

# **KANTIPUR ENGINEERING COLLEGE**

**(Affiliated to Tribhuvan University)**

**Dhapakhel, Lalitpur**



**[Subject Code: CT755]**

**A MAJOR PROJECT FINAL REPORT ON  
"PROJECT X : A BUSINESS INTELLIGENCE FOR  
BIKE INSURANCE"**

**Submitted by:**

**Amit Dhoju [091/BCT/2071]**

**Kaushtup Bista [103/BCT/2071]**

**Roshi Maharajan [117/BCT/2071]**

**A MAJOR PROJECT SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE  
OF BACHELOR IN COMPUTER ENGINEERING**

**Submitted to:**

**Department of Computer and Electronics Engineering**

**July, 2018**

# **”PROJECT X : A BUSINESS INTELLIGENCE FOR BIKE INSURANCE”**

**Submitted by:**

**Amit Dhoju [091/BCT/2071]**

**Kaushtup Bista [103/BCT/2071]**

**Roshi Maharajan [117/BCT/2071]**

**Supervised by:**

**Ajay Mani Paudel**

**External Director**

**A MAJOR PROJECT SUBMITTED IN PARTIAL  
FULFILLMENT OF THE REQUIREMENT FOR THE DEGREE  
OF BACHELOR IN COMPUTER ENGINEERING**

**Submitted to:**

**Department of Computer and Electronics Engineering**

**Kantipur Engineering College**

**Dhapakhel, Lalitpur**

**July, 2018**

## ABSTRACT

There is no alternative for intelligence. What if intelligence used in gambling? the result is success. In short this is our project, title "**Project X:A Business Intelligence For Bike Insurance**". Our project describe a predict policy for bike insurance by visualization towards the history of bike insurance data. Our system will train for our intelligence and will be able to predict the possible claim. The insurance industry worldwide is facing the challenges of deregulation, consolidation and convergence of financial services. There is today a pressing demand for cutting edge services of insurance business management and enriched customer experiences at a significantly lower cost. Insurance is gambling and we are trying to give technique to win this gamble. Our system will determine the possible accident of bike so that management of insurance in-charge policy for new client whether to insurance or not. If system predicts the possible claim then system will suggest the insurance company not to make that insurance otherwise insurance is made. Based upon CC of Bike, Bike Manufacturer, Lote, Zone, and Type Cover, our system will predict the possible claim.

**Keywords**— Business Intelligence, Prediction, Analysis, Insurance

# ACKNOWLEDGMENT

Write Acknowledgment Here. Ea his munere torquatos, quidam essent luptatum cu pro.  
Ei duo scaevola

Second para of Acknowledgment. Sed veri aequae persecuti ut. Ut accusam mediocrem  
accusamus eos, quis

to display members name under Acknowledgement

Amit Dhoju	[091/BCT/2071]
Kaushtup Bista	[103/BCT/2071]
Roshi Maharajan	[117/BCT/2071]

# TABLE OF CONTENTS

<b>Abstract</b>	<b>i</b>
<b>Acknowledgment</b>	<b>ii</b>
<b>List of Figures</b>	<b>v</b>
<b>List of Abbreviations</b>	<b>vi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Insurance . . . . .	2
1.2.1 Definition of Non-Life Insurance . . . . .	2
1.2.2 Principle of Insurance . . . . .	2
1.2.3 Insurance Risk . . . . .	3
1.3 Problem Statement . . . . .	3
1.4 Objectives . . . . .	3
1.5 Application . . . . .	4
1.6 Project Features . . . . .	4
1.7 Feasibility Analysis . . . . .	4
1.7.1 Economic Feasibility . . . . .	4
1.7.2 Technical Feasibility . . . . .	4
1.7.3 Operational Feasibility . . . . .	5
1.8 System Requirement . . . . .	5
1.8.1 Software Requirement . . . . .	5
1.8.2 Hardware Requirement . . . . .	5
<b>2 Literature Review</b>	<b>6</b>
2.1 Sisense . . . . .	6
2.1.1 How Sisense work . . . . .	7
2.2 Qlik Sense . . . . .	8
2.2.1 How does Qlik sense work . . . . .	8
<b>3 Theory</b>	<b>9</b>
3.1 Bayes' Theorem . . . . .	9
3.2 Classification . . . . .	10
3.3 Bayesian Classification . . . . .	10

3.3.1	Learning Models . . . . .	11
3.4	Naïve Bayesian Classifier . . . . .	11
3.5	Data Pre-Processing . . . . .	12
3.6	Implementing Algorithm . . . . .	12
<b>4</b>	<b>Methodology</b>	<b>14</b>
4.1	E-R Diagram . . . . .	14
4.2	Flow Chart . . . . .	15
4.3	Use-Case . . . . .	16
4.4	Class Diagram . . . . .	17
4.5	Sequence Diagram . . . . .	18
4.6	Collaboration Diagram . . . . .	20
4.7	Software Development Process . . . . .	21
<b>5</b>	<b>Result and Discussion</b>	<b>23</b>
5.1	Works Completed . . . . .	23
5.2	Works Remaining . . . . .	23
5.3	Problems Faced . . . . .	23
5.4	Work Schedule . . . . .	24
	<b>References</b>	<b>25</b>

## LIST OF FIGURES

1.1	Insurance Risk . . . . .	3
3.1	Structure of Naïve Bayesian Network . . . . .	11
4.1	Entity Relationship Diagram . . . . .	14
4.2	Flow Chart Of Project . . . . .	15
4.3	Use-case Diagram . . . . .	16
4.4	Class Diagram . . . . .	17
4.5	Sequence Diagram: Training . . . . .	18
4.6	Sequence Diagram: Testing . . . . .	18
4.7	Sequence Diagram . . . . .	19
4.8	Collaboration Diagram . . . . .	20
4.9	Incremental Development Model Chart . . . . .	21
5.1	Gantt Chart . . . . .	24

## **LIST OF ABBREVIATIONS**

**BI** Business Intelligence

**CC** Cubic Centimeter

**CSV** Comma Seprated Value

**NN** Neural Network

**SQL** Server Query Language

**SVM** Support Vector Machine



# CHAPTER 1

## INTRODUCTION

### 1.1 Background

**Business Intelligence(BI)** is defined as the ability for an organization to take all its capabilities and convert them into knowledge. This produces large amounts of information which can lead to the development of new opportunities for the organization. When these opportunities have been identified and a strategy has been effectively implemented, they can provide an organization with a competitive advantage in the market, and stability in the long run.[1]

Business Intelligence in Insurance provides the information to insurance company on claim management of motorbike which helps in gain insight and visibility on the cost of claims; analyze claims breakdown by CC of Bike, Bike Manufacturer, Lote , Zone and Type Cover. To predict and analyze the insurance historical data based on the claims, data can identify suitable policies for risk modelling , reinsurance, profitability analysis, loss analysis, claim analysis, claim estimation by using the Back propagation algorithm to find the claim pattern according to CC of Bike, Bike Manufacturer, Lote, Zone, and Type Cover. Some insurers have gone for non-scalable temporary solutions, which often fail to leverage the ever-increasing volumes of data. Hence, recognizing the need for an effective business a data warehouse cannot be the answer to all the information requirements hence it is also very important to set clear business objectives for the business intelligence solution with total top management support.

## **1.2 Insurance**

A practice by which a company provides a guarantee of compensation for specified loss, damage, illness, or death in return for payment.

### **1.2.1 Definition of Non-Life Insurance**

Non-life insurance, also called property and casualty insurance, is a type of coverage that is very common and covers businesses and individuals. It protects them, monetarily, from disaster by providing money in the event of a financial loss. Before you purchase this type of insurance or if you already own any kind of non-life insurance, you should understand what it is.

### **1.2.2 Principle of Insurance**

1. Insurance transfers the financial consequences of an existing risk.
2. Law of Large Numbers
3. The more predictable the outcome
4. Pricing and risk management
5. Considerations in Setting Premium Rates

### 1.2.3 Insurance Risk

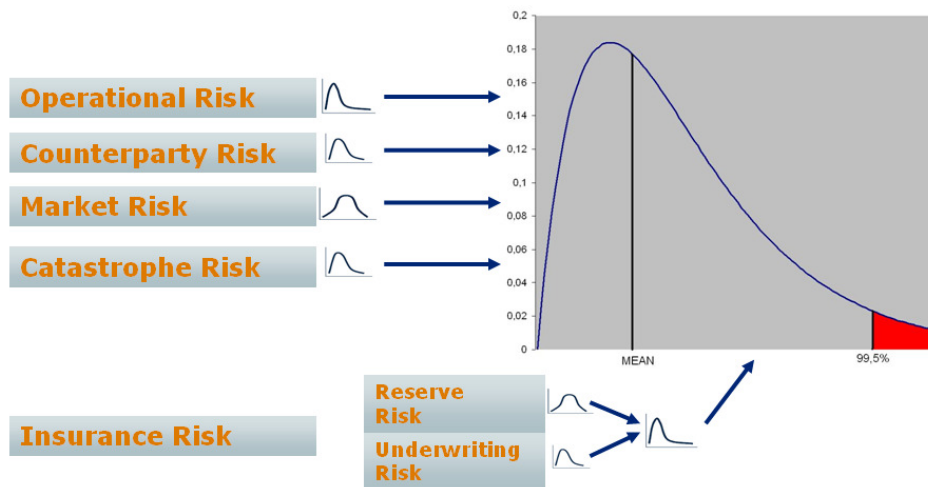


Figure 1.1: Insurance Risk

Source: <https://www.pinterest.co.uk/pin/534521049520207946/>

## 1.3 Problem Statement

The insurance industry worldwide is facing the challenges of deregulation, consolidation and convergence of financial services. There is today a pressing demand for cutting edge services of insurance business management and enriched customer experiences at a significantly lower cost. Our project aims to predict the risk behind bike insurance. It relates the information of the current customer and bike to the past data determining whether to make insurance or not.

## 1.4 Objectives

The objectives of this project are:

- To predict the possibility to claim of bike insurance.
- To visualize the insurance performance.
- To generate the claimed and unclaimed report.

## **1.5 Application**

Project-X is a web application which analyze and visualize the user data. The application of this is help the management level for policy making in insurance company. It provides visualization to the company. The visualization generate the report for claimed and unclaimed report separately with respective attributes. The report provides the management level with overall glance of the insurance to claim policy. The generated report is printable. Due to this , it required less human resource. In simple meaning it is decision support system for bike insurance company.

## **1.6 Project Features**

- Claimed and unclaimed report generation based on various attributes.
- Insurance performance can be visualized.

## **1.7 Feasibility Analysis**

### **1.7.1 Economic Feasibility**

The economic feasibility of the software refers to the cost of its creation, cost required to launch the system and its efficiency in terms of cost after its launch to real platform. It determine whether the project is economically and logically possible or not. This project reduces the operations and maintenance cost of IT infrastructures

### **1.7.2 Technical Feasibility**

This feasibility includes the software and hardware used for project.As this project will be developed using 'C sharp' as programming language and simply using the logical concept, it won't require sophisticated equipment. A PC with Windows Operating System will be able to run the software. However it may face some technical issues in data processing if the computer running this software doesn't meet the minimum hardware requirements such as processing power and RAM.Thus watching overall ,this project is

technically feasible.

### **1.7.3 Operational Feasibility**

A feasibility study aims to objectively and rationally uncover the strengths and weaknesses of a project. We will be gathering data from the insurance Company for the previous data related to the bike insurance. Besides gathering data from the company, our project provides user friendly platform, people with simple knowledge of computer can operate easily and the system can perform the rest operation.

## **1.8 System Requirement**

### **1.8.1 Software Requirement**

- i. Operating System : Windows OS
- ii. Framework : Microsoft .net framework
- iii. Database : Microsoft SQL Server 2014

### **1.8.2 Hardware Requirement**

- i. Processor: 1 GHz or faster (32-bit or 64-bit)
- ii. RAM: 1 GB

## CHAPTER 2

### LITERATURE REVIEW

Insurance is a practice by which a company provides a guarantee of compensation for specified loss, damage, illness, or death in return for payment. Bike Insurance is insurance for bike, motorcycle. Its primary use is to provide financial protection against physical damage or bodily injury resulting from traffic collisions and against liability that could also arise there from. Bike Insurance also provide financial support in case of bike stolen, damaged to bike from events other than traffic collisions , and keying and collision with stationary object. [2]

#### 2.1 Sisense

Sisense is Business Intelligence software by Sisense Inc., the industry in for complex data-easily prepare analyze and explore growing data from multiple sources. Sisense is awarded with **Best Business Intelligence Software for 2016**. It is the business analytics software that lets you easily prepare and analyze big, scattered datasets. it have 3 types of user

- i. Admin : Can access all features of Application.
- ii. Designer : Mainly to design the elastic cubes and all admin function except Serve and user management.
- iii. Viewer : Can only view data result and visualization.

The unique about Sisense are :

- 1. Single-Stack architecture: A single tool that helps you collect, prepare, organize and analyze data.
- 2. In-chip Engine and Proprietary technology.
- 3. Optimal use of computational resources.
- 4. A 90-minutes real test-drive for prospective client.

### **2.1.1 How Sisense work**

Sisense is webapp that work on both Computer or in smartphones. It works on the dashboard system with Elastic Cubes which is the proprietary analytical tools database that enables to connect multiple data sources and run complex queries in split of second. The data sources that can be used with sisense are file-based data sources (Excel ,CSV), Traditional database (SQL, MySQL, Oracle) and Online Web-Services(Google adwords). Data sources are added in elastic cube manager. Different types of data sources can be added in same cube. Any two field with same type of data on different data sources and be connected by drag and drop facility of sisense. At last, by building elastic cubes the data will be pull form data sources into elastic cube. When the result is ready it is ready for calculation and the result can be visualize in multiple form.

## **2.2 Qlik Sense**

QlikSense is software launch by Qlik software company. QlikSense lets you discover insights that query-based BI tools simply miss. One can freely search and explore across all his data, instantly pivoting your analysis when new ideas surface.[3]  
The unique about QlikSense are :

1. Full spectrum of visual analytic.
2. Externally governed assets and data modules.
3. Data compression.
4. Policy-based security rules

### **2.2.1 How does Qlik sense work**

Qlik Sense is platform of data analysis. It analyze data and make discovery on your own. One can easily share knowledge and data in organization using Qlik sense. It generates views on information for the user. It does not require predefined and static reports. One do not have to dependent on other resources just click and learn. Each click sends instant responds updating every visualizations and view in the app with new data and visualizations. App model help by asking question and user can answer the next question on their own without going back to expert for new report or visualizations for the data .



## CHAPTER 3

### THEORY

#### 3.1 Bayes' Theorem

Bayes Theorem is a statement from probability theory that allows for the calculation of certain conditional probabilities. Conditional probabilities are those probabilities that reflect the influence of one event on the probability of another event. The term generally used in Bayes theorem are prior probability and posterior probability. The prior probability of a hypothesis or events the original probability obtained before any additional information is obtained. The posterior probability is the revised probability of the hypothesis using some additional information or evidence obtained.[4]

Bayes Theorem can be written as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Where,

- i.  $P(A)$  is the prior probability of A ,
- ii.  $P(B)$  is the prior probability of B ,
- iii.  $P(A|B)$  is the posterior probability of A given B &
- iv.  $P(B|A)$  is the posterior probability of B given A

Bayes' Theorem for n independent variable:

$$P(B|A) = P(B_1|A) * P(B_2|A) * P(B_3|A) * ... * P(B_n|A)$$

## 3.2 Classification

Classification in data analysis is the task of assigning a class to instances of data described by a set of attributes. Classification or supervised classification includes the construction of a classifier which is trained on a set of training data that already has the correct class assigned to each data point. This builds a concise model of the distribution of class labels. It is then used to classify new data where the values of features are known but the class is unknown.

Many algorithms have been developed for supervised classification based on

1. Artificial Intelligence
2. Perceptron-based Techniques (Single and Multi-Layered Perceptron)
3. Statistical Learning Techniques (Bayesian networks, Instance based techniques)

## 3.3 Bayesian Classification

Bayesian Network provides a powerful graphical method for encoding the probabilistic relationship among a set of variables and hence and naturally be used for classification. It learned by using likelihood to achieve classification accuracy. It learned by supervised learning, where a training data set of instances with labels representing instance of classes is used to train as classifier.[4]  
Likelihood is calculated by:

$$C_{LL}(G|D) = \sum_{i=1}^N \log P(C_i|V_i)$$

### 3.3.1 Learning Models

Types of Learning models:

#### 1. Generative Model

Generative models summarize data probabilistically and are more flexible, since the user can bring in conditional independence assumptions, priors, and hidden variables. Generative classifiers learn a model of the joint probability of the variables and the related class label, and use Bayes theorem to compute the posterior probability of the class variable and make predictions.

#### 2. Discriminative Model

Discriminative models (e.g. NN and SVM) only learn from data to make accurate predictions by directly estimating the class posterior probability or via discriminant functions, and thus offer the user less flexibility in data representation and inference.

### 3.4 Naïve Bayesian Classifier

A Naïve Bayes Classifier is a probabilistic classifier based on applying Bayes theorem with strong independence assumptions. When represented as a Bayesian network, a Naïve Bayes classifier has the structure depicted in Figure below. [5] It shows the independence assumption among all features in a data instance.

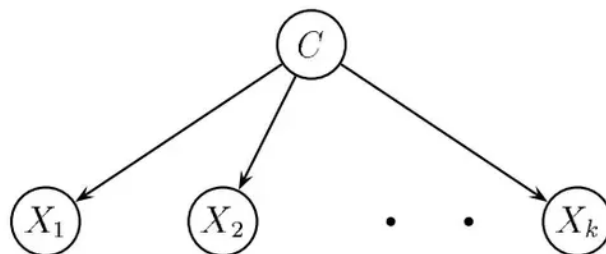


Figure 3.1: Structure of Naïve Bayesian Network

Source: <http://www.mdpi.com/1099-4300/19/6/247>

Naïve Bayes Classifier is simple probabilistic classifier that calculates a set of probability by counting the frequency and combinations of value in a given set of data. The algorithm uses Bayes theorem and assumes all attributes to be independent given the value of the class variable. Naïve Bayesian classifier is based on Bayes theorem and the theorem of total probability.

### 3.5 Data Pre-Processing

- Pre-process Steps
  - Data Cleaning
  - Data Integration and Transformation
  - Data Reduction
- Data in real world is dirty
  - incomplete: lacking attribute values, lacking certain attributes of interest, or containing only aggregate data
  - noisy: containing errors or outliers
  - inconsistent: containing discrepancies in codes or names
- No quality data, no quality mining result
  - Quality decisions must be based on quality data
  - Data warehouse needs consistent integration of quality data
  - Reduced Accuracy

### 3.6 Implementing Algorithm

- Problem Statement
  - Given Features :  $X_1, X_2, X_3, \dots, X_n$ .
  - Predict a label Y.
- Consider each attribute and class label as random variables.
- Given a record with attributes  $(A_1, A_2, A_3, \dots, A_n)$ .
  - Goal is to predict class C.
  - Specifically find value that maximizes  $P(C|A_1, A_2, A_3, \dots, A_n)$
- Compute posterior probability  $P(C|A_1, A_2, A_3, \dots, A_n)$  for all value of C using

Bayes theorem.

$$P(C|B) = \frac{P(A_1 * A_2 * A_3 * ..... * A_n|C)P(C)}{P(A_1 * A_2 * A_3 * ..... * A_n)}$$

- Choose value that maximizes  $P(C|A_1, A_2, A_3.....A_n)$
- Equivalent to choosing value of C that maximizes.  $P(A_1, A_2, A_3.....A_n|C)P(C)$
- Assume independence among attributes  $A_i$  when class is given:
  - $P(A_1, A_2, A_3.....A_n|C)P(C) = P(A_1|C_j)P(A_2|C_j)P(A_3|C_j)...P(A_n|C_j).$
  - Estimate  $P(A_i|C_j)$  for all  $A_i$  and  $C_j$ .
  - New point is classified to  $C_j$  if  $P(C_j) * P(A_i|C_j)$  is maximum.

# CHAPTER 4

## METHODOLOGY

### 4.1 E-R Diagram

The Entity Relationship Diagram of the project:

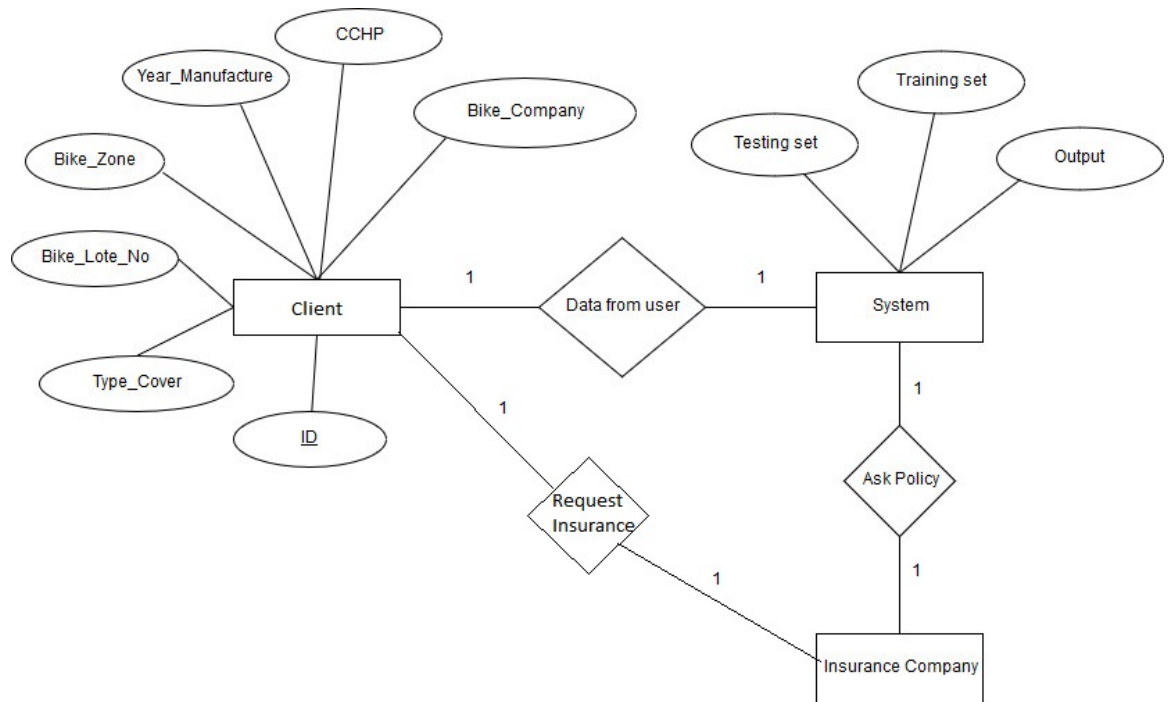


Figure 4.1: Entity Relationship Diagram

## 4.2 Flow Chart

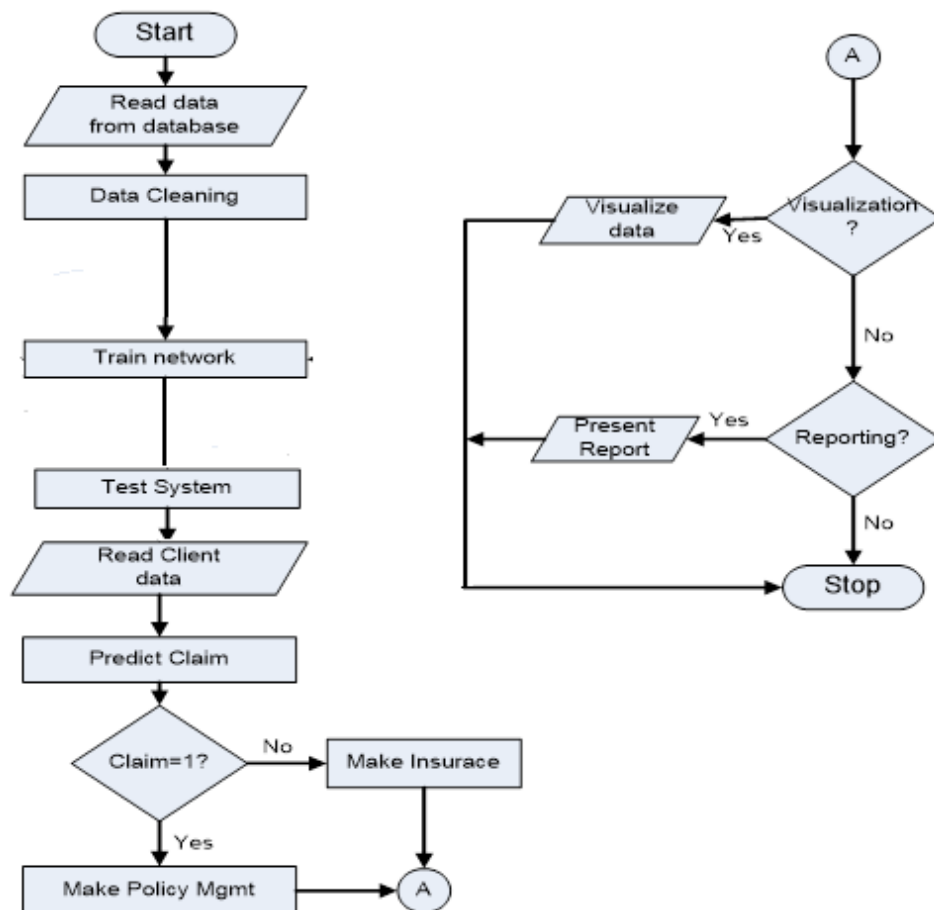


Figure 4.2: Flow Chart Of Project

### 4.3 Use-Case

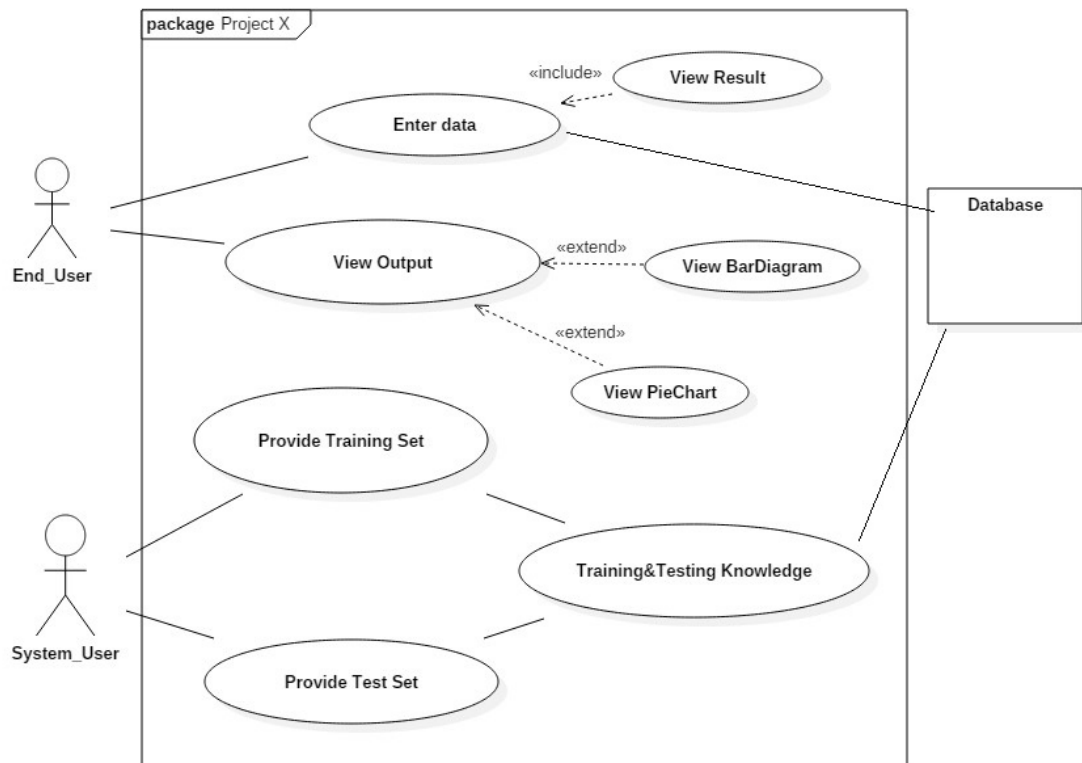


Figure 4.3: Use-case Diagram

In this project there will be three options. The customers information will be taken and based on which insurance to be done is decided. Report and Visualization of data based on various attributes can be done which helps to know the progress of an insurance company.



## 4.4 Class Diagram

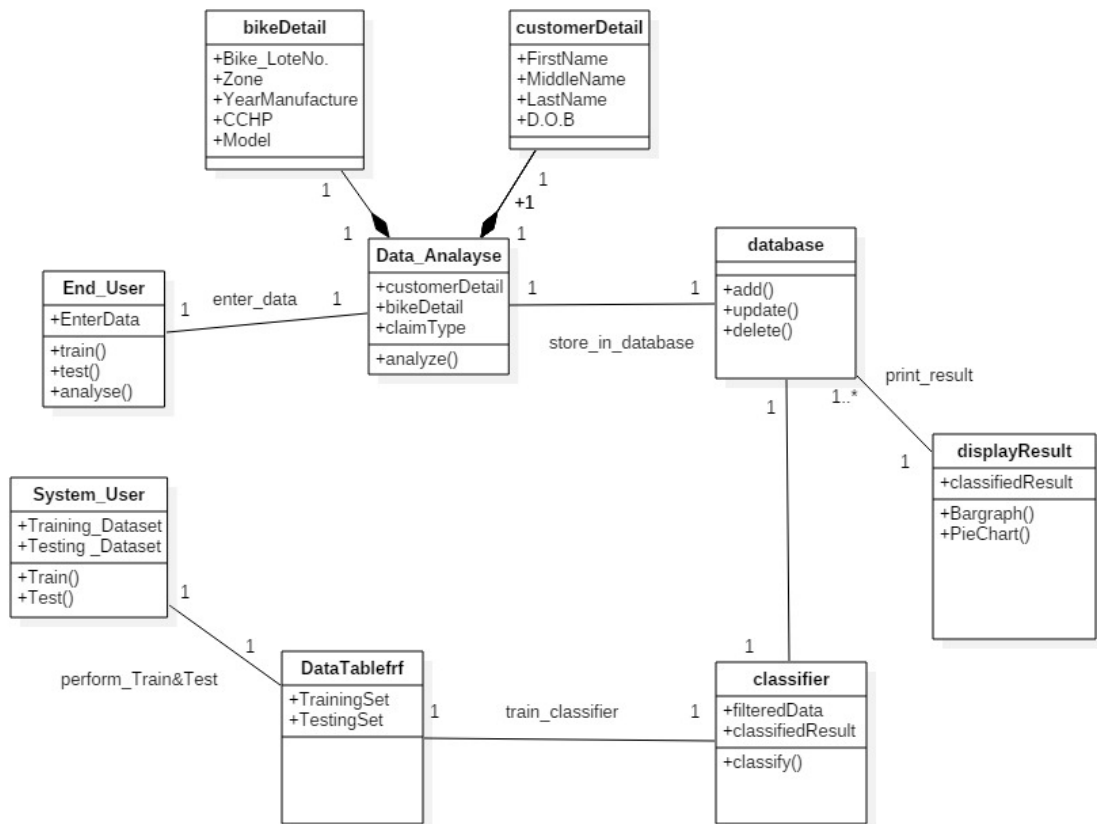


Figure 4.4: Class Diagram

## 4.5 Sequence Diagram

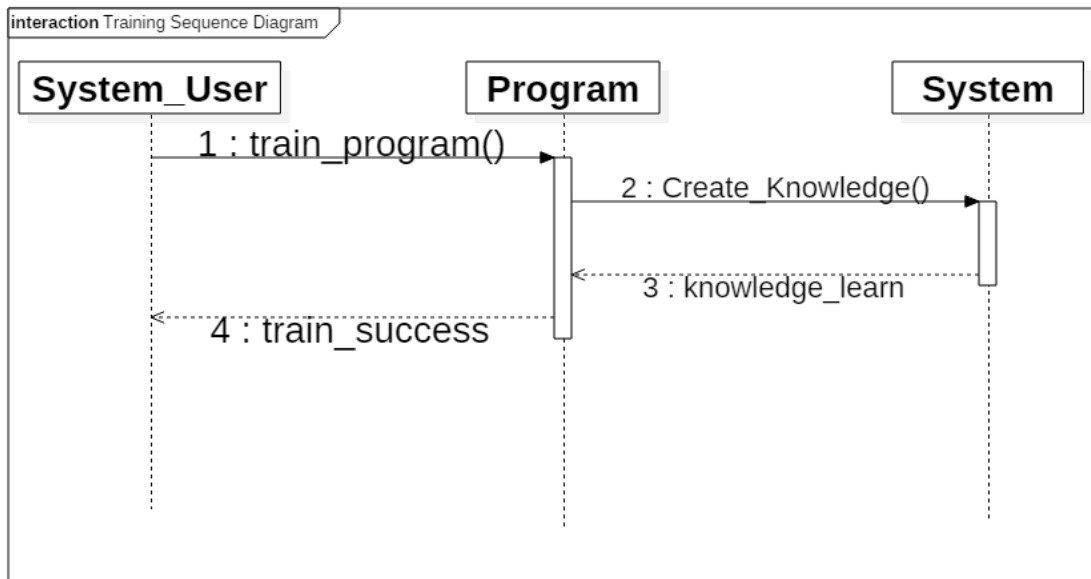


Figure 4.5: Sequence Diagram: Training

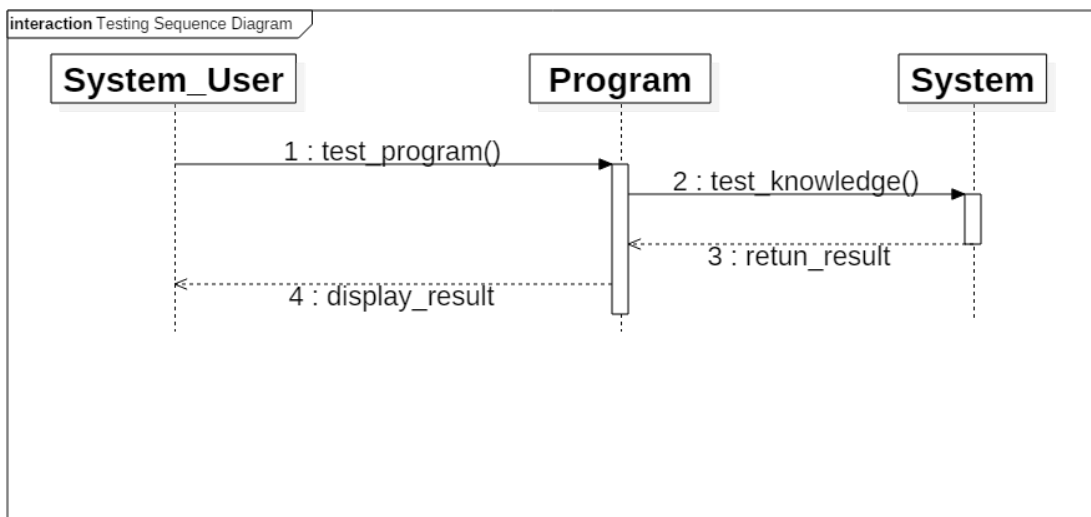


Figure 4.6: Sequence Diagram: Testing

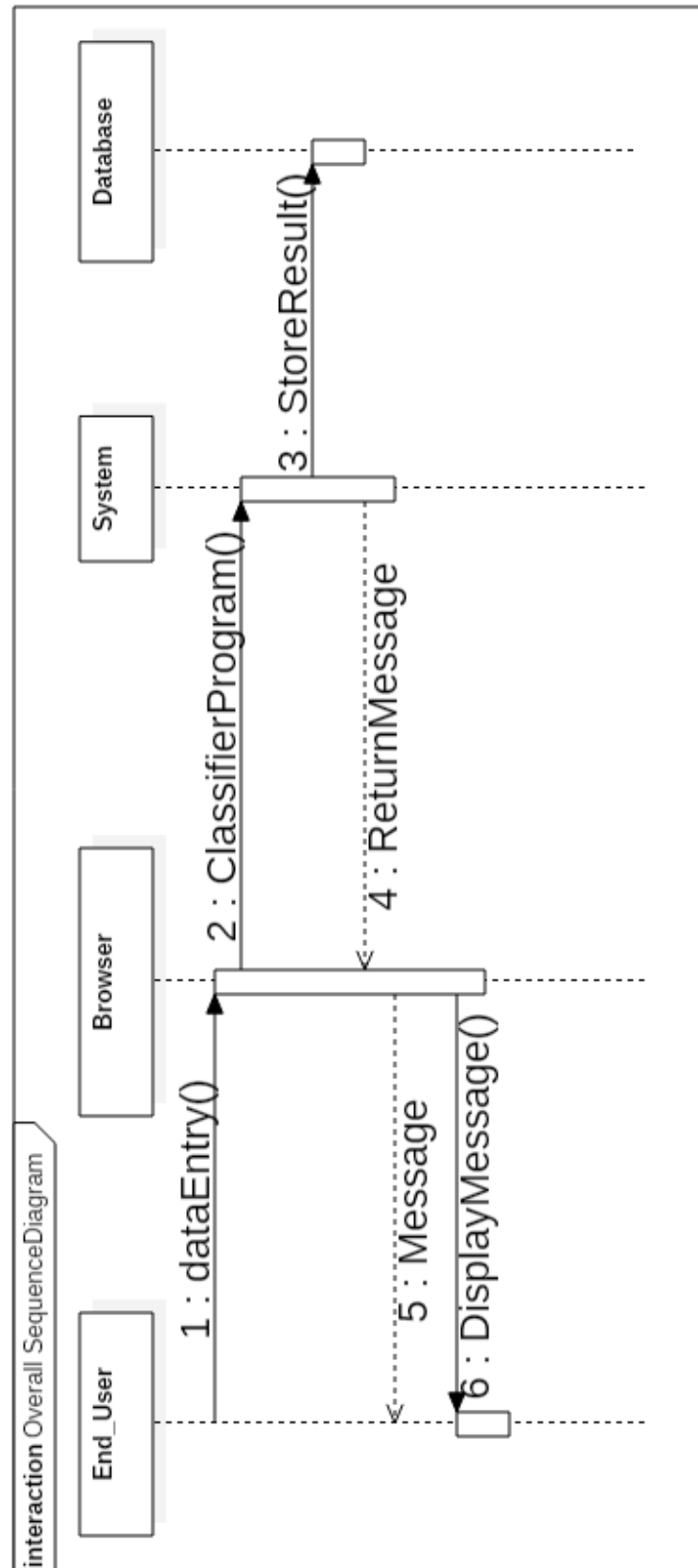


Figure 4.7: Sequence Diagram

## 4.6 Collaboration Diagram

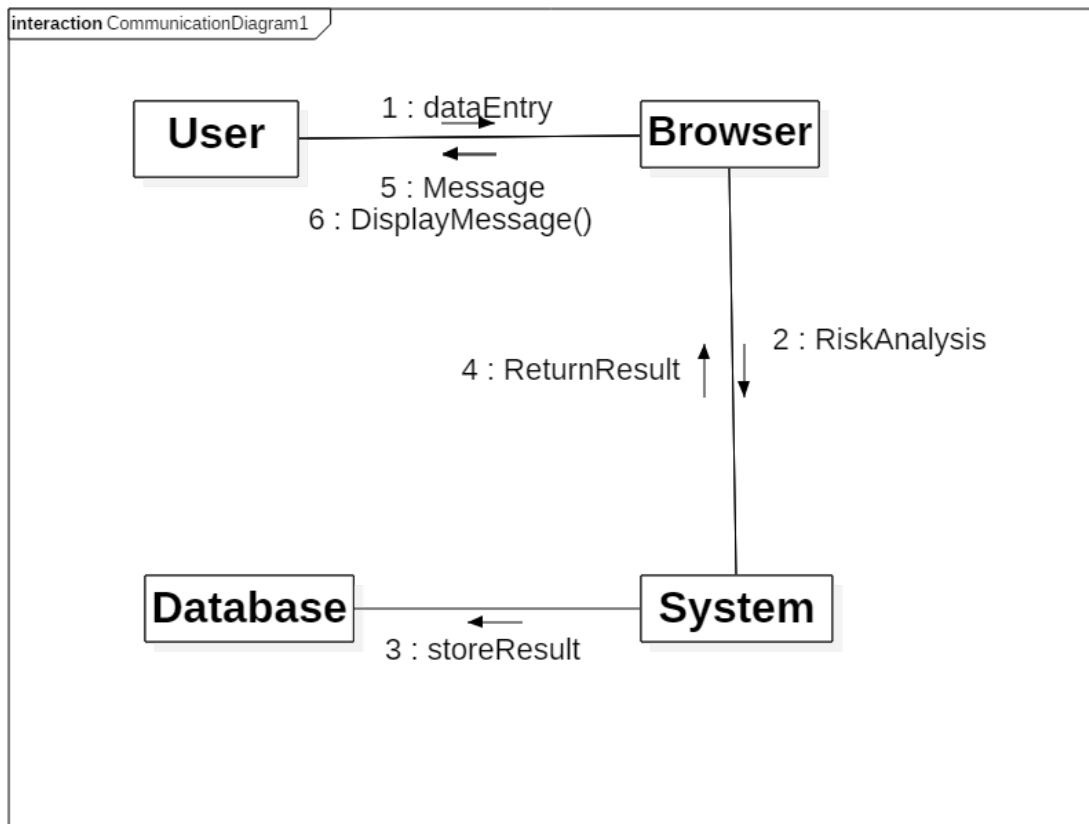


Figure 4.8: Collaboration Diagram

## 4.7 Software Development Process

Project-X is prepared by following approaches based on incremental software development model. The objective of following incremental model, it provides flexibility of starting project by available requirements, such that requirements can be added on the basis of later added projects objectives. One of the main feature of the model, it provides designing, testings phase in each increment.

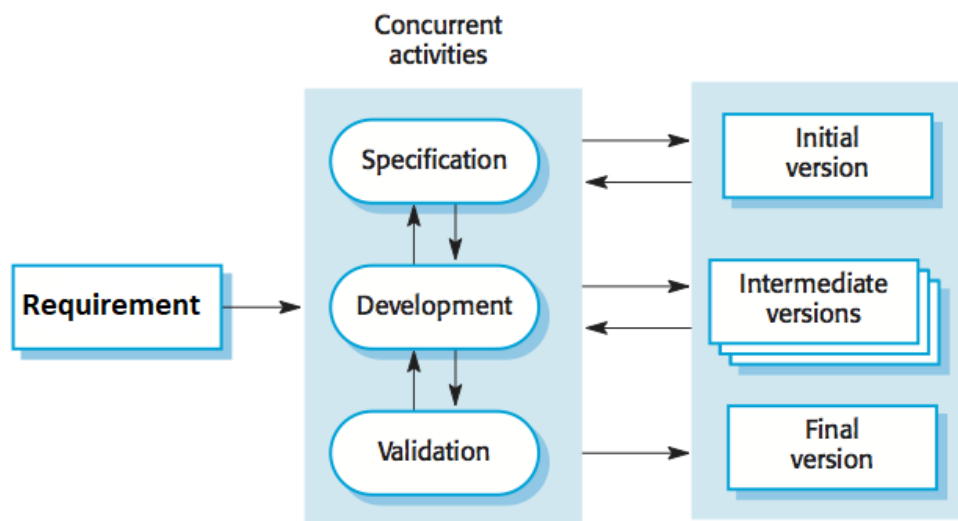


Figure 4.9: Incremental Development Model Chart

Source: <https://istqbexamcertification.com>

Project-X is application based on Naïve Bayesian Classifier for classifier data and predict the result for user. We mainly divide our project into 3 phase.

### 1. 1st Incremental Phase

As Initial phase, we mainly involved in researching about our project basic idea. We also visit company to get data insurance data that is vital for our project. We have done most of front end.

### 2. 2nd Incremental Phase

We started with implementing of algorithm and training and testing of data was done to find accuracy of data from collected data. We also get more data from company for further analysis of program.

### **3. 3rd Incremental Phase**

We completed almost of front end and back end. We tested our program with 70% accuracy and crystal report

## **CHAPTER 5**

### **RESULT AND DISCUSSION**

#### **5.1 Works Completed**

We have completed the User Interface of our project. We have also collected the data from Nepal Insurance Company located at Kamaladi, Kathmandu.[6] Data from the company is not as per our requirement so we have manually filtered the data. We have completed the testing and training of data. We also completed to analyze the data from user on basis of our training and testing set.

#### **5.2 Works Remaining**

We have yet to generate crystal report depending upon our data. And also to optimize the algorithm for better accuracy and front end.

#### **5.3 Problems Faced**

- Data filtering through various process.
- Limited data.
- Implementing the Algorithm.

## 5.4 Work Schedule

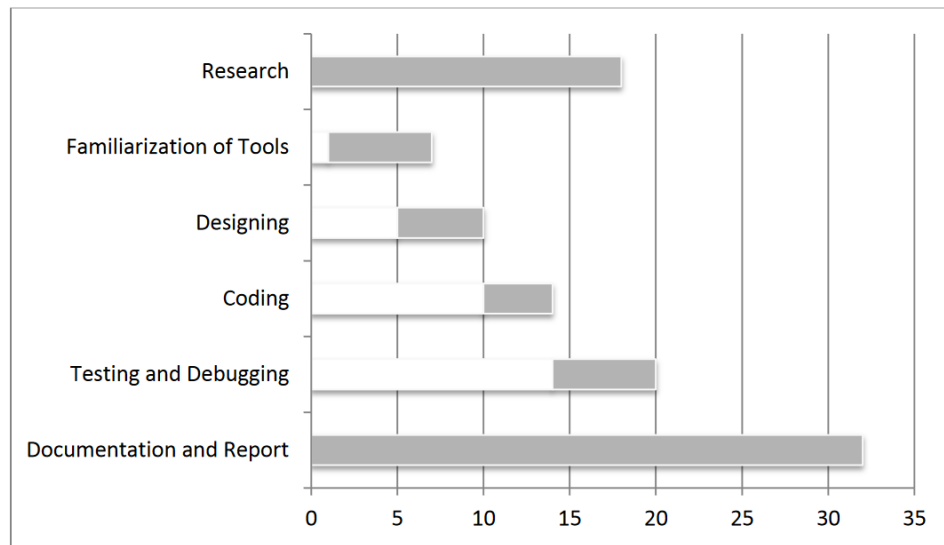


Figure 5.1: Gantt Chart



## REFERENCES

- [1] H. Chen, R. H. Chiang, and V. C. Storey, “Business intelligence and analytics: From big data to big impact.” *MIS quarterly*, vol. 36, no. 4, 2012.
- [2] S. R. Insurance, “Bike insurance,” 2015.
- [3] O. Troyansky, T. Gibson, and C. Leichtweis, *QlikView Your Business: An Expert Guide to Business Discovery with QlikView and Qlik Sense*. John Wiley & Sons, 2015.
- [4] B. Garg, “Design and development of naive bayes classifier,” 2013.
- [5] N. Friedman, D. Geiger, and M. Goldszmidt, “Bayesian network classifiers,” *Machine learning*, vol. 29, no. 2-3, pp. 131–163, 1997.
- [6] N. I. Company, “Insurance data,” Kamaladi, Kathmandu, Tech. Rep., 2017.