



Daffodil
International
University

**Predicting the death of road accidents In Bangladesh
Using Artificial Neural Network**

Supervised By

Ms. Nusrat Jahan (Assistant Professor)

Department of Software Engineering

Daffodil International University

Submitted By

Mst.Sinhanaznin Eti

ID:181-35-2427

Department of Software Engineering

Daffodil International University

**The Thesis report has been submitted in fulfillment of the requirements for the
Degree of Bachelor of Science in Software Engineering.**

APPROVAL

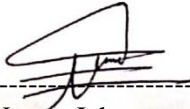
This thesis is being titled as “Predicting the death of road accidents In Bangladesh Using Artificial Neural Network” submitted by Mst. Sinhanaznin Eti, 181-35-2427 to the Department of Software Engineering, Daffodil International University has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of Bachelor of Science in Software Engineering and for approval as per it’s inner styles and contents inside

BOARD OF EXAMINERS



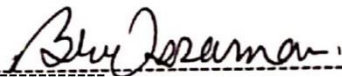
Dr. Imran Mahmud
Associate Professor and Head
Department of Software Engineering
Daffodil International University

Chairman



Nusrat Jahan
Assistant Professor
Department of Software Engineering
Daffodil International University

Internal Examiner 1



Khalid Been Badruzzaman Biplob
Senior Lecturer
Department of Software Engineering
Daffodil International University

Internal Examiner 2



Professor Dr M Shamim Kaiser,
Professor
Institute of Information Technology
Jahangirnagar University

External Examiner

DECLARATION

It's been declared that this thesis including all the research-based experimental works has been completed by me under the supervision of Ms. Nusrat Jahan (Assistant Professor), Department of Software Engineering of Daffodil International University. I also declare that neither this thesis nor any part of this whole research-based experiment has been submitted elsewhere for award of any degree.

Sinha

Mst.Sinhanaznin Eti
181-35-2427
Department Of Software Engineering
Daffodil International University



Nusrat Jahan
Assistant Professor
Department of Software Engineering
Daffodil International University

ACKNOWLEDGEMENT

I am grateful to the Almighty Allah (SWT) for allowing me to complete this Bachelor of Science study with this research. I am thankful to my Parents who have always been a support for me throughout my life. They always let me have faith and belief that I can achieve something and have been supporting me. I want to thank my well-wisher who always inspires me for doing something creative.

I would also love to show my respect and gratitude to my Supervisor, **Ms. Nusrat Jahan**, supporting me in completing this research and encouraging me in research and implementation of this work.

Besides, I would like to thank **Dr. Imran Mahmud**, Associate Professor & Head In Charge of the Department of Software Engineering, Daffodil International University for motivating us for quality research. I also want to thank my teacher who has been continuously supporting us throughout this undergrad. Last but not the least, I want to thank all the staff of the Department of Software Engineering and Daffodil International University for continuously working hard to provide us with the best facilities in classrooms and laboratories.

Table of Contents

APPROVAL	ii
CHAPTER ONE	1
INTRODUCTION.....	1
1.1 Background	1
1.2 Motivation of the research	2
1.3 Problem Statement.....	4
1.4 Research Question.....	4
1.5 Research Objectives	4
Research Scope.....	5
Research Organization	5
CHAPTER TWO	6
LITERATURE REVIEW	6
CHAPTER THREE	16
RESEARCH METHODOLOGY	16
Figure 2.1: Proposed Research Methodology Diagram	17
3.1 Data Requirement	17
3.2 Data Collection	18
3.3 Data Pre-processing	18
3.4 Algorithm Used	19
Figure 3.1: Architecture of Artificial Neural Network	22
CHAPTER FOUR.....	27
RESULT ANALYSIS AND DATASET DISCUSSION	27
Table 4.1 : Result Analysis	27
CHAPTER 5.....	29
5.1 FINDINGS AND CONTRIBUTIONS	29
5.2 LIMITATION	30
5.3 RECOMMENDATIONS FOR FUTURE WORK	30
5.4 References	31

ABSTRACT

Road accidents are now a common occurrence in our country. Road accidents are increasing by leaps and bounds in Bangladesh every day. Every day countless people are losing their lives in road accidents. The effect of a road accident is pitiful. This research paper has done to analyze the previous year's accidents records in Bangladesh deeply to predict the death of road accidents by using deep learning machine-learning approach. In this study, we have applied four models such as Artificial Neural Network, Decision Tree, K-Nearest Neighbor, Logistic Regression. This research would be helpful for the policymakers and stakeholders related to the road to take the future steps with the highest accuracy of 98% in Artificial Neural Network algorithm.

Keywords: Deep Learning, Machine Learning, Road Accident.

CHAPTER ONE

INTRODUCTION

1.1 Background

Road accidents in Bangladesh have become widespread nowadays. This not only damages our economies but also affects many families as well. In recent years, the road accident has become a global problem and marked as the ninth prominent cause of death in the world. Every day when we read the newspapers, we can see a series of accidents and death of so many people owing to accidents with buses, trucks, motorcycles and other vehicles. Road accidents are regular phenomena in our daily life. These accidents occur for various reasons. The increase of the number of vehicles and inefficient drivers on the road, the poor conditions of road and maintenance of the roads are highly responsible for the crisis. The roads in our country is very narrow. Also, population growth, Reckless driving is responsible for bringing about accidents. The basic factors which increase the severity of probable accidents are vehicle features, highway characteristics, environmental effects, poor driving skills, health problem, psychological problems, drivers' behavior. Most of the drivers in our country are totally careless about the traffic rules and regulations. Most of the drivers are not concern about the rules that's why they recklessly drive their vehicles and they have not adequate knowledge of the traffic system. Pedestrians are also responsible for street accidents. Sometimes pedestrians do not have concentration when crossing the road while drivers can no longer control themselves to protect them from accidents. Defective roads often cause the car to overturn, leading to major accidents. According to the record of police (2015), in last 21 years 84,000 road accidents occurred, 56,000 died and 63,000 injured [1]. But the actual number is higher

than the aforementioned figure because all the people who suffer injury don't make a general diary at the police station. And it is not possible that police keep records of all accidents. A report published in Prothom Alo sources from BRTA, reported that there was an accident statistic which suggested that the damages due to RTA were worth 40 thousand crore taka in the last 3 years. It also showed that Bangladesh is losing 2% - 3% (Accident Research Institute, BUET,2019). gross domestic product (GDP) in each year for the causal impact of road accidents [3]. The number of accidents remained almost the same from the previous year but the death toll increased by 8.07 percent (Jatri Kalyan Samity, annual report, 2020) [4]. The Samity also claimed 5,514 road accidents occurred in 2018 and where 77,221 lives death and left 15,466 injured in the country. In 2019, at least 5,227 people had killed while 6,953 had injured in 4,702 road accidents across the country (Nirapad Sarak Chai, annual report, 2019). The loss of life and property in the accident and the traffic jam caused by the accident are hampering the economy. Overall, this is a significant loss for a country like ours. Because the death toll from road accidents are actually the highest among low-income country. And Bangladesh is one of them (WHO, Global Status Report on Road Safety, 2015) [5]. It causes loss of life and property due to accidents and traffic jams caused by accidents are disrupting the economy. Moreover, reducing this loss has become a huge concern for our country now. Now, therefore, it is time to do research on road accidents.

1.2 Motivation of the research

Researchers are trying to figure out how to reduce this by analyzing the severity of previous accidents. Many researchers are trying to find out the right results through the traditional research such as surveys and questionnaires. Using this type of data for predicting factors of

road accidents are not suitable. Because road accidents are unpredictable. And also, this is very difficult to get 100% accurate data. And also, many researchers are use some machine learning algorithm. To overcome this challenge, researchers should follow any advance method such as machine learning and deep learning algorithm. Implementation of an advanced method that can give better analysis result is a crying need here. Machine Learning and deep learning is one of the most advanced scientific fields that can be applied here to get a better result.

Solution. Firstly, the data about the accidents have to analyze very well through advanced machine learning and deep learning algorithms to determine the severity of an accident.

In this research paper, the author analyzed data of road accidents through three supervised learning algorithm techniques such as Logistic Regression, Decision tree, K-Nearest Neighbors (KNN) and one deep learning algorithm that is Artificial neural network. The prime goal of this research paper is to analyze the road accidents and determines the severity of an accident by applying those algorithms. We use those algorithms because of their proven accuracy in this sector. Here we focus Artificial neural network because It's a deep learning algorithm and using deep learning algorithm there are many kinds of benefits. Neural networks have the ability to learn by themselves and produced the output that are not limited to the inputs we provide. The ANN architecture is flexible to be adapted to new problems in the future.

We collected the dataset from GitHub Author classified this dataset into two classes for death prediction (Yes or No) and took four features to get the appropriate result such as Injured, Accident History, Responsible car and Victim car. Analysis of last seven years accidents data (2013-2019) and applied these techniques to achieve the best solutions. This research paper has done the analysis of t seven years accidents data (2013-2019). Among these four techniques we got the best result in Artificial neural network with 98% accuracy. Which will later help policymakers to reach better conclusions.

1.3 Problem Statement

The task of accurate road accident prediction is difficult mainly due to the complex nature of traffic accidents. Another problem with accident prediction/detection is the scarcity of accidents in both space and time. Another problem with accident prediction/detection is the scarcity of accidents in both space and time. Due to the limited number of samples, it is challenging to precisely predict the occurrence of individual accidents. A large number of existing works on traffic accident detection or prediction are applying classical models such as classification or regression on limited data. This leads to unsatisfactory performance. The classification models mainly create a standard predicting model based on the information learned from the training set.

1.4 Research Question

How accurately ANN can be used for road accident prediction compared to machine Learning algorithms?

1.5 Research Objectives

1. identify the factors that are causing the problem.
2. In-depth analysis of previous year's accident records in Bangladesh.
3. Predict the death based on these features by using deep learning approach.
4. Improve the accuracy from existing paper.

Research Scope

We highlight our Scope as follows:

1. Helping policymakers and BRTA take future actions
2. Raise awareness among road users and stakeholders
3. Help reduce the number of deaths.
5. Help the investigative team determine the cause of death.

Research Organization

The rest of this paper is organized as follows. Chapter 2 presents a sort description of some previous work related to this study. That discusses research into predictive models, and into the needs of projects which could leverage a framework like this, to help motivate the design of the framework. Chapter 3 presents the brief description of research methodology. Chapter 4 describes the dataset features and results by comparing these four methods based on their accuracy. Finally, Chapter 5 presents the future directions of this research and finishes the paper with some important suggestions.

CHAPTER TWO

LITERATURE REVIEW

Author	Year	Paper	Method	Findings
Md.Farhan Labib,Ahmed Sady Fifat,Md.Mosabbir Hossain,Amir Kumar Das,Faria Nawrine	2019	Road Accident Analysis and prediction of accident severity by using Machine learning In Banglade sh	Decision Tree, K-Nearest Neighbors (KNN), Naive Bayes and AdaBoost	Authors classified the severity of accident into fatal, Grievous, Simple Injury and Motor collision these four categories. Low Accuracy In KNN (67%)

Author	Year	Paper	Method	Findings
Joy Paul,Joy Paul,Kazi Fahim Lateef,Md. Robiul Islam,Sagor Chandro Bakchy	2020	Prediction of Road Accident and Severity of Bangladesh Applying Machine Learning Techniques	Decision tree, Random Forest, Multilayer perceptron's, categorical naive bayes	building a complete accident prediction model showed the comparative results of various machine learning methods

Author	Year	Paper	Method	Findings
P Sumanth1, P Sai Anudeep2 and S Divya3	2020	Analysis of Machine Learning Algorithm with Road Accidents Data Sets	Logistic Regression, Random Forest, Support Vector Classifier (SVM), Decision Tree, K- Nearest Neighbors (KNN), and K- Means	The aim of the authors of this paper is to implement AI in various accident- related fields such as traffic flow forecasting, accident forecasting and tourist location, etc. low accuracy in KNN

Author	Year	Paper	Method	Findings
Ali J. Ghandour, Huda Hammoud and Samar Al-Hajj	2020	Analyzing Factors Associated with Fatal Road Crashes: A Machine Learning Approach	Sequential minimal optimization (SMO), Random Forest, Artificial neural network (ANN), Logistic regression. Naïve Bayes	Proposed a model that adopts a hybrid ensemble machine learning classifier structured from sequential minimal optimization and decision trees to rebutting to fatal road injuries Big difference in various sources data.

Author	Year	Paper	Method	Findings
Arun Venkat1, Gokulnath M1, Guru Vijey K.P1, Irish Susan Thomas1, D.Ranjani2	2020	MACHINE LEARNING BASED ANALYSIS FOR ROAD ACCIDENT PREDICTION	Random forest, Decision tree, Logistic regression, Nearest Neighbors Algorithm	Presents a survey of various existing work related to accident prediction using machine learning area.

Author	Year	Paper	Method	Findings
Salahadin Seid Yassin1 · Pooja	2020	Road accident prediction and model interpretation using a hybrid K-means and random forest algorithm approach	K-means techniques, Random forest	The authors were studying using a traditional statistical-based approach for model building. These techniques help to get insights and identify the underlying cause of vehicle accidents and related factors on road safety

Author	Year	Paper	Method	Findings
Yash Kumar Arora, Santosh Kumar, Umesh Kumar Tiwari, Shubhank Singhal, Vijay Kumar	2019	Accident Severity in India	Polynomial Regression, R2 Score, Road Accidents	Polynomial regression algorithm to predict accident severity in India. They collected the data set from a government site of the Ministry of Road Transport and Highway (MORTH)

Author	Year	Paper	Method	Findings
Vipul Rana, Hemant Joshi, Deepak Parmar ³ , Pradnya Jadhav, Monika Kanojiya	2020	Road Accident Prediction using Machine Learning Algorithm	Logistic Regression	Analyze the previously occurred accidents in the locality which will help us to determine the most accident-prone area and help us to set up the immediate required help for them.

Author	Year	Paper	Method	Findings
Annie Racheal Rajkumar, Srihari Prabhakar, A Meena Priyadharsini	2020	Prediction of Road Accident Severity Using Machine Learning Algorithm	Logistic Regression, Naive Bayes Classification, XGBoost Classification	<p>Accident severity is determined by the damage resulting from the accident in terms of bodily harm (fatal accidents being the most severe), and helps categorize the collision.</p> <p>Accident severity is directly related to the speed of the vehicle at the time of the collision.</p> <p>Cannot perform well in small dataset</p>

Author	Year	Paper	Method	Findings
G Pavan Karthik, Sneha, Sudalaimuthu	2020	Analysis of Road Accidents using Machine Learning	K-means, Random Forest algorithm	The authors aim to develop machine learning-based models to appropriately determine injury severity and help spread awareness

Author	Year	Paper	Method	Findings
Sanjay Misra, Sanjay Misra, L. Fernandez- Sanz	2020	An Artificial Neural Network Model for Road Accident Prediction: A Case Study of a Developing Country	Multi-Layer Perceptron Neural Network (MLPNN)	Author produce a design Of an artificial neural network model for the analysis and prediction of road accidents rates in a developing country. And sigmoid and linear functions are used as an activation function.

Author	Year	Paper	Method	Findings
Aklilu Elias Kurika ¹ , Irfan Ahmad Ganie ² , Yuliyanti Kadir ³ , Patrick D. Cerna ⁴ , Frice L. Deseis	2020	Predicting Factors of Vehicular Accidents using Machine Learning Algorithm	Decision tree, random forest, Bayesian classifiers (Naïve Bayes and Bayesian Network)	Machine Learning approaches have been applied for data analysis and prediction of car traffic accident datasets to explore important features and pattern relationships to car traffic accident occurrences.

Author	Year	Paper	Method	Findings
Miao Chong ¹ , Ajith Abraham ² and Marcin Paprzycki ^{1, 3}	2004	Traffic Accident Analysis Using Machine Learning Paradigms	Artificial Neural Network Using Hybrid Learning, Decision Tree, Support Vector Machine	Author develop machine learning based intelligent models that could accurately classify the severity of injuries (5 categories). This can in turn lead to greater understanding of the relationship between the factors of driver, vehicle, roadway, and environment and driver injury severity.

Author	Year	Paper	Method	Findings
Adnan Majeed	2021	Research Paper on Road Accident of UK Traffic (1979-2019) Data Sets Analysis	Decision Tree, KNN, Naïve Byas	The objective of this research is detecting the different statistical analysis to achieve the better road accident prediction. Many machines learning algorithm has been tested and analyzed. Time series data forecasting has been tested.

Author	Year	Paper	Method	Findings
Md Shakhawat Hossain*, Md Omor Faruque	2019	Road Traffic Accident Scenario, Pattern and Forecasting in Bangladesh	Augmented Dickey-Fuller (ADF) test (a unit root test) and Philip Perron test. Smoothing techniques.	Author find out most accident porn locations. see the RTA scenario, patterns and its frequency. forecast the number of RTA for the future.

Author	Year	Paper	Method	Findings
Md. Ebrahim Shaik ^{a,*} , Md. Milon Islam ^b , Quazi Sazzad Hossain ^a	2021	A review on neural network techniques for the prediction of road traffic accident severity	Convolutional neural network, Recurrent neural network.	It completed a data mining structure to recognize, analyze and choose credits adding to road accidents.

Author	Year	Paper	Method	Findings
P. Sundar Kumar	2019	Analysis of Road Accident		The paper focused the main idea of this project the reduce road accidents by dissecting the accident prone areas and taking security methods to diminish them

Author	Year	Paper	Method	Findings
Nancy.P, Dilli Rao, Babuaravind.G, Bhanushree.S	2021	Highway Accident Detection and Notification Using Machine Learning	Convolutional neural network, Recurrent neural network.	In this paper, authors proposed a system that using the Image Processing technique, accidents can be detected and information will be collected from the CCTV footage. The information would be processed by Machine learning tools to detect possible accidents. In this paper, the authors aimed to analyze vehicles and their state where the accident is concluded. There were 800 images in the dataset, were 400 for the mild crash, and 400 for the severe crash. The clustering algorithm has performed

CHAPTER THREE

RESEARCH METHODOLOGY

We used supervised classification Machine Learning algorithms such as Logistic regression, k-Nearest Neighbors and Artificial Neural Network in Deep learning Algorithms. These algorithms have applied based on some features like Injured, Accident History, Victim Car and responsible car. Our proposed model includes the collection of appropriate data sets and pre-process the data set in just a few steps. Later, preprocessed data is divided into training and testing datasets. The training set carried a known output and the model learned about that data to be generalized later on to other data, also had test set to test prediction of the model on that sub-set. Then the training set was fitted by various algorithms which in return gave us various models. Evaluating the test dataset multiple times yields the desired prediction model.

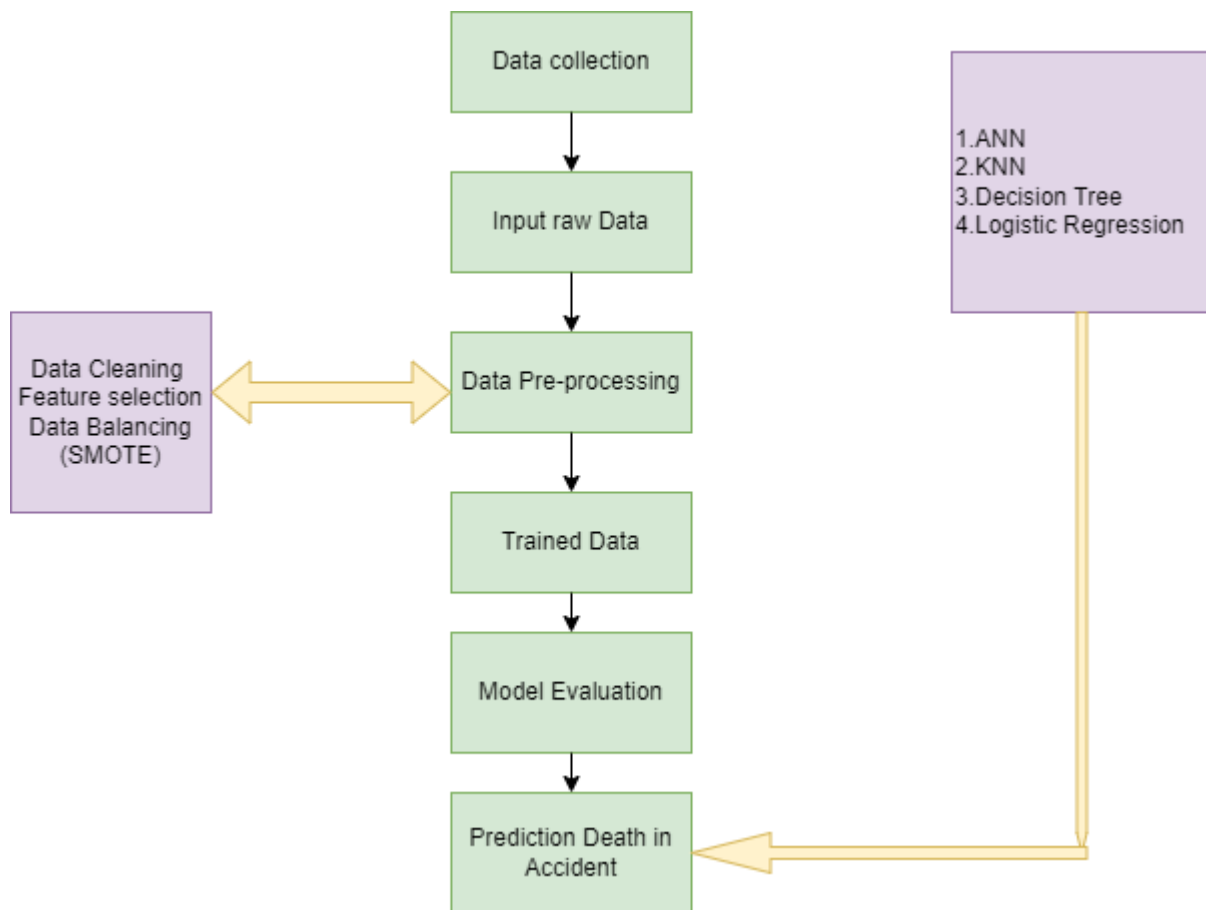


Figure 2.1: Proposed Research Methodology Diagram

3.1 Data Requirement

A minimum set of data is required, including outcome data (including the number of persons killed and injured by type of road users, type of roads, vehicles, etc.). based on our previous studies we are deciding to collect this information.

3.2 Data Collection

The most essential and one of the most difficult tasks was to collect the correct dataset. Appropriate predictions are possible only through accurate datasets. From our data requirement understanding step we got to know about the data we require to collect. First, we have collected the data from GitHub. So first we used Google's Search Engine to find the sites. Then we collected the data. And we got it as a CSV file as it is a well-accepted format for data science experiments. The dataset contains 1237 road accidents data. We split our entire dataset into two parts Training Dataset and Test Dataset. 70% of the whole dataset has been chosen randomly by using a python library as a training data set and the remaining 30% has been used as our test dataset. We have used the 70-30 ratio for splitting the dataset because of its proven accuracy.

Link of the dataset: <https://github.com/DipuSiddik/Road-Accident-Data>

3.3 Data Pre-processing

In this step the data cleaning, data reduction, feature selection and data transformation has been made to prepare the best quality datasets for further analysis. We have found four major factors that are directly involved in road accidents and help to predict the death in accidents. Our data set was an imbalanced data set. Then we used oversampling technique "SMOTE" for balancing the data. After balancing the data, we used machine learning and Deep learning algorithms based on these features such as Injured, VictimCar type, Responsible Car type and Accident History to classify the death (Yes or No) in an accident.

SMOTE (synthetic minority oversampling technique) is one of the most commonly used oversampling methods to solve the imbalance problem. It aims to balance class distribution by randomly increasing minority class examples by replicating them. SMOTE synthesizes new minority instances between existing minority instances.

The steps of SMOTE algorithm are:

1. Identify the minority class vector.
2. Decide the number of nearest numbers (k), to consider.
3. Compute a line between the minority data points and any of its neighbors and place a synthetic point.
4. Repeat step 3 for all minority data points and their k neighbors, till the data is balanced.

3.4 Algorithm Used

In this section we will discuss each of the algorithms we have used throughout our model training.

3.4.1 K-NEAREST NEIGHBOR (KNN)

One of the simplest Supervised Machine Learning algorithms is K-Nearest Neighbor (KNN) and it is based on feature similarity. It analyzed the distance of all the remaining data points from unknown data points and sorted the data points in small to large sizes (Ascending Order) according to the value of distance. The first K-number of points would take from the sorted data point. Out of these K-numbered data points, the class with the greatest number of points have to identify the unknown data point as that class. For this research paper, we used

the Euclidean distance formula to measure the distance.

3.4.2 LOGISTIC REGRESSION

Logistic regression is another technique borrowed by machine learning from the field of statistics. It is the go-to method for binary classification problems (problems with two class values). In this post you will discover the logistic regression algorithm for machine learning.

Logistic regression is named for the function used at the core of the method, the logistic function. The logistic function, also called the sigmoid function, was developed by statisticians to describe properties of population growth in ecology, rising quickly and maxing out at the carrying capacity of the environment. It's an S-shaped curve that can take any real valued number and map it into a value between 0 and 1, but never exactly at those limits.

When the dependent variables are a binary number 0 and 1, then Logistic Regression can be an appropriate regression analysis to conduct. Logistic regression explains the relationship between one dependent binary variable and one or more nominal, ordinal, interval, or ratio-level independent variables. Multinomial logistic regression is required for classification between the three classes. Logistic regression only works when the data is linearly separate in higher dimensions. As a result, the author has used this algorithm here. Our goal in logistic regression would be to keep the output from the hypothesis within 0 to 1. If the output is less than 0.5, then the output of our logistic regression will be 0, and if it is equal to or greater than 0.5, then the last output of the logistic regression will be 1.

3.4.3 Decision Tree

Decision tree Algorithm is a supervised learning algorithm. Decision tree algorithm is used for solving classification problems. By learning simple decision rules inferred from prior data, a training model has been created using a Decision Tree for predicting the class or value of the target variable. To predict a class label in a Decision tree record we start from the root of the tree.

For the construction, a root node is required, and the best attribute is chosen using the gain approach. Sub nodes are then formed using the decision made in relation to the quality status determined at each node. When each of the nodes has been evaluated for its gain status, A leaf is determined at the conclusion of the Root node and is referred to as a class. This action has taken place successively worked until each node has a class specified at the conclusion. The motive for applying a decision tree is to create a training model Through which we can predict the target variables by learning decision rules inferred from training data. Among the classification algorithms, the decision tree algorithm is much easier to understand than the others.

3.4.4 Artificial Neural Network

Artificial neural network is a computational model that is inspired by the human (in particular the brain) central nervous system. Artificial neural network is capable of machine learning and pattern recognition. Artificial neural network is usually a highly interconnected system with neurons from which it can compute values from input. Artificial neural networks are clusters of primitive artificial neurons. This creates clustering layers that are connected to each other. So, all these neurons are connected to form a huge and complex structure called

neural network. The output of a single neuron depends on the input of thousands of interconnected neurons.

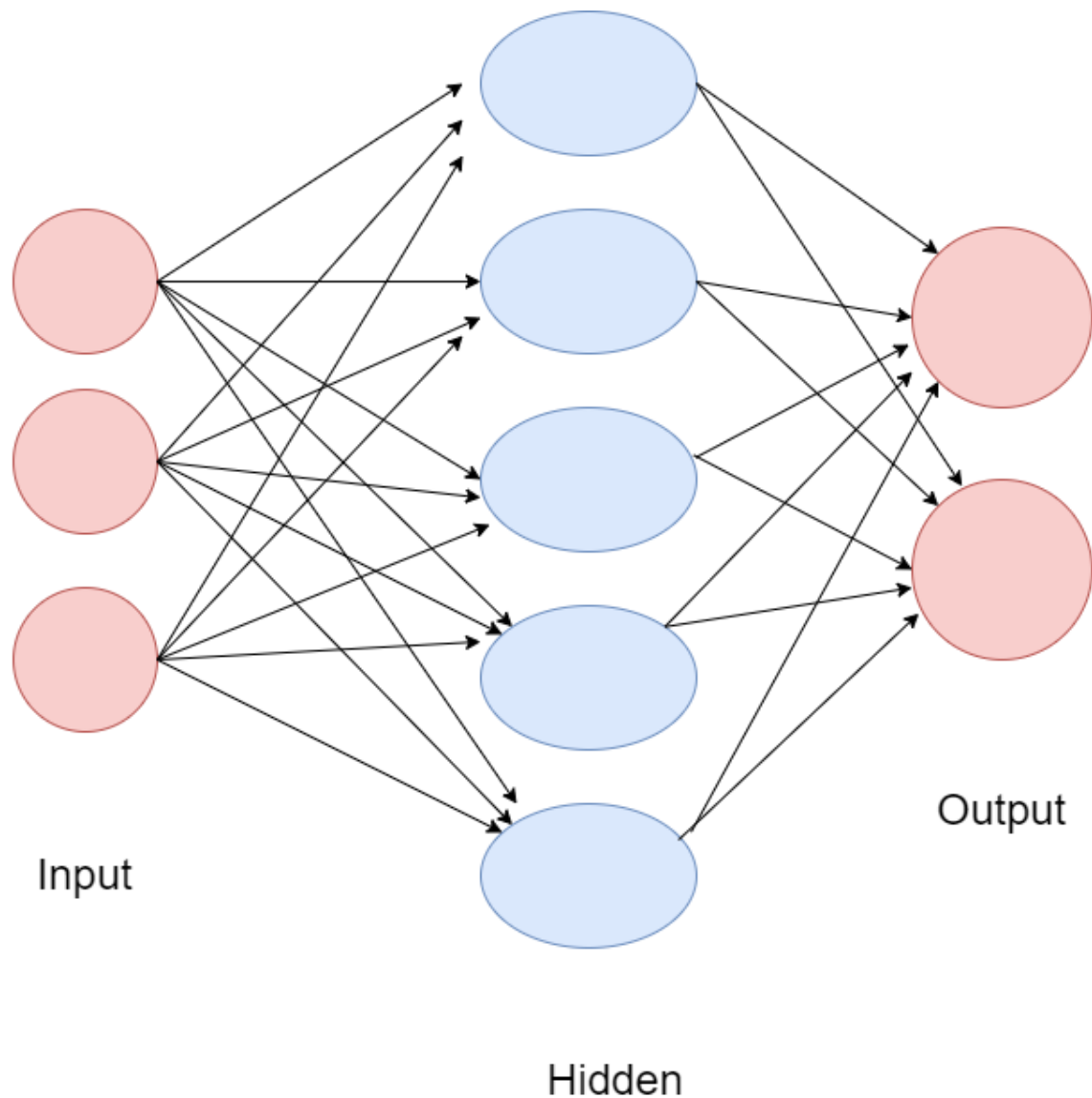


Figure 3.1: Architecture of Artificial Neural Network

ANN is made of three layers. Input layer, Hidden layer, output layer. Input level nodes must have a connection to hidden level nodes. from each hidden layer node with the nodes of the output layer. The input layer takes the data from the network. The input layer accepts the

inputs after that hidden layer processes the inputs, Finally the output layer produces results.

Input Layer : Preceding information for neural networks

Hidden Layer: intermediate layer between input and output layer and place where all the computation is done.

Output Layer : Generates results for the given input layer

Here [figure:2] each circular node represents an artificial neuron and the arrows represent a connection from the output of one artificial neuron to the input of another. Each connection can transmit a signal to other neurons, such as the synapses of the biological brain. An artificial neuron receives a signal and then performs a process and can signal connected to it. A connecting signal is a real number and output of each neuron is calculated by some non-linear function of the sum of its inputs. The connection is called edges. Neurons and edges have a normal weight that corresponds to the progress of learning. Weight increases or decreases the strength of a connection signal. Neurons can have a threshold so that a signal is sent only if the overall signal exceeds that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signal starts from the first layer (input layer), to the last layer (output layer), possibly after traversing the layers multiple times.

The equation for the neural network is a linear combination of the independent variables and their respective weights and bias (or the intercept) term for each neuron. The neural network equation looks like this:

$$Z = \text{Bias} + W_1X_1 + W_2X_2 + \dots + W_nX_n$$

Where,

* Z is the symbol for denotation of the above graphical representation of ANN

* W_i s, are the weights or the beta coefficients

* X_i s, are the independent variables or the inputs, and

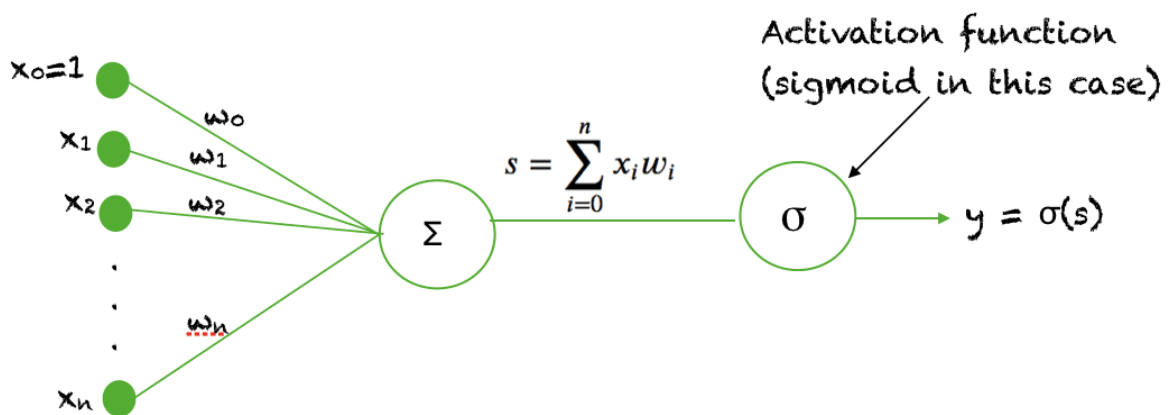
*Bias or intercept = W_0

There are three steps to perform in any neural network:

1. First of all take the input variables and the above linear combination equation of $Z = W_0 + W_1X_1 + W_2X_2 + \dots + W_nX_n$ to compute the output or the predicted Y values, called the Y_{pred} .
2. Calculate the loss or the error term. The error term is the deviation of the actual value from the predicted value.
3. Minimize the error term.

Here we use two activation functions. Sigmoid and relu.

Sigmoid Activation Function: Another name for the sigmoid activation function is Logistic function and also called a squashing function. Sigmoid activation function is a very popular activation function for neural networks. Sigmoid function is the key to learning how a neural network understands complex problems. The most important reason we use the signal function is that **it** exists between (0 to 1). So it is especially used for models where our output is probably to be predicted as the probability of anything exists only between the range of 0 and 1, Sigmoid is the right choice. If the input to the function is either a very large negative number or a very large positive number, output is always the same for any number between 0 and 1.



The figure below shows the role of an activation function at a level of the neural network. A weighted sum of the inputs is passed through an activation function and the output acts as an input at the next level. When sigmoid is used as the activation function for a neuron, it guarantees that the output of this unit will always be between 1 and 0. a sigmoid that assigns a sigmoid function to an activation function is called a sigmoid unit. The sigmoid function can be used in neural networks for learning complex decision functions.

Rectified Linear Unit

The Rectified Linear Unit is the most commonly used activation function in deep learning models. The function returns 0 if it receives any negative input, but for any positive value x it returns that value back. So, it can be written as

$$f(x) = \max(0, x)$$

$$f(x) = \max(0, x).$$

The rectified linear activation function or ReLU for short is a piecewise linear function that will output the input directly if it is positive, otherwise, it will output zero. It has become the default activation function for many types of neural networks because a model that uses it is

easier to train and often achieves better performance. Relu is popular for activation functions of deep learning multi-layer networks.

Artificial neural networks are used to get a model that is the best for predicting road traffic accidents, as ANN always learns from past experience after it completes its first training, and so, it becomes an appropriate methodology for prediction. The ANN technique shows some tolerance to a good extent over errors that may exist within the training set. It has the ability to show the veils and dependencies that are not linear and still learn from its past experience after completing its first training, and this makes it an appropriate method that is suitable for prediction.

CHAPTER FOUR

RESULT ANALYSIS AND DATASET DISCUSSION

We do an experiment on our dataset to find out their effect on a traffic accident. Injuries in road accidents occurred 906 times out of 1237 times. Did not happen in the remaining 331 times.43% accidents have occurred by passenger cars and the remaining 57% accidents have occurred by freight cars out of 1237 road accidents. That 59% of passenger cars were victims of these accidents and the rest 41% of freight cars.85% of accidents were occurred due to collision and the remaining 15% were overturned and others. We have tried to find out the death ratio from our dataset that among 1237 accidents in our country death happened 1209 times. Therefore, it is alarming for our country.

Different algorithms were applied for our work.

In this research paper, we have classified our dataset into two classes Yes or No in death based on four approaches of deep learning and machine learning algorithms. Among these four machine-learning approaches, we got maximum 98% accuracy in the Artificial Neural Network.

Algorithm	Result
Artificial Neural Network	98%
k-Nearest Neighbors	97%
Decision Tree	97%
Logistic Regression	79%

Table 4.1 : Result Analysis

We observe that deaths occurred in most of the accidents in our dataset. We classified our dataset into two classes in death (Yes or No) based on four features. Though our dataset was an imbalanced dataset so we used the oversampling technique “SMOTE” for balancing this dataset and then we split it for the train and test data set. 70% for train data and 30% of test data. Firstly, we applied the K-Nearest Neighbor Algorithm and got better performance with the accuracy of 97% and then we applied decision tree and got again 97% accuracy and applied Logistic regression we got 79% accuracy. Therefore, we applied Artificial Neural Network algorithm where we got the best performance among all of the algorithms 98%.

Although we have almost the same accuracy of ANN, KNN and decision tree. But the best one is Artificial Neural Network. Because using Artificial Neural Network there are many kinds of advantages. ANN uses past experience to learn how to deal with new and unexpected situations. The statistical distribution of the data does not need to be known when developing an ANN model. There is no need for prior knowledge about the relationships amongst the variables being modeled. ANN has the ability to model complex, nonlinear relationships without previous assumptions of the nature of the relationship, like a black box. The most important key element of ANN paradigm is the novel structure of the information processing system. So, Artificial Neural Network is best for predicting road accidents for this dataset.

The above result shows that Artificial Neural Network is significantly best by performance than all other Classification Algorithms. Decision Tree and KNN are significantly good by their performance and the last one Logistic Regression are poor by performance when compared to other algorithms with the dataset.

CHAPTER 5

CONCLUSIONS AND RECOMMENDATIONS

5.1 FINDINGS AND CONTRIBUTIONS

In developing countries like Bangladesh, road accidents are an awful problem. Not only Bangladesh road accidents are the major problem around the world. Recently researchers predict that by 2030, road accidents will be the fifth leading cause of death worldwide. Road accidents simultaneously adversely affect social and economic spheres. Losses in road accidents are unbearable, to the society as well as a developing country like us. So, it has become an essential requirement to control and arrange traffic with an advanced system to decrease the number of road accidents in our country. Moreover, it's a primary need for our country now, to tackle this situation where every day so many people were killed in a traffic accident and day by day this rate is getting increased. Artificial neural network is a functional and a great approach to take an accurate decision with the experience to manage the current situation and the findings of the analysis part can be suggested to traffic authorities for reducing the number of accidents.

Here, the author has analyzed the road accident dataset and shown a possibility of death in road accidents with the help of Deep learning and machine learning algorithms. After splitting the dataset into 70% for training and 30% for testing, the best accuracy is achieved by the Artificial Neural Network. Here we use four features for predicting accuracy that is Injured, Responsible car, victim car and Accident History.

5.2 LIMITATION

We know accidents are unpredictable so it is difficult to predict the death by using the previous year's accidents data.

5.3 RECOMMENDATIONS FOR FUTURE WORK

In this paper, further features can be added to identify the road accident and severity more specifically. Moreover, in future it will be very useful and beneficial.

5.4 References

- [1] Md. Farhan Labib, Ahmed Sady Rifat, Md. Mosabbir Hossain, Amit Kumar Das, Faria Nawrine. "Road Accident Analysis and Prediction of Accident Severity by Using Machine Learning in Bangladesh" 7th International Conference on Smart Computing & Communications (ICSCC), 2019.
- [2] Ali J. Ghandour, Huda Hammoud, Samar Al-Hajj. "Analyzing Factors Associated with Fatal Road Crashes: A Machine Learning Approach" National Council for Scientific Research (CNRS), 2020.
- [3] Sahil Dabhade, Sai Mahale, Avinash Chitalkar, Pushkar Gawhad, Vicky Pagare "Road Accident Analysis and Prediction using Machine Learning." International Journal for Research in Applied Science & Engineering Technology (IJRASET), 2020, Vol 8, Issue I, pp 100-103.
- [4] Arun Venkat, Gokulnath M, Guru Vijey K. P, Irish Susan Thomas, D.Ranjani "Machine Learning Based Analysis Based Analysis for Road Accident Prediction." International Journal of Emerging Technology and Innovative Engineering, 2020, Vol 6, Issue 02, pp 31-37.
- [5] <https://www.afro.who.int/publications/global-status-report-road-safety-2015>
- [6] N. Sridevi, M. V. Keerthana, Monisha V. Pal, T. R. Nikshitha, P. Jyothi "Road Accident Analysis Using Machine Learning." International Journal of Research in Engineering, Science and Management, 2020, Vol-3, Issue-5, pp 859-861.
- [7] P Sumanth, P Sai Anudeep and S Divya "Analysis of Machine Learning Algorithm with Road Accidents Data Sets." International Journal of Engineering and Management Research, 2020, Vol-10, Issue-2, pp 20-25.
- [8] Md. Sharif Hossen "Analysis of Road Accidents in Bangladesh." American Journal of Transportation and Logistics, 2019, pp 1-19.
- [9] Nicholas Fiorentini and Massimo Losa "Handling Imbalanced Data in Road Crash Severity Prediction by Machine Learning Algorithms" Multidisciplinary Digital publishing Institute (MDPI), 2020, pp 1-24
- [10] Paul, J., Jahan, Z., Lateef, K. F., Islam, M. R., & Bakchy, S. C. (2020). Prediction of Road Accident and Severity of Bangladesh Applying Machine Learning Techniques. 2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC). doi:10.1109/r10-htc49770.2020.9356987

