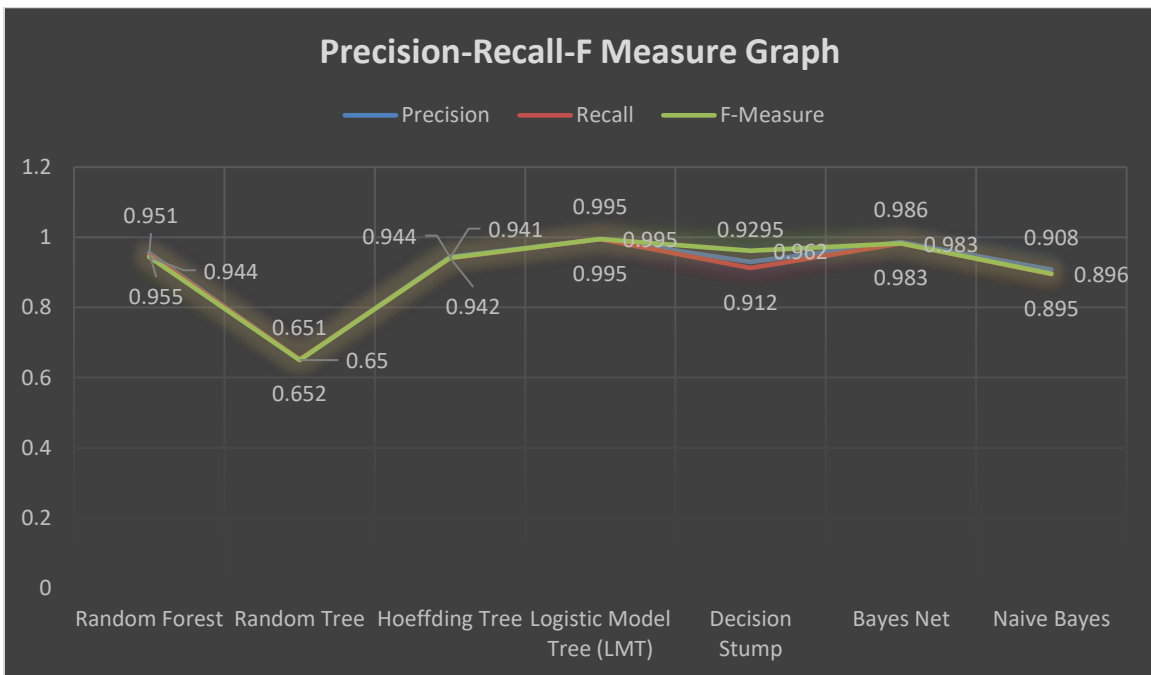


Figure 4.3: Precision-Recall-F Measure Graph of Different Classifiers

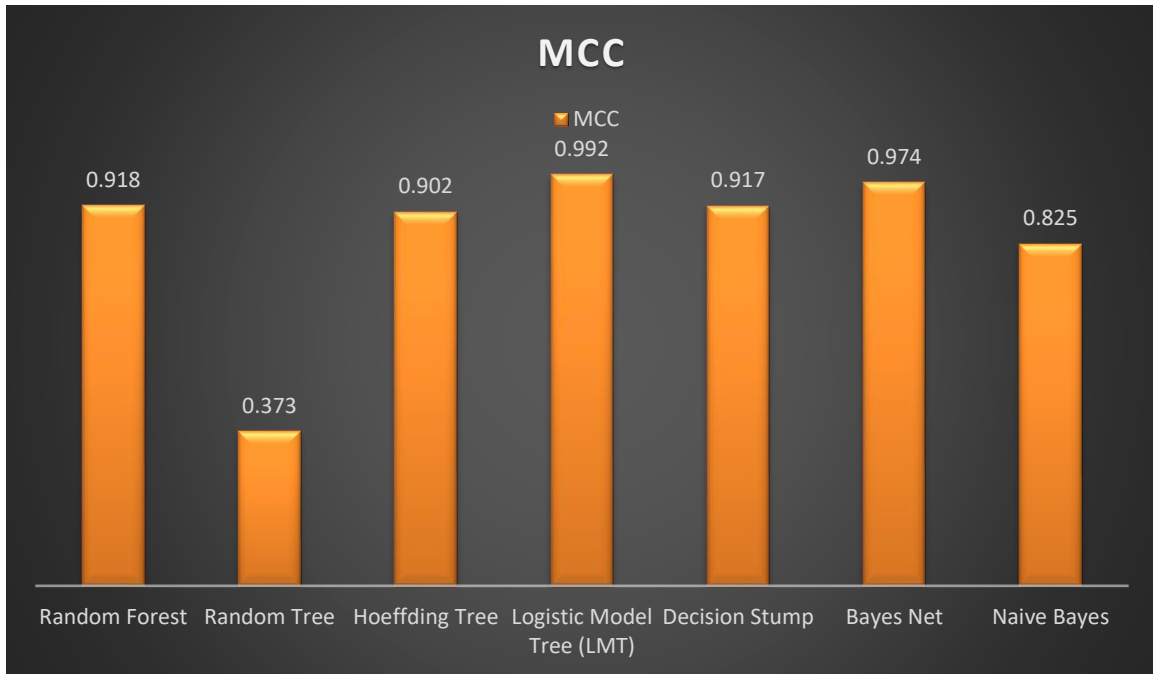


Figure 4.4: MCC Values of Different Classifiers

Table 4.2: Kappa Statistics, MAE and RMSE of Different Classifiers

Algorithm Name	Kappa Statistics	MAE	RMSE
Random Forest	0.9096	0.2047	0.2509
Random Tree	0.3745	0.2389	0.4726
Hoeffding Tree	0.8971	0.0529	0.1928
Logistic Model Tree (LMT)	0.9913	0.0106	0.0545
Decision Stump	0.8316	0.1011	0.2248
Bayes Net	0.97	0.0132	0.0893
Naive Bayes	0.82	0.0901	0.2337

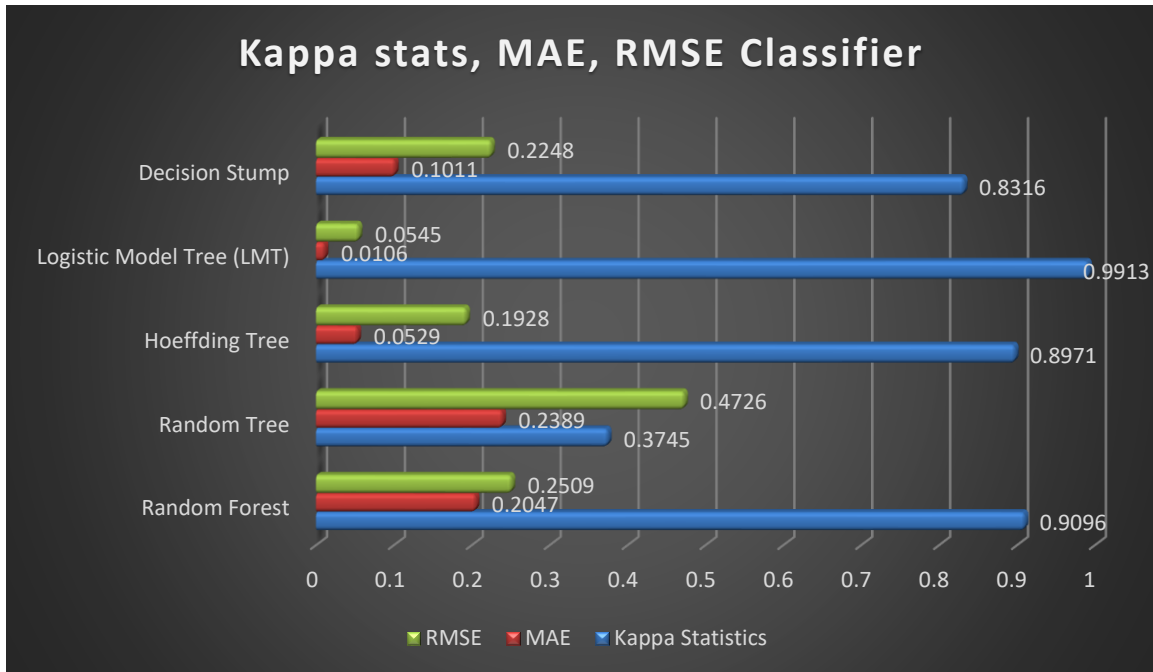


Figure 4.5: Kappa statistics, MAE and RMSE classifier

4.4 DISCUSSION

From the "performance metrics of different algorithms" table we have seen that Logistic Model Tree (LMT) algorithm got 99.5098% accuracy to predict the dataset.

On the other hand, from the "kappa statistics, MAE and RMSE" table we have seen that Logistic Model Tree (LMT) got the highest score because its MAE and RMSE values are near to Zero (0).

Here, kappa statistics shows that how closely the algorithm classify the dataset. Kappa statistics are measured from 0 to 1. The closer to the 1 means more accurate output.

Mean absolute error (MAE) measures the common importance of the errors in a hard and fast of forecasts, without considering their direction. It measures accuracy for non-stop variables. And Root mean square error (RMSE) is the degree of the way well a regression line fits the records factors. RMSE can also be construed as a Standard Deviation within the residuals

Why is a machine learning algorithm shows the result with 100% accuracy?

Ans: In WEKA, it is using a part of the training data for testing purposes. So, at the time of training, the machine learns about the data and gains knowledge about the dataset. And when it is tested again at that time the machine gives the same value.

CHAPTER 5

IMPACT ON SOCIETY, ENVIRONMENT AND SUSTAINABILITY

5.1 IMPACT ON COMMUNITY

An entrepreneur is a vital man or woman in society. Entrepreneurs increase monetary boom with the aid of introducing progressive thoughts, technologies, merchandise, and services. Due to marketers, each day wishes in special sectors are met. This paper will undoubtedly impact in the society. These studies will assist create more new marketers. Entrepreneurs' capabilities may be more advantageous thru complete guidance and suitable educational assistance. Through this, they'll get guidance on the way to gather entrepreneurial talents.

Improving business enterprise key fitness calls for in the path of the meant work business which in addition will affect work in the eyes of the general public. This effect has been described via collaboration in intentional paintings no matter whether it exists. This work may also consist of volunteering at college, in close by networks, representative chip-in, or herbal chip-in. Interest in non-commercial enterprise paintings additionally the male undergraduate elegance is extra frequently worried in interest to begin a non-commercial commercial enterprise out of doors in their work example, beginning an entertainment activity club.

5.2 IMPACT ON ECONOMY

Entrepreneurship is considered essential for a dynamic financial system. Entrepreneurs create employment opportunities now not simplest for themselves but additionally for others. Entrepreneurial interest can affect a country's economic overall performance by means of bringing new products, techniques, and manufacturing methods to the marketplace and promoting productivity and opposition extra widely.

Most students can't pick out an appropriate challenge to study in order that they cannot make their careers higher. If a scholar knows he has the skills to be an entrepreneur, he can construct his career hence. He would not watch for a job anymore. He will able to deliver jobs to much greater. The unemployment charge among university graduates in Bangladesh is 38.6 percent. Entrepreneurs provide new activity opportunities. And this research allows

the creation of a greater number of new entrepreneurs. The unemployment rate will be lower in our country.

5.3 ETHICAL PERSPECTIVE

One of the aspects of any proper deed is morality. The end result of the work is obtained simplest whilst the moral elements of the works are maintained. It is discussed and defined in elements in diverse articles. It can bring steerage in anybody's existence in any case. One of the areas in which ethics has become increasingly substantial is the wide variety of languages, business, and so on. Businesses are small and medium-sized businesses. There are advantages to using it each financially and socially. A two-part query that introduces itself. In this paper, I actually have attempted to recollect the grievance of the ongoing enthusiasm for commercial enterprise and a part of the moral elements. As well as the social and exact area that consists of the financial aspects as nicely. It is the situation of a morally mirrored image. We can say that the thrilling component of that is the moderately little communication approximately commercial enterprise ethics. I would like to distinguish and guard the idea of a business that can without problems meet as the focus of the investigation of the ethics of a residential company. I even have tried a short initiative to be able to lead to ethical inquiries at the required level. The moral aspects of a commercial enterprise and conflict that I have targeted in this paper, there are three basic businesses of ethical troubles in enterprise to this degree. The business with which decentralization, growth, and financial system are linked has to paintings with increase. There are many statistics that emphasize the benefits of the employer. All assumptions need to be divided on both facets. I am not trying to give all of the ethical assumptions of the agency, however, to provide a framework that we can additionally examine diverse moral and treasured problems of the enterprise.

5.4 SUSTAINABILITY PLAN

Building a marketing strategy that makes the affiliation budgetary, cultural, and ecologically justifiable. The scale of climate trade and social equity always escalates into frustration and complexity most effectively while the global financial system is depressed. Legitimate development positions high at the political association. But the boundaries seem nevertheless unreasonably complicated to cope with the sports plans vehemently. Cutting-edge composting unequivocally helps the relationship between the feasible unforeseen improvement and practical commercial enterprise mission, the initiative is recognized as the quality title for cultural and monetary development and as a consequence of radical alternatives. I at the beginning discover the ebb and flow concerns of big enterprise and prudent development with the association to agree upon a definition so one can fill in as a purpose for the accompanying discussions and examination. Along those traces, a chart of the crucial subjects tended to by means of the composing suggesting efficient enterprise mission identification given. New experience appearances, creative drivers, execution, and foundations are a few of the vital topics I investigated. Reasonable undertaking insinuates the disclosure, advent, and maltreatment of venturesome open entryways that upload to practicality via making social and characteristic alternatives for others in the public subject (Hockerts and wustenhagen, 2010; Pacheco et al., 2010; shepherd and patzelt, 2011). The divided conviction among business endeavor and acceptability is the opportunity of life expectancy, making sure reliable items, characteristics, or businesses: securing current assets for individuals in a while sensibility and making specific responses for the because an extended-time again run enterprise. Regardless, I want to override, for the conservative commercial enterprise area, the possibility of lifestyle expectancy with one in-all impact. Usually, fashionable enterprise centers around existence variety, i.E., making suffering matters or agencies.

CHAPTER 6

SUMMARY, CONCLUSION, RECOMMENDATION, AND IMPLICATION FOR FUTURE RESEARCH

6.1 SUMMARY OF THE STUDY

Entrepreneurship is the dynamic procedure of creating incremental wealth. The main significance of entrepreneurship is the creation of activity opportunities, innovation, and improve the economy. It provides civilization with a massive number of goods and offerings and enhances the growth of social welfare. This is why an entrepreneur is so important for our community, society, nation, and country.

A successful entrepreneur desires to have entrepreneurial qualities in addition to some character tendencies. In my paper, I mapped persona trends to the entrepreneurial characteristics of each of the scholars and the usage of numerous forms of machine-getting to know algorithms. Random Forest, Random Tree, Hoeffding Tree, Logistic Model Tree (LMT), Decision Stump, Bayes Net, and Naive Bayes algorithms were implemented on my dataset. The system was trained and tested by the 408-survey data that I have collected. I got the lowest accuracy from the Random tree and it was 65.1961%. Where the average accuracy was around 90%. The logistic Model Tree (LMT) performed well among them and it achieved 99.5098% accuracy.

6.2 CONCLUSION

The reason for our paper is to locate capability entrepreneurs thru the application of machine learning algorithms. Seven different kinds of machine learning algorithms classifiers have been applied right here. Among them, the Logistic Model Tree (LMT) gives optimal and comparatively better accuracy which is 99.5098%. This research is This research might be beneficial for all ages of people because they'll recognize approximately their abilities. Also, assist our country to expand by means of growing new entrepreneurs. Entrepreneurship education has a more influence on the financial literacy in the growth aspect consciousness of commercial enterprise opportunities and might enhance entrepreneurship abilities. This has implications that the importance of growing awareness

of commercial enterprise possibilities in entrepreneurship education. By using the predicting tool, we can know about our drawbacks. We can invest more time to our disability for changing them.

6.3 FURTHER STUDY

I have plans to go further with this research topic. In this research paper, I am working with 408 people's data. In the future, I will work with more machine-learning algorithms with a good amount of data to get more accurate outcomes. The barrier is I am working with WEKA, in the future I will train and test my dataset manually with python program for getting better accuracy.

REFERENCE:

- [1] Swasti Singhal, Monika Jena, “A Study on WEKA Tool for Data Preprocessing, Classification and Clustering”. International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-2, Issue-6, May 2013
- [2] Eshwari Girish Kulkarni, Raj B. Kulkarni, “WEKA Powerful Tool in Data Mining”. International Journal of Computer Applications (0975 – 8887) National Seminar on Recent Trends in Data Mining (RTDM 2016)
- [3] Batta Mahesh, “Machine Learning Algorithms -A Review”. International Journal of Science and Research (IJSR) ISSN: 2319-7064 ResearchGate Impact Factor (2018)
- [4] Xiaogang Wang, “Machine Learning Basics”. xgwang@ee.cuhk.edu.hk. January 5, 2015
- [6] Matthias Schonlau, Rosie Yuyan Zou, “The random forest algorithm for statistical learning”. The Stata Journal (2020)
- [7] Yonghui Xu, Xi Zhao, Yinsheng Chen and Zixuan Yang, “Research on a Mixed Gas Classification Algorithm Based on Extreme Random Tree”. 26 April 2019
- [8] Asmah Muallem, Sachin Shetty, Jan W. Pan, Juan Zhao, Biswajit Biswal, “Hoeffding Tree Algorithms for Anomaly Detection in Streaming Datasets: A Survey”. Computational Modeling and Simulation Engineering Faculty Publications, 2017
- [9] Arvind Kumar, Parminder Kaur and Pratibha Sharma, “A Survey on Hoeffding Tree Stream Data Classification Algorithms”. CPUH-Research Journal: 1(2), 28-32 ISSN (Online): 2455-6076, 2015
- [10] Bernhard Pfahringer, Geoffrey Holmes, and Richard Kirkby, “Handling Numeric Attributes in Hoeffding Trees”. T. Washio et al. (Eds.): PAKDD 2008, LNAI 5012, pp. 296–307, 2008. c Springer-Verlag Berlin Heidelberg 2008
- [11] NIELS LANDWEHR, MARK HALL, EIBE FRANK, “Logistic Model Trees”. Machine Learning, 59, 161–205, Springer Science + Business Media, Inc. Manufactured in The Netherlands: 2005
- [12] Viet-Ha Nhu, Ataollah Shirzad, Himan Shahabi, Sushant K. Singh, Nadhir Al-Ansari, John J. Clague, Abolfazl Jaafari, Wei Chen, Shaghayegh Miraki, Jie Dou, Chinh Luu, Krzysztof Górski, Binh Thai Pham, Huu Duy Nguyen and Baharin Bin Ahmad, “Shallow Landslide Susceptibility Mapping: A Comparison between Logistic Model Tree, Logistic Regression, Naïve Bayes Tree,

Artificial Neural Network, and Support Vector Machine Algorithms”. International Journal of Environmental Research and Public Health: 16 April 2020

[13] Ayinde, Dr Adetunji, Bello and Odeniyi, “Performance Evaluation of Naive Bayes and Decision Stump Algorithms in Mining Students’ Educational Data”. JCSI International Journal of Computer Science Issues, Vol. 10, Issue 4, No 1, ISSN (Print): 1694-0814 | ISSN (Online): 1694-0784, July 2013

[14] Taram Nayab Shah, Muhammad Zakir Khan, Mumtaz Ali, Bilal Khan, Nauman Idress, “CART, J-48graft, J48, ID3, Decision Stump and Random Forest: A comparative study”. University of Swabi Journal (USJ). Uni. J. Swabi., Vol.2, Issue, 1. April 2018

[15] Junbiao Pang, Qingming Huang, and Shuqiang Jiang,” Multiple Instance Boost Using Graph Embedding Based Decision Stump for Pedestrian Detection”. D. Forsyth, P. Torr, and A. Zisserman (Eds.): ECCV 2008, Part IV, LNCS 5305, pp. 541–552, 2008. c Springer-Verlag Berlin Heidelberg 2008

[16] Remco R. Bouckaert,” Bayesian Network Classifiers in Weka”. September 1, 2004

[17] Jie Cheng, David A. Bell, Weiru Liu,” An Algorithm for Bayesian Belief Network Construction from Data”.

[18] Ida Bagus Adisimakrisna Peling, I Nyoman Arnawan, I Putu Arich Arthawan, and IGN Janardana, “Implementation of Data Mining to Predict Period of Students Study Using Naive Bayes Algorithm”. International Journal of Engineering and Emerging Technology, Vol. 2, No. 1, January—June 2017

[19] Mohamed EL KOURDI, Amine BENSALD, Tajje-eddine RACHIDI,” Automatic Arabic Document Categorization Based on the Naïve Bayes Algorithm”.

[20] Masud Karim, Rashedur M. Rahman,” Decision Tree and Naïve Bayes Algorithm for Classification and Generation of Actionable Knowledge for Direct Marketing”. Journal of Software Engineering and Applications, 2013, 6, 196-206

[21] Seth Lloyd, Masoud Mohseni, Patrick Reberntrost,” Quantum algorithms for supervised and unsupervised machine learning”. 4 November 2013

[22] Ouafae EL AISSAOUI, Yasser EL ALAMI EL MADANI, Lahcen OUGHDIR and Youssouf EL ALLIOUI,” Combining supervised and unsupervised machine learning algorithms to predict

the learners' learning styles". Second International Conference on Intelligent Computing in Data Sciences (ICDS 2018)

[23] S. Rothmann, E. P. Coetzer, "The big five personality dimensions and job performance". *SA Journal of Industrial Psychology*. 29. doi:10.4102/sajip. v29i1.88. Retrieved 27 June 2013.

[24] Holte, Robert C. (1993). "Very Simple Classification Rules Perform Well on Most Commonly Used Datasets". CiteSeerX 10.1.1.67.2711.

[25] Svetlana S. Aksenova, "WEKA Explorer Tutorial for WEKA Version 3.4.3". 2004

[26] Esra Bal, Özlem Kunday, "Investigating the Role of Entrepreneurial Traits and Human Capital on the Internationalization Degree of Small and Medium Enterprises in Turkey". August 2014

[27] Nicos Nicolaou, Scott Shane, "Can genetic factors influence the likelihood of engaging in entrepreneurial activity?" *Journal of Business Venturing* 24 (2009) 1–22

[28] Harry Matlay, Marina Solesvik, "Cultural factors and entrepreneurial intention: The role of entrepreneurship education". Article in *Education and Training* · November 2014

[29] Dulcenombre Madrid, Umair Ahmed, Rajendra Kumar, "EXAMINING THE IMPACT OF CLASSROOM ENVIRONMENT ON ENTREPRENEURSHIP EDUCATION: CASE OF A PRIVATE UNIVERSITY IN BAHRAIN". *Journal of Entrepreneurship Education*. Volume 22, Issue 1, 2019

[30] "Big Five Personality Test | IPIP NEO | AidaForm," [aidaform.com. https://aidaform.com/templates/big-five-personality-test-ipip-neo.html#:~:text=The%20Big%20Five%20personality%20traits](https://aidaform.com/templates/big-five-personality-test-ipip-neo.html#:~:text=The%20Big%20Five%20personality%20traits)

[31] Wikipedia Contributors, "Big Five personality traits," Wikipedia, Mar. 06, 2019. https://en.wikipedia.org/wiki/Big_Five_personality_traits

[32] "Entrepreneurial Self-Assessment Survey." [Online]. Available: <https://s3-us-west-2.amazonaws.com/oerfiles/WMBusiness/Entrepreneur-Self-Assessment-Survey.pdf>

[33] "Supervised vs Unsupervised Learning Explained," Seldon, Oct. 16, 2021. <https://www.seldon.io/supervised-vs-unsupervised-learning-explained#:~:text=Supervised%20machine%20learning%20is%20generally>

[34] D. Johnson, "Supervised vs Unsupervised Learning: Key Differences," Guru99.com, Sep. 21, 2019. <https://www.guru99.com/supervised-vs-unsupervised-learning.html>

[35] "Supervised and Unsupervised learning - GeeksforGeeks," GeeksforGeeks, Apr. 17, 2019. <https://www.geeksforgeeks.org/supervised-unsupervised-learning/>

[36] "python - Should my model always give 100% accuracy on Training dataset?," Stack Overflow. <https://stackoverflow.com/questions/62186526/should-my-model-always-give-100-accuracy-on-training-dataset>

ORIGINALITY REPORT

23 %
SIMILARITY INDEX

11 %
INTERNET SOURCES

16 %
PUBLICATIONS

8 %
STUDENT PAPERS

PRIMARY SOURCES

1	"Innovative Data Communication Technologies and Application", Springer Science and Business Media LLC, 2022 Publication	9 %
2	dspace.daffodilvarsity.edu.bd:8080 Internet Source	3 %
3	Ahmed A. Marouf, Adnan F. Ashrafi, Tanveer Ahmed, Tarikuzzaman Emon. "A Machine Learning based Approach for Mapping Personality Traits and Perceived Stress Scale of Undergraduate Students", International Journal of Modern Education and Computer Science, 2019 Publication	1 %
4	www.researchgate.net Internet Source	1 %
5	Marina Solesvik, Paul Westhead, Harry Matlay. "Cultural factors and entrepreneurial intention", Education + Training, 2014 Publication	1 %
