Human Guided Machine Learning Framework for Making Better Prediction

Aziza Ashrafi

ID: 2013-1-60-008

Muktadir Ahmmad

ID: 2013-3-60-021

A thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science and Engineering

Department of Computer Science and Engineering
East West University
Dhaka-1212, Bangladesh

September, 2018

Declaration

We, hereby, declare that the work presented in this thesis is the outcome of the investigation performed by us under the supervision of Amit Kumar Das , Lecturer, Department of Computer Science and engineering, East West University. We also declare that no part of this thesis has been or is being submitted elsewhere for the award of any degree or diploma.

Countersigned

. . . .

(Amit Kumar Das)

Supervisor

Signature

. . . .

(**Aziza Ashrafi)**

(2013-1-60-008)

Signature

. . . .

(**Muktadir Ahmmad)**

(2013-3-60-021)

Letter of Acceptance

This thesis report entitled *"Human Guided Machine Learning Framework for Making Better Prediction"* submitted by Aziza Ashrafi (ID: 2013-1-60-008) and Muktadir Ahmmad (ID: 2013-3-60-021) to the Department of Computer Science and Engineering, East West University is accepted by the department in partial fulfillment of requirements for the Award of the Degree of Bachelor of Science and Engineering on August, 2018.

Supervisor

. . . .

(Amit Kumar Das)

Lecturer, Department of Computer Science and Engineering, East West University

Chairperson

. . . .

(Dr. Ahmed Wasif Reza)

Chairperson and Associate Professor,

Department of Computer Science and Engineering, East West University

Abstract

Machine-made prediction is a key tool to gain faster performance rather than rely on human prediction. Artificial Intelligence showed better performance than human in many situations. That's why thousands of research have been taking place to develop more advanced Artificial Intelligence which can not only perform faster but also predict better than human. But a human has some quality which can never be gained by a machine. Emotion, empathy, sensing, feeling are the characteristics where a machine cannot overcome human. These characteristics made human adaptive to take a decision over unstructured information, identify unusual circumstances and its consequences. So, we can say that a human and a machine both have their specific quality in the certain scenario. In this paper, we will try to figure out if a machine can predict punishment like a human judge. We will implement a machine learning algorithm to create a system where human and machine can perform together to improve decision with less time. Additionally, the machine's performance will be checked by increasing the number of observations in a different range. Through the study, we will try to evaluate the system whether the system can develop a new way to implement Artificial Intelligence in the judicial system. In this research paper, we present a punishment prediction system where a human can give their decision also if necessary. We apply several machine learning algorithms such as *Naïve Bayes*, *Logistic Regression*, *Support Vector Machine*, *Multiple Linear Regression*, *Artificial Neural Network for Regression and Classification.* By calculating both the test accuracy and the predictive power of the models, we observe which algorithm performs better and stable than the other models. We will also try to demonstrate what will be the impact of implementing such technology and what will be its technological, social, ethical and economic effect. In the end, we will try to state if a combined approach can produce a far more fruitful result than our regular judicial system. This combined model is a new way to represent an artificially intelligent agent as a judge.

Acknowledgments

We especially thank our supervisor, “Amit Kumar Das”, who helped us with heart and soul, who gave us the opportunity to work with him in this thesis, who gave us proper instruction through the research, helping us in our learning approach, instructing us to use different approaches of Artificial Intelligence and Machine learning.

We would also like to thank our instructor from the Department of Law, “Adity Rahman Shah” who helped us with every detail to understand the “Law of Bangladesh”, who gave her valuable time and also guiding us to fulfill our research. Also, we have to mention that the data set we have made was impossible to build without her guidance.

There are few other honorable mentions who helped in our research, “Mridul Bepari”, “MD. Monirul Islam”, “Aakib Sadat” from the Department of law and also few other experts gave us valuable information.

Aziza Ashrafi

July, 2018

Muktadir Ahmmad

July, 2018

Table of Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

1.1 Introduction

As research fields are increasing, implementation of artificial intelligence is reaching to a higher limit. Recently, a combinational approach of both human and machine has shown its higher performance potential for business companies. Hybrid models are proving their efficiency in various research sectors but still, there are so many fields where both human and AI-based system can be implemented. In this research, our work is to initialize Human and AI combinational technology to a field where AI didn't show enough sign before. One such field is court and justice. To specify we can say using machine intelligence in the decision-making system will amplify the working progress of a human judge. With the help of machine learning mechanism, we will try to prove if there is a possibility to make the intervention of humans with the machine in a critical decision-making process. Probably most of the higher research focuses on creating better performance, higher accuracy, less cost and less time consuming than a human or machine alone can do.

1.2 Differences in decision making capability

With the recent growth of faster industrial development, we observe that machines are replacing humans in many fields. Machine intelligence is more preferred for its time efficiency, availability, cost-effectiveness and risk-taking capability. Although it doesn't mean that humans are being ignored completely. Smart inventions are opening another doorway for humans to participate with the machine. From the top view, humans are always unique in their adaptability, creativity, sensation, wisdom and especially emotion.A research showed that human have dynamic capabilities and it can be upgraded to different stage according to the complexity of the task. Enhancing crowd workers performance was one experimental success [1]. Countless fields are being integrated with machine intelligence and humans. But when it comes to decision making there is always a competition going on. This research is based on a particular type of decision making where machines are yet haven't been introduced largely or too young in this sector. The study is about Judgment and Punishment. Whenever we see the participation of a machine to simplify a human's task, it always concludes with a massive change in output from that human alone. Because while machines are doing the time-consuming part of a job, a human can get more time to focus on other complex issues.

1.3 Human with Machine Intelligence

When we think about the job of a judge, we often think about a human who is working with so many rambling and unstructured data and cases. Probably we estimate humans to be the best performer in that field. We can try to increase his performance with the help of a machine intelligence. Considering a machine as a judge might not be a dream that people want to see. But if there is a possibility of integrating machine intelligence to enhance humans performance, that might be revolutionary. Although a human's thought on a particular case might not be implemented into a machine, a machine can bring change to a humans thought. We can't train a machine to have common sense but we can teach it rules. When we are thinking about performance, a machine can perform better if there is structured data, on the other hand, a human is more adaptive to work with complex and unstructured data. If we think of industrial development, we can see that machine-based intelligence is always been there to complete the complex and dangerous work where humans can't reach. A research demonstrated that, combined forecasting is always better and less risky than individual forecasting [2].

1.3.1 Performance improvement

A research in 2005 from Intel showed that market-developed forecast is beating traditional forecasting with accuracy and better result [3]. Recent research over Automated driving, combined intelligence for business prediction, Hybrid intelligence, Human computation and Task marketing with multi-agent system proves that machine intelligence can be improved to the maximum successful state with the collaboration of human participants [4][5][6][7]. On the other hand, there is also a human performance improvement observed in some research when machine learning algorithms were guiding humans in task marketing [7][8]. Our work is to implement a predictive mechanism for a criminal offense and examine if a machine can participate in judgment to enhance the workflow for the human.

1.3.2 Lowering risk

One advantage of implementing machine is, a machine can take any kind of risk and also a machine can run one task continuously forever. Finding a similar data among thousands of record might be hard for human but it is nothing for a machine. When industrial development only measures performance and cost, the only solution is to give human the control over final decision and keep the repeating task for the machine. So maximum utilization of time and effort will be possible through the hybrid process. We can teach modern AI technology to learn from new input. So there is a chance to get better prediction with less error in future.

1.4 Machine Intelligence in Judgment

Making an AI-powered machine that will replace a judge may not be appreciated by our society. However, we may exercise the machine intelligence to create an assistant of a judge. Artificial intelligence has large-scale potential to change our lifestyle and job. With the success of artificial intelligence, many visions are established like the machine will replace human and grasp all of their work. However, in reality, the machine can take over some places of human but not the all. Instead, it can be used to enhance the human abilities. Both of them have different competency. The computer is fast at solving complex mathematical problems, its programs can repeat forever.

Whereas, the human has the power of creativity, empathy, intuition and so on. Human and machine have merely different potency, neither of them is objectively better. With this sooth about both humans and machine, a novel paradigm for the judicial system can be anticipated by aggregating best sides of both. It will yield the most accurate and trustworthy system. This system will reside as an example that humans will never be irrelevant. Human will be the part of the equation that quest empathy, common sense, intuition, sentiment, etc. In this paper, we will claim that it will be beneficial to link human and machine intelligence for judging a delinquent. In fact, it might be practiced to make the judicial system more productive and error free.

1.5 Model design

In the next page, we have given the model about our model. The model is combining both human and machine in different situation. Machine prediction comes from machine learning algorithm models which is used on training data. After the machine prediction is completed, the decision is taken to human participant. If the human participant accept the machine predicted decision then it directly goes for the final result. If not, then the human predictor can provide his decision and the newly added punishment will be sent to the training data for further improvement.

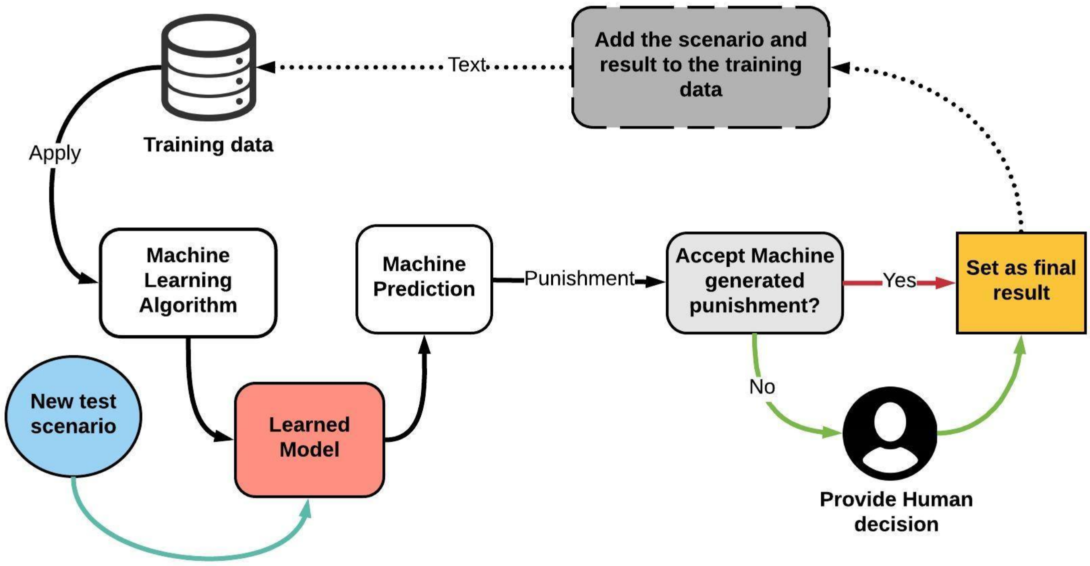

Figure 1.1: System Diagram for machine and human participation in decision making process for a criminal offense

1.6 Our Contribution

Within this paper, we will discuss the possible implementation of a punishment prediction system where both Artificial Intelligence and Human can interact. We are implementing Machine Learning Algorithms to determine machine prediction. From some criminal case, we will create our dataset and analyze it with different machine learning models and see if a machine can predict punishment correctly. Then we will provide a system for the human judge to receive the suggested information from the machine and then determine if he is going to keep the existing punishment or apply on his own.

Our contribution is summarized below:

- We collect information about cases related to "Women and Children Repression Prevention Act, 2000" ("Nari o Shishu Nirjatan Daman Ain, 2000") and it's punishment. Also, understand what are the criteria maintained specifically to determine a punishment.
- Under the guidance of some law experts, we identify both increasing and decreasing factors that have an influence on the punishment.
- From the unstructured information, we made our own dataset.
- Few machine learning algorithms have been implemented to check which algorithm works better for our system and provide the highest accuracy.
- A prediction model which can predict most accurate punishment is implemented in 'Python'.
- Finally, we implement the combined system where a machine will be trained over some training data and include a Graphical User Interface (GUI) where a human can take part.
- Each new scenario will be added to the training dataset with the human verification.

1.7 Organization of this book

In chapter II, we will describe some background related to our research idea. Chapter III will describe about problem statement and proposed work. Chapter IV will be about methodology . Result analysis and evaluation of the model will be described in chapter V. Effects and Issues with the system will be described in chapter VI and chapter VII will be the conclusion.

1.8 Conclusion

There are possibilities for further research on the decision-making process of a judge with Artificial Intelligence in future. Until researchers create error free machine predicting mechanism, there is no chance of making fully machine dependent model that works better than human and machine hybrid intelligence. But there are possibilities to research over the different criminal case and describe it into a structured system where a machine can participate. In the end, taking a decision from a machine intelligence system might seem odd for the human, But if that system reduced biased and error result, it can be considered as a supporting system for the human. Also, it opens a new way to implement Artificial Intelligence for the human's development.

Chapter 2

Related Works

2.1 Introduction

Combined prediction models are the most emerging feature in modern Artificial intelligence technology. Developing a way to give humans the ability to teach machine when it is wrong is mostly effective in general case. So, researchers proved that a human expert can participate with a machine only when a machine specially needs to learn a new label. With such technique, both time and cost can be reduced. Through the next section, we will describe most of these research where machine and human have been brought together.

2.2 State of the Art

In the past few years, the implementation of artificial intelligence have reached to a higher level. Modern Medical System, Task marketing, Automated driving, Market forecasting are some of the example where researches for linking human with artificial intelligence is being held. Developing a way to utilize machine ability is the most effective way to get faster production. Thus in most of it's case, putting together a machine had gave enough improvement than general. In next few scenario, we will discuss about some of these implementation and what was their outcome.

2.2.1 Hybrid AI system in Medical Research

Combined intelligence of both human and machine have been used for analyzing medical related data such as recognizing various skin diseases from images. Both of the intelligence can describe the result of the patient's ECG (Electrocardiography) test. However, questions may arise when a result generated from AI system conflicts with the result generated from human one. In such a situation, it is difficult to trust AI generated diagnosis and act upon it. To escape from this problem, in the paper [9] a method of synthesizing many diagnoses to scrutinize health data are proposed using both machine and human intelligence.

Many systematic biases or even human error have occurred during formal medical diagnoses when a homogeneous group of people makes the diagnosis. Sometimes it is preferred to use "wisdom of crowds." The "wisdom of crowds" effect results when integrating a group of forecasts yields a better prediction than any single forecast (Galton, 1907; Surowiecki, 2005). However, it is still very much difficult to find a better way of aggregating forecasts. In paper [9] a statistical technique is used to determine which prediction or decision should be considered next and when combined prediction will outperform a premeditated accuracy threshold. Humans are always best at assessing and resolving novel and ambiguous situation. They can make ethical decisions that are beyond the capabilities of a machine. Differently, a computational machine can respond faster than human when the situation is known and resides within its learning boundary.

2.2.2 AI in Automated Driving

Paper [5] primarily focuses on making a partnership between a computer and a driver for automated driving that will undoubtedly escalate safety. They have concentrated their idea on the decision-making process. A more optimal solution will be achievable if drivers are engaged in the decision making step. This joint cognition system will evaluate the environment by using the sensing capabilities of the vehicle. The computer will decide when or how to engage the driver by assessing the driver's state, time to collision and many other parameters. Some parameters like heart rate variation, eye movement and emotion are used to determine driver's state. Advanced sensors will be required to appraise the physiological output of the driver [5]. Driver's senses of vision and hearing are also used as valuable input for this system when the machine asks for taking the decision. In this way, human will expand the capabilities of the automated system and machine will enhance human capabilities and reduce human failings [5]. Combined intelligence used as a joint cognition will eliminate or at least lessen the rate of the accidents and help us to achieve a safety envelope.

2.2.3 Interactive Teaching Technique

Paper [6] also focuses on hybrid intelligence. The alliance of both human and AI intelligence is referred to as hybrid intelligence. This hybrid system will help to lessen the shortcomings of existing machine intelligence system. Introducing human to this hybrid system will complement the AI system. Even with the advancement of machine learning, we cannot pledge that the machine is perfect for the ambiguous situation. Human is still far better than machine when an environment has many unknown variables. Till now the machine cannot reach the level of reasoning that the human has. In the hybrid system, machine and human can work as a partner of each other. A machine will perform vast and critical computational work in one go, where human will participate to detect error or failure and even can give feedback. If this process continues to run in such a circular way, then it will become more intelligent and will make fewer mistakes.

However, the primary challenge is that how we rightfully integrate both intelligence that they will become better gradually. One constraint is that human intelligence is associated with costs. If the system does not require human intelligence in the loop (that means in real time program execution) for instant decision making, then another easy way of accessing human is crowdsourcing platform. An artificial system needs reasoning capabilities to identify the time and situation where it can be benefited from the complementary human intelligence. This paper [6] describes the human as a teacher of the machine that will teach the AI system how to act. The unrealistic atmosphere may be created when teacher always monitor student that whether he makes the right decision or not and provide advice. As it is tough to monitor every single act of the student, in the paper [6], it is advised to take an interactive teaching technique where the teacher will only help students when they ask for advice. This approach will indeed minimize cost while maintaining the same learning goals. The primary focus of paper [6] is to find out reasoning methods that will optimize access to human intelligence.In another research, combining a teacher and a student in advising opportunity concluded with reducing the amount of attention required by the teacher [10].

2.2.4 Machine generated translation

A recent discovery from Dafna Shahaf and Eric Horvitz (2010) shows that machine-generated translation can be improved to an acceptable stage by human translator's assistance. Their study based on Generalized Task Market (GTM) which contains participation of human and machine helpers side by side for solving problems like translating between different languages. In this mechanism, machines were upgrading its translating mechanism from the optimal solution suggested by human agents. To achieve an optimal solution, they assigned the task to agents in a way that prioritizes order [7]. When machine learns from human translator it reaches to a higher accuracy every time. Another research on computing with human and machine joined solvers gained classification accuracy on visual identification of a machine by using human aided vision system [11].

2.2.5 Hybrid prediction in Marketing Forecast

Although statistical models invariably generate more accurate prediction than experts, humans are better at recognizing the anomalous situation. In paper [4], it is visible that combine prediction from artificial neural net agents and real humans gives more strong result than any one of the groups alone. It links knowledge management and business intelligence using this combined knowledge from both of the groups. A prediction market is used to forecast the action of a football team by integrating predictions from both artificial neural net agents and experts. Predictions from the three groups (only group of humans, the group of artificial neural net agents, and hybrid group) are compared to determine which group is best at prediction. The primary task of each group is to predicts, what would be the next play (pass or run) in a football game based on the current situation of the game. However, to assess the quality of the prediction, understanding the trade-offs when comparing the predictions is a must. Three common criteria - accuracy, discrimination, and reliability were used to understand the trade-offs of three groups of predictors

2.2.6 Enhance human performance

A research by Andrew Mao, David C. Parkes, Ariel D. Procaccia and Haoqi Zhang (2011) over Human Computation managed to achieve better performance gain using algorithmic workflow. They used Amazon Mechanical Turk (AMT) platform to demonstrate participant's performance over Graph Coloring problem. However, the research was not to prove that humans can perform better than a computer. Instead, they used machines as a guide for human participants to achieve the result within minimum time [8].

2.3 Conclusion

Through the background research , we can jump to a conclusion that there is possibilities to include a human participant into critical decision making task. Also a machine can be used as an action taker when human expert is busy over other task. For the project, we will use the procedure to include human decision only when needed. In some of the scenario in above we have seen that a machine algorithm can help a non expert human participant to predict better with less error.

Chapter 3

Problem Statement and Proposed Work

3.1 Introduction

In this section, we will describe why AI-powered judging system is necessary and can be proved as an effective addition to the judicial system. There is also one thing to be noted that the total dataset in this system is based on the Law of Bangladesh. Also, the court and law enforcement system in Bangladesh is totally non-digital, most crime case and data were kept in a handwritten file here. So there is very little chance of finding any digital data which can be included in the training dataset. Through the problem statement, we will discuss why this research is important. In the next section, we are arranging the whole problem scenario into multiple points.

3.2 Problem Statement

Trying to construct a data set from thousand of unstructured data from law report is hard. Also without proper guidance, there is a chance of imputing wrong data into the dataset. In next few sections, we will discuss the various difficulty of solving a pending case into sections.

3.2.1 Statistics about the Increasing Number of Unsettled Cases

According to the Bangladesh Supreme Court, there are over 1383591 civil and 1784860 criminal and 86049 others cases pending, which leads to a total of around 33 lakhs pending cases [12].

Table 3.1: Number of Ongoing Criminal and Civil Cases by the Divisions of Bangladesh [12].

Division	No. of ongoing cases
Dhaka	11,02,710
Chattogram	5,24,563
Rajshahi	3,40,064
Khulna	3,63,730
Barishal	1,64,326
Sylhet	1,41,192
Rangpur	2,25,600
Total	28,61,185

Table 3.2: Comparison between the Number of Filed Cases and the Number of Settled Cases [12].

Type of case	No. of cases	No. of cases settled
Family case	66170	6735
Loan default	30614	1847
property claims	90275	4132
Land survey	2,80,419	3986
Women and children abuse	1,65,400	7976
Special tribunal	92801	1416
Narcotics	1,25,390	2313

The percentage of unsolved cases is above 90% in most of the instances, as new pending cases outnumber the settled cases. If we look at the above statistics, we will understand that women and child are becoming more vulnerable in our society and the number of settled cases is low. So, we like to work in this area as an example for our system. As the number of unsettled cases increases day by day, criminals have become more confident, and it will substantially increase their tendency to commit more crime.

3.2.2 Causes behind the Slow Deliberation of the Cases

The shortage of required workforce, lack of infrastructural facilities, deferment of case hearing and taking a long time for deliberation are the main reasons behind this pile of cases. In 2017, the number of judges in the High Court was 90, and the number of active judges in the lower court was 1400 only [13]. In between 2008 to 2015, pending cases increased by 75% [13]. A UNDP forecast has claimed that Bangladesh may scale up to the 5 million unresolved cases by 2020 [13]. If this problem does not settle early, it will be an unbearable burden for the judicial system that will make it stand still. This situation will deter people from coming to the court. They may find out other ways such as money or muscle power to solve their problems. These extra judicial means may enhance the cost of justice and even may become the cause of occurring more severe crimes.

To get rid of the case logjam and for the deliberation of the justice, a legitimate atmosphere is required. So, we propose a method that will expedite to dispose of unsettled cases. The appointment of a computer or an artificial agents as an assistant to a judge for providing logistical support will assure the smooth settlement of the cases. It will favor the judicial system to lessen the number of unresolved cases.

3.2.3 Influences of Human biases on the decision-making process

Substantial evidence from the previous act of the artificial intelligent agents in different complex scenarios proofs that machine learning models provide better predictions than human because the several factors can bias human's judgment. Past experiences in life (Juliusson, Karls-son, & Garling, 2005) and different cognitive biases (Stanovich & West,

2008) can change the decision [14]. Results from the previous decisions can influence cognitive biases. Even age, socioeconomic status may influence in taking an imperfect decision [14]. Inaccurate judgment and false logic can be generated due to memory error and changes in thinking patterns (Evans, Barston, & Pollard, 1983; West, Toplak & Stanovich, 2008). As cognitive function may decline along with age, older people may make a wrong decision by becoming overconfident and forget to apply necessary strategies (de Bruin et al., 2007). Even fatigue can change the decision.

3.2.4 Issues with Machine Prediction

At first we will cite the acceptance issue. General people may not admit an artificial agent as their judge. Many ethical and legal questions may arise. Secondly, the initial model does not have enough data to deliver an excellent prediction. In fact, it might generate an incorrect prediction for exceptional cases without human supervision. Even from the ethical perspective, 1% error will be considered as a huge mistake. The original notion of jurisprudence is that the ten criminals may let go free, but no innocent should be punished [15].

3.3 Proposed Work

Although many factors influence human decisions, the human is still better than the machine in some scenarios because of its ordinary sense and the power of using unstructured information. Therefore, in this paper, we argue that combining artificial intelligence agent and judge as a human will procreate a better justice system. We will briefly review previous works related to this field. Then we will develop a sample system where both human and machine will participate in the decision-making process. To complement and reduce the flaws of each other in the decision-making process, we will combine both human and agent to reproduce a fair judicial system. It will undoubtedly minimize the time required for judging an offense and reduce human biases in the decision-making process. As we will integrate human in this system, it will be more reliable to solve the fuzzy and inscrutable situation.

3.4 Conclusion

In this research, we can try to make a machine learning system where human and machine can participate together and reduce the decision-making time. Although we can also see that there are some limitations of the machine and one of the limitation is making a wrong decision. These problems can be overcome by establishing a connection with a data scientist for labeling data when a machine shows less confidence. But it is costly and if we want to train people to understand the labeling, we will need to spend enough time. So with the next few chapters, we will discuss the possibility of the improvement of machine prediction by adding the judge's decision. Moreover, this method will also reduce the cost of the data labeling.

Chapter 4

Methodology

4.1 Introduction

In this section we will fully describe our methodology as well as different algorithms for our model. Also we will try to include proper structure of the model. To test our hypothesis, we consider three situations. First, we will gather data from law experts on several crime scenes. Secondly, we will develop a system that will predict punishment using a machine learning algorithm. Thirdly, we will engage human or judge in our system by offering an opportunity to give feedback on machine-generated prediction. If machine prediction is similar to a judge's decision on the same crime scene, then the judge does not need to change machine prediction. Whereas machine prediction is different from the judge's decision on the same plot and if the judge evaluates that machine's prediction is not acceptable then the judge can give his feedback in the system. This system will store every feedback and gradually learn more. By using the judge's feedback, this system will be more efficient day by day.

4.2 Steps throughout the development of the system

Here is the process to develop the model with machine. We implemented machine learning models for machine predictions, We used 'Python' as our developing tool. Through the next few section we will show our developing process with proper details of the algorithm.

4.2.1 Dataset Development

In the first step, we face a problem that we do not have any suitable dataset for punishment prediction. For this reason, we consider developing a dataset with the help of the law experts. For constructing our dataset, we envisage cases on women and child abuse. We focus on cases re-lated to 'Nari o Shishu Nirjatan Daman Ain, 2000 (Women and Children Repression Prevention Act, 2000)' formulated by 'Ministry of LAW, Bangladesh' which is authenticated and practiced by Judicial Court of Bangladesh. Judging an offender depends on many different parameters. Various factors can increase the punishment of the offenders. Same as the increasing parameters, there are also some parameters that can decrease punishment. For instance, we consider the age of an offender. If the offender's age is below 16, he or she will be considered as a juvenile, and his or her punishment will be minimized usually. However, exceptions can occur. If the juvenile's mental state reach to the level of the adult's mental state and the crime that he or she has performed is considered as fulsome for the humankind, then the judge can take a drastic decision and can increase his or her punishment. Similar to this situation, many different or exceptional cases can arise. It might be tough for a machine to judge an exceptional case. So, if we want that machine will be able to predict like a judge or at least near to a judge in regular cases, then we have to consider some factors for developing our initial dataset.

For simplicity, we consider five parameters for judging a case. These are the age of the offender, offense against the offender, mental state of the offender while committing the crime, the level of brutality towards the victim and the victim condition. These all are the categorical variable. Again, we mention that more parameters can be added in the future for better prediction. As a first time approach, we keep it simple.

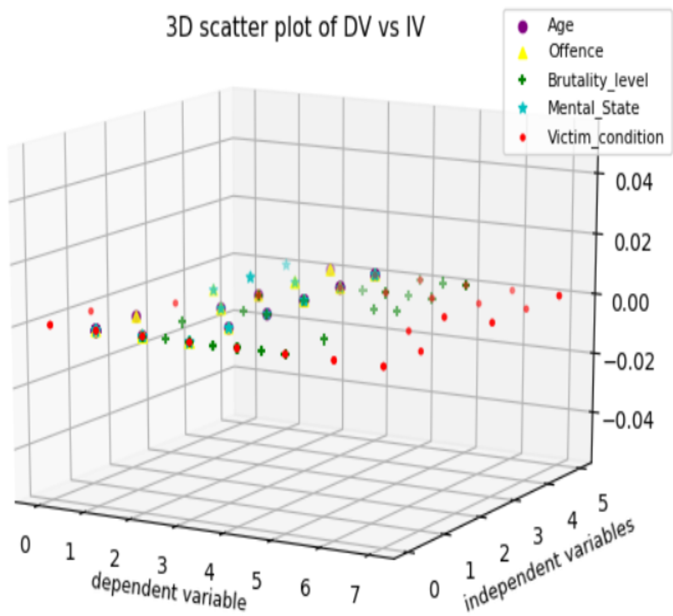

Figure 4.1: 3D scatter plot of Dependent Variable (DV) vs Independent Variable (IV)

4.2.2 Model Selection

choosing a machine learning algorithm for a dataset depends on several factors like size, quality, and nature of the data. Moreover, it is equally important for model selection that what we want to do with the data. It is tough even for the experienced researchers to tell which machine learning algorithm will work best on a specific dataset without trying them all. As we want to train our model based on examples, we are supposed to use supervised learning algorithms. Supervised learning algorithms use labeled data to find patterns on the training dataset. Among the different supervised learning models, we need to choose the one which will fulfill our purpose. We can use classification algorithms like Naïve Bayes Classifier because they predict a category like email spam, sentiment analysis, document categorization, financial fraud, etc. The classification algorithms are adapted to predict a class instead of predicting a real number. Different tree based algorithm like a decision tree, random forest usually depends on the wisdom of the crowd. These models are fast to train but can be slow during the time of making a prediction. Decision tree algorithm learns by recurrently hierarchically splitting the dataset for maximizing information gain. It learns a non-linear relationship.

Moreover, Random forest subdivides dataset and variables randomly to predict each subset. Then it makes the final prediction based on each prediction from the subsets. It is good when the dataset has many features and many observations. Support vector machine filters data into categories, when a new value is assigned it places the value in one of those categories. It does not directly provide probability estimates. SVM goods at higher dimensional space and provides a clear margin of separation. Now we have another option, and that is a regression. Regression technique is useful for solving the cause-effect type of relations. Linear regression is used when one needs to determine the value of the dependent variable based on independent variables.

Logistic regression acts like a binary classification problem. For example, if a model need to predict the presence of cancer within a patient, then we might need to use logistic regression. However, if we want to forecast how many years the patient will live, then we need to use linear regression. As we need to predict the punishment (dependent variable) based on some independent variables (like age, offense, victim condition etc.), we can also use multiple linear regression. We cannot use simple linear regression because the dataset has more than one independent variable. It seems like how can we apply multiple linear regression whereas dependent variable “punishment” is not numeric in the dataset, rather it is a string. We will encode it to a range of numeric values, where higher the value of the variable will indicate more severe punishment.

Hence, without applying several supervised machine learning algorithm, we can’t specify the well suited one for the proposed system. For simplicity, we will try multiple linear regression first where a threshold value will be used for predicting target value or class. After that, we will also apply and evaluate others supervised classification algorithms and identify the right model for our system.

4.2.3 Preprocessing of Dataset

For applying machine learning algorithm to make the prediction, we have to train the model first with the dataset. So that the model can predict rationally on a new testset. However, many machine learning models cannot process categorical data directly. As they are based on mathematical functions, so it is not desirable to apply calculation on the categorical data. For quantifying, developers need to use LabelEncoder and OneHotEncoder in python.

4.2.3.1 LabelEncoding and OneHotEncoding

LabelEncoder labels categorical variables. For example, It will label the value of the categorical variable "student belongings"- [pen, pencil, pen, eraser, pencil, pen] as [0,1,0,2,1,0]. It imposes "ordinality" automatically. It means that the average of pen and eraser is pencil, which is not desirable for this case. This happens because the label encoder does not understand the weight of the input value for each categorical variables. If we use this, the computer will think that data with value 1 is greater than value 0. Here, the variable "student belongings" is not ordinal that means that we cannot rank the values. So, only using LableEncoder will lead to false prediction. We need to use dummy variables after applying LableEncoder. OneHotEncoding is used for creating dummy variables. For example, the original dataset is in table 4.1. After applying label encoder on student dataset, it is converted as table 4.2 . Now, we will create dummy variables in table 4.3.

Table 4.1: Student's Id and their belongings

Student id	Student belongings
10	pen
12	pencil
17	pen
20	eraser
22	pencil

Table 4.2: Applying Label Encoding on Student Dataset

Student id	Student belongings
10	0
12	1
17	0
20	2
22	1

Table 4.3: Student Dataset after Creating Dummy Variables

Pen	Pencil	Eraser	Student id
1	0	0	10
0	1	0	12
1	0	0	17
0	0	1	20
0	1	0	22

Here, observation 1 indicates that student with id 10 has a pen. Observation 5 indicates that student with id 22 has a pencil. Using dummy variables we can represent the categorical variable with only 0 and 1. However, we can optimize this more by removing the dummy variable trap. Dummy variable trap happens when the dataset has more than one dummy variables which are related to each other. In the above case, it is clear that if it is not pencil and eraser that it must be a pen. So we can cut any of the three newly generated columns for removing dummy variable trap. Suppose, we remove pen column. Still, all the possible observations can be expressed by the rest of the columns. After removing the dummy variable trap, dataset will be like table 4.4.

Table 4.4: Student Dataset After removing the Dummy Variable Trap

Pencil	Eraser	Student id
0	0	10
1	0	12
0	0	17
0	1	20
1	0	22

4.2.3.2 Map Categorical Variable to Numeric

However, dummy variables can be used for non-ordinal variables. In our proposed system, all of our categorical variables are ordinal. Because we can rank age, offense, mental state and brutality levels. So, we do not need to use a dummy variable. We only need to convert text into the numerical value which will represent the weight of each unique value of the variables for predicting punishment. With the help of the law experts, we assign an appropriate weight for each unique value of the variables. We rank the value of the categorical variables according to their influence on the punishment. For doing this, we use a python dictionary. Then we use map function from pandas library which map values using input correspondence. The whole dataset will be mapped when the program will execute. Therefore, we codify it in a way that the textual data will be converted to the numerical values. Now our dataset is ready for applying machine learning algorithm.

4.2.4 Applying Several Supervised Machine learning Algorithms

We apply several supervised machine learning algorithms on our current dataset to find the right one for our system. First we use multiple linear regression. We also use Support Vector Machine, Logistic Regression, Naïve Bayes classifier algorithm. We also apply artificial neural network for both as regression and classification model.

4.2.4.1 Multiple linear regression:

Multiple linear regression (MLR) is the modified version of simple linear regression that is usually practiced on dataset having more than one independent variables . Simple linear regression is like a one-to-one relationship. One is the independent variable, and another one is a dependent variable. Multiple linear regression is like the many-to-one relationship. It shows the relation-ship between one dependent and multiple independent variables. Explanation of Simple and Multiple Linear regression respectively using diagram 4.2 and 4.3.

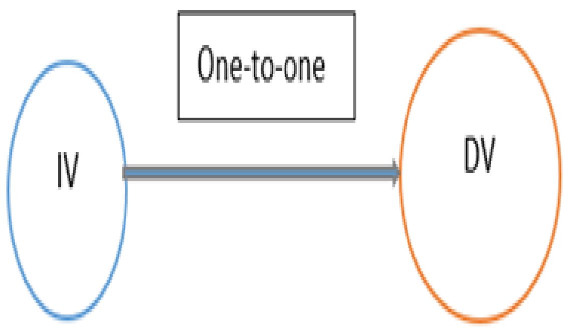

Figure 4.2: One-to-one relationship between independent and dependent variable in simple linear regression

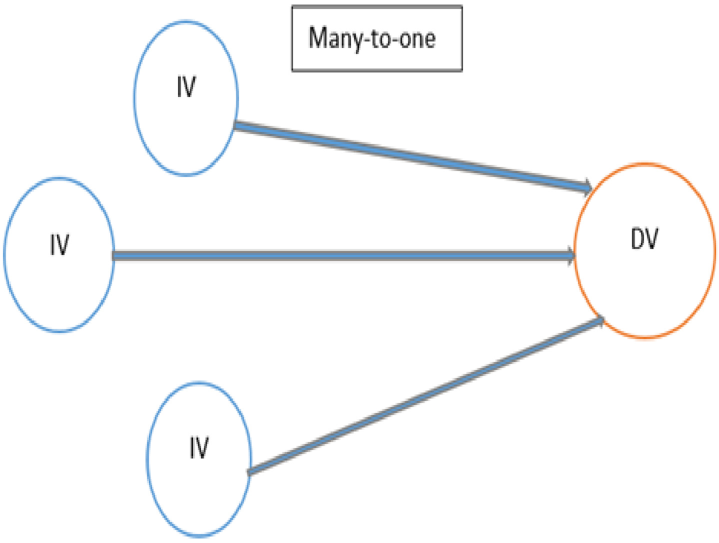

Figure 4.3: Many-to-one relationship between independent variables and dependent variable in multiple linear regression

Multiple regression model,

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n + \epsilon \tag{4.1}$$

Where, $y = dependent\ \ variable$

$\beta_0 = intercept$

$\beta_(1 \dots n) = coefficient$

$x_(1 \dots n) = independent\ \ variables$

$\epsilon = error$

In multiple linear regression, error term is assumed to be zero. So, multiple regression equation will be,

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n x_n \tag{4.2}$$

Each coefficient indicates how the target variable (y) will change due to one unit of change in the predictor variable when all other predictor variables assumed to be constant. Before applying multiple linear regression, overfitting and multicollinearity problems need to be taken into consideration.

Overfitting:

Overfitted model fits all points during training. If a model works better on the training set than the test set, then the model might have overfitting problem. For ex-ample, a model might show 96% accuracy on the training set but only 70% on the test set due to overfitting problem. To prevent overfitting problem several techniques like increasing the training set, regularization, cross-validation, early stopping, feature re-moving etc. are generally used by the researchers.

Multicollinearity:

It cannot be guaranteed that independent variables are only correlated with the dependent variable. Independent variables can have relations with other independent variables. This is called multicollinearity. If we increase the number of independent variables, it will also increase the chance of having multicollinearity.

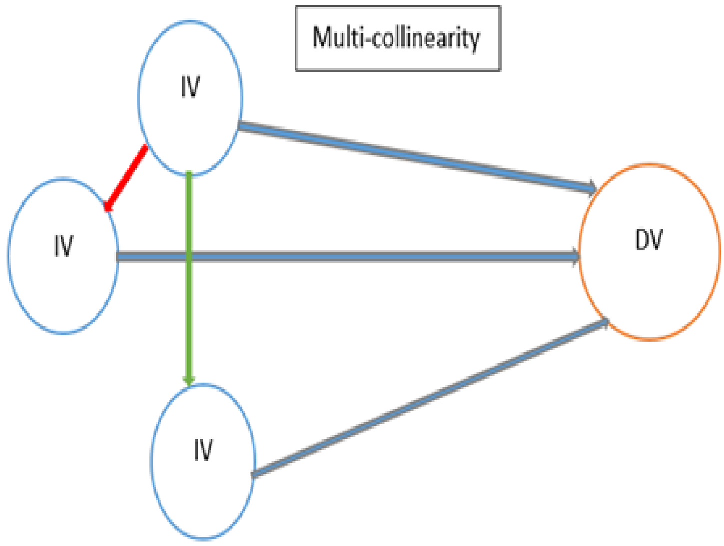

Figure 4.4: Multi-collinearity problem

The ideal situation is that the independent variables will only correlate with the dependent variable not others independent variable.

We use ordinary least squares linear regression method. In a linear regression model, ordinary least-squares estimates the unknown parameters of a function which has a set of explanatory variables. It uses least squares principles. Least-squares is a method to approximate the solution of overdetermined systems. A system is overdetermined when there are more equations than the unknowns. Least squares ensures that the sum of squares of residual will be minimized. For finding the minimum of the sum of squares, the gradient is set to zero. If our model contains n parameters, there will be n gradient equations. Gradient takes the place of the derivative for the functions which have several variables, but it is a vector-valued function, not a scalar-valued.

4.2.4.2 Naïve Bayes Classifier:

Naïve Bayes (NB) is widely used for text classification problems. It provides great result in natural language processing. Although our original dataset contains text, it seems like we can use Naïve Bayes. But actually, each unique value of each independent variable indicates significant weights that has an influence on punishment prediction. We use Gaussian Naïve Bayes algorithm for classification. In Gaussian Naïve Bayes algorithm it is assumed that the features follow a Gaussian distribution. For example, if some features have bigger values compare to other features, it will cause serious difficulties in classification. For this one need to normalize the data. Naïve Bayes can perform badly if there is a lack of independence of the explanatory variables to identify the correct target class. So, if independent variables in the dataset are correlated with each other Naïve Bayes will perform poorly. Naïve Bayes works well when the attributes don't affect each other likelihood. Naïve Bayes algorithm assumes that all the features are not related to each other. That means the presence or absence of an attributes doesn't influence the presence or absence of other attributes. It doesn't learn the relationship between features.

4.2.4.3 Logistic regression:

Logistic regression (LR) is also a classification algorithm. It doesn't provide probabilities. It is used on the problem like email classification to check whether an email is spam or not. If we consider each unique value of the target variable as a unique class, then we have to use the one-vs-rest method of logistic classifiers. It actually defines the problem as several binary classifiers. If we have n classes, then n separate logistic regression classifier is required for the model, then the probability of each class is predicted over the rest of the class combined. In multiclass case, we use one-vs-all method which fits a binary problem for each label. Model will choose the class which has the higher probability. It uses maximum likely hood approach for training the model which will help to best fit the data. But logistic regression is sensitive to outlier as it diverges to its loss very quickly. In our logistic regression model, we use L2 regularization technique to reduce the overfitting problem. Overfitting occurs when our model performs better in training but performs badly in the testing session. Regularization technique can reduce overfitting without bringing any change in the dataset. L2 regularization technique basically adds an extra term to the cost function for minimizing error. It is also known as weight decay because it prefers to learn small weights and minimize the cost function.

4.2.4.4 Support Vector Machine:

Support Vector Machine (SVM) is mostly used in classification problems. It plots each data point in N-dimensional space and tries to segregate the classes in a best possible way. N is the number of features in the dataset. It takes margins that maximize the distance between each nearest class. If the data points have N-dimension, SVM separate data point (N-1) dimensional hyperplane. SVM ignores outliers for finding the right margin. SVM performs better for binary classification problem. For multiclass classification problem, it is used like a set of binary classification problem. It is good for small to medium dataset. For larger dataset, training becomes slow in SVM models. It performs badly when the dataset has more noise. Because target classes become overlapped due to the noise. SVM is useful for the dataset with the higher dimension that means the number of features are very large.

4.2.4.5 ANN for Regression:

We also develop a Artificial Neural Network (ANN) model for regression. We build a simple model that has a fully connected hidden layer with 20 neurons. In the input layer, we have five neuron which is same as the number of features or independent variable. We use a rectifier activation function in the hidden layer. As it will solve a regression problem, we don't use any activation function in the output layer. The cost function mean squared error is optimized using "ADAM" optimization algorithm. Hyperparameters like epochs are set to 100 and batch size is set to five. Then we tune the topology of the neural network twice. One by increasing the number of the hidden layer which is represented as a deeper network and another one is represented as a wider network by increasing the number of neuron in the existing hidden layer. We have done this to check which neural network is better. When we add one extra hidden layer to our model with 15 neurons which is the half the number of neurons compares to the first hidden layer; we didn't see any significant change in our model accuracy. But when we make our model wider by adding more neurons in the first hidden layer, its accuracy increases. But adding more than 30 neurons didn't bring any change in the accuracy rate.

4.2.4.6 ANN for Classification:

Further, we also develop a neural network architecture for multiclass classification problem. But it is important to reshape the output attribute when using the neural network for classification problem. That means we need to create dummy variables for the target variable which is a one-hot encoded binary matrix. We create a simple fully connected neural network with one hidden layer which has 30 neurons in the hidden layer. We use a rectifier activation function in the hidden layer. As we create dummy variables for the target variable and we have 6 class, so the output layer must have 6 neurons. This output layer generates 6 output values, one for each class. The largest output value among the 6 class will be considered as the predicted class by the model. In the output layer, we use the softmax activation function to ensure that the output values will be in the range of 0 and 1 and will be used as predicted probabilities. Softmax function predicts the probabilities of each class over all other possible target classes. Sum of the probability of all class will be equal to 1.

$$softmax = \frac{exp^(inputs)}{(sum(exp^(inputs)))} \tag{4.3}$$

$$F(x_i) = \frac{exp(x_i)}{\sum_{j=0}^{k} exp(x_j)} \quad i = 0, 1, 2, \ldots, k \tag{4.4}$$

We use "ADAM" gradient descent optimization algorithm for minimizing the cross entropy cost function. We set the number of epoch to 150 and batch size to 5 for training the model.

4.2.4.7 Artificial Neural Network Architecture

Figure 4.5 will be used for 6 class classification problem. This artificial neural network has 30 neurons in the hidden layer. It has only one hidden layer. Figure 4.6 will be used for 11 class classification problem. This artificial neural network has 40 neurons in the hidden layer. It also has only one hidden layer.

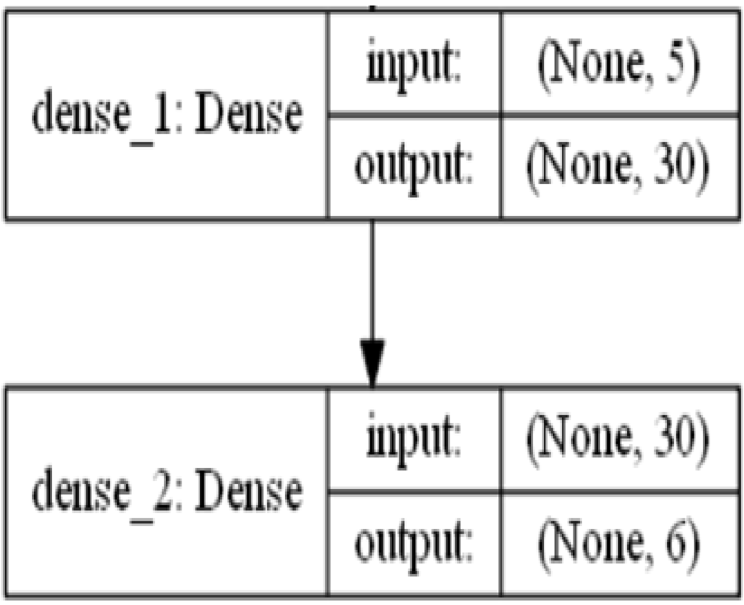

Figure 4.5: Artificial Neural Network Architecture for 6 class classification problem

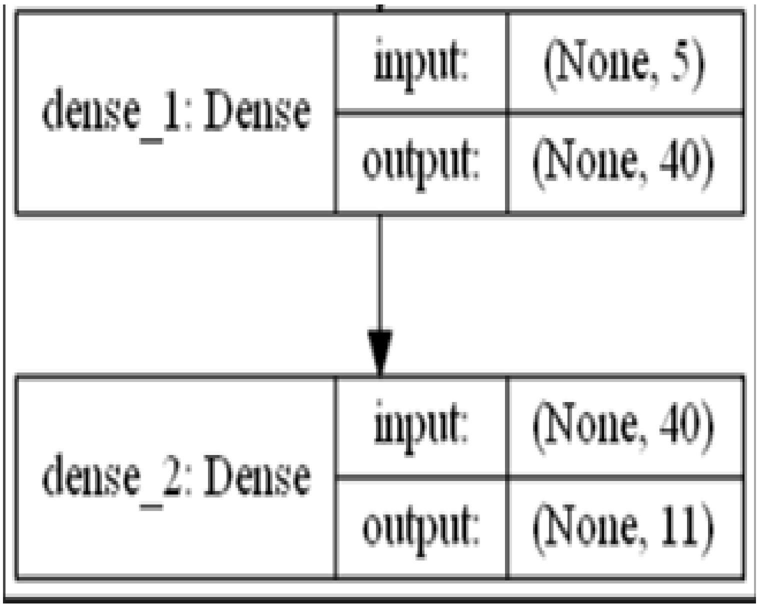

Figure 4.6: Artificial Neural Network Architecture for 11 class classification problem

Reason behind using ADAM Optimization Algorithm:

A good optimization algorithm can improve the accuracy of the model within minutes, hours or days. An improved version of the stochastic gradient descent (SDG) algorithm is the "ADAM" optimization algorithm. It has been recently used in many computer vision and natural language processing applications. It is highly beneficial to use on non-convex problems. It is computationally efficient and requires very less memory. It is well suited for problems with large dataset or parameters. It can easily handle very noisy gradients and requires little tuning for hyper-parameters. In SGD, learning rate doesn't change during training but Adam changes the learning rate for each network weight from the estimation of the first and second moments of the gradients. So, it is also known as "adaptive moment estimation". Adam combines the benefits of both Adaptive Gradient Algorithm (AdaGrad) and Root Mean Square Propagation (RMSProp). Adam outperforms these techniques when gradients become sparser [16].

4.2.5 Cost Function

A cost function is a measure of judging a model. It tells us whether the model goes wrong or not to estimate the relationship between dependent and independent variables. It compares the estimated prediction against the "ground truth". The model will find the appropriate weights that will minimize the cost function. For multiple linear regression, Mean Square Error (MSE) is the cost function and we use ordinary least squares (OLS) method to minimize this function. OLS estimates the unknown parameters. It chooses the parameter by minimizing the sum of the squares of differences between the actual and predicted value of the dependent variable.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^{'})^2 \tag{4.5}$$

$Where, y_i = actual\ \ value$

$y_i^{'} = predicted\ \ value$

$n = number\ \ of\ \ observations$

Logistic regression uses a different cost function. Logistic regression model works on the sigmoid activation function.

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}} \tag{4.6}$$

$z = \theta^T x$. $\theta' s$ are actually the weights for particular features. The goal of machine learning is to estimate θ from the given (x,y). It indicates the importance of the variable for the output.Theta transpose x,$\theta^T x$ is a product of the $1 \times n$ matrix and $n \times 1$ matrix i.e. a 1×1 matrix which is a scalar value.

If output $y = 1$, then $h_\theta(x)$ need to be close to one. It means that $\theta^T x$ must be larger than zero. Conversely, If output $y = 0$, then $h_\theta(x)$ need to be close to zero. It means that $\theta^T x$ should be less than zero.

$$For \;\; y = 1, h_\theta(x) \approx 1, \;\; \theta^T x \gg 0 \tag{4.7}$$

$$For \;\; y = 0, h_\theta(x) \approx 0, \;\; \theta^T x \ll 0 \tag{4.8}$$

So, cost function for logistic regression is $-(y \log h_\theta(x) + (1-y)\log(1-h_\theta(x)))$. For $y = 1$, only the first term will be considered and second term will be canceled out. So, for $y = 1$, cost function will be $-(y \log h_\theta(x))$. Again, when y=1,

$$-(y \log h_\theta(x)) = -(\log h_\theta(x)) \;\; = -(\log \frac{1}{1+e^{-(\theta^T x)}}) \;\; = -(\log \frac{1}{1+e^{(-z)}}) \tag{4.9}$$

Same as for $y = 0$,

$$cost \;\; function = -(\log(1 - \frac{1}{1+e^{(-z)}})) \tag{4.10}$$

For, $y = 1$, when z or $\theta^T x$ is big, cost is low. However, if z is zero or negative, cost is massive and gives an exponential curve. Same for the situation $y = 1$, when z or $\theta^T x$ is small cost is low. If z is large then cost is high which is also exponential. Cost function for logistic regression considering all values and regularization term,

$$J(\theta) = \min_\theta \frac{1}{m}[\sum_{i=1}^{m} y^{(i)}(-\log h_\theta(x^{(i)}))+(1-y^{(i)})(-\log(1-h_\theta(x^{(i)})))]+\frac{\lambda}{2m}\sum_{j=1}^{n}(\theta_j)^2 \tag{4.11}$$

For support vector machine,

$$For \;\; y = 1, \;\; \theta^T x \geq 1 \tag{4.12}$$

$$For \;\; y = 0, \;\; \theta^T x \leq -1 \tag{4.13}$$

This range builds a safety margin for SVM classification than logistic regression. For SVM, the cost function is redefined. For y=1, it is represented as $cost_1(z) \;\; or \;\; cost_1(\theta^T x)$ because when $\theta^T x$ reaches to 1 or lower numbers, cost will grow. From 1 onwards, the cost is flat. Same for y=0, it is $cost_0(z) \;\; or \;\; cost_0(\theta^T x))$. When $\theta^T x$ reaches to -1 or greater value, cost will grow. Below -1, cost is flat. This modified cost function gives a computational advantage and makes SVM as an easier optimization problem. The cost function for SVM with regularization term,

$$J(\theta) = \min_{\theta} C \sum_{i=1}^{m} [y^{(i)}(cost_1(\theta^T x^{(i)})) + (1 - y^{(i)})(cost_0(\theta^T x^{(i)}))] + \frac{1}{2} \sum_{j=1}^{n} (\theta_j)^2 \quad (4.14)$$

Logistic regression uses logistic loss which diverges faster than hinge loss used by SVM. So, it is sensitive to outliers. In logistic regression, there might be a minor degradation in accuracy as logistic loss never becomes zero even if it classified the point successfully. Hinge loss,

$$l(y) = max(0, 1 - (\hat{y} - y)) \quad (4.15)$$

$Where, \hat{y} = intended\ \ output$

We can somewhat relate neural network to the logistic regression. Logistic regression can be assumed as one layer of the neural network. Even we can use a sigmoid activation function in the hidden layer of the neural network which is used by logistic regression. One benefit of logistic regression is that the logistic cost function is convex which guaranteed to find the global minimum of the cost. When we use this logistic activation function in a multilayer neural network its convexity will be lost. However, backpropagation helps to find a powerful predictive model. Neural network uses the cross-entropy cost function. Neural network cost function is the generalization of the logistic regression cost function. For multiclass classification, it calculates a separate loss for each class label for each observation and sums the result. The cost function for the neural network,

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} \sum_{k=1}^{k} y_k^{(i)} \log(h_\theta(x^{(i)}))_k + (1 - y_k^{(i)})(\log(1 - (h_\theta(x^{(i)}))_k))] \quad (4.16)$$

Here,m= number of training data,k= number of class and $h_\theta(x)$ is k dimensional vector.

In a simpler form, the cross-entropy cost function for a single neuron can be written as,

$$C = -\frac{1}{n}\sum_{x}[y \ln a + (1-y)\ln(1-a)] \tag{4.17}$$

$Here, n = number\ \ of\ \ items\ \ in\ \ training\ \ data$

$y = desired\ \ output$

$a = actual\ \ output$

$a = \sigma(z)$

$z = \sum_j w_j x_j + b$

$w = weight$

$b = bias$

$\sum_x$indicates summation over all training data.Cross-entropy cost function for many-neuron multilayer networks will be,

$$C = -\frac{1}{n}\sum_{x}\sum_{j}[y_j \ln a_j^L + (1-y_j)\ln(1-a_j^L)] \tag{4.18}$$

$\sum_j\ \ means\ \ summing\ \ over\ \ all\ \ output\ \ neurons.$

4.2.6 Punishment Prediction Framework

Algorithm 1 *Punishment Prediction Framework*

INPUT: *Structured data set = dataset, User input=NewCase*

OUTPUT: *Predicted Punishment*

Apply Machine Learning Model to Predict Punishment

1. Load dataset, DataFrame[][]= load (dataset).
2. Set numeric label for only each unique value of the each categorical variable, EncodedValue =Set(categoricalVariable).
3. **for** each categorical variable : **do**
4.   Map(categoricalVariable,EncodedValue).
5. **end for**
6. Slice dataset into independent variables X and dependent Variable Y.
7. Create object of the machine learning model, Model=CreateModelObject().
8. Train the model, Model.Train(X,Y).
9. TestCase= input(NewCase).
10. Map the test case to the encoded value, TestX=MapTestCase (TestCase,EncodedValue).
11. Predicted the punishment, prediction= Model.predict(TestX).
12. Invert the numeric value of the predicted punishment to the categorical value, punishment= invertMap (prediction,EncodedValue).
13. Show (punishment).

Add the new case to the dataset along with the punishment

14. ResponseOfJudge= AskJudgeDecision(TestCase, punishment).
15. **if** ResponseOfJudge == "agree": **then**
16.   Add test case to the dataset, Add(TestCase, punishment).
17. **else**
18.   Add test case to the dataset, Add(TestCase, ResponseOfJudge).
19. **end if**

4.2.7 System's layout

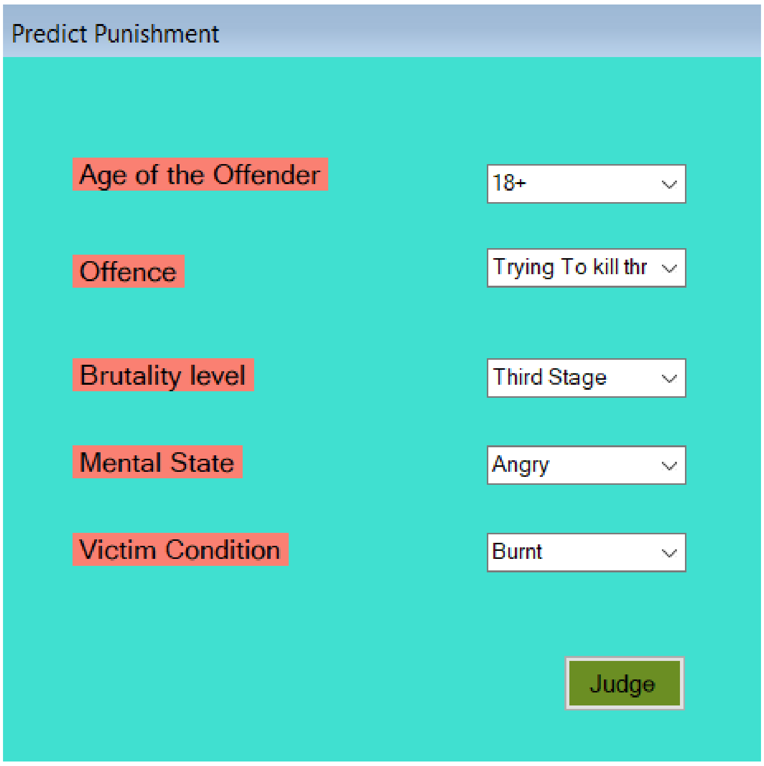

Figure 4.7: Input Layout of the system

Figure 4.8: Output Layout of the system

4.3 Conclusion

As different algorithms will have different accuracy in their own rules. It is possible to utilize a algorithm to it's highest accuracy by changing the dataset in a way that only the preferred algorithm will work good enough for making the decision. There are some algorithms where the performance grows higher with increasing data, for many other algorithms, it might not work in the same way. As we are going to make a sample software and the dataset was not used before in any other research we apply several algorithms to determine which one will be good fit for the current dataset.

Chapter 5

Result and Analysis

5.1 Introduction

In this result and analysis section, we will analyze our predictive model's efficiency in the different algorithms. Making a decision with the lowest error will always have higher priority. Through the process, we will compare both the test accuracy and predictive power of the different algorithms. Finally, we will choose the model with the highest accuracy and predictive power. Then we will check if there are any overfitting or underfitting problem associated with the model using the test and training accuracy curve.

5.2 Comparative analysis of the performance of different algorithms

First, we use test-train split method. We set the ratio of train-test split to 0.33 which means that 33% of the dataset has been used as test set and 67% has been used as training set.

5.2.1 Models with 6 class

Table 5.1: Accuracy, MSE, RMSE on the test dataset when the dataset has only 48 observations (6 class)

ML Algorithm	Accuracy on test set	MSE	RMSE
Multiple Linear Regression	81.25%	0.19	0.44
Naïve Bayes	38%	2.88	1.70
Logistic Regression	81%	0.56	0.75
Support Vector Classifier	69%	0.69	0.83
ANN Regression	37.5%	1.0	1.0
ANN Classifier	75%	0.44	0.66

Table 5.2: Accuracy, MSE, RMSE on the test dataset when the dataset has 142 observations (6 class)

ML Algorithm	Accuracy on test set	MSE	RMSE
Multiple Linear Regression	65.95%	0.66	0.81
Naïve Bayes	55%	1.34	1.16
Logistic Regression	68%	1.15	1.07
Support Vector Classifier	79%	0.53	0.73
ANN Regression	63.83%	0.53	0.73
ANN Classifier	72.34%	0.64	0.80

Table 5.3: Accuracy, MSE, RMSE on the test dataset when the dataset has 577 observations (6 class)

ML Algorithm	Accuracy on test set	MSE	RMSE
Multiple Linear Regression	66.49%	0.51	0.71
Naïve Bayes	65%	0.63	0.79
Logistic Regression	78%	0.64	0.80
Support Vector Classifier	86%	0.19	0.44
ANN Regression	83.24%	0.34	0.58
ANN Classifier	92.67%	0.28	0.53

5.2.2 Models with More Class

First, we work on three types of offense. When we add more new offenses in the dataset, it increases the number of class. After adding 5 new offenses, we get total 11 class in the target variable.

Table 5.4: Accuracy, MSE, RMSE on the test dataset when the dataset has 1291 observations and number of class is 11

ML Algorithm	Accuracy on test set	MSE	RMSE
Multiple Linear Regression	23.65%	7.36	2.71
Naïve Bayes	61%	2.38	1.54
Logistic Regression	59%	2.38	1.54
Support Vector Classifier	78%	1.16	1.08
ANN Regression	24.36%	6.47	2.54
ANN Classifier	91.10%	0.98	0.99

5.2.3 Charts Comparing Test Accuracy

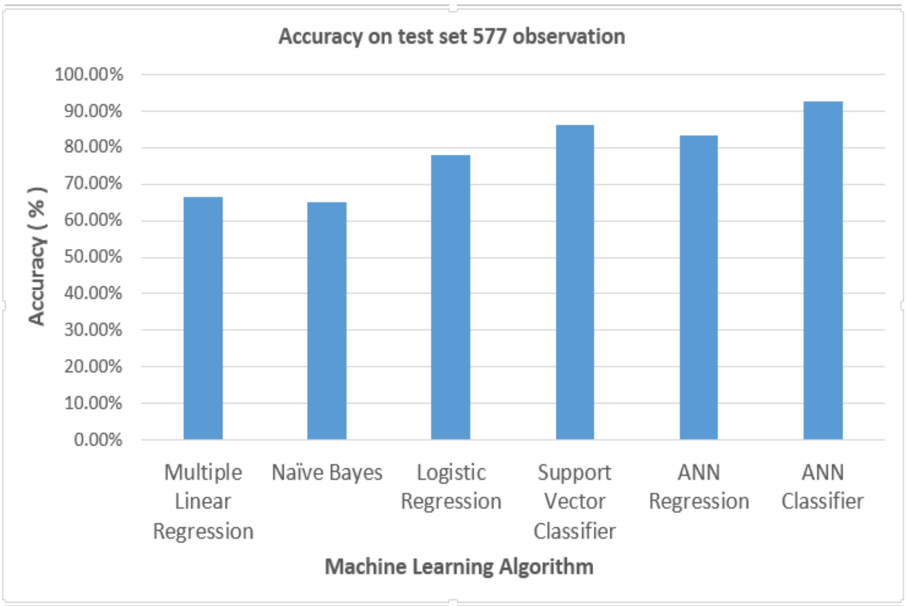

Figure 5.1: Compare Accuracy of different models on test set for 577 observations with 6 class

The next diagram is for comparing test accuracy of 142 observations and 577 observations.

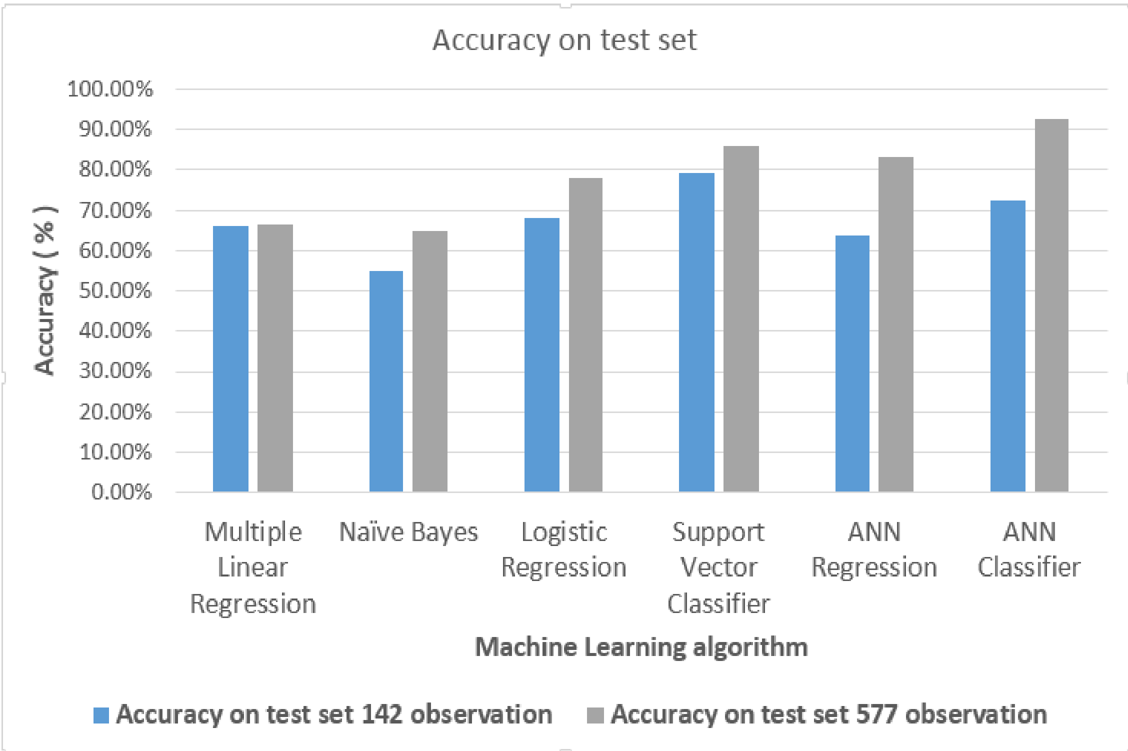

Figure 5.2: Compare Accuracy of different models on test set for 142 observations and 577 observations with 6 class

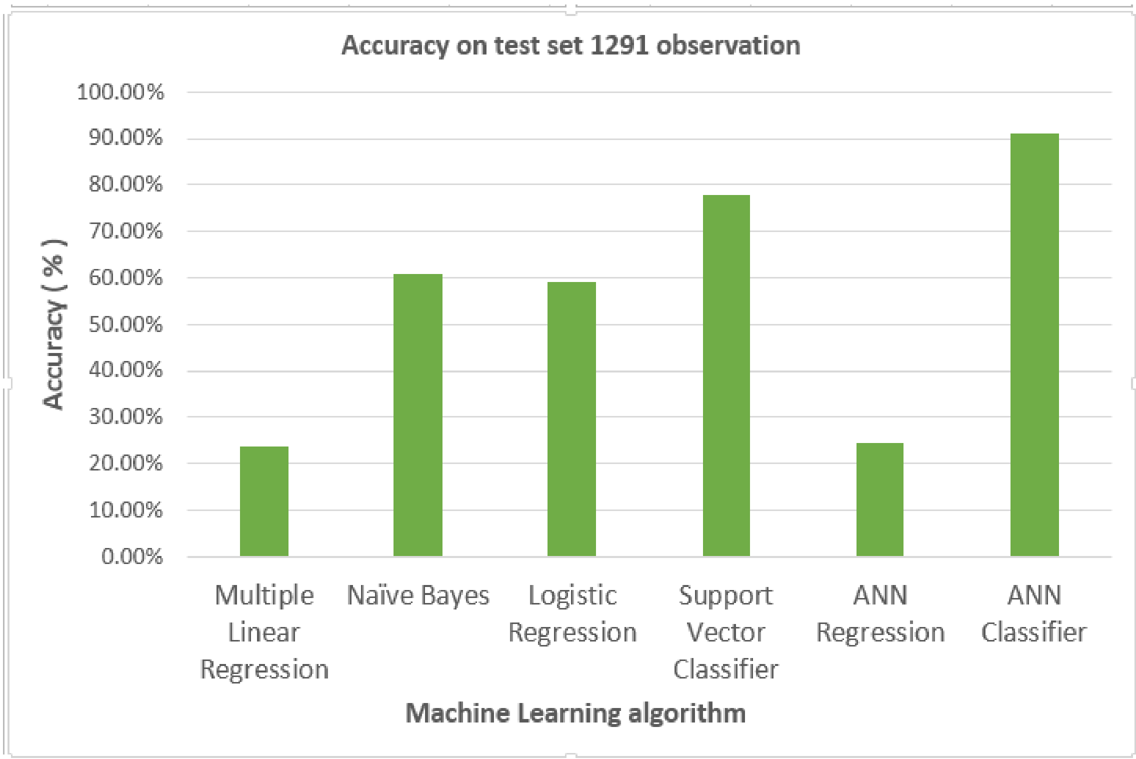

Figure 5.3: Compare Accuracy of different models on test set for 1291 observations with 11 class

5.3 Cross Validation

Cross-validation is generally used to compare and select a model from various models. Here, we use k-folds cross-validation for assessing the predictive power of our models on a new dataset. When we have a limited dataset, it actually uses resampling procedure. K is the number of groups a given dataset will be split into. We apply 10-folds cross-validation, so our dataset will be split into 10 groups. It shuffles the dataset randomly first, then splits it into 10 groups. For each unique group, it takes the group as test data and uses the rest of the nine groups as training data. 10-folds cross-validation is mostly used for its low bias and modest variance range. There is always a bias-variance trade-off for choosing the value of k.

5.3.1 Predictive Accuracy of Models for 6 Class

Table 5.5: Predictive accuracy of different models using 10-folds cross validation for 48 observations

ML Algorithm	Predictive Accuracy
Multiple Linear Regression	63.50%(+/- 30.50%)
Naïve Bayes	38%(+/- 37%)
Logistic Regression	70%(+/- 11%)
Support Vector Classifier	68%(+/- 30%)
ANN Regression	35% (+/- 28.72%)
ANN Classifier	68.50 %(+/- 10.97%)

Table 5.6: Predictive accuracy of different models using 10-folds cross validation for 142 observations

ML Algorithm	Predictive Accuracy
Multiple Linear Regression	71.05% (+/- 11.60%)
Naïve Bayes	56%(+/-8%)
Logistic Regression	76%(+/- 9%)
Support Vector Classifier	89%(+/- 10%)
ANN Regression	70.05% (+/- 12.08%)
ANN Classifier	90%(+/-9.69%))

Table 5.7: Predictive accuracy of different models using 10-folds cross validation for 577 observations

ML Algorithm	Predictive Accuracy
Multiple Linear Regression	63.58% (+/- 5.95%)
Naïve Bayes	65%(+/-8%)
Logistic Regression	76%(+/- 5%)
Support Vector Classifier	89% (+/- 5%)
ANN Regression	80.24% (+/- 4.38%))
ANN Classifier	93.94%(+/-3.08%))

5.3.2 Predictive Accuracy of Models for 11 Class

Table 5.8: Predictive accuracy of different models using 10-folds cross validation for 1291 observations

ML Algorithm	Predictive Accuracy
Multiple Linear Regression	26.11% (+/-10.33%)
Naïve Bayes	60%(+/-6%)
Logistic Regression	59%(+/- 5%)
Support Vector Classifier	76% (+/- 4%)
ANN Regression	39.97% (+/- 4.24%)
ANN Classifier	94.11% (+/- 3.22%)

After adding more classes, we observe that ANN classifier outperforms all other algorithms to a great extent. It seems that linear regression models, Naïve Bayes will be very bad fit when we have many classes. The graph of comparing predictive power of 577 observations and 1291 observations is given in Figure-5.4.

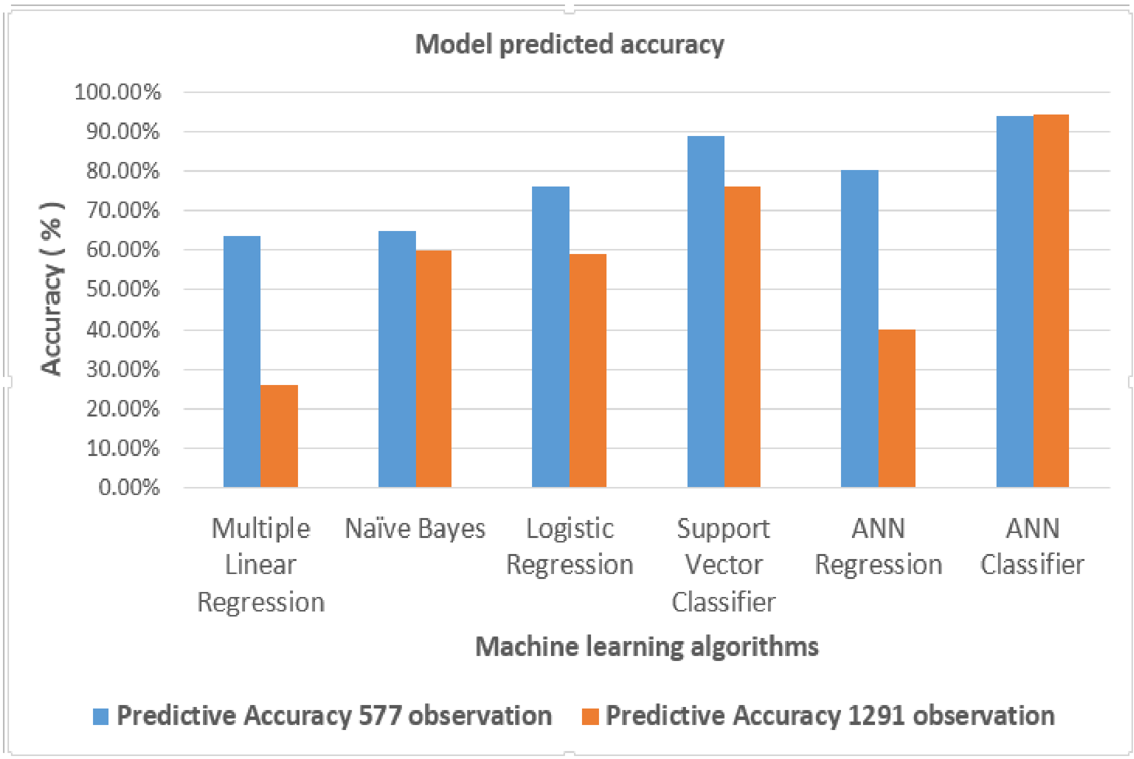

Figure 5.4: Compare Predictive Accuracy of different models for 577 observations with 6 class and 1291 observations with 11 class

5.4 Predicted Y vs Actual Y Scattered Plot

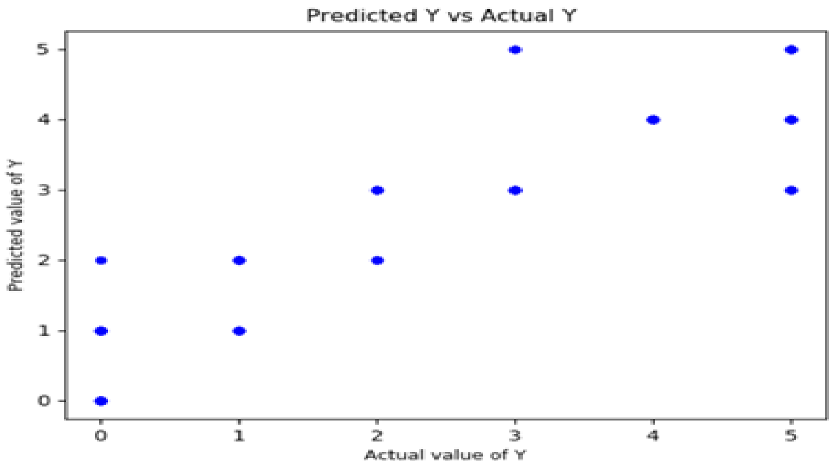

Figure 5.5: Scatter Plot of Predicted Value Vs Actual Value using Multiple Linear Regression

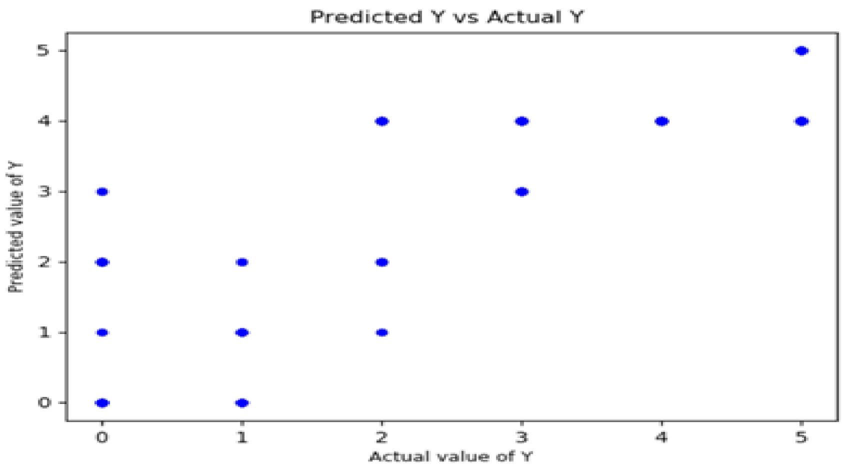

Figure 5.6: Scatter Plot of Predicted Value Vs Actual Value using Naïve Bayes Classifier

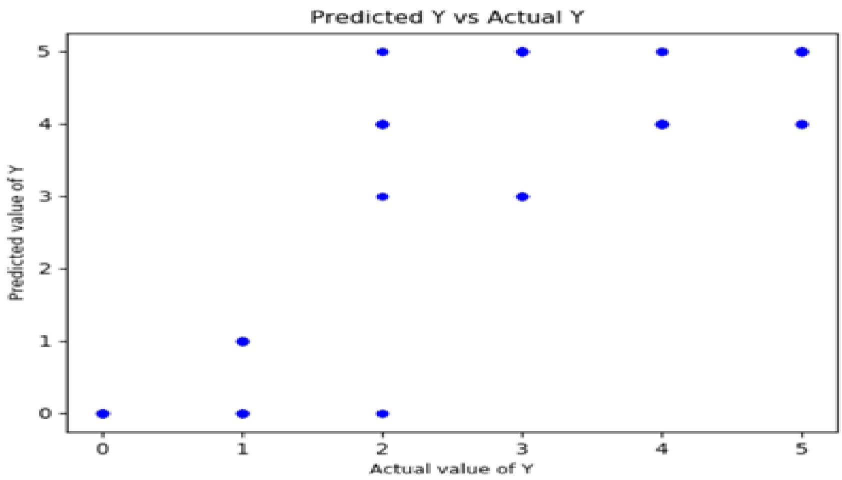

Figure 5.7: Scatter Plot of Predicted Value Vs Actual Value using Logistic Regression

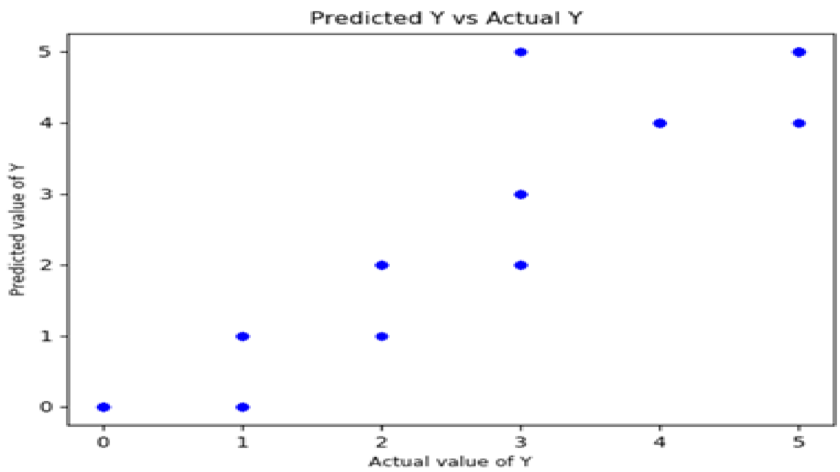

Figure 5.8: Scatter Plot of Predicted Value Vs Actual Value using SVM

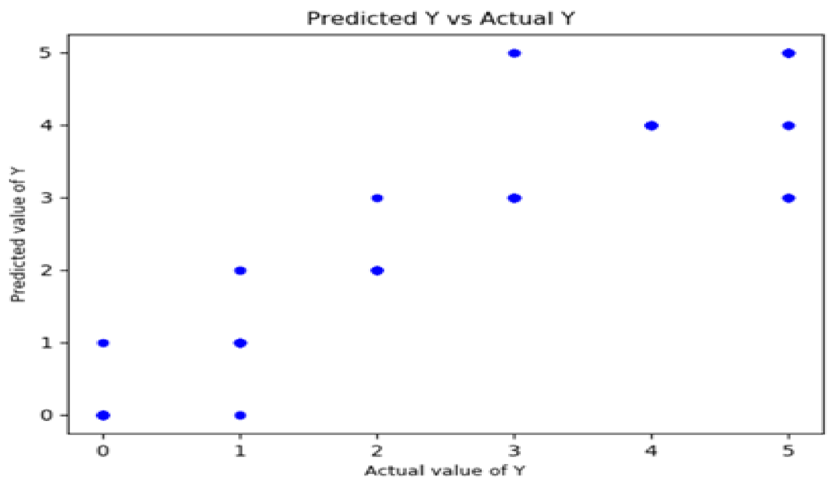

Figure 5.9: Scatter Plot of Predicted Value Vs Actual Value using ANN Regression

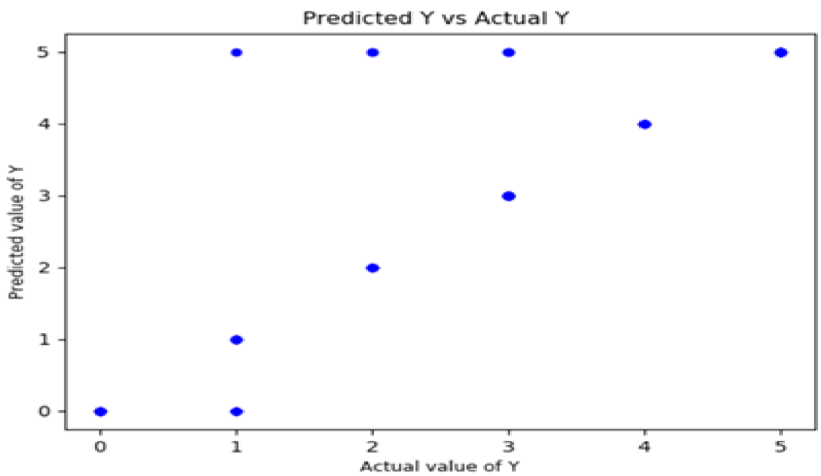

Figure 5.10: Scatter Plot of Predicted Value Vs Actual Value using ANN Classifier

From the Figure-5.5 to Figure-5.10, we observe that how different models make error in classification on different points. In figure 14, it is clearly visible that the predicted values of the target variable indicate correctly classified classes in most of the cases. Moreover, it has less off-diagonal points than the other models. Off-diagonal points indicate misclassification. Even in neural network classifier, some classes are also misclassified by the model.

5.5 Decision Region Boundary

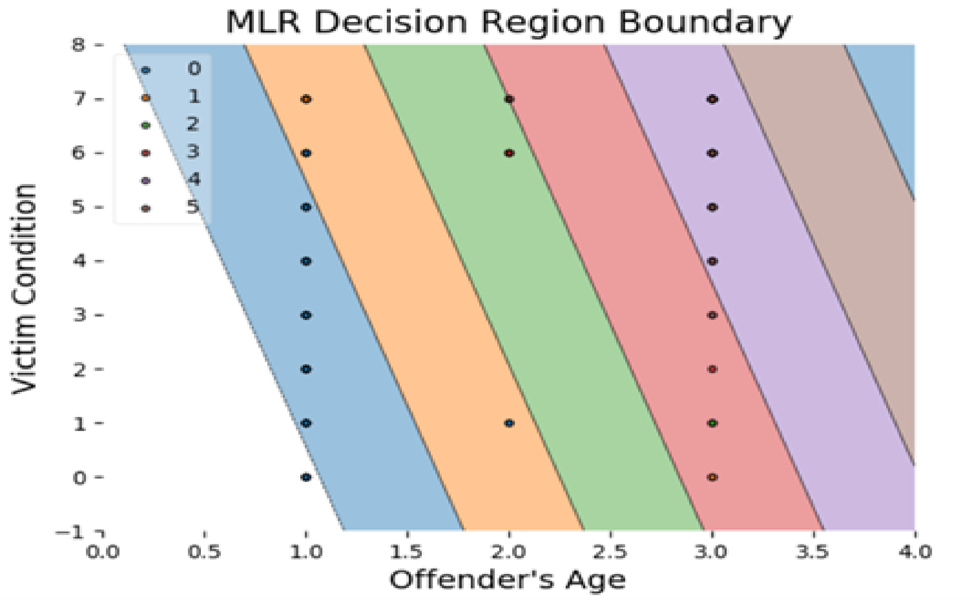

Figure 5.11: Decision Region Boundary using Multiple Linear Regression for 6 class on 142 observations

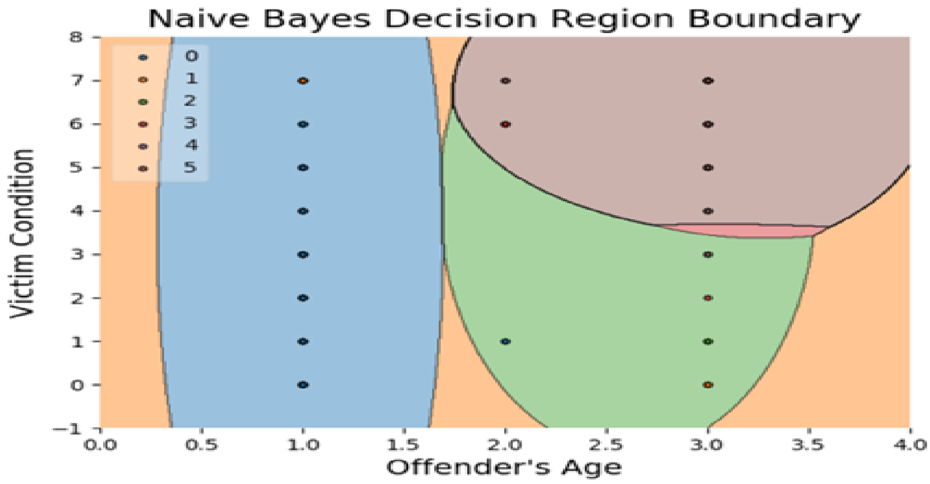

Figure 5.12: Decision Region Boundary using Naive Bayes Classifier for 6 class on 142 observations

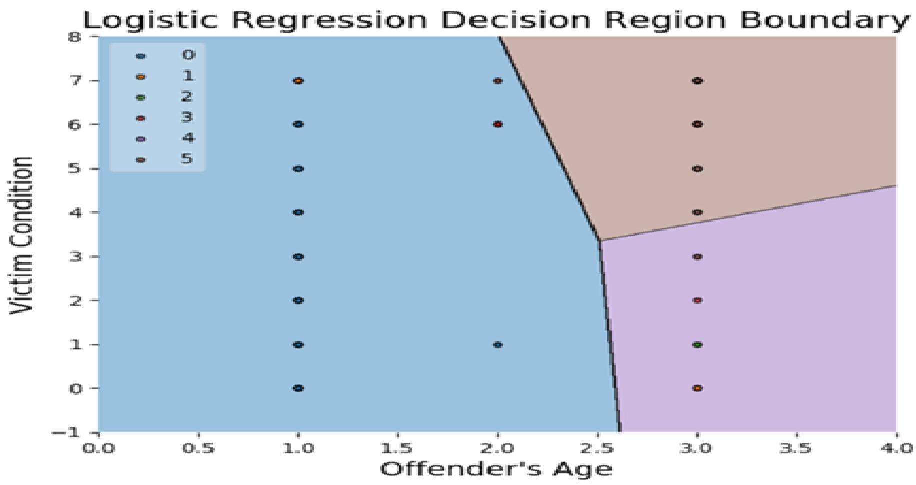

Figure 5.13: Decision Region Boundary using Logistic Regression for 6 class on 142 observations

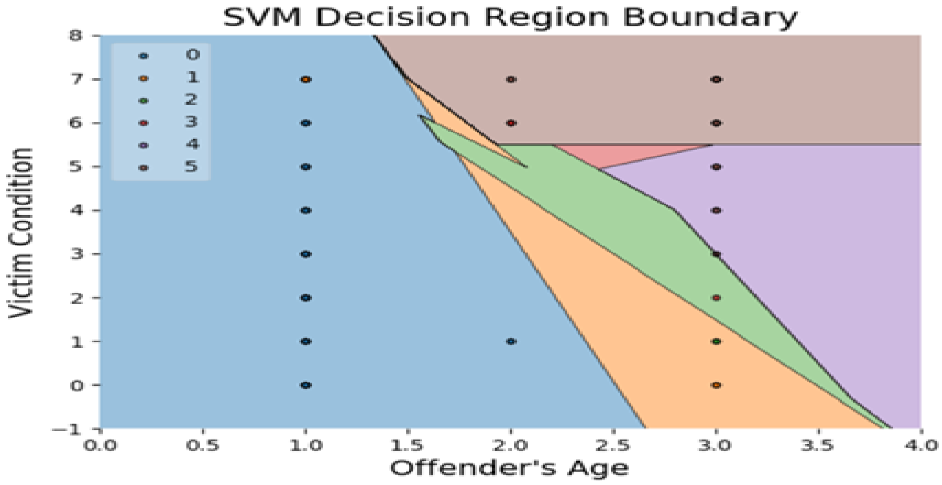

Figure 5.14: Decision Region Boundary using Support Vector Machine for 6 class on 142 observations

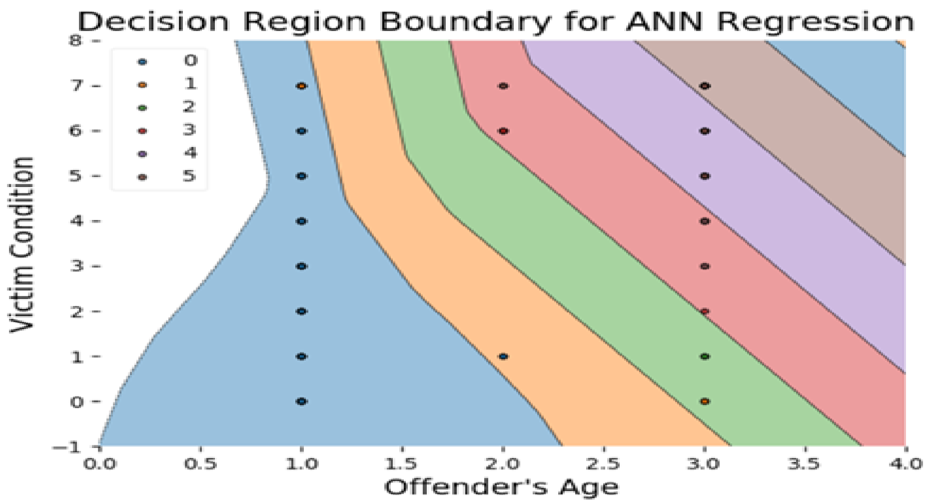

Figure 5.15: Decision Region Boundary using ANN Regression for 6 class on 142 observations

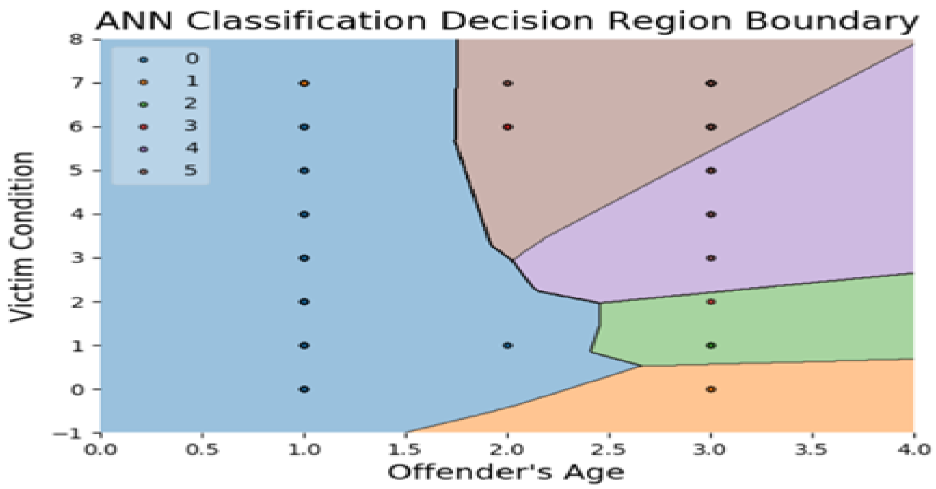

Figure 5.16: Decision Region Boundary using ANN Classifier for 6 class on 142 observations

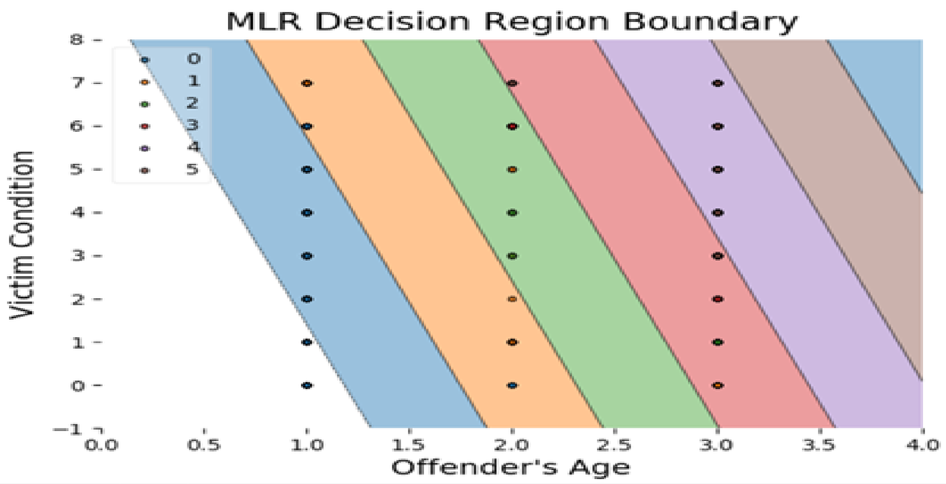

Figure 5.17: Decision Region Boundary using Multiple Linear Regression for 6 class on 577 observations

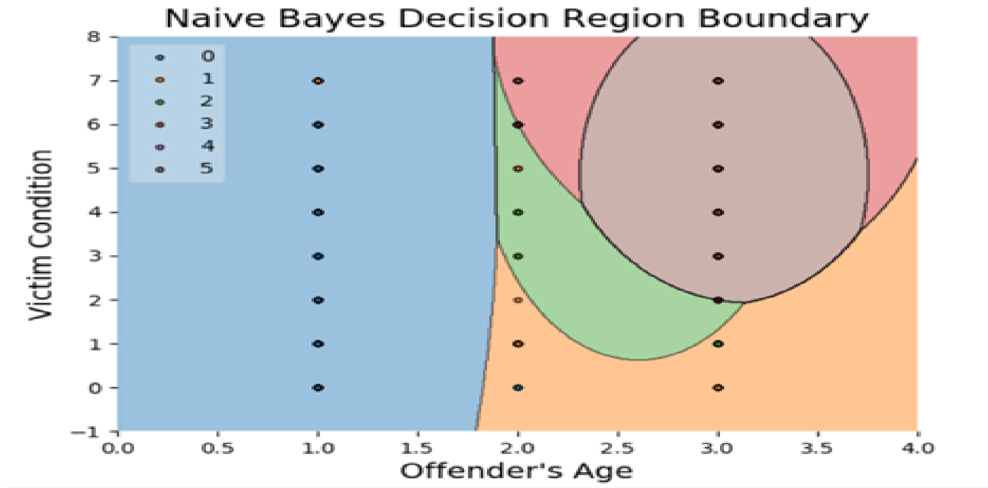

Figure 5.18: Decision Region Boundary using Naive Bayes Classifier for 6 class on 577 observations

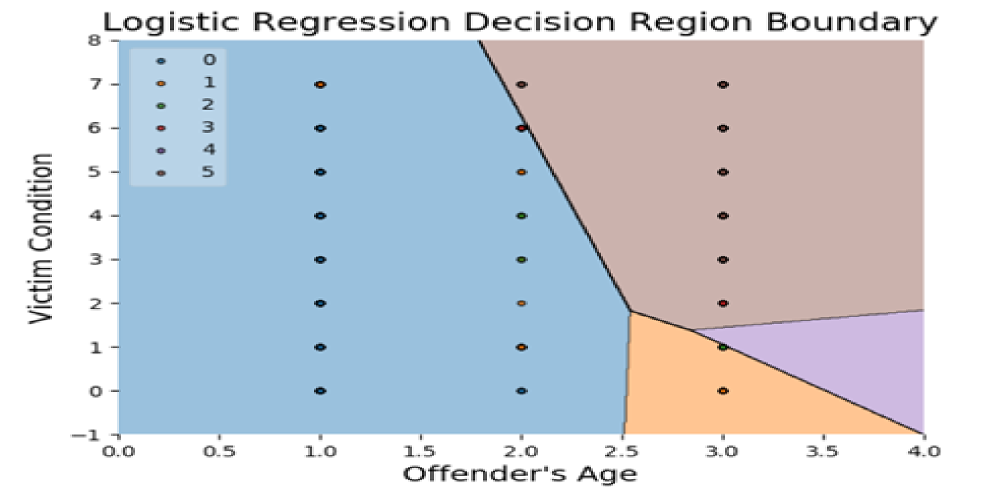

Figure 5.19: Decision Region Boundary using Logistic Regression for 6 class on 577 observations

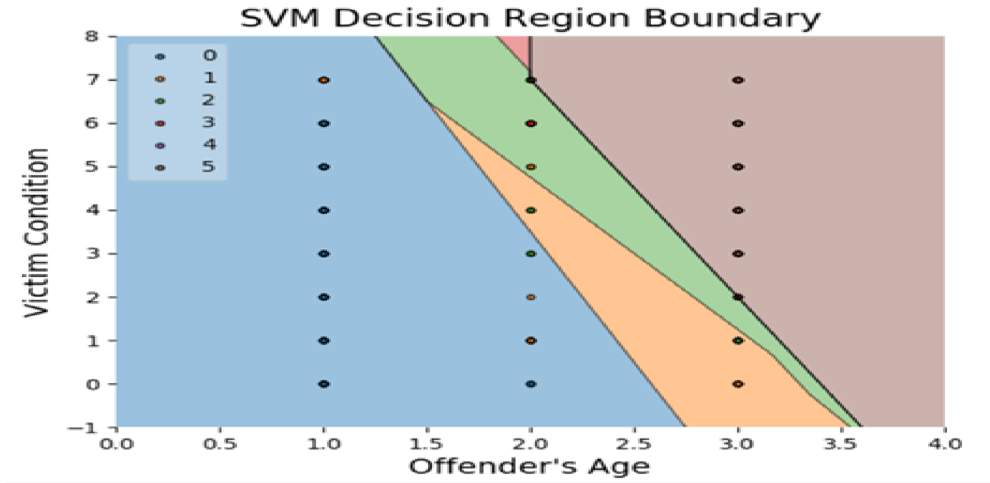

Figure 5.20: Decision Region Boundary using Support Vector Machine for 6 class on 577 observations

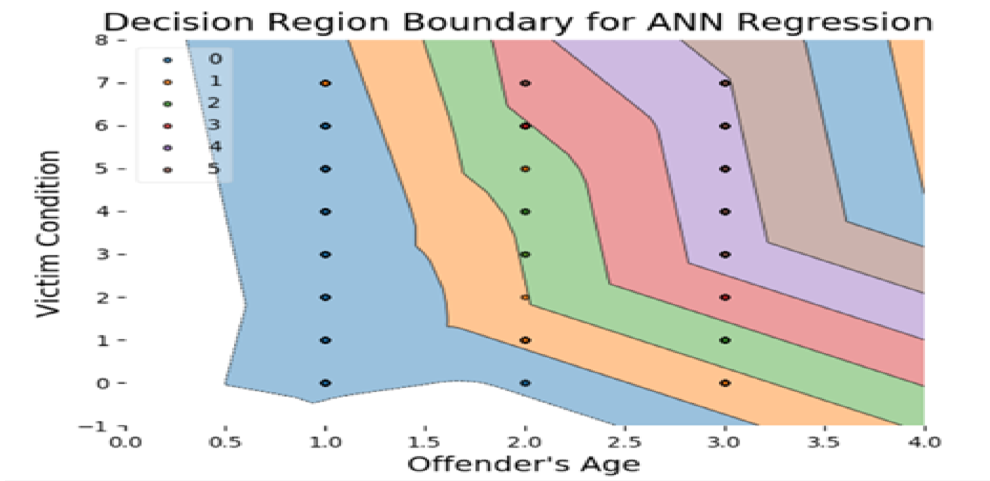

Figure 5.21: Decision Region Boundary using ANN Regression for 6 class on 577 observations

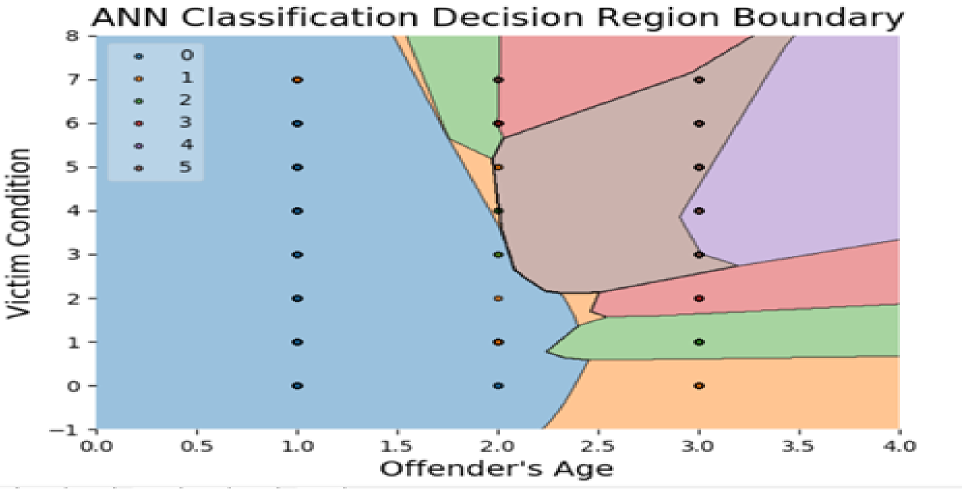

Figure 5.22: Decision Region Boundary using ANN Classifier for 6 class on 577 observations

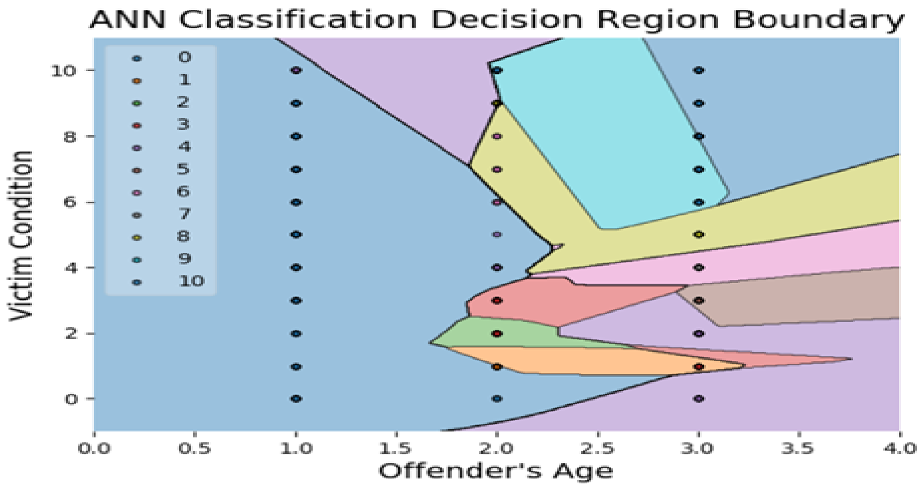

Figure 5.23: Decision Region Boundary using ANN Classifier for 11 class on 1291 observations

5.6 Evaluating Models

In this section, we will discuss how much effective our model is for prediction. Classification accuracy, confusion matrix, and classification report are widely used evaluating matric for a model. We will compare our models with these metrics. Moreover, training and test accuracy curves will be plotted together to identify overfitting or underfitting problem.

5.6.1 Classification Accuracy

Through the next table, we will try to compare our accuracy measurement with various classifier models both for 6 class and 11 class classification problem. We observe that the neural network classifier performs better on the test dataset than the other models both for 6 class and 11 class classification problem.

$$Classification\ \ Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{5.1}$$

$WhereTP = True\ \ Positive, FP = False\ \ Positive, TN = True\ \ Negative, FN = False\ \ Negative$

Table 5.9: Accuracy on the test dataset when the dataset has 577 observations (6 class)

ML Algorithm	Accuracy on test set
Multiple Linear Regression	66.49%
Naïve Bayes	65%
Logistic Regression	78%
Support Vector Classifier	86%
ANN Regression	83.24%
ANN Classifier	92.67%

Table 5.10: Accuracy on the test dataset when the dataset has 1291 observations and number of class is 11

ML Algorithm	Accuracy on test set
Multiple Linear Regression	23.65%
Naïve Bayes	61%
Logistic Regression	59%
Support Vector Classifier	78%
ANN Regression	24.36%
ANN Classifier	91.10%

5.6.2 Confusion Matrix and Classification Report

Confusion matrix is one of the easiest metrics for finding the correctness and accuracy of the model. It describes the correctness of a model on a test dataset. It is a table with two dimensions. Both of the dimensions have classes. One is for the actual class and another one is for predicted class. The diagonal elements represent the number of correctly classified classes. The off-diagonal elements are misclassified by the classifier. Our diagonal values are higher which indicates many correct predictions. From the confusion matrix, we get the true positive (TP), true negative (TN), false positive (FP), false negative (FN) rate. These values are used to create a classification report. Both from the result of test-train split and cross-validation, we observe that both SVM and neural network classifier works well. However, the neural network classifier also outperforms SVM.

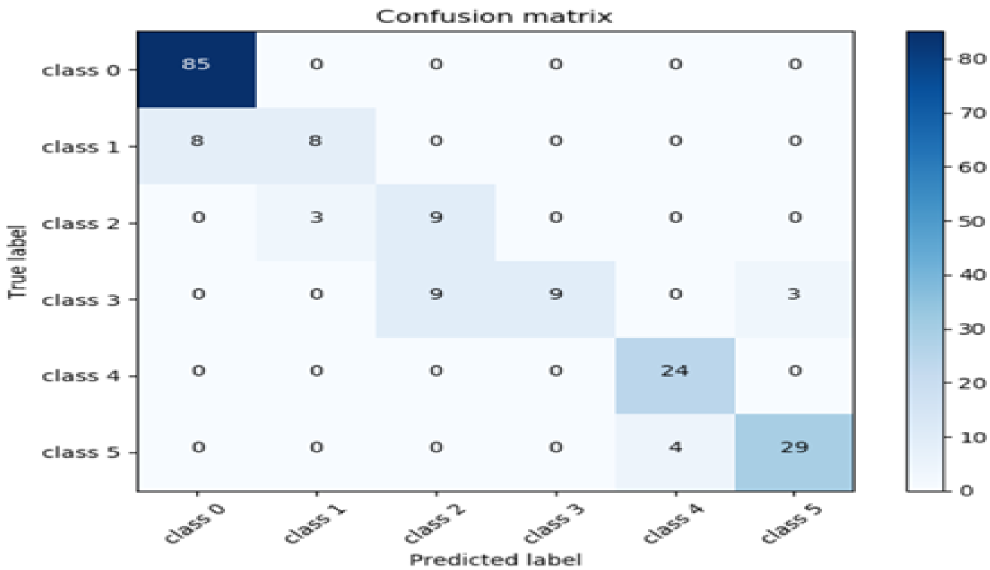

Figure 5.24: Confusion matrix using SVM for 6 class

	precision	recall	f1-score	support
class 0	0.91	1.00	0.96	85
class 1	0.73	0.50	0.59	16
class 2	0.50	0.75	0.60	12
class 3	1.00	0.43	0.60	21
class 4	0.86	1.00	0.92	24
class 5	0.91	0.88	0.89	33
avg / total	0.87	0.86	0.85	191

Figure 5.25: Classification report using SVM for 6 class

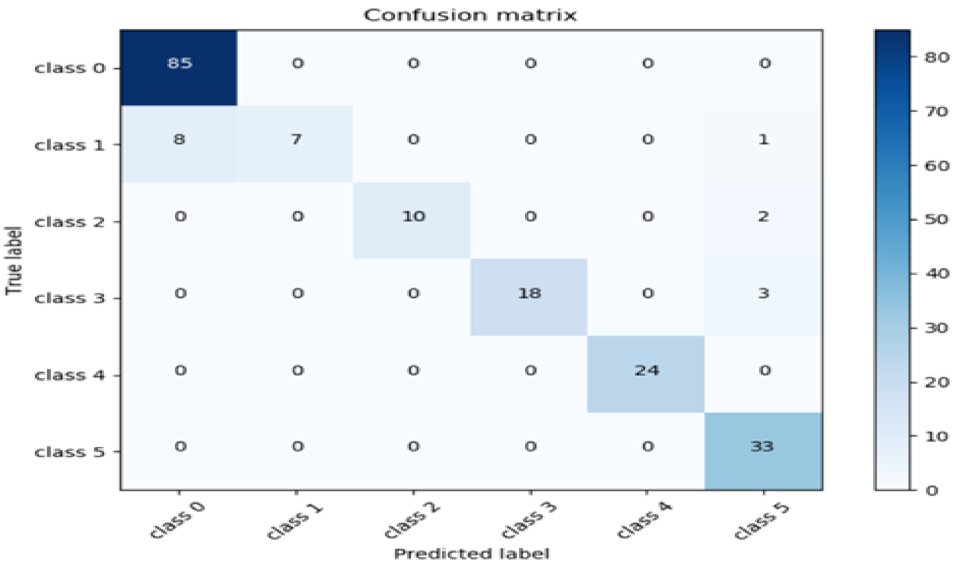

Figure 5.26: Confusion matrix using ANN Classifier for 6 class

	precision	recall	f1-score	support
class 0	0.91	1.00	0.96	85
class 1	1.00	0.44	0.61	16
class 2	1.00	0.83	0.91	12
class 3	1.00	0.86	0.92	21
class 4	1.00	1.00	1.00	24
class 5	0.85	1.00	0.92	33
avg / total	0.94	0.93	0.92	191

Figure 5.27: Classification report using ANN Classifier for 6 class

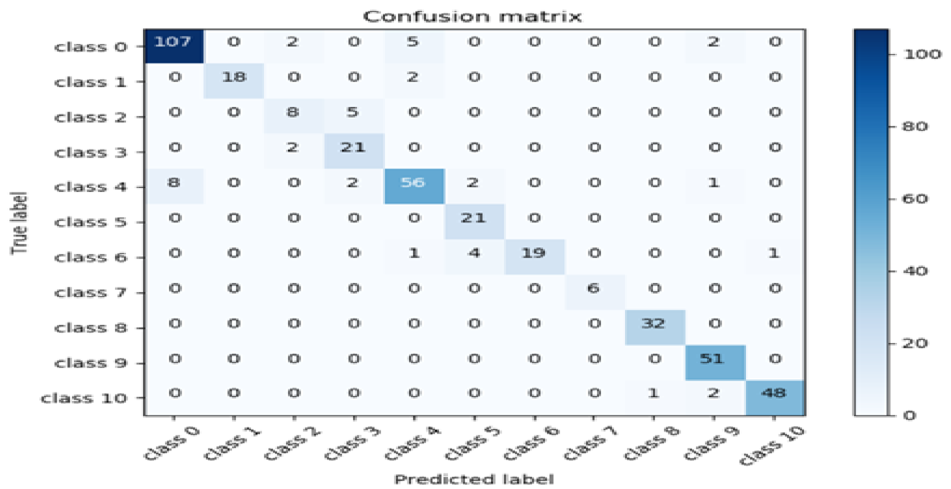

Figure 5.28: Confusion matrix using ANN Classifier for 11 class

	precision	recall	f1-score	support
class 0	0.93	0.94	0.94	116
class 1	1.00	0.90	0.95	20
class 2	0.80	0.62	0.70	13
class 3	0.64	0.91	0.75	23
class 4	0.95	0.81	0.87	69
class 5	0.78	1.00	0.88	21
class 6	1.00	0.76	0.86	25
class 7	1.00	1.00	1.00	6
class 8	0.97	1.00	0.98	32
class 9	0.91	1.00	0.95	51
class 10	0.98	0.94	0.96	51
avg / total	0.92	0.91	0.91	427

Figure 5.29: Classification report using ANN Classifier for 11 class

Recall tells us how often the model predicts yes when it is actually yes. Precision tells us how often it is actually correct when the model predicts it as yes. F-measure or F-1 score represents both recall and precision. It uses a harmonic mean instead of the arithmetic mean because it punishes the extreme values. When the values of precision and recall are different, it gets closer to the smaller one. Best score for the F-1 score is '1' and the worst score is '0'. F-1 can be used to compare the classifier models. From the classification report, we observe that the F-1 score is 0.85 for SVM (6 class). For our neural network classifier model, the F-1 score is 0.92 (for 6 class) which is pretty close to '1'. So, we can claim that the neural network classifier will perform better than SVM and it will be a good fit for the system. Even when we increase the number of class, we observe that the F-1 score of ANN classifier is 0.91. So, ANN classifier will be a good choice for our system.

$$Precision = \frac{TP}{TP + FP} \tag{5.2}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.3}$$

$$F1Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5.4}$$

$WhereTP = True\ \ Positive,$

$FP = False\ \ Positive,$

$TN = True\ \ Negative,$

$FN = False\ \ Negative$

5.6.3 Training and Testing Accuracy Curve

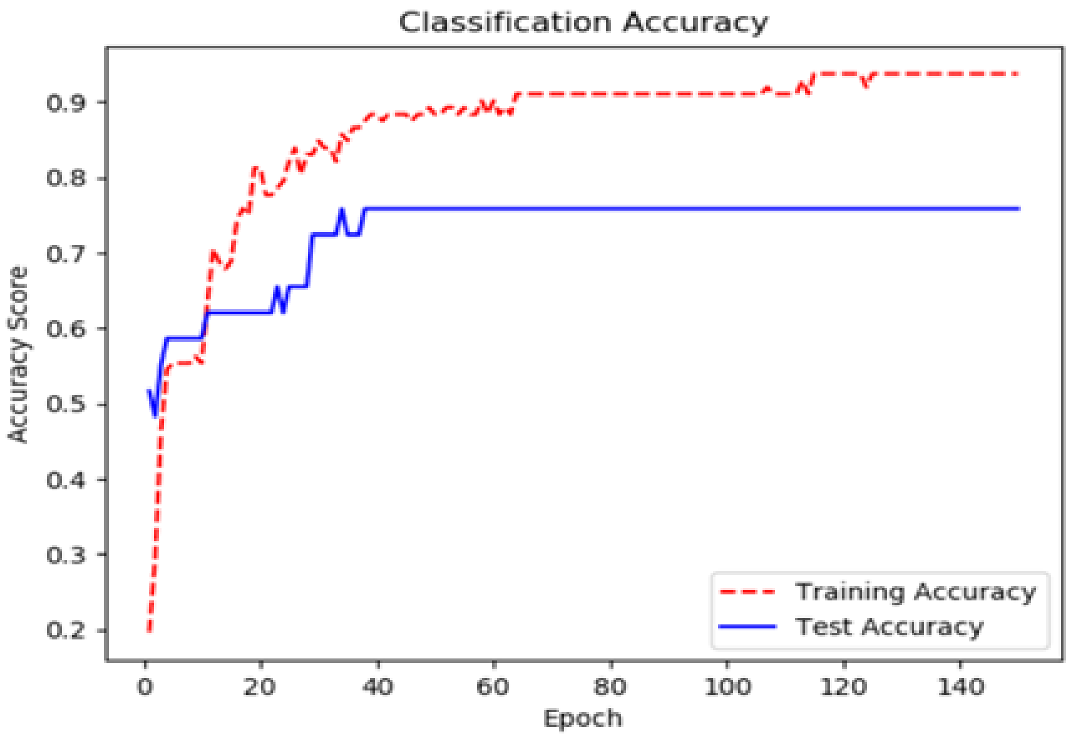

Figure 5.30: Training and Testing Accuracy Curve Vs Epochs for ANN Classifier on 142 observations (6 class)

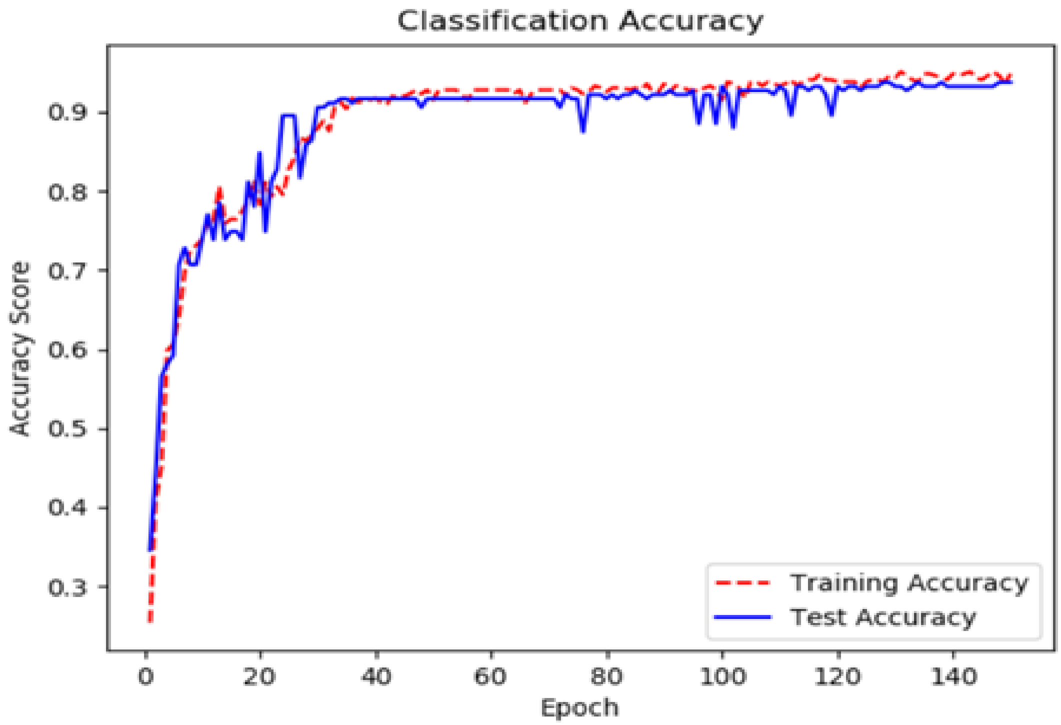

Figure 5.31: Training and Testing Accuracy Curve Vs Epochs for ANN Classifier on 577 observations (6 class)

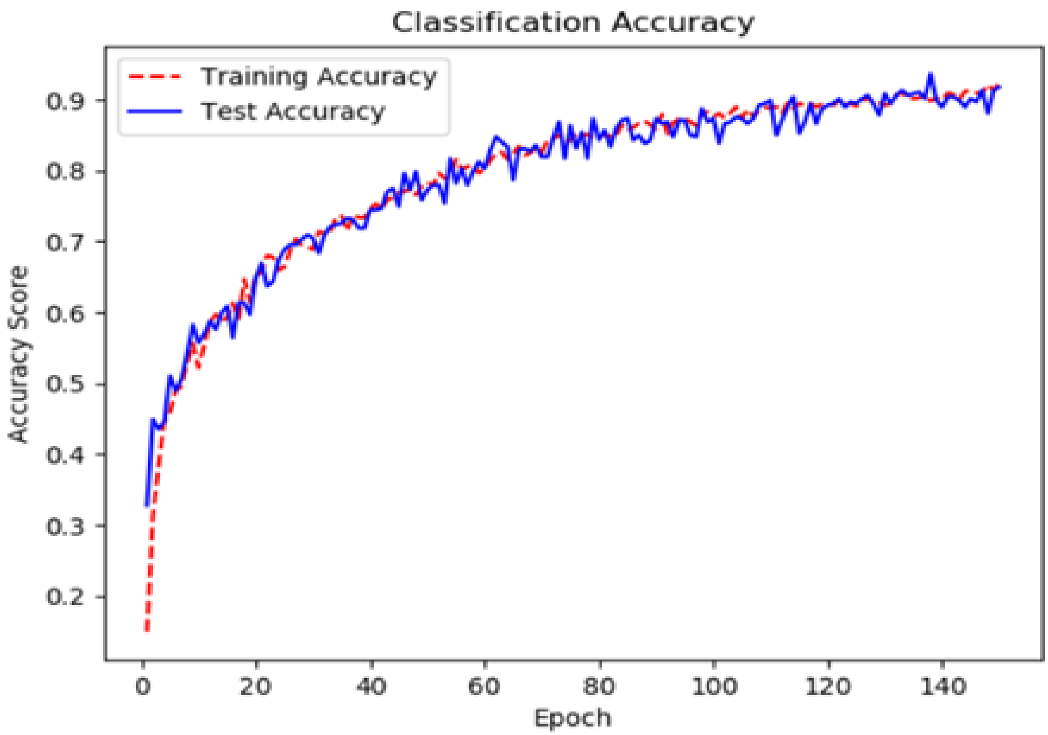

Figure 5.32: Training and Testing Accuracy Curve Vs Epochs for ANN Classifier on 1291 observations (11 class)

Here, we observe that for 142 observations testing accuracy is much lower than training accuracy or testing error is much higher than training error. So, it is a high variance problem. The model for 142 observations suffers from overfitting problem i.e. training set fits the model well but generalizes poorly. One way to solve this high variance problem is to increase the number of training data. Next, we use 577 observations. From the accuracy curve of both training and testing, we observe that after almost 40 epochs testing accuracy curve is slightly below the training accuracy curve. That means this model will perform better on test set than the previous one. Hence, increasing the number of training example reduces overfitting problem for our model. Our classification accuracy for 577 observations is up and close to 93% and the gap between the training and testing accuracy is very small. Even when we increase the number of class and observations to 1291, we observe that test accuracy curve goes along with the training accuracy curve.

5.7 Conclusion

Different algorithm might create different result, our job is to highlight that only a machine alone can not always produce a better outcome. If machine prediction goes wrong it will create huge controversy. Although human predictor can also produce error result but their error decision is better and comparatively relevant. But if human developers help a machine to correct it's error that might reduce time and keep the machine optimal.

Chapter 6

Effects and Issues

6.1 Introduction

Within the next few sections, we will argue with some of the issues that can occur while developing our system for combining human and artificial intelligence. Many technological issue, social effect, economic effect and ethical issues may rise against the system. It's true that removing every issue from a developing system in the first place is impossible. But most issues can be minimized if we discover what will be the possible solution to the technological and social drawback. So future research can lead to a new destination.

6.2 Technological Issues

Implementing machine generated punishment for crime might create some impact. First, Artificial intelligence is a hard stuff and in most of the scenario, we might fail to implement such technology. Also, it may cause some good amount of time and resource. As there is not a huge amount of digital data is present, we might face difficulty to implement machine generated mechanism. The involvement of human judge with a machine is also dependent on what kind of case is being held at the court. If there is a totally new case scenario, the data labeling should be done according to the rule. In this case, a labeler with enough knowledge is required. In other case, training a judge to become habituated with the AI system may require handful of time. The difficulty level goes maximum when there is too much complex data to analyze.

When companies try to implement big data and machine learning mechanism, there is always a limitation. Most of the data needed to be analyzed that come as unstructured or as text data. The time required to analyze the data is so high that concludes why most of the companies analyze only 1% of the data [17]. One solution for such issue is to create a machine learning model where Human and Artificial Intelligence can be combined. In other word making human involvement for labeling the training dataset. In this case train, tune and testing are completed by human involvement. But the process is still costly even if institutions use inexperienced human agent. Also, there is always a risk if institutions decide to choose involvement of third-party outsourcing. Although it is realized that combining human with artificial intelligence always gives the most optimal solution but it is still cost ineffective. Few organizations have a key idea to optimize this combined procedure to its maximum level. The idea is to minimize human intervention through a 'labeling engine'. It invites a human agent only when it needs to reach its maximum performance. In this case, no data scientist or expert is needed. A participant heaving minimum amount of domain knowledge for the intervention will be enough for this procedure.

For our Machine Learning purpose, the technique we will follow is close to a technique called 'Human In The Loop' (HITL). The machine will have access to the relevant case data with appropriate labeling. In the HITL process, it is called 'computer confidence'. If the computer confidence is lower than the specified value, only than a human agent will be asked to label the data so that the machine can train based on human predicted data and reach a better performance. Every time a human participate to correct machine generated labeling, the machine prediction accuracy is changed to a better position. Continuous human interaction with the system will not only increase the performance for the better decision but also it will keep the system up to date. However, we don't use confidence for our system. It will always give prediction and ask human decision.

6.3 Social Issue

Machine dependency over judgment might create negative reaction to general people. Also it is a bit hard to make people feel comfort with machine based decision. If further research is occurred and more advanced technology is made for judicial decision there is still a chance to occur fault decision from machine. And if that occurs even once, that is more than enough to create a huge impact on human mind to think completely negative about machine prediction. Although combining human and machine is not easy because every case has it's own variation and if there is new case the human predictor must have an option to insert it into the training data. The biggest question for social issue is that, "Is it trust worthy?". The reality is we can not prove that it will be accurate 100% every time. But with a human predictors involvement we can consider to get much less error with the lowest amount of time. So to train people to be adaptive with such technology will be a complex task always.

6.4 Economic Effects

As artificial intelligence (AI) is changing our way of living day by day, it is also important to know how much impact artificial intelligence will make to business in future. The global economy is rising fast with the rapid growth of industrialization and plenty use of AI-based technologies. The increasing of higher consumer demand is also forcing firms to choose an automated production system. To achieve a higher amount of products and services, most of the industries ended up with augmenting labor force with AI technologies. A recent site showed that The World Bank is estimating the economic growth from 2017 to 2018 is 3.1 percent which is far higher than expected (2.2 percent). Researchers expect that within next 10 to 15 years, impact on GDP will be much more noticeable with AI technologies [18]. It is also suggested that the dynamic firms which use both human and machine intelligence for fast production and services, will produce a far more affordable product for the consumer within less time. Massive GDP growth over using AI and machine technologies is also noticeable in the geographical economy in North America and China. North America is developing their economic growth with a massive implementation of AI giving them the leading stance. On the other hand, China is making a huge impact of AI on their mass production rate while heaving lower labor cost.

Machine predicted judgment might open a new way for people to save countless time and money. Because if we manage to give case data scenario to a machine to find out a similar case from the previous collection it will save a countless amount of time. If a judge tries to classify a case that will be also easier when it is done by a machine. A machine can filter the key material from a case scenario to simplify it. The rest of the punishment is only based on evidence. As it is human and machine combined model, a case with different structure can be also labeled by a human so the machine can train over new case data. On the negative issue, an expert who has enough knowledge of both machine learning and law is required to transfer raw case data to machine codifying data and this could be costly. But unless a machine have a technical issue it can be ensured to work without corruption or other social influence. Justice without corruption is always a heart for economic progression. The cost of the system management can be high at first because there was no such thing implemented in court and justice before. While implementing this system jobless market in some sector might occur but also there is a chance for new job opportunity.

6.5 Ethical View

From ethical measurement there is some limitation of the project. First of all, it feels little odd when we think about a machine helping in justice. Because a machine can't be controlled with emotion which is possible in case of human. So when a machine combined with a human that may cause a proper outcome without any other influence. A question that, "What will happen if a machine completely takes over the part of the job of a judge?" there is no possibility for a machine completely taking over this system because the machine prediction might fail if there is enough new case scenario with no primary result. For that reason, a human experienced agent will have to participate to label those cases for the machine. Another ethical question may rise that, "Will it be acceptable for an offender to accept his fate based on this hybrid system?". The truth is we can't change human's mind for accepting a judgment which has both judge and machine participation. Only time will tell. Human was not adaptive with cell phone before it was invented.

6.6 Conclusion

Some ethical and social issue can never be avoided. Also some technological issue might create a bigger trouble in odd situations. More or less, research over a new idea will generate negative impact in some scenario but that doesn't mean that an experiment is useless. Appropriate investment and proper development can open a new era for people in the future. New industrial system is always offering machine intelligence to contribute with human which is also a door way for making new job opportunity.

Chapter 7

Conclusion and Future Work

7.1 Introduction

This paper describes a combinational system with Human and Artificial Intelligence in the decision-making process. As many research showed that a machine based prediction is far more superior to the human in many situations. As business perspective, depending on machine creates far more time efficiency in rather than depending on the human. But when we can't depend on machine prediction only in that point human reliability increases. In other words, human intelligence is always a key process to amplify machine intelligence. Through the design, we can demonstrate that with the help of the human, machine intelligence will become more accurate in making the prediction. Which we showed through this combinational approach. In this paper, we implemented some predicting algorithm like Naïve Bayes, Artificial Neural Network, Linear Regression and Support Vector Model and showed their outcome for our dataset.

7.2 Future Work

We believe there is also some real-world situation where both human and Machine contribution can be valuable. In the future, there is a chance of adding the new features to our dataset. We will add more crimes and punishment in the training dataset in the future. We will also bring some changes to our system. We will save our trained model so that it doesn't need to be trained for every new test case. Next, we will not add the new response directly to the dataset. Rather, we will store new test cases and response in a separate database and after a certain limit of the newly stored data; we will evaluate that response and then will add to the dataset. After adding new data, we will check the accuracy of our model to confirm that it will make a good prediction. In further progression, we will also try to implement a text analysis system for our model.

7.3 Conclusion

We believe that human can be a part of the system when simpler details can bring a change to the decision. For example, any wrong decision which may conclude by punishing the offender will never be accepted by mankind. For that purpose, a human can be a part of the job to correct the machine or enhance the accuracy of the machine-generated prediction. We believe that there are many additional works need to be done to properly describe the circumstance where the hybrid model can leverage existing judicial system in a new way. We hope our initial work will encourage others to work on the similar idea in future.

Bibliography

[1] S. Doroudi, E. Kamar, E. Brunskill, and E. Horvitz, "Toward a learning science for complex crowdsourcing tasks," in *ACM CHI*, 2016.

[2] M. Hibon and T. Evgeniou, "To combine or not to combine: selecting among forecasts and their combinations," in *INSEAD*, 2004.

[3] Hopman and J. W, "Using forecasting markets management demand risk," in *Intel Technology Journal*, 2007.

[4] Y. Nagar and W.Malone, "Making business predictions by combining human and machine intelligence in prediction markets," in *Thirty Second International Conference on Information Systems*, 2011.

[5] D. Miller and W. Ju, "Joint cognition in automated driving: Combining human and machine intelligence to address novel problems," in *AAAI Spring Symposium Series*, 2015.

[6] E. Kamar, "Directions in hybrid intelligence: Complementing ai systems with human intelligence," in *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 2016.

[7] D. Shahaf and E. Horvitz, "Generalized task markets for human and machine computation," in *Association for the Advancement of Artificial Intelligence*. AAAI Press, 2010.

[8] A. Mao, D. C. Parkes, A. D. Procaccia, and H. Zhang, "Human computation and multiagent systems: An algorithmic perspective," in *Association for the Advancement of Artificial Intelligence.* AAAI Press, 2011.

[9] E. Duhaime, "Combining human and artificial intelligence for analyzing health data," in *AAAI Spring Symposium Series*, 2016.

[10] O. Amir, E. Kamar, A. Kolovob, and B. Grosz, "Interactive teaching strategies for agent training," in *AIJCAI*, 2016.

[11] A. Kapoor, D. Tan, P. Shenoy, and E. Horvitz, "Complementary computing for visual tasks: Meshing computer vision with human visual processing," in *IEEE*, 2008.

[12] N. TV, "Govt-moves-to-settle-long-pending-case," http://en.ntvbd.com/bangladesh/166973/Govt-moves-to-settle-long-pending-cases. , May 2018, accessed on 25- Jun- 2018.

[13] P. Alo, "5m pending cases by 2020," http://en.prothomalo.com/bangladesh/news/136201/5m-pending-cases-by-2020. , Jan 2017, accessed on 25- Jun- 2018.

[14] C. Dietrich, "Decision making: Factors that influence decision making, heuristics used, and decision outcomes," in *Inquiriesjournal*, 2018.

[15] W. Blackstone and W. Browne, *Commentaries on the laws of England.* St. Paul: West Pub. Co, 1897.

[16] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *3rd International Conference on Learning Representations*, 2015.

[17] D. Somers, "How hitl(human-in-the-loop) will benefit machine learning applications," https://www.comparethecloud.net/articles/jitl-benefits-machine-learning/, 2018, accessed: 24-Jul-2018.

[18] J. Gillham, L. Rimmington, H. Dance, G. Verweij, A. Rao, K. Roberts, and M. Paich, *The macroeconomic impact of artificial intelligence.* PwC, 2018.

Appendix A

List of Acronyms

AdaGrad	Adaptive Gradient Algorithm
AI	Artificial Intelligence
AMT	Amazon Mechanical Turk
ANN	Artificial Neural Network
CV	Cross Validation
DV	Dependent Variable
ECG	Electrocardiography
FN	False Negative
FNR	False Negative Rate
FP	False Positive
FPR	False Positive Rate
GDP	Gross Domestic Product
GTM	Generalized Task Market
GUI	Graphical User Interface
HITL	Human In The Loop
IV	Independent Variable
LR	Logistic Regression
MLR	Multiple Linear Regression
MSE	Mean Square Error
NB	Naive Bayes

NN	Neural Network
OLS	Ordinary Least Squares
RF	Random Forest
RMSE	Root Mean Square Error
RMSProp	Root Mean Square Propagation
SGD	Stochastic Gradient Descent
SVM	Support Vector Machine
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
UNDP	United Nations Development Programme

Appendix B

List of Notations

$\beta_1 ... n$	Coefficient
$\in$	Error
$=$	Equal
σ	Sigma
β_0	Intercept
θ	Theta
T	Transpose
$\leq$	Less then or equal
$\geq$	Greater then or equal
$\sum$	Summation
exp	Exponential
$==$	Equals to
TP	True Positive
TN	True Negative
FP	False Positive
FN	False Negative
y	Dependent Variable
x	Independent Variable
$h_\theta(x)$	Sigmoid Function
w	Weight

b	Bias
log	Logarithm
λ	Lambda
y_i	Actual value
$y_i^{'}$	Predicted value
n	Number of observations
m	Number of training data
k	Number of class
$\hat{y}$	Intended output
$l(y)$	Hinge Loss
$J(\theta)$	Cost Function
C	Cross Entropy Cost Function
$\approx$	Approximate
$\sum_j$	Summing over all output neurons
$\sum_x$	Summation over all training data