



**Daffodil**  
*International*  
**University**

**Thesis Title: An Ensemble method based architecture using  
random forest importance for employee's turn over prediction**

By

**Emran Hossain  
Student ID: 161-35-1609**

**Submitted to the Department of Software Engineering of Daffodil Internation  
University in partial fulfillment of the requirements for the degree of B.Sc.  
(Honors).**

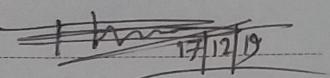
**Department of Software Engineering  
DAFFODIL INTERNATIONAL UNIVERSITY**

**Fall, 2019**

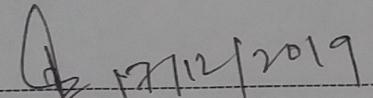
## APPROVAL

This thesis titled on "**An Ensemble method based architecture using random forest importance for employee's turn over prediction**", submitted by Emran Hossain, ID: 161-35-1609 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 approval as to its style and contents.

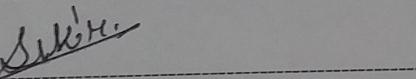
### BOARD OF EXAMINERS

  
17/12/19  
\_\_\_\_\_  
**Prof. Dr. Touhid Bhuiyan**  
Professor and Head  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

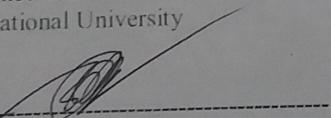
Chairman

  
17/12/2019  
\_\_\_\_\_  
**Dr. Md. Asraf Ali**  
Associate Professor  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

Internal Examiner 1

  
\_\_\_\_\_  
**Asif Khan Shakir**  
Lecturer  
Department of Software Engineering  
Faculty of Science and Information Technology  
Daffodil International University

Internal Examiner 2

  
\_\_\_\_\_  
**Prof Dr. Mohammad Abul Kashem**  
Professor  
Department of Computer Science and Engineering  
Faculty of Electrical and Electronic Engineering  
Dhaka University of Engineering & Technology, Gazipur

External Examiner

---

## DECLARATION

hereby declare that this thesis has been done by **Emran Hossain** under the supervision of **Mr. Md. Anwar Hossen, Lecturer (Senior Scale)**, Department of Software Engineering, Daffodil International University. It also declare that nither this thesis nor any part of this has been submitted elsewhere for award of any degree.

Emran

**Emran Hossain**

Student ID: 161-35-1609

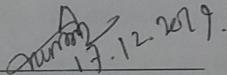
Batch: 19

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

Certified by:

  
17.12.2019.

**Mr. Md. Anwar Hossen**

Lecturer (Senior Scale)

Department of Software Engineering

Faculty of Science & Information Technology

Daffodil International University

©Daffodil International University

## Daffodil International University

Faculty of Engineering

Department of Software Engineering

N B Please fill up this form and get clearance from accounts department before attending the final defense and this is mandatory

Accounts Confirmation Form	
Student name and ID	1. Emran Hossain 2. 161-38-1609
Is he paid his/her money?	Yes ..... No .....
<p style="text-align: right;">Signature Authority of the Accounts Daffodil International University</p> <p style="text-align: right;">Paid Up to Internship DIU Finance &amp; Accounts</p> <p style="text-align: right;">Emran Hossain /21212</p>	



# Daffodil International University

Daffodil International University Library  
Daffodil Tower, (DT)-3, 3<sup>rd</sup> Floor  
102/1, Shukrabad, Dhanmondi - 1207  
Tel: 9116774 (Ext-123, 150,151)

Library Confirmation Form		
Name of the Student	Emran Hossain	
Student ID	161-35-1609	
Group	<input type="checkbox"/> Yes	<input checked="" type="checkbox"/> No
Group IDs		
1. Project Title	An Ensemble method based architecture using random forest info for employee turn over Prediction	
2. Internship Title		
Submission of Soft Copy of Reports	<input type="checkbox"/> Yes	<input type="checkbox"/> No
Name and Designation of the Project Supervisor	Signature of the Project Supervisor	
Md. Anwar Hossen Senior Lecturer		
Name and Designation of the Internship Supervisor	Signature of the Internship Supervisor	

This is for your kind information that the management of DIU has decided to receive students' Project/Thesis Reports by DIU Library through this email ([projectreport@diu.edu.bd](mailto:projectreport@diu.edu.bd)) to check Plagiarism by Turnitin Software before submitting to the departments. Students have to submit a plagiarism checking report provided by the DIU library with their Project Report/ Thesis to the respective departments.

Acceptable range of plagiarism at DIU has been settled as follows:

- a) Project/ Thesis report of undergraduate students – 50%
- b) Project/ Thesis report of Masters students – 40%

Only the acceptable reports will be submitted for further processing.

Actual plagiarism: ~~38~~ %

Name and Signature \_\_\_\_\_  
Concerned Library Officer \_\_\_\_\_  
Daffodil International University

---

## Abstract

Employee turnover means we understand the departure of an employee from a company and is defined as the number or percentage of employees who are replaced by new employees. Companies face many problems as employees leave and they are financially disadvantaged in hiring new employees. When the employees leave, the company has to train new employees, and the company has to face many problems.

The main purpose of this paper is to predict the turnover of employees with the help of a logic machine. Currently turnover of staff has become an important issue. Therefore, it is possible to solve this problem using machine learning method. Employee turnover has become an important issue these days due to heavy workload, low pay, low job satisfaction, poor working environment

We have determined employee turnover selection factors using some prediction models. We first pre-processing the data and run it with all the features of the model, in which our prediction comes 99.03. Secondly, sequencing has been used to reduce features from a high number to a relatively small signal-canton using a Sequential backward selection algorithm (SBS). Then use Chi2 and Random Forest important algorithms to determine the common key features. From which employee turnover can be regarded as the most salient feature.

After our second steps, we basically get three features that are responsible for the employee's departure. The features we have selected have been tested with machine learning art algorithms such as Random Forest, Decision Tree, Support

Vector Machine(SVM), K-Nearest Neighbor(kNN), Multi-Layer Perceptron(MLP).

**Keywords:** Component. Machine learning. SBS, Chi2, Random Forest, Decision Tree, SVM, KNN, MLP.

---

## Acknowledgment

First of all, I thank God Almighty who created me, and to this day, has given me hope and saved me from all kinds of danger. My parents have played a big role in bringing me this far. They have supported me through tough times and help me build my career in Software Engineering Sector. I have been able to stand here today with God's undeserved kindness and the blessings of my parents.

I am grateful to Md. Anwar Hossain Sir who has helped me so much since coming to Varsity. He helped me a lot in my thesis. He always gave us the guide line so that by the grace of Allah I was able to finish my work. He would tell me how to work when needed.

Many thanks to my friends and older brothers because they helped me with all kinds of work. They encouraged me in all kinds of work.

Emran Hossain

December, 2019

---

# Table of Contents

<b>Abstract</b>	v
<b>Acknowledgment</b>	vii
<b>Chapter 1 Introduction</b>	1
1.1 Overview . . . . .	1
1.2 Motivation . . . . .	1
1.3 Problem Statement . . . . .	2
1.4 Research Question . . . . .	2
1.5 Research Objectives . . . . .	3
1.6 Contributions . . . . .	3
1.7 Organization of the Thesis . . . . .	4
<b>Chapter 2 Background Study</b>	5
2.1 Overview: . . . . .	5
2.2 Data Collection . . . . .	6
2.3 Data Pre-Processing . . . . .	6
2.3.1 Label Encoding . . . . .	7
2.3.2 Missing Value Imputation . . . . .	7
2.3.3 Feature Scaling . . . . .	8
2.3.4 Correlation Matrix . . . . .	9

<b>Chapter 3 Literature review</b>	<b>10</b>
3.1 Introduction . . . . .	10
3.2 Motives of Employees Turnover . . . . .	10
3.2.1 Deficiency of Employee Motivation . . . . .	10
3.2.2 Work Environment . . . . .	12
3.2.3 Render . . . . .	15
3.2.4 Job Dissatisfaction . . . . .	17
3.2.5 Deficiency of Training and Development . . . . .	18
3.2.6 Job Satisfaction on Employees . . . . .	20
<b>Chapter 4 Methodology</b>	<b>23</b>
4.1 System Model . . . . .	23
4.1.1 Feature Selection . . . . .	24
4.1.2 Sequential Backward . . . . .	24
4.1.3 Chi Square . . . . .	25
4.1.4 Random Forest Important . . . . .	25
4.1.5 Feature Selection . . . . .	26
4.2 Build predictive models Using algorithm . . . . .	28
4.2.1 Random forest. . . . .	28
4.2.2 Decision Tree classifier . . . . .	29
4.2.3 support vector classifier . . . . .	29
4.2.4 Multi layer Perception: . . . . .	30
4.2.5 Gaussian Naive Bayes . . . . .	30
4.2.6 k-Nearest neighbor : . . . . .	31
4.2.7 Hyper Parameter Tuning . . . . .	32
4.2.8 K-fold cross-validation . . . . .	32
4.2.9 Ensemble method . . . . .	33
4.2.9.1 Bagging . . . . .	33

4.2.9.2    Boosting . . . . .	34
<b>Chapter 5 Experiments Analysis</b>	<b>35</b>
5.1 Preprocessing Result: . . . . .	35
5.2 Feature Selection Result: . . . . .	36
5.3 Experiments . . . . .	38
<b>Chapter 6 Conclusions and Future Work</b>	<b>43</b>
6.1 Conclusion : . . . . .	43
6.2 Future Work : . . . . .	44

---

## List of Figures

2.1	Dataset before the label encoding . . . . .	7
2.2	After applying the label encoding, the Height column will be converted: . . . . .	8
4.1	random Forest Feature Importance . . . . .	26
4.2	System Model . . . . .	27
5.1	Dataset before the label encoding. . . . .	35
5.2	Dataset after the label encoding. . . . .	36
5.3	The number of features after applying SBS. . . . .	37
5.4	ROC curve before all feature selection. . . . .	39
5.5	ROC curve before all feature selection. . . . .	40

---

## **List of Tables**

4.1	Summary of the Data . . . . .	28
5.1	Random forest feature importance scores . . . . .	36
5.2	Chi2 scores . . . . .	37
5.3	Measure performance using algorithms by selecting all features . . .	38
5.4	Measure performance using algorithms by selecting reduce features	41
5.5	Measure performance using algorithms by 10 fold Cross validation .	42
5.6	Accuracy performance measurement of ensemble algorithms . . . .	42

# **Chapter 1**

---

## **Introduction**

### **1.1 Overview**

Employee turnover means we understand the departure of an employee from a company and is defined as the number or percentage of employees who are replaced by new employees. Companies face many problems as employees leave and they are financially disadvantaged in hiring new employees. Companies face many problems as employees leave, if the experienced employees leave the company, then there is a lot of work disruption and the company has to recruit new staff and train them, causing financial loss to the company. At present, turnover of employees very common phenomenon. Conversely, employee's retention is a vital issue and challenge to all organisation currently.

### **1.2 Motivation**

When I read the thesis paper and the I see the HR data set I could see that each paper has only prediction. Then I thought it would be much better if I could build apps with him and increase his prediction. Thinking about this, I was motivated to do a thesis paper.

### 1.3 Problem Statement

Employee turnover at high rates is a serious challenge for the organization. As a person based, the firm's employee turnover rate has grown at a ruthless rate of 30 percent in the current year. Worker turnover can be detrimental to the efficiency of an organization if skilled workers often leave the organization and have a high percentage of highly educated workers in the working population (Armstrong, 2009).

Basically, employee turnover damages an organization's earnings rate, which, in turn, affects its profitability . The are various statement is the factors that results into increase in employee turnover and with the increase in the turnover it is denote that employees are not satisfied with the job as well as with the company

According to Maxwell (2010), one of the key issues for employee turnover is that the company has lost the most experienced and skilled employees who were heavily invested in training on the task functions of different firms. Every year, organizations face many problems due to employee turnover. When employees leave, the company faces many problems that overcoming it becomes a major challenge for a company. By doing so, new employees have to be recruited in the companies and they have to re-train, which is a huge cost for the establishment. As a result, the company has a significant decay effect(Premeaux, 2000). On the other hand, there are some specific reasons why an employee should leave lack of opportunities and challenges for career development, job dissatisfaction, conflict with other staff members or management and heavy workload, poor working environment, low pay etc. For these reasons employees are frustrated.

### 1.4 Research Question

- Why is it related to employee turnover and each of the selected factors?
- Why do employees leave the company and what are the reasons?

- Why does the company do a lot of damage to employees when they leave, and what kind of problems do companies face when they leave?

## 1.5 Research Objectives

- The purpose of the proposed study is to demonstrate that the organization's culture, wage scale, fair value, career developers' challenges are directly or indirectly related to employee turnover among these factors.
- Choosing the reasons why high employees are headed towards turnover. Employee turnover has become an important issue these days due to heavy workload, low pay, low job satisfaction and poor working environment.
- Companies face many problems as employees leave, if the experienced employees leave the company, then there is a lot of work disruption and the company has to recruit new staff and train them, causing financial loss to the company.

## 1.6 Contributions

- Employee turnover is now a major problem for companies. This study uses machine learning to explain why employees leave the company. Employees leave the company due to some kind of problem. The company will be able to solve the problem very quickly so that they can understand causing of employee leave. We will use some of the tools in this study that will allow us to accurately calculate the percentage of employees leaving. So that the company does not have to face any problem to decide.
- Companies face many threats due to high turnover and low productivity improvements. Employers aware of these symptoms support job development, job training,

and leadership development before applying the strategy: This study is designed to address the primary causes of contributing to dissatisfaction.

In this study we will create an application to solve employee turnover very quickly.

## 1.7 Organization of the Thesis

The treatise is organized as follows:

- • **Chapter 1 Introduction.** In this chapter an introduction to the periodic patterns mining researches is presented. The definition, importance and existing approaches are clearly introduced. Next, the research contribution focuses.
- • **Chapter 2 Background Study.** in this chapter comes background study where we discuss our Learning techniques and algorithms with related works.
- • **Chapter 3 Literature review on Employee turnover.** In this chapter , we will discuss how different researchers have expressed their views on employee turnover.
- • **Chapter 4 Method.** in this chapter we will get into the machine learning training session. In this section we will train our agent and analyze all the training data and optimize our training technique
- • **Chapter 5 Experiments Analysis.** In this chapter, it has been shown the effectiveness and efficiency of our proposed method.
- • **Chapter 6 Conclusion and Future Work.** Finally, this chapter concludes the dissertation indicating the limitations and future works.

## **Chapter 2**

---

### **Background Study**

#### **2.1 Overview:**

Employee turnover is the number of permanent employees leaving the company during the reference period with respect to the number of actual permanent employees active on the last day of the previous reporting period.

- **There are five reasons to which employee leave the organization:** 1.Dissatisfaction at work is one of the main causes of absenteeism. Robbins and Decenzo (2001) state that a person with high job satisfaction has positive attitudes towards work. People with negative attitudes are absent. In a study conducted at a military health care center in Mullins (2005), it was discovered that there was a positive relationship between job dissatisfaction, staff turnover and levels of absenteeism.

2. According to Moureen (2004), the work environment is also a major cause of employee turnover. Employees prefer to work in an environment that suits them. This is the most common reason why they jump from one company to another in a few months. If they find a suitable work environment in a specific company, they can work in the same organization for several years.

3. The important reasons for turnover are "exit rate" and "employment stability". These studies were conducted in the beginning of the twentieth century and since then employee turnover has become a widely discussed topic (Heavey et al., 2013).

4. Another reason may be that employees are not satisfied with the organization. There are some instances when employees leave the company even when they pay competitive salaries. It is the responsibility of the organization itself. Employees do not approve of management style and they are not satisfied with the organization's culture. Because of this, the employees leave the company. Chen et al. (2010) have investigated the effect of improving job satisfaction among employees, and the research has proven that each of these measures has a positive relationship with decreasing employee turnover rate.

5. Man's nature is to seek growth. If you feel that there is no room to grow in a particular organization, you will look for a place where you can grow more. However, this is not a factor for everyone. There are those who are happy to do the same even after ten years

## 2.2 Data Collection

Data collection method can either be quantitative and qualitative data, quantitative is used as a “synonym for any data collection techniques (such as a questionnaire or data analysis procedure such as graphs and statics ) that generates or uses numerical data “ (Saunders et al., 2009) in contrast qualitative is used as a synonym for any data collection techniques such as an interview or data analysis procedure such as data that generated or use non numerical data” (Saunders et al., 2009).

We took the research data set we used from Kaggle and then we processed the data set.

## 2.3 Data Pre-Processing

Data processing is the process by which raw data is converted into meaningful information. Data is technically manipulated to produce accurate results that either

lead to a problem solving or an existing task to improve.

Data preparation is referred to as pre-processing where raw data is cleared. Raw data from the data set is closely examined to address any errors. The purpose of this step is to eliminate bad data (unnecessary, incomplete or incorrect data). Creating high-quality data through it.

### 2.3.1 Label Encoding

We call the label encoding the designation of converting labels into numbers. So that the machine can convert labels to readable forms. This allows machine learning algorithms to better understand how those labels should be handled in a better way. Finally, label encoding is an important pre-processing step for the dataset.

#### Example:

Suppose in some data sets we have column height.

Height
Tall
Medium
Short

Figure 2.1: Dataset before the label encoding

### 2.3.2 Missing Value Imputation

Data is rarely clear on datasets in the real world. Typically, such data is incomplete, noisy, and irrelevant. The important task of any data scientist is to redistribute

Height
0
1
2

Figure 2.2: After applying the label encoding, the Height column will be converted:

data by meeting incomplete, noisy values. Using these data can lead to inaccurate prediction or classification for any model. The goal of this article is to fill in the missing values in the dataset.

To remove the missing values, we will first check whether there is a missing value in the dataset. If there is a missing value, we will calculate it and add a value. If we do not fit a missing value, my algorithm will not be properly classified.

### 2.3.3 Feature Scaling

Dataset properties, units, and ranges can have very different features. To overcome this problem we need to bring all the features to the same level. This can be obtained by scaling the properties. Feature Scaling is a technique to change the unique properties of data present in a particular range. The results can vary greatly when using the feature scaling algorithm. The feature scaling method is used to handle data values or units. If feature scaling is not done, then a machine learning algorithm considers higher values as lower and larger values as higher values.

- Youn, E. Jeong, M. K. (2009). "Class dependent feature scaling method using naive Bayes classifier for text data mining".

Example: If we do not use the algorithm feature scaling method, it can hold 500

grams value greater than 5 kg but this is not true and in this case, the algorithm will give incorrect prediction. So we can solve this problem by using feature scaling to bring all values to the same level.

### 2.3.4 Correlation Matrix

Data within a dataset can be related for several reasons.

- One variable may or may not depend on the values of the other variables.
- One variable may be mildly associated with another variable.
- The two variables can depend on any of the third variables.

We remove the correlation matrix in order to eliminate if one variable has a high or low quality relationship between the other variables.

# **Chapter 3**

---

## **Literature review**

### **3.1 Introduction**

Employee turnover is an important case-study phenomenon. There is a huge literature on the causes of employee turnover. Many models contribute to turnover by developing, and researchers are predicting different means for why employees leave the organization. Researchers who have examined are based on only one or more variables that result in small changes in employee turnover. Another criticism of the turnover study is because they do not adequately predict the Psychological processes involved in individual employee turnover decisions.

### **3.2 Motives of Employees Turnover**

#### **3.2.1 Deficiency of Employee Motivation**

Armstrong, (2003) thinks that motivating an employee can be one of the biggest challenges for a company. Employee motivation is the key to a company's effectiveness. Fair promotion based on performance encourages employees greatly. If the company fails to do so, then the risk of employee turnover increases.

George, (25), states that clearly defining appropriate job descriptions can impact employees 'development of responsibilities, thus increasing employees' motivation.

### **3.2 MOTIVES OF EMPLOYEES TURNOVER**

---

DuBois (25) argues that the lack of rewards for hard-working employees due to the management of organizations that are not committed to praising and recognizing qualified employees reduces the morale of most employees.

Gollweiser, (1999) Team building enhances employees' motivation because it provides employees with a pleasant environment that helps employees interact and learn with each other. There are many theories related to motivation such as essential theory, expectation theory, goal setting theory. Each theory motivates to rise to a higher level. When employees believe they can accomplish a task, they believe that they are capable of performing the task at a high level.

Maureen, (2004) hypothesized that the application of good communication channels helps in establishing strong relationships and interconnections among employees of the organization and serves as a good motivating factor.

Trevor(25) states that if a company's equity theory is advised by someone and provides a high level of motivation by ensuring that employees believe in those results. Goal setting theory suggests that specific and difficult goals lead the employee towards higher motivation and success. Thomas (24) affirms that Abraham Maslow's theory is that people have desires that influence their behavior.

According to Samwel S and Chipuza C (2009) they conducted a study on employee retention and rotation: the use of the motivational variable as a panacea. A study was conducted in South Africa and two public sector organizations and two private sector organizations were examined to investigate to what extent the identified intrinsic and extrinsic motivation variables influenced the maintenance and reduction of employee turnover in industry organizations public and private. . The aim of the research was to achieve the following objectives: to identify and establish the main intrinsic and extrinsic motivational variables used by selected organizations in the public and private sector to retain their employees; determining to what extent the intrinsic and extrinsic motivational variables identified influence employee retention and rotation in selected organizations; and making rec-

### **3.2 MOTIVES OF EMPLOYEES TURNOVER**

---

ommendations to the management of selected organizations on how to effectively retain employees and reduce turnover. The study adopted the research project of the transversal survey, investigating to what extent the motivational variables selected influence the decision of the employees to stay or leave an organization.

Murphy (2009) stated that the application of ineffective performance evaluation systems influences the reward of suffering employees, since they are classified as the best and this discourages most of the employees whose efforts are not recognized. In the motivation equation, contribution, performance and result are key factors that contribute to high motivation. Entries are all that an employee contributes to work or organization, such as time, effort, education and experience. The results are all that an employee gets from a job or an organization, such as payments, job security and benefits. Organizations take on input. High levels of performance contribute to the efficiency, effectiveness and overall objectives of the organization. Managers use the results to motivate people to contribute with contributions.

#### **3.2.2 Work Environment**

Employees should have the necessary equipment to perform their duties. These include proper accessories, equipment and computer technology, as well as adequate lighting, designated workplaces and proper seating. Due to these factors, poor working environment leads to lower productivity and job dissatisfaction. In the event that employees fail to function properly, they eventually end up (Bratton, 2003).

According to Moureen (2004), the work environment is also a major cause of employee turnover. Employees prefer to work in an environment that suits them. This is the most common reason why they jump from one company to another in a few months. If they find a suitable work environment in a specific company, they can work in the same organization for several years.

### 3.2 MOTIVES OF EMPLOYEES TURNOVER

---

Homer, (2007) argues that a safe work environment leads to a higher level of employee satisfaction at work and helps creation retain employees for a long time. Unsafe work environments, such as non-standard work environments, have a wider range of types and styles. Some non-standard environments have a lot in common with standards. For example, in the case of a construction site, there is no activity without workers and is generally empty when no one works. On the other hand, this environment is often external and temporary, two factors that separate it from a common office.

Horton (2007) has argued that the work environment is often described as good or bad. A good environment is a place where workers feel comfortable and appreciated. Workers in these types of environments tend to be more productive and happy. A bad working environment is a place where the worker feels undervalued, threatened or restless. Due to the nature of these environments, there is often a high turnover rate of workers and workers generally fail to reach their potential.

Spencer (2001) Inadequate furniture and work equipment reduces the level of employee satisfaction and this plays an important role in achieving higher employee turnover rates. Company policy relating to employee behavior expectations must be created and applied in accordance with the law. Any complaint by employees or customers must be dealt with promptly and in accordance with the law. Employees who behave in such a way as to endanger any person should be warned or dismissed, if necessary, both to follow the legal management of the situation and to ensure a safe work environment for all (Spencer, 2001).

Breugh (2000) has argued that the lack of open office design leads to an isolated work environment that reduces the ability of employees to interact freely with others and, therefore, affects cases of staff turnover. Furthermore, in order for the work environment to be adequate, the organization must be prepared in the event of an emergency situation. A first aid kit and emergency food and water supplies must be stored in an accessible way with periodically changed items to ensure fresh-

### **3.2 MOTIVES OF EMPLOYEES TURNOVER**

---

ness. Emergency preparation exercises and meetings can help your company work towards a safe work environment that would be possible in the event of natural disasters such as floods.

Trevor (2004) stated that electrical and other risks should be avoided in the workplace through proper maintenance. The team should be regularly maintained and your company should apply precautions such as wearing safety glasses or other equipment, including helmets, gloves or work boots with steel spikes. A safe work environment is something that the company and its staff must constantly focus on achieving. The existence of a well-equipped first-aid kit illustrates the organization's ability to handle emergencies such as workplace accidents and this instills trust among employees.

According to Martin (2005), human resources are very essential in the management of any business, without which it would not have happened. However, employers face the dilemma of rotation as the average stay of employees in the company is 23-24 months. according to the 2006 Bureau of Labor Statistics (Izzack, 2010) According to Harrie (2002), the Employment Policy Foundation states that it costs the employer an estimate of 15,000 per person, including the costs of separation, paperwork, unemployment; vacancy fees, including overtime or temporary employees and replacement expenses that include advertising, interview time, transfer, training and efficiency reduction when colleagues leave (Horton, 2007). Providing a stimulating work environment that promotes happy, motivated and authorized people reduces employee turnover and absence rates. Promoting a work environment that fosters personal and professional growth promotes harmony and stimulation at all levels of the organization (Martin, 2005).

According to Linda (2002), one of the most important factors determining staff turnover rates is the atmosphere in the workplace. Employees must feel respected and important to the company, rather than being expendable. The way in which managers and supervisors behave with employees is of particular importance, as

### **3.2 MOTIVES OF EMPLOYEES TURNOVER**

---

many employees will have to deal with the management on a daily basis and can rely on them for workplace reviews, increases and holidays or requests. of holidays. due to illness The training of managers to treat employees with respect and conduct business transparently can be an important factor in keeping good workers at work.

In Kenya, a study by Edwin (2005) suggested that to manage employee turnover, organizations should analyze and modify their work procedures and policies so that employees can exploit their full potential and even gain significant experience working. There are many cases where employees have left the company due to projects or activities that do not require their full potential. Employees would surely leave if they had no experience and simply put them in the "bench".

These are some of the main causes of employee turnover that can surely be avoided by organizations after taking some necessary steps to improve their internal employee services. However, there are many other causes that contribute to employee turnover; such as lack of employee motivation, pressure at work, stress at work, prejudice and favoritism, ego and employee attitudes, bad management of employees, etc. (Edwin, 2005).

#### **3.2.3 Render**

The main objective of a compensation system is to attract and retain high quality staff. To retain staff, the rewards that employees receive for their compensation must be satisfactory. Grobler, Warnich, Carrell, Elbert and Hartfield (2002) affirm that compensation refers to all forms of tangible financial services benefits that these employees receive as part of their compensation package. Robbins (2003) agrees that the classic goals of any compensation system are to attract, retain and motivate employees. Low pay is one of the reasons for staff turnover. If employees are not satisfied with what they get paid, they will usually look for greener pastures.

### 3.2 MOTIVES OF EMPLOYEES TURNOVER

---

Furthermore, working conditions, motivation, recognition and the opportunity to grow must also be feasible. An employee can still leave, even if another job offers a lower salary, as long as he enjoys job satisfaction. Although financial incentives encourage employees to stay short-term, they need long-term growth opportunities (Tyani, 2001).

The other main factor that causes employee turnover is the pay scale. Today, inflation prevails on the market, so that no person meets their salary, so there is a conflict and the result is that the employee leaves the organization on the other side, every organization wants to get more work. of the employee at a lower cost. It is often believed that the causes of staff turnover can be prevented through growth and development opportunities, positive organizational relationships, nature of work performed and respect for the organization and its values. While these environmental factors are important, a lever for employee turnover is monetary compensation; a point that employers often miss When employers only establish a direct correlation between traditional motivational theories and employee turnover, they underestimate the complexity of the relationship between employer and employees. The salary can accentuate the positive or negative aspects of other contributing factors: it is much easier to support a micro-management manager if it is paid very well. (Carroll, 2009) Performance evaluation and feedback is considered an essential management activity. While common opinion suggests that performance comments have an impact on employee motivation and self-esteem, many studies of how it works have convincing evidence.

The last factor that influences the turnover of the personnel is the evaluation of the employee, this is a very delicate situation because after the evaluation, most of the employees are demotivated and do not behave well in the future or perhaps feel disappointed by their evaluation and eventually they leave the organization. This is the most important problem that all organizations face when the staff works, works hard and thinks that I am the best employee and that his boss expected a good eval-

uation, but when the evaluation was disappointed at some point, c 'is a real reason for this attitude because some leaders show partiality towards their employees, so the evaluation is not right and this has been damaged for those employees who work hard and leave the organization. (Chen, January 2013).

A study by Nelson (2009) found that the wage scale is the most common cause of high employee turnover rates in many microfinance institutions in Africa. Employees look for jobs that pay well when the microfinance institutions they work for do not offer good salaries, they tend to look for jobs that pay them fairly well. To solve this problem, employers should offer competitive enough wages to retain and attract qualified and talented personnel. Unsatisfactory performance assessments are also one of the reasons employees leave a company.

#### **3.2.4 Job Dissatisfaction**

Dissatisfaction at work is one of the main causes of absenteeism. Robbins and Decenzo (2001) state that a person with high job satisfaction has positive attitudes towards work. People with negative attitudes are absent. In a study conducted at a military health care center in Mullins (2005), it was discovered that there was a positive relationship between job dissatisfaction, staff turnover and levels of absenteeism.

Another study is that of Tiamiyu L. and Disner J (2009), who conducted a survey on "A study on external voluntary rotation of internal auditors" which aimed to determine which factors are relevant to the voluntary external rotation of internal auditors . The research uses knowledge of the audit literature and voluntary rotation of employees where it indicates that job dissatisfaction, organizational commitment and alternative job opportunities are relevant to the voluntary rotation of employees. In this survey, members of the San Luis Chapter were examined to determine the relevance of these factors for the voluntary external rotation of internal auditors over

the past five years.

Maintaining a high level of job satisfaction is important both to maintain and to keep employees involved. If an employee is not satisfied, it can cause withdrawal from work. Retirement from work generally means that the employee avoids the work situation physically, mentally or emotionally. Withdrawal from work usually occurs when the employee is not satisfied with the nature of the work, salaries, supervisors, employees or the personality of the employees (Noe et al., 2010).

The result suggested that the availability of alternative job opportunities is more related to the voluntary rotation of internal auditors with respect to their job satisfaction and that the commitment of the organization is not significantly relevant for the voluntary rotation of the internal auditor. Therefore, management should pay attention to the results of the internal auditors' annual salary survey conducted by the entities to ensure that the salary of its auditors is consistent with their counterparts in the organizations. This will reduce the tendency for auditors to leave voluntarily due to the availability of alternative work. This study could serve as a useful management tool to promote employee retention or to facilitate a functional rotation of internal employees.

#### **3.2.5 Deficiency of Training and Development**

According to Ichniowski (2004), another tool for employee retention is employee training. Training can help employees be more competent and appreciate their work even better. They can gain confidence in their work, which in turn can increase their morale. Reducing employee turnover will benefit companies and save money (Mullins, 2005). Furthermore, the replacement of employees is expensive and valuable skills are lost. By providing employees with training and development, they are more likely to be willing to take on more appropriate responsibilities for career advancement or professional progress.

### **3.2 MOTIVES OF EMPLOYEES TURNOVER**

---

Trevor (2004) stated that the best strategies for maintaining the workforce generally include a combination of positive reinforcements, creating a satisfactory work environment, adequate salaries for employees and offering incentives and financial bonuses. On the other hand, Bratton (2003) stated that the best employee retention strategies include pension plans such as medical insurance, paid holidays and holidays. Anthony (2006) supported these views and said that employees are more satisfied with their work and less likely to leave, as a better work environment and employee benefits allow them to influence positive changes within a 'organization.

According to Paul (2004), training and development are a function of human resource management related to organizational activity aimed at improving the performance of individuals and groups in organizational contexts. It has been known by various names, including "human resource development" and "learning and development. The induction of new employees and staff development are two essential requirements for job preparation that are mandatory for employers. It starts with recruitment (Bratton, 2003). Those employees who start their work without any guidance probably do not know the workplace policies and procedures from which they can benefit. Additional training during the employment relationship helps the employee update the their ability and to be more productive and productive and efficient in the workforce: when employees do not have the necessary training to be more productive, their performance is affected and they will go alone for jobs that provide training and support to employees or will be fired for poor performance (Armstrong, 2003).

According to Dubois (2004), employee retention has a direct and causal relationship with the needs and motivation of employees. The application of a motivation theory model, like Maslow's Hierarchy of Needs, is an effective way to identify an effective conservation protocol.

Moureen (2004) stated that a solid workforce retention plan can create a productive atmosphere, reduce turnover rates and contribute to the success of any organi-

zation. Employees who feel appreciated and appreciated will generally go beyond the duty to ensure the success of their employers. Many employees who decide to leave their current employers and seek work elsewhere do so because they feel they are micro-managed. To avoid this common problem, employers must authorize staff members by encouraging their contributions and allowing them to take the lead role in projects. This, in turn, often leads to innovative ideas that can increase efficiency, improve morale and increase a company's final result (Georgi, 2004).

#### **3.2.6 Job Satisfaction on Employees**

Many surveys have been conducted that relate work satisfaction to employee work performance and employee turnover, where the results concluded that these three variables are significantly correlated also in contemporary organizations (Schuler and Jackson, 2007; Redman and Wilkinson, 2011; Thomson and Phua, 2012). Furthermore, job satisfaction is beneficial not only for the individual performance of employees, but also for organizational performance. The study by Bender and Heywood (2006) explored various factors related to job satisfaction and specifically identified workplace motivation which was also affirmed by a previous study by House Wigdor (1967) which postulates that The motivation is related to job satisfaction and therefore to the desirable performance of employees. Modern studies such as Mankin's (2009) have provided the link between motivational theories and theories of job satisfaction. Furthermore, the studies identified in the area of human resource management have shown a strong connection between job satisfaction and job performance over the last few decades and according to Bhatti and Qureshi (2007), job satisfaction has become a fundamental component for human resource strategies in employee participation, commitment and productivity. As previously identified, job satisfaction was considered a determining factor in increasing employee performance, regardless of company size (Coomber and Barriball, 2007;

Tansel, 2014).

The hygiene factors linked to job dissatisfaction are salary, supervision, interpersonal relationships, company policies and administration, working conditions, job security and the balance between work and personal life. The factors that lead to job satisfaction or to so-called motivational factors are praise, recognition, opportunities for promotion, responsibility, a sense of accomplishment and work itself. According to the theory, a reduction in employee dissatisfaction will only reach a state of dissatisfaction, but will not lead to greater job satisfaction and, consequently, if job satisfaction decreases, the employee will only achieve a state of non-working satisfaction rather than lead to job dissatisfaction (Herzberg et al., 1959).

Derek et al. (2007) have suggested that job satisfaction and emotional commitment are the basic variables with rotation intentions. Furthermore, he argues that no support was found in the results for the dedication of continuity as a variable with rotation intentions. The results also concluded that significant positive correlations were found for the workload of the distal variables, work stress, employee salaries, job satisfaction and conflicts between work and family. Magnus in al. (2001) it was discovered that job insecurity is correlated mainly with the reactions of exit and devotion, but not with the voice. In terms of production, it is surprising. The model in our results refers to the connection obtained between job insecurity and organizational rotation intention. Although the size of the effect sizes differed from country to country, there was a positive relationship between job insecurity and propensity to leave the organization.

The feeling of dissatisfaction at work can lead the worker to withdraw from work often thinking of abandoning or participating in other forms of retirement from work, such as absenteeism (Mobley, 1977). Retirement from work was classified as a change in behavior, a physical pension from work and a psychological pension from work (Noe et al., 2010). The physical retirement from work includes absenteeism, both in the form of delay and illness, for employees who request a

### **3.2 MOTIVES OF EMPLOYEES TURNOVER**

---

transfer or simply leave work. Employees can even retire while they are at work, not really working. All these behaviors are expensive for the employer (Noe et al., 2010).

Maertez and Campion (2004) have suggested that, in the first place, quitting smoking is an important and important event in life and people remember the details surrounding those events with relative ease. Secondly, quitting smoking tends to be associated with emotional excitement, both negative and positive; Such events are easier to remember than those with little emotional arousal. Moreover, people often remember such events better after a long period rather than a short period of time. Furthermore, the time elapsed since he stopped smoking was not related to the type of decision; decreasing the possibility that memory loss has an effect on results. Another possibility is that discovered relationships could be due in part to dependent patterns or implicit rotation theories. However, if these patterns reflect reality or are widely used by employees, they may not threaten the validity of our results. Praise to al. (1999) concluded that the turnover of employees in schools and universities is partly due to the low morale of teachers, which in turn has increased the turnover rate of the system.

## **Chapter 4**

---

## **Methodology**

### **4.1 System Model**

We first collected the dataset in our study. We received the dataset from kaggle. After collecting the dataset, we encode the categorical variables from the dataset. Second, we checked in the dataset to see if there were any missing values, because we could not properly teach our model if there were missing values. Third, we have scaling the feature, whose function is to adjust if the value is greater or less in a word, so that the result is better. Fourth, we checked whether there was any correlation column in the dataset.

First I did the preprocessing within the data set. The process of preparing the dataset is given below:

- I have omitted the co-relation column using the co-relation matrix.
- Finds missing values and replaces values there.
- Level encoding is used to convert labels to numbers.

Then an algorithm is used to separate useful features from the data set, which is discussed below:

- Important features are extracted using Chi2 test.Important features are selected to get the best Score.
- Important features are extracted using Random forest important.Important features are selected to get the best Score.

- Good results are obtained by selecting the Chi 2 test and random forest implant matching feature.

In this study I used six algorithms for machine learning:

- Random forest
- Decision Tree
- Multi layer Perception(MLP)
- Gaussian Naive Bayes
- k-Nearest neighbor(KNN)
- SVM

After all the algorithms finish than all models have been traning and test using 10-Fold cross validation.

In this study I used some assembly method:

- Bagging
- Boosting

### 4.1.1 Feature Selection

There are three Feature selection used this Thesis.

- 1.Sequential Backward Selection(SBS)
- 2.Chi Square
- 3.Random forest

### 4.1.2 Sequential Backward

Selection(SBS):The sequential backward selection algorithm (SBS) eliminates the incremental features of the subset of complete features until the new feature has the desired number of features.To decide which feature will be eliminated in each step, we need to define a J function that we want to minimize. The criterion calculated

by the function can simply be the difference in the performance of the classifier after and before eliminating a particular characteristic. Sequential Backward Selection works greatest when the optimal subset of functionality is large, as Sequential Backward Selection spends most of its time visiting large subgroups. The main restriction of Sequential Backward Selection is its disability to reassess the effectiveness of a feature after it has been rejected.

### 4.1.3 Chi Square

The Chi2 algorithm is mainly used to determine the highest value characteristics from the Chi2 test statistic which consists of training and test sets. To achieve this, we have separated the features into training sets and test sets. Our training set consists of non-negative features that are in Boolean form or frequencies because the Chi2 algorithm behaves on these characteristics.

Feature selection is a process in which features are automatically selected in the data that most contributes to the forecast or output variable you're interested in. The advantage of selecting features before modeling data is:

- Avoid over-sizing: less redundant data increases model performance and leads to fewer opportunities to make decisions based on noise.
- Improves accuracy: Less distracting data means modeling accuracy improves.
- Reduction of training time: less data means that the algorithms train faster

### 4.1.4 Random Forest Important

To collect all the characteristic values in one, the Scikit-Learn Random Forest library was used. We used the importance of the random forest feature to access the weights of each feature to be able to work with the most weighty features. To get the importance of the random forest, we have created a random forest classifier with

the entropy criterion and the estimator number 1000. Here, the estimator indicates the number of trees in the forest. We set the function and training set on Random Forest Classifier. Finally, we had access to the importance of the feature with the properties of importance of the characteristics of the classifier. We have ordered the values of importance and we have gained importance together with the names of the features.

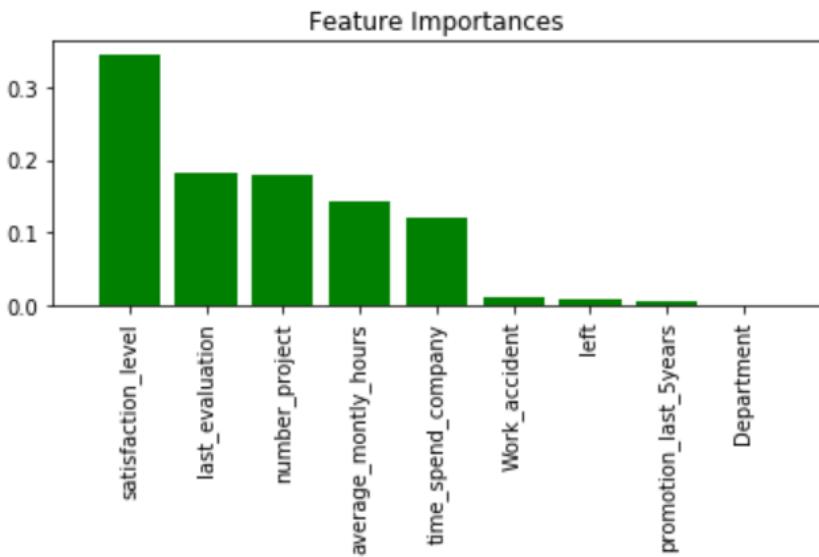


Figure 4.1: random Forest Feature Importance

### 4.1.5 Feature Selection

We compared the results values from Chi 2 and Random Forest significance and took the top three variables. Then we obtained a good value by modeling them with training and testing. But in our study we have the highest value by doing model training and testing with just five important values from random forest. Using these five variables, we created the final attribute of the dataset. Subsequently, the training and test set were separated and the test set contained 20 percent of the entire dataset.

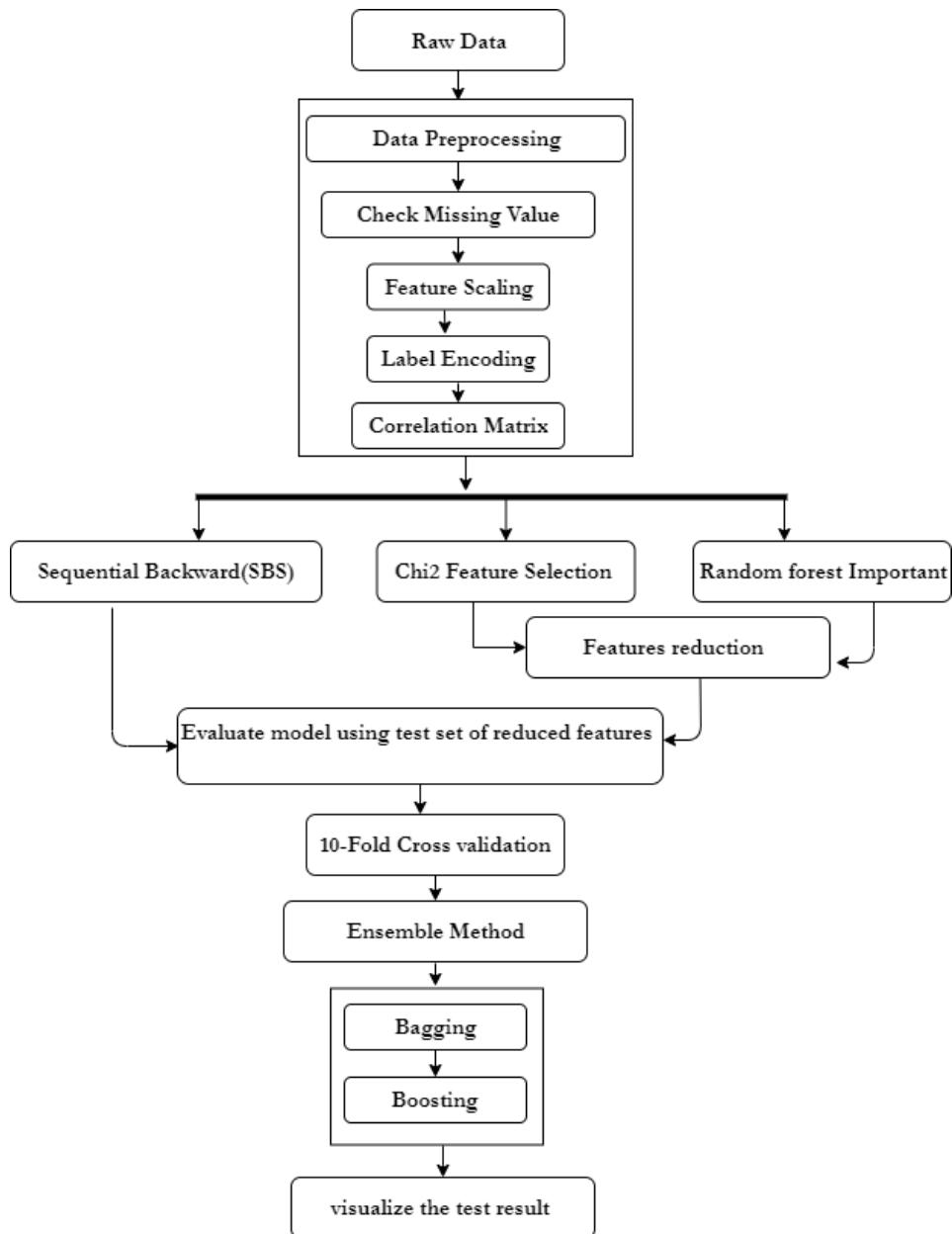


Figure 4.2: System Model

Table 4.1: Summary of the Data

Variable	Description
satisfaction_level	Satisfaction Level of the employees
Last_evaluation	Last evaluation of employees
number_project	Number of projects the employee has done
average_montly_hours	The average monthly hours of the employee
time_spend_company	The total number of times the employee has spent in the company
Work_accident	If any employees have done work experience or not
left	Whether the employee has left the company or not
promotion_last_5years	Whether the employee has any promotion last 5 year
type	The type of the job
salary	Salary level of the employee

## 4.2 Build predictive models Using algorithm

First we make the predictions using the algorithm, then we divide the training set and test sets into 70:30. Then, the Random Forest, Decision Tree, MLP, SVM, Gaussian NB and KNN algorithm training sets have been used to construct predictive models. A test set has been used to evaluate model performance

### 4.2.1 Random forest.

First we will train the random forest algorithm with all the features. Random Forest (RF) decision is a combination of trees where each tree depends on random quality independently. These trees depend on the sub-spacing of the dataset and apply it for precision calculations. Random forests (Breiman, 2001) are a gathering method that can also be thought of as the nearest neighbor prediction.

### 4.2.2 Decision Tree classifier

We have created the Decision Tree classifier using the decision tree classification, attribute selection method, and Python Scikit-Learn package. Less time is spent on decision trees and faster pace of categorization, which can avoid decision errors and all kinds of deviations. Through a Case Study (Sikarodi et al., 2015)

Here is how we did the classification by Decision Tree:

We first select the best attribute using Attribute Selection Measures (ASM) to split the records. Make that decision a node feature and split the dataset into smaller satellites. Repeat this process until each child's condition is matched and I start to re-create the tree. We have developed the criterion parameter as "entropy" in which the term is responsible for determining the quality of the division. "Entropy", which we use to get information

### 4.2.3 support vector classifier

In general, carrier support machines are considered a classification system, but can be used in both classification and resistance problems. It can easily handle multiple continuous and categorical variables. SVM creates a hyper-plane in a multidimensional space to separate different classes. SVM produces an optimal hyper-plane recursively, which is used to reduce errors. The dataset contains continuous labels so we used the support vector classifier to our advantage. In addition to performing the linear classification, SVM can efficiently perform a non-linear classification, explicitly mapping their inputs to the high-dimensional feature system, so we used the support vector classifier here. The vector support machine is a supervised machine learning algorithm subject to the creation of hyper planes from the vector support line to determine to which class the data set belongs. The left side of the margin of the carrier vector is the negative hyper visor and the right side of the margin of the carrier vector is called the positive hyperbolic lane. The main objective is to train the

SVM to learn the structure of the data while tracking the exact class labels. In general, the hyper plane provides the best results, with the maximum distance from the location of the training data closest to any class. Wenzel, Florian; Galy-Fajou, Theo; Deutsch, Matthäus; Kloft, Marius (2017). “Bayesian Nonlinear Support Vector Machines for Big Data”. Machine Learning and Knowledge Discovery in Databases (ECML PKDD).

### 4.2.4 Multi layer Perception:

Ciresan and colleagues (2010) have shown that despite extinct gradient problems, GPUs make it possible for multi-layered feed forward neural networks to expand. Multi layer perception is a simple ANN that follows a feedback strategy. There are three levels completely connected in MLP. The input is hidden and the output level is connected to the next level. In addition, these layers contain nodes or neurons. In addition, the output level depends on the number of target attributes on which the input levels depend primarily on the target attributes. About 10-20 nodes are in hidden layers as inputs. In addition, only the hidden and output levels have activation functions. Each node has a weight function that is multiplied and passed. In addition, weight is added before each activation function. Consequently, subsequent propagation is another characteristic of the MLP with which he learned the part and updates the weights based on the requirements to handle fewer errors in the ANN. We create the MLP classifier using the default parameters provided by the Scikit-Learn library.

### 4.2.5 Gaussian Naive Bayes

Naive Bayes is the simplest and fastest classification algorithm, suitable for a large amount of data. The Naive Bayes classifier has been used successfully in various applications, such as clear filters, text classification, sensitivity analysis and rec-

ommendation systems. Use the Bayesian probability theorem to predict unknown classes. The basic idea of Naive bias is to calculate the probability of each condition occurring when a categorization is given (Russell, 1995).

Bayes theorem, the theorem commonly used in automatic statistical learning, in particular, when we need to find out the probability of an event considering the event that has already occurred. The Naive Bayes algorithm uses the Bayes theorem concept that can be executed in classified data. The algorithm predicts the membership probabilities of each class in which the record belongs to a specific class and is considered the most likely class, called Maximum to Posteriors (MAP).

### 4.2.6 k-Nearest neighbor :

It is called k nearest neighbor, which uses k “closest” points (nearest neighbor) for performing classification (k is a positive integer, typically small) (Altman, 1992). The closest k-Nearest neighbor(KNN) can be applied to classification and regression problems, but in most cases applies to the k-Nearest nearest classification problems. KNN is widely used during calculations and for the simplistic interpretation of a powerful predictive power. The data used for the KNN can be a basic, quantitative or nominal measurement scale that has many types that will be used for classification. In kNN where  $k = k$  (rule of the k-Nearest neighbor), the algorithm closest to the k-Nearest neighbor and the voting majority. Who would normally be curious if the number of classes is 2? In the case of higher K, there will be examples of variants and KNN should label the new point as the maximum number of instances. When there is more than one class, the weighted KNN is applied at the point where each point is weighted and the weight is calculated using its Euclidean distances. However, we apply this algorithm to the system model with  $k = 2$ , which means that we are allowing the 2 closest neighbors for the classification process, as it provides satisfactory results.

### 4.2.7 Hyper Parameter Tuning

Each machine learning algorithm has a set of variables that capture values from different sources. Sometimes they are fulfilled by user input, sometimes they are learned by data. This variable is called a parameter. Whose values are learned by the data known as a parameter. However, there are a set of parameters whose values are set before the learning process begins, this variable is called hyper parameter. The Python Hyper parameter is passed as an argument to the Model Constructor class in the scikit-Learn Machine Learning Library.

Here are the hyper parameters that were followed in these research models:

- Random Forest Classifier: random\_state=42, n\_estimators=100, max\_features =’auto’, class\_weight=None, n\_jobs=None, verbose=0, warm\_start=False.
- Decision Tree Classifier: random\_state=0, criterion = ”gini”.
- K-Neighbors Classifier: n\_neighbors=1.

### 4.2.8 K-fold cross-validation

In K-fold cross-validation, the original sample is randomly divided into subsets of equal size. Thus between the subfamilies K, a single sub-undercut model is retained as validation data for the test and the remaining K-1 samples were used as training data. The cross-validation process is repeated several times, with the subset K used exactly once as validation data. The results for K are then plotted as the average of the individual estimates. The advantage of the method in the cross validation of K folds through repeated random sub sampling is that all observations are used for both training and validation, and each observation was used once for validation. 10 times cross-validation is commonly used, but K usually starts from an unpublished parameter. In the cross-validated stratified K folds, the partitions are selected so that the average response value between the partitions is approximately equal. In the

case of binary classification, this means that each partition has approximately equal proportions of the two types of labels.

### 4.2.9 Ensemble method

The methods set are meta-algorithms that combine different machine learning techniques in a predictive model to reduce variance (bagging), distortion (reinforcement) or improve predictions (stacking). The set methods can be divided into two groups: sequential set methods in which the basic students are generated in sequence (for example, AdaBoost) and parallel set methods in which the base students are generated in parallel (for example, random forest). The basic motivation of sequential methods is to exploit the dependency among basic students, since overall performance can be increased by weighing examples previously labeled with greater weight. The basic motivation of parallel methods is to exploit the independence among the basic students, since the error can be drastically reduced by the average.

#### 4.2.9.1 Bagging

We used a bagging classifier, a combined meta-estimate that adapted each base classifier to a random satellite from the original database and then combined them into separate forecasts (by grade or average) to form the final forecast. This national meta-estimator is generally used as a way to reduce the variance of a black box estimate by introducing randomization into a compilation set method and then creating the representation. Fit many large trees to bootstrap-resampled versions of the training data, and classify by majority vote. (Breiman, 1996)

#### 4.2.9.2 Boosting

The increase is a sequential process, in which each successive model attempts to correct the errors in the previous model. Later models depend on the previous model. Next we will discuss how the Boosting works. Fit many large or small trees to re-weighted versions of the training data. Classify by weighted majority vote. (Freund Shapire, 1996).

1. A subset is created from the original data.
2. Initially, all data points received the same weight.
3. A basic model was created in this subset.
4. This model was used to predict the entire data set.

## Chapter 5

# Experiments Analysis

### 5.1 Preprocessing Result:

Here I have shown some of the preprocessing results:

- Label Encoding

left	promotion_last_5years	Department	salary
1	0	sales	low
1	0	sales	medium
1	0	sales	medium
1	0	sales	low
1	0	sales	low
1	0	sales	low
1	0	sales	low
1	0	sales	low
1	0	sales	low
1	0	sales	low

Figure 5.1: Dataset before the label encoding.

left	promotion_last_5years	Department	salary
1	0	7	1
1	0	7	2
1	0	7	2
1	0	7	1
1	0	7	1
1	0	7	1
1	0	7	1
1	0	7	1
1	0	7	1
1	0	7	1

Figure 5.2: Dataset after the label encoding.

## 5.2 Feature Selection Result:

Here I have extracted the most useful features from all the properties using Chi2 and Random Forest.

- SBS
- Chi2 Feature selection
- Random forest Important

Table 5.1: Random forest feature importance scores

Feature name	Importance
satisfaction_level	0.345032
last_evaluation	0.182678
number_project	0.180379
average_montly_hours	0.143579
time_spend_company	0.121260

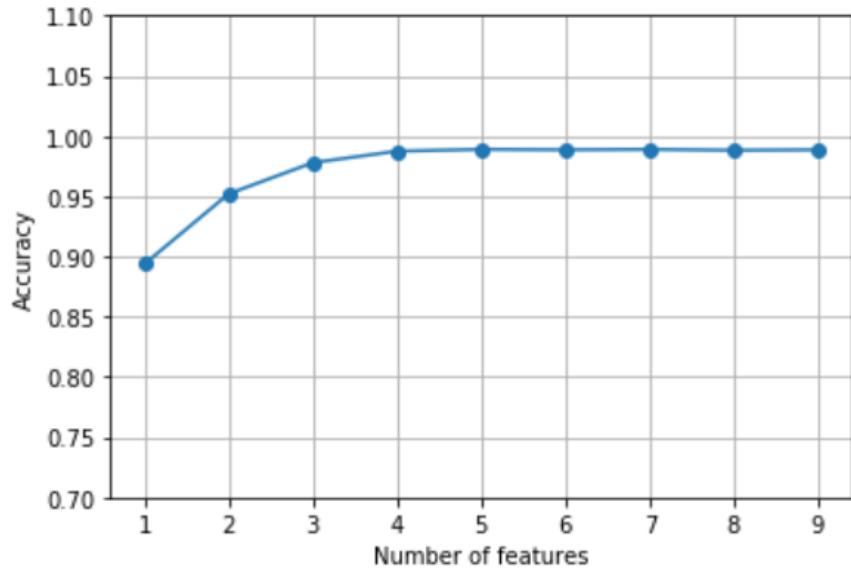


Figure 5.3: The number of features after applying SBS.

Table 5.2: Chi2 scores

Feature name	Coefcient
average_montly_hours	945.588933
satisfaction_level	228.193183
time_spend_company	191.708547
promotion_last_5years	56.044865
number_project	3.390205

## 5.3 Experiments

In this chapter I will show all the results. I will show all the results separately. Which I divided into 4 steps.

The steps are:

- Prediction All feature
- Prediction Reduce Feature
- 10\_Fold Cross Validation
- Ensemble Learners
  - Bagging
  - Boosting

After post-processing the data to get accuracy forecasts, I train all the properties and test them with algorithms, from which I get the best accuracy predictions using random forest 99.3.

Table 5.3: Measure performance using algorithms by selecting all features

Path	Random forest	Decision tree	SVM	MLP	GaussianNB	KNN
Training accuracy (All variables)	1.0	1.0	96.11	93.61	0.779778	1.0
Test accuracy (All variables)	99.3	97.91	94.87	93.43	77.98	95.02

Initially, I trained our model with all the features other than a work accident and got a receiver operating characteristic curve, which is provided in Figure: 5.4. The ROC curve provides the results for each of the above mentioned classifiers. Then I looked at the ROC curve to get a better insight. When I reduce the number of features I'm paying close attention to what happens.

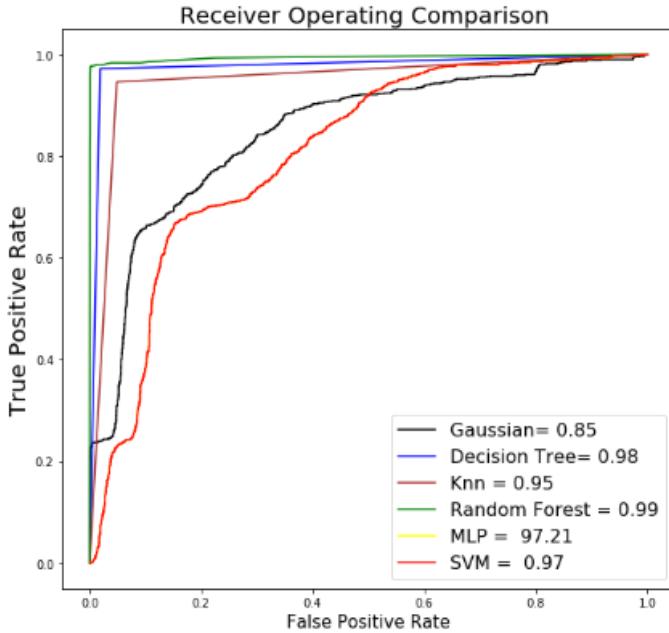


Figure 5.4: ROC curve before all feature selection.

Our feature is comprised of nine variables for which I was creating an employee leave frame, and our label as the left column of the dataset to indicate whether the employee was gone. It is clearly seen that when the number of considered features has been increased from five to nine using SBS, the model shows maximum accuracy (98.81%). The accuracy is illustrated compared to the number of Figure: 5.3. Last seen that Random Forest works better than all other algorithms in Classifier.

I picked one more stage to reduce the feature and also applied the importance of Chi2 and Random Forest to perform. By doing this, I can access their prominent features and their respective scores.

### 5.3 EXPERIMENTS

---

The results of Chi 2 and Random Forest are given in Tables 5.1 and 5.2 for calculating the feature importance. From Tables 5.1 and 5.2 we can clearly see the importance of random forest features and score of Chi 2. According to Chi2 satisfaction level, average monthly hours , promotion last 5 years, Time Spend company and number of project is the five highest ranked factors.

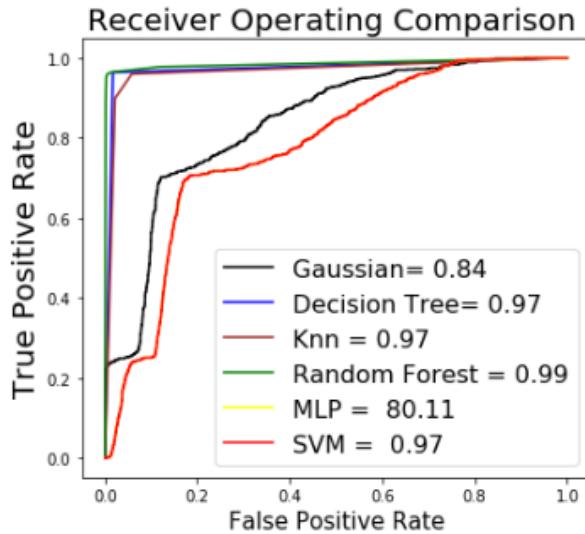


Figure 5.5: ROC curve before all feature selection.

On the other hand, in the Random Forest, these five important factors are select satisfaction level, last evaluation, number of project, average monthly hours, Time Spend company. I experimented with Chi 2 and Random Forest. But I chose the features of Random Forest.

Subsequently, it was run with different classifications built into the selected algorithms. We visualize the ROC curve after the second phase characteristic reduction, which is provided in Figure 5.5. In Figure 5.5, after applying the second step feature reduction, it is randomized that the random forest classifier offers the best accuracy in both cases only and it is 0.99. SVM 0.97 and decision tree 0.97. Further, the KNN model did not change much (0.97). Comparing with Figure 5.4, we

### 5.3 EXPERIMENTS

---

can detect the unrecognized accuracy of the MLP model (0.81) after applying the feature reduction which previously had an accuracy of 0.97. Here, the false positive value was detected more than the previous result. The accuracy of the Gaussian NB has not changed much, though it has been reduced to 0.81, which was 0.83 earlier. In the ROC curve, we try to evaluate the best-classified model according to the highest true-positive value and the low-false positive value, and we also compared all the classified models that are controlled by the ROC curve. Table 5.4 shows the results of the accuracy of the different classifiers after the second-level feature reduction process. Table 5.4 shows the results of the accuracy of the different classifiers after the feature reduction process

Table 5.4: Measure performance using algorithms by selecting reduce features

Path	Random forest	Decision tree	SVM	MLP	GaussianNB	KNN
Training accuracy (Reduced variables)	99.75	99.98	96.35	93.37	80.32	99.03
Test accuracy (Reduced variables)	98.64	97.93	95.27	93.67	79.33	95.93

From the ROC curve, we can clearly see that the random forest still offers better model functionality as the region under the value holds more value.

after applying the cross validation, it is randomized that the random forest classifier offers the best accuracy in both cases only and it is 99.40. Table 5.5 shows the results of the accuracy of the different classifiers after the Cross Validation process

In this study I used the bagging and boosting of the ensemble algorithm. after applying the Bagging, it is randomized that the random forest classifier and decision tree offers the best accuracy in cases only and it is 0.98 and 0.98. Table 5.6 shows the results of the accuracy of the different classifiers after the Bagging process.

### 5.3 EXPERIMENTS

---

Table 5.5: Measure performance using algorithms by 10 fold Cross validation

Path	Random forest	Decision tree	SVM	GaussianNB	KNN
K-Fold Cross Validation score	99.40	98.32	95.90	80.21	97.26

after applying the Boosting, it is randomized that the random forest classifier and decision tree offers the best accuracy in cases only and it is 0.99 and 0.98. Table 5.6 shows the results of the accuracy of the different classifiers after the Boosting process.

Table 5.6: Accuracy performance measurement of ensemble algorithms

Path	Random forest	Decision tree	KNN
Training accuracy (Bagging)	99.20	99.3	95.3
Test accuracy (Bagging)	98.0	98.0	94.1
Training accuracy (Boosting)	100	100	100
Test accuracy (Boosting)	98.5	98.3	98.3

# **Chapter 6**

---

## **Conclusions and Future Work**

### **6.1 Conclusion :**

This chapter discusses the main reasons for the study. The impact of employee turnover from the company has been reviewed. Employee turnover is a big issue for any company. Maintaining a skilled workforce is essential for the continued productivity of an organization. But most of the time, it is very difficult to control employee turnover rates within the organization. The purpose of this study was to determine the significant causes of employee turnover. The survey found that many reasons employees leave the company. Among them, our study selected some of the most important attributes that are the key factors in employee turnover. In our study, we worked on nine features, reduced them to three, and built a predictive model with five features. They are: satisfaction level, final evaluation, how many projects the worker has done, average monthly hours, time spent.

These factors must be taken into account when an employee is in the company; an employee's satisfaction level should be kept high by providing them with the compensation they need when dealing with additional workload burden. Newly hired employees must be well-behaved so that they can adapt to the company environment. Skilled employees must be rewarded for allowing them to stay in the company.

## **6.2 Future Work :**

In the future, we plan to work with in-depth features to see if there is any hidden reason why employees are leaving, and I will work to create an app for that. In the future, we plan to work with in-depth features to see if there is any hidden reason why employees are leaving, and I will work to create an app for that.