PREDICTING EMPLOYEE TURNOVER INTENTION IN IT&ITES INDUSTRY USING MACHINE LEARNING ALGORITHMS

By

S.K. MONISAA THARANI (19MBA180)

Under the guidance of

S N Vivek Raj
Assistant Professor

A PROJECT REPORT

Submitted
In partial fulfilment of the requirements for the award of the
Degree of

MASTER OF BUSINESS ADMINISTRATION
Kumaraguru College of Technology
(An autonomous institution affiliated to Anna University, Chennai)
Coimbatore - 641 049

August 2020

BONAFIDE CERTIFICATE

Certified that this research report titled "**PREDICTING EMPLOYEE TURNOVER INTENTION IN IT&ITES INDUSTRY USING MACHINE LEARNING ALGORITHMS**" is for course completion of Major Project is the bonafide work of MONISAA THARANI S K (19MBA180) who carried out the project under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form part of any other project report or Internship on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

Faculty guide
S. N. Vivek Raj
Assistant Professor
KCT Business School

Head of the Department
Dr. Mary Cherian
Professor & Head
KCT Business School

Softcopy Submitted for the Project Viva-Voce examination held on 15.09.2021

Internal Examiner

(Signature)

External Examiner

(Signature)

DECLARATION

I hereby declare that this industry research entitled as, **"PREDICTING EMPLOYEE TURNOVER INTENTION IN IT&ITES INDUSTRY USING MACHINE LEARNING ALGORITHMS"** has been undertaken for academic purpose for the course submitted to Anna University in partial fulfilment of requirement for the award of degree of Master of Business Administration. The Internship report is the record of the original work done by me under the guidance of Prof. S.N. Vivek Raj, Assistant Professor, KCT-BS during the academic year 2020.

I, also declare hereby, that the information given in this report is correct to the best of my Knowledge and behalf.

ACKNOWLEDGEMENT

I express my sincere and heart-felt gratitude the **Management of KCT Business School**, for their prime guidance.

I express my thanks to **Dr. Mary Cherian, Head of the department, KCTBS** for implementing this internship and providing under the supervision in its execution. I am indebted to my Institution and my faculty members without whom this internship would have been a distant reality.

I also offer sincere thanks to my Project guide **Prof. S.N. Vivek Raj** for giving me support and guidance for this internship from inception to closure.

I also declare hereby, that the information given in this report is correct to the best of my knowledge and behalf.

TABLE OF CONTENTS

CHAPTER	CONTENTS	PAGE NO
I	**INTRODUCTION**	**10**
	1.1 Statement of the problem	12
	1.2 Research questions	12
	1.3 Objective of the study	12
II	**INDUSTRY PROFILE**	**13**
	2.1 Background of the industry	13
	2.2 Market size	14
	2.3 Major players	15
	2.4 Government initiatives	15
	2.5 Recent trends in the industry	16
	2.6 Challenges faced by the industry	17
III	**REVIEW OF LITERATURE**	**19**
	3.1 Research gap	25
IV	**RESEARCH METHODOLOGY**	**26**
	4.1 Introduction	26
	4.2 Population Size	27
	4.3 Sample Size	28
	4.4 Research design	28
	4.5 Sampling Design	28
	4.6 Tools for data collection	28
	4.7 Features of the data	28
	4.8 methods of data collection	28
V	**DATA MODELLING AND COMPARISION**	**28**

	5.1 Logistic regression (forward method)	31
	5.2 Logistic regression (backward method)	33
	5.3 Naïve bayes	36
	5.4 K nearest neighbour	37
	5.5 Kernel SVM	38
	5.6 Random forest	40
	5.7 XG Boost	41
	5.8 Artificial neural networks	42
	5.9 Decision tree	45
	5.10 Comparison of classification results	51
	5.11 Text mining	51
VI	**CONCLUSION FINDINGS AND SUGGESTION**	**55**
	REFERENCES	**56**
	APPENDIX 1	**60**
	APPENDIX 2	**73**

LIST OF TABLES

TABLE NO.	PARTICULARS	PAGE NO.
2.1	Leading IT companies	15
5.1	Demographic factors distribution	29
	Logistic regression (forward)	30
5.2	Model summary	31
5.3	Confusion matric of forward logistic regression	32
5.4	Variables in the equation	32
	Logistic regression (backward)	33
5.5	Model summary	33
5.6	Confusion matric of backward logistic regression	34
5.7	Variables in the equation	34
5.8	Confusion matrix of naïve bayes	36
5.9	Confusion matrix of K nearest neighbour	37
5.10	Confusion matrix of kernel SVM	39
5.11	Confusion matrix of Random forest	40
5.12	Confusion matrix of XG Boost	41
5.13	Confusion matrix of Artificial neural networks	43
	Decision Tree	45
5.14	Confusion matrix	47
5.15	Classification results of algorithm	**51**

LIST OF FIGURES

FIG NO.	PARTICULARS	PAGE NO.
2.1	Market size of IT industry	14
4.1	CRISP DM diagram	26
5.1	Age Distribution	30
5.2	Companies switched based in job role	30
5.5	ROC curve of naïve bayes	37
5.6	ROC curve of K nearest neighbour	38
5.7	ROC curve of kernel SVM	39
5.8	ROC curve of random forest	41
5.9	ROC curve of XG Boost	42
5.10	ROC curve of artificial neural networks	44
5.11	Neural network diagram	44
5.12	Optimal score	45
5.13	Decision tree diagram	46
5.14	Relatively important variables	47
5.15	ROC curve of decision tree	48
5.16	Gain chart	49
5.17	Lift chart	50
5.18	10 frequently used words	51
5.19	Word cloud	52
5.20	Survey sentiments	52
5.21	Emotions in text	53

ABSTRACT

Employee's determination to leave the organisation is one of the significant factors impacting the performance of the organisations since it affects the overall profitability. Organizations need to strategize to reduce the turnover goals of the workers to have a competitive advantage over other organizations. By understanding the factors impacting the employee's intent to leave the organization, the management can intervene with strategic policies and decisions so that intent of the employees to leave the organization will be reduced substantially and thus increasing the employee's engagement towards work. This research paper uses machine learning algorithms to predict employee's intention to leave the organization in the near future and identifies the significant features impacting the employee's intention to leave the organization. Data has been collected from 416 employees working in IT and ITES companies using convenience sampling and structure questionnaire. The Research also used text mining to analyse the open-ended questionnaire filled by the employees there by mining the frequently used words and employee sentiments. From the study it is found that among the Classification algorithms used for predicting employee's turnover intention, XG boost performed relatively better with high accuracy, recall, precision and f score. Using Logistic Regression, it is found that alternative job opportunity, gender, education, willing to relocate from work place, alternative job opportunity, job stress and attitude towards covid affects the employee's intent to leave the organization to a greater extent.

Keywords: machine learning, classification, logistic regression, prediction, intent to leave.

CHAPTER 1

INTRODUCTION

The information technology (IT) sector in India has grown to a great extent to cover several aspects of technology and computing. The Indian IT/ITeS industry also contributes towards the economic growth of the nation by employing about 10 million people. Moreover, the IT industry has also played a leading role in the Indian economy by promoting exports, improving standards of living and generating revenues. In recent years, there has been a massive increase in the competition among companies in sustaining in the business. In spite of the industry's good performance, it faces a systemic issue of high employee turnover, which in turn affects the industry's performance. Most of the employees leave their current organizations for learning new skills and increasing their competencies. Turnover intention can be described as the rate to which a member of staff is willing to leave a particular organisation; it affects organisational sustainability and rating. Turnover intention is a process whereby an employee decides to quit or leave a particular organisation for another one for some reasons. It implies an employee's personal anticipated likelihood that he or she has a deliberate intention to quitting the establishment in the near future. It can also be described as employee's consideration or thinking to quitting a job. Employee's turnover intention has been a serious problem of organisations regardless of their size, locations or nature of business as the effect of high turnover intention on organisational objectives affects negatively the quality of organisational products or services. Turnover intention may arise as a result of some factors or features directly present in an organisation such as, organisational policies, motivational strategies and organisational culture among others. An employee would choose to join or depart an organization depending on many causes i.e. work environment, work place, gender equity, pay equity and many other. The rest of the employees may think about personal reasons for instance relocation due to family, maternity, health, issues with the managers or co-workers in a team. Employee turnover is a major problem for the organizations particularly when trained, technical and key employees leave for best opportunities from the organizations. This finally results into monetary loss to substitute a trained employee. The employee turnover identification helps in predicting and resolving the issues of intention. We can use this data to stop the turnover rate of the employees.

On the other side, turnover intention may be defined as the intention of employees to quit the organization. Intentions are a statement about a specific behaviour of interest. Turnover intent is the probability that an individual will change his or her job within a certain time period and thus it leads to actual turnover. It is the individual's intention to voluntary quit the organization or the profession. Turnover intention has been acknowledged as the best predictor of actual turnover. Actual Turnover is expected to increase as the intention increases. The measurement of Turnover Intention can determine the likelihood of the staff leaving the organization. This helps to determine how one can find opportunities to reduce the overall turnover. In view of this trend, companies announced several training and development programmes with an aim of encouraging and hence retain them. Therefore, companies are focusing on career planning and development of the employees in order to retain them, which have become a critical success strategy for the Indian IT industry. The profits of the company can be improved by company efficiency. Staff retention is more important than acquisition of new staff. Employee turnover is something that many businesses wish to minimize as it helps to keep a cohesive, experienced team with the company. The purpose of this study is to determine the turnover intention of the employees which case the huge lose to the organization. Turnover intention is a complex phenomenon that depends on various factors. Employee turnover intention is affected largely by employees' stress, recognition, organization support, alternative job opportunities etc., where there is lack of these factors that leads to the less satisfaction and commitment to the organization that leads to the high turnover intention of the employee. Organizations have to take strategic steps to reduce the turnover intentions of the employees. In order to have a competitive edge over the other organizations, the turnover has to be controlled by taking measures favourable for the employees which may lead to increase in their commitment level. Effective employee engagement depends on the successful connect between employees and organizational representatives including supervisors, senior leaders, HR personnel. Many of the young generation tend to hop their jobs more than the older generations. The problem is critical because it affects not only the sustainability of work but also the continuity of enterprise planning and culture. Training and adaption of employees are time and money consuming. So, by considering these factors the employee turnover intention in the Coimbatore city is predicted by using supervised machine learning algorithm and bringing insights and decisions about the reduction of turnover intention of employees

1.1 STATEMENT OF THE PROBLEM

- Labour turnover is inevitable part of any business, but if turnover intention rate increases that affects the quality of the service and productivity of the company.
- It is important to predict employee turnover intention and analyse retention decision because turnover acquires the worthy cost, both in terms of direct cost and indirect cost which includes replacement, recruitment, training etc.,
- The turnover intention of the employee will affect the productivity of the company where there is a lack of organization commitment
- The lack of recognition and organization support will also lead to the job stress and make the employee to look for external job opportunity.

1.2 RESEARCH QUESTIONS

- How does the demographic factors affect the intention of employee to leave the organization?
- How does the factors like job stress, Organization commitment, recognition, Perceived organization support and job satisfaction affect the intention of the employee to leave the organization?
- Using supervised machine learning algorithm, how to classify employees into two different categories having intent to leave the company and having no intent to leave the company?

1.3 OBJECTIVE OF THE STUDY

Primary objective:

- To Predict employee turnover intention in IT & ITeS industry using machine learning algorithms

Secondary objective:

- To classify employees into different categories based on factors
- To extract features that affect the turnover intention of the employees
- To identify the best classification algorithm in predicting turnover intention
- To derive sentiments out of data using text mining

CHAPTER 2

INDUSTRY PROFILE

2.1 BACKGROUND OF THE INDUSTRY

India's IT Services industry was born in Mumbai in 1967 with the establishment of the Tata Group in partnership with Burroughs. Highly skilled Indians immigrated to the western countries for taking up jobs from the 1970s onwards as India's universities and colleges produced more engineers than the Indian industries and factories could absorb. India's growing significance in the information technology enabled it to form a good relationship with the United States of America and several European Countries. The first software export zone, SEEPZ – the precursor to the modern-day IT park – was established in Mumbai in 1973. More than 80 percent of the country's software exports were from SEEPZ in the 1980s. The Indian Government bought the EVS EM computers from the then Soviet Union, which were used in large companies and research laboratories. The immigration laws in the United States of America were relaxed in year 1965 which attracted a large number of skilled Indian professionals aiming for research. The Indian economy was state-controlled and the state remained hostile to the software industry through the 1970s. Import tariffs were as high as 135% on hardware and 100% on software and software industry was not considered an "industry", so that exporters were ineligible for bank finance.

The National Informatics Centre was established in March 1975, the starting up of The Computer Maintenance Company (CMC) in October 1976. During the same period 1977-1980 the other Information Technology companies of India such as Tata InfoTech, Patni Computer Systems and Wipro had become visible. The 'microchip revolution' of the 1980s had influenced both Smt. Indira Gandhi and her successor Shri. Rajiv Gandhi that electronics and telecommunications were vital to our country’s growth, development and prosperity. During 1986 -87, the Indian government worked upon the formation of three wide-area computer networking schemes: INDONET (intended to serve the IBM mainframes in India), NICNET (the network for India's National Informatics Centre), and the academic research-oriented Education and Research Network (ERNET). In 1988 the World Market Policy and the establishment of the Software Technology Parks of India (STP) scheme helped to attract foreign direct investment, the Indian Government permitted foreign equity of up to 100 percent and duty-free import on all inputs and products. The share of IT industry (software) exports raised from 1 percent of the total exports in 1990 to 38 percent of the total exports in 2011. Bangalore is known as the Silicon Valley of India and contributes around 33% of Indian IT Exports.

2.2 MARKET SIZE

The information technology (IT) industry in India consists of two major components: IT services and business process outsourcing (BPO). The sector has increased its contribution to India's GDP from 1.2% in 1998 to 7.5% in 2012. According to NASSCOM, the sector aggregated revenues of US$147 billion in 2015, where export revenue stood at US$99 billion and domestic at US$48 billion, growing by over 13%. The growth in the IT sector is attributed to increased specialisation, and an availability of a large pool of low-cost, highly skilled, fluent English-speaking workers – matched by increased demand from foreign consumers interested in India's service exports, or looking to outsource their operations. The share of the Indian IT industry in the country's GDP increased from 4.8% in 2005–06 to 7% in 2008. In 2009, seven Indian firms were listed among the top 15 technology outsourcing companies in the world. The business process outsourcing services in the outsourcing industry in India caters mainly to Western operations of multinational corporations. As of 2012, around 2.8 million people work in the outsourcing sector. Annual revenues are around $11 billion, around 1% of GDP. Around 2.5 million people graduate in India every year. Wages are rising by 10–15 percent as a result of skill shortages. India’s IT & ITeS industry grew to US$ 181 billion in 2018-19. Exports pfrom the industry increased to US$ 137 billion in FY19 while domestic revenues (including hardware) advanced to US$ 44 billion. IT industry employees 4.1 million people as of FY19.

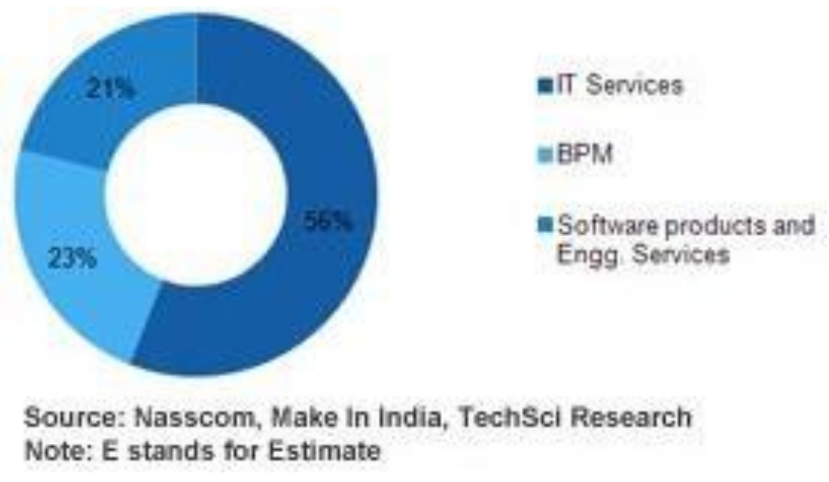

Fig 2.1: Market size of IT industry

2.3 MAJOR PLAYERS

Table 2.1: The following the leading IT companies with their market capitalization.

COMPANIES	MARKET CAPITALIZATION (in cr)
Tata consultancy services	**845,337**
Infosys	**282,028**
HCL Technologies	**153,370**
Wipro limited	**153,043**
Tech Mahindra Ltd	**70,141**
Larsen & Toubro infotech Ltd	**29,302**
Mphasis Ltd	**17,738**
Mindtree Ltd	**11,855**
Hexaware Technologies Ltd	**10,165**
Redington India ltd	**4,438**

2.4 GOVERNENT INITIATIVES

Some of the major initiatives taken by the government to promote IT and ITeS sector in India are as follows:

- On May 2019, the Ministry of Electronics and Information Technology (MeitY) launched the MeitY Start-up Hub (MSH) portal.
- In February 2019, the Government of India released the National Policy on Software Products 2019 to develop India as a software product nation
- The government has identified Information Technology as one of 12 champion service sectors for which an action plan is being developed. Also, the government has set up a Rs 5,000 crore (US$ 745.82 million) fund for realising the potential of these champion service sectors.
- As a part of Union Budget 2018-19, NITI Aayog is going to set up a national level programme that will enable efforts in AI and will help in leveraging AI technology for development works in the country.
- In the Interim Budget 2019-20, the Government of India announced plans to launch a national programme on AI and setting up of a National AI portal.

- National Policy on Software Products-2019 was passed by the Union Cabinet to develop India as a software product nation.

2.5 RECENT TRENDS IN THE INDUSTRY

- **IOT software**
  Internet of Things technology can include any sensor, electronics or software that is connected to the internet and can be utilized remotely and exchange data. Often the technology works together for enhanced functionality
- **IOT hardware**
  IoT Hardware includes a wide range of devices such as devices for routing, bridges, sensors etc. These IoT devices manage key tasks and functions such as system activation, security, action specifications, communication, and detection of support-specific goals and actions.
- **Saas/paas**
  PaaS: hardware and software tools available over the internet. SaaS: software that's available via a third-party over the internet. On-premise: software that's installed in the same building as your business.
- **IOT connectivity**
  IoT Connectivity technologies provide the network infrastructure and communication capabilities required by IoT devices to collect, transport, and exchange data over the internet and to be remotely monitored and controlled.
- **Robotics/drones**
  Robots and drones play an important role in agriculture, horticulture and nature conservation. They replace the senses of farmers, and come into action with Swiss precision. With the use of a robot or other forms of Smart Farming, less energy and raw materials are required, unattractive working conditions are avoided and revenues are higher and of better quality.
- **AR/VR**
  Augmented reality (AR) adds digital elements to a live view often by using the camera on a smartphone. Examples of augmented reality experiences include Snapchat lenses and the game Pokemon Go. Virtual reality (VR) implies a complete immersion experience that shuts out the physical world. Using VR devices such as HTC Vive, Oculus Rift or Google Cardboard, users can be transported into a number ofreal-world

and imagined environments such as the middle of a squawking penguin colony or even the back of a dragon.

- **Big data /analytics**
  Big data analytics is the often-complex process of examining large and varied data sets, or big data, to uncover information such as hidden patterns, unknown correlations, market trends and customer preferences that can help organizations make informed business decisions.
- **Enterprise social software**
  Enterprise social software includes corporate intranets and other software platforms aimed at maximizing productivity, improving communication, saving time and promoting collaboration. Whether the software has been written specifically for corporate communication or it is integrated with more mainstream social media software, the goal of implementing enterprise social software is to improve transparency by making information more accessible in spite of organizational boundaries.
- **Next gen security**
  Next-generation endpoint security uses modern artificial intelligence (AI), machine learning, and a tighter integration of network and device security to provide more comprehensive and adaptive protection than traditional endpoint security solutions.
- **Artificial intelligence**
  Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. Specific applications of AI include expert systems, natural language processing (NLP), speech recognition and machine vision.

2.6 CHALLENGES FACED BY THE INDUSTRY

- **Customer service** Improve customer service by listening to and meeting the client's needs. Make customer service job number one.
- **Human resource:** Develop creative ways to minimize stress, satisfy employee needs, and match corporate needs to employee goals.
- **Productivity:** Make the best use of new technologies like cloud and mobile computing but search out additional ways to increase productivity.
- **Complexity:** Manage and tame the complexity beast.
- **Obsolescence:** Manage and tame the complexity beast. Increase the productive life of systems, software, and equipment.

- **Budget:** Accomplish more with budgets similar to last year.
- **Marketing/public relations:** If you don't have the expertise, hire marketing and PR experts who can get it right.
- **Multinational operations:** Install a culture of teamwork among international team members with diverse backgrounds and varying ethnicities.
- **Mobile generations:** Make use of mobile technology without tearing down the virtual wall between work and family and leisure time.
- **Data storage and retrieval:** Determine what data, if any, is susceptible to bit rot and transfer to new media before it becomes a problem.

CHAPTER 3

REVIEW OF LITERATURE

Tanya Attri(2018) predicts why an employee leaves the organization using data mining techniques and found that by using various models it was found that SA-SVM which was tuned with Bayesian optimization which sows the sensitivity of about 80.43% but low accuracy. The model with good accuracy shows a sensitivity of about 20.6%. so, the feature selection is more important.

Rohit Punnoose (2016) examines about the Prediction of Employee Turnover in Organizations using Machine Learning Algorithms and found that the XG boost is a superior algorithm hat shows a high accuracy with low run time has an efficient memory utilization to predict the employee turnover and it helps in retention of the employee.

Sandeep Yadav et al (2018) examines about the Early Prediction of Employee Attrition using Data Mining Techniques and it was found that salary or other financial aspect like promotions are not the main reasons behind the attrition of employees.

Rachna Jain et al (2018) did a study on Predicting Employee Attrition using XG Boost Machine Learning Approach predicts the model has accuracy less than 30% bit the model XG Boost shows accuracy 90%.

Huey-Ming Tzeng et al (2004) study Predicting nurses' intention to quit with a Support Vector Machine: A new approach to set up an early warning mechanism in human resource management and found that the prediction has the accuracy rate of about 89.2%. Thus, the Support Vector Machine can predict nurses' intention to quit the organization by without asking these nurses whether they have an intention to quit organization.

Amir Mohammad Esmaieeli Sikaroudi el al (2015) studied data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing). The results show that SVM, PNN and KNN are sensitive to parameters. In contrast, Naive Bayes is the most user-friendly model that has a good performance in classification.

Shikha N. Khera (2019) Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques shows that the model accuracy results about 52 percent to 97 percent. The SVM model shows confusion matrix result about 85 percent and misclassification error of about 14%.

Saranya et al (2015) studies the impact of perceived organisation support and organisation commitment on turnover intention of women employees in IT industry. This shows the relationship between the between perceived organization support and organization commitment which decreases the turnover intentions. Organization commitment has the highest effect on turnover intentions among the IT professional.

Jesse W. Campbell et al (2014) studied about the Internal Efficiency and Turnover Intention: Evidence from Local Government in South Korea and found that public service motivation is related to the job performance and other behaviours. Decease in organization efficiency and least motivation result in turnover intention. Therefore, increase overall performance.

Ramachandran et al (2011) examines about an Interactive Mining Approach to find the job Satisfaction and Staff Turnover Intentions and found that job satisfaction is explained by the effect of and miss leading of control, where the effect of job stress on job satisfaction is found to be insignificant in the research.

Priyada Sudhakaran et al (2019) examines about the Understanding the relationship between work variables and voluntary turnover intentions of software professionals in India the purpose of identifying the relationship if any between the number of jobs changed by an employee and his present designation with his voluntary turnover intentions by using anova technique

Vidya v. Iyer (2011) studies about the Understanding turnover intentions and behaviour of Indian information systems professionals: a study of organizational justice, job satisfaction and social norms. It used Discriminant validity to find the turnover intentions and behaviours of Indian IS professionals using a theoretical framework that was most relevant in the Indian context.

Anuruddhika Jayasundera (2017) explains about the effect of perceived organizational support on turnover intention among Gen Y employees while also examining the impact of leader member exchange on the relationship between perceived organizational support (POS) and Turnover intention (TI). By using structured equation modelling the relationship between POS and TI is mediated by JS and OC. Hence, it was verified that JS and OC, which can be considered as outcomes of POS, also contribute in reducing turnover intention.

Anupama sharma et al (2015) studied about Job-Leisure Conflict, Turnover Intention and the Role of Job Satisfaction as a Mediator: An Empirical Study of Indian IT Professionals. By using correlation test the job-leisure conflict influences the IT professionals' turnover intention

Vidya V. Iyer et al (2008) studied about the Turnover intentions of Indian is professionals and as a result Job Satisfaction, Organizational Satisfaction, and Social Norms as the main determinants of Turnover Intentions among Indian IS Professionals.

Abdulmajeed Saad Albalawi et al (2019) studied about the Perceived organizational support, alternative job opportunity, organizational commitment, job satisfaction and turnover intention: a moderated-mediated model and by using contemporary variance-based structural equation modelling they predict the organizational commitment mediates the association between perceived organizational support and turnover intention, perceived alternative job opportunities and turnover intention.

Yuting Li et al (2019) examined about the Empirical analysis of factors impacting turnover intention among manufacturing workers by using SPSS and structural equation modelling they predict that job satisfaction and organizational commitment negatively and significantly affected manufacturing workers' turnover intentions, while work-family conflict positively and significantly affected turnover intentions

Antonio Frian (2018) studied about the Millennials employee turnover intention in Indonesia and by using multiple regression analysis the author found that millennial employee turnover intention significantly affected by perceived alternative employment opportunity and employee development system.

Jacobs E et al (2008) examined about the Organisational culture of hospitals to predict turnover intentions of professional nurses by using general linear modelling the results are concluded that the organisational culture has a significantly negative correlation with turnover intentions. Organisational culture also interacted with job satisfaction, knowledge sharing, and the white professional nurses' category to decrease turnover intentions and with Organisational Citizen Behaviours to increase turnover intentions in a final predictive model

Fasanmi samuel Sunday (2016) studied about the Organizational citizenship behaviour and turnover intent: a path analysis of Nigeria bankers' behavioural variables. By using multivariate multiple regression analysis, he concluded that affective commitment, procedural justice and psychological empowerment have direct effects on the negative relationship between citizenship behaviour and turnover intent.

Caroline Arnoux-Nicolas et al (2016) examined about the Perceived work conditions and turnover intentions the mediating role of meaning of work by using multiple regression the results show that adverse working conditions were positively and significantly associated with turnover intentions

Hemdi et al (2006) studied about the Predicting turnover intentions of hotel employees: the influence of employee development human resource management practices and trust in organization and by using multiple regression, Principal component factor analyses and as a result It is suggested that to enhance employees' trust in organization and subsequently to reduce turnover intentions, hotels need to continue to provide training and development programs to their employees, conduct fair and formal appraisal system, and provide ample and clear career advancement to their employees

Suhaidah Hussain et al (2019) studied about the Factors affecting employees' turnover intention in construction companies in Klang, Selangor and by using Multiple Regression Analysis the results are concluded that communication and organizational politics had a negative relationship with employees' turnover intention

Mehmet Nurettin Ugural et al (2020) analysed Determinants of the turnover intention of construction professionals: a mediation analysis and by using Confirmatory Factor Analysis (CFA), mediation analysis the results are concluded that individual difference in the self-construal is related to turnover intention indirectly by virtue of employees' perceptions of organizational prestige.

Biyan wen et al (2020) studied about the Role stress and turnover intention of front-line hotel employees: the roles of burnout and service climate and by using exploratory factor analysis the result is concluded that that role stress as a four-dimensional construct (i.e., conflict, ambiguity, qualitative overload and quantitative overload) has a statistically significant impact on burnout, which leads to turnover intention.

Everd Jacobs et al (2007) analysed the development of a knowledge sharing construct to predict turnover intentions and by using general linear modelling the results are predicted that significant negative relationship was found between knowledge sharing behaviour and turnover intentions

Orhan Uludag et al (2011) analysed that the effects of job satisfaction, organizational commitment, organizational citizenship behaviour on turnover intentions and by using multiple regression the results are predicted that that job satisfaction is positively related to organizational citizenship Behavior and negatively related to turnover intention

T. Rathakrishnan et al (2016) examined that Turnover intentions of lecturers in private universities in Malaysia by using multiple regression. The results are predicted that compensation satisfaction, job autonomy, KPI achievability, and job satisfaction explained turnover intention.

Bandhanpreet Kaur et al (2013) examined that Antecedents of turnover intentions: a literature review and the result from the analysis has been concluded as quality of work life, job stress, job satisfaction and organizational justice have an impact on the turnover intentions. As turnover intentions are the antecedent of the turnover of the employees

Archana Singh et al (2017) examined that Antecedents of turnover intention: testing a conceptual model in the context of professionals in India and by using Discriminant validity, factor analysis the results concluded that the organizations can suitably modify their HR policies and programs which in turn will help in retaining professionals, high potentials and those possessing critical skills which will reduce their human capital costs, thus providing them with a competitive advantage in the marketplace.

Alvia Santoni et al (2018) studied about the model of turnover intentions of employees and by using structured equation modelling found that the turnover intention employees, especially the intention to move but for fear not getting better job can be lowered if employees feel satisfied with the work itself that is reinforced by work environment extern/internal factors that in the form of selfish abstinence.

Belete ak (2018) analysed about the Turnover intention influencing factors of employees: an empirical work review and by the analysis it has been concluded that the job satisfaction, job

stress, organizational culture, organizational commitment, salary, organizational justice, promotional opportunity, demographic variables, leadership styles, and Organizational Climate.

3.1 RESEARCH GAP

There were a number of gaps left by different reviewed theoretical and empirical literature ranging from geographical, methodologies used, time as well as the nature of organization studied. There is lack of studies based on the intention of the employee to leave the organization. Most of the studies was based on the turnover rate of the employee within the organization. It is important to know the factors that affect the intention of the employee that the turnover rate of the employees. The literatures were based on turnover rate of the certain organizations and particular geography but less focused on the intention of the employee that act as a main factor to leave the organization. If the intention of the employee is identified the organization can make required changes in the strategies to minimize the turnover rate of the employees. This study aimed to fill gaps left by the previous researches specifically in assessing the factors influencing employee turnover intention in IT & ITeS industry with special reference to Coimbatore using supervised machine learning algorithms.

CHAPTER 4

RESEARCH METHODOLOGY

4.1 INTRODUCTION

The methodology involved in the study was CRISP-DM and classification algorithm was used to predict the turnover intention of the employee where the data is collected by google form which consists of quantitative information's which focus on the various aspects of turnover intention in Coimbatore, Chennai and Bangalore districts at IT & ITeS industry. the phases of CRISP-DM are shown diagrammatically and also listed below.

- Business Understanding
- Data Understanding
- Data Preparation
- Modelling
- Evaluation
- Deployment

Business understanding:

understanding the goals and prerequisites from a business viewpoint, and convert this information into a data mining problem definition and a fundamental arrangement intended to accomplish the targets. The problem being defined in the study was to analyse the major factors that influence the turnover intention of the employees in IT and ITeS industry and to predict the reason behind the intention of the employees.

Data understanding:

The data has been started collecting and then get familiar with the data then to identify the problems and to discover the basic insights that has been view in the data collected. With the help of the charts obtained from the data's that has been collected from the respondents the basic insights can be gathered and insights can be understood.

Data preparation

By includes all activities required to develop the final data set the initial raw data has been processed. Tasks include table, case, and attribute selection as well as transformation and cleaning of data for modelling tools.

Modelling

The selected modelling techniques has been applied and calibrate the tool parameters to obtain the values. Typically, there are several techniques for the same data mining problem type. Some techniques have specific requirements on the form of data. The classification algorithm has been adopted to obtain the final results and also text mining gas been performed to understand the data in clear.

Evaluation

After evaluating the model, the steps has been executed to construct the model, to be certain it properly achieves the problem objectives to determine important issue that has not been sufficiently considered.

Deployment

Organize and present the results of data mining. Deployment can be as simple as generating a report or as complex as implementing a repeatable data mining process. To evaluate the best prediction model and the major factors that affect the turnover intention of the employees.

4.2 POPULATION SIZE

The target population of IT and ITeS employees are in the cities of Chennai, Bangalore and Coimbatore which has population about 1.5 million in Bangalore, 5 lakhs in Chennai and 1 lakh in Coimbatore.

4.3 SAMPLE SIZE

All IT & ITeS employees of Coimbatore, Bangalore and Chennai districts in Tamil Nadu in India. The sample size obtained for the study was 416 which was collected from the respondents.

4.4 RESEARCH DESIGN

Predictive research: Predictive Design is a commonly used statistical technique to predict future behaviour. Predictive Design solutions are a form of data-mining technology that works by analysing historical and current data and generating a model to help predict future outcomes. In this research the predictive design is used to predict the turnover intention of the employees due to the factors such job satisfaction, stress, recognition, organization commitment, alternative job opportunities and perceived organizational support.

4.5 SAMPLING DESIGN

Convenience sampling were adopted where the respondent population was selected based on the convenience and it has been collected which is convenient to hand

4.6 TOOLS FOR DATA COLLECTION

The questionnaire was structured comprising of 25 questions consist of demographic variables and predictor variables. The variables have been selected based on the review of literature.

4.7 FEATURES OF DATA

The aim of the study is to find the predicting employee turnover intention in IT & ITeS industry with special reference to Coimbatore using supervised machine learning algorithms. It consists of demographic variables like age, gender, educational status, marital status, experience, salary etc., and predictor variables like job satisfaction, job stress, recognition, organization commitment, alternative job opportunities and perceived organizational support.

4.8 METHOD OF DATA COLLECTION

The data collection method adopted for the study was online survey. Google from has been created for the survey which consists of 33 variables. From that 7 variables contains subset of questions. The main demographic variables are collected along with the factors that affect the turnover intention of the employees.

CHAPTER 5

DATA MODELLING AND COMPARISION

Table 5.1: demographic factors distribution:

Variables	Number of respondents (416)	Percentage of respondents
Gender		
Male	229	55%
Female	187	45%
Education		
Under Graduate	329	79%
Post Graduate	83	20%
Doctorate	4	1%
Marital status		
Married	320	77%
Unmarried	96	23%
Salary		
Under ₹41,000	287	69%
₹41,000-₹60,000	50	12%
₹61,000-₹80,000	42	10%
₹81,000-₹100,000	37	9%
Wish to Relocate from workplace		
Yes	200	48%
No	216	52%
Working in home town		
Yes	104	25%
No	312	75%

The age of the respondents (fig.5.1) are sparsely distributed where 81(21%) respondents are in the age of 22, 97(23%) respondents are in the age of 23, 52(13%) of employees are in the age of 25. In average 57% of respondents are between the age of 21-25 with the under graduate degree (79% respondents) remains unmarried (77% respondents) and receive the salary under

₹41,000 (69% respondents). Employees are not willing to relocate from the workplace even though they are not working in their home town. The major factors that determine the intention of the employees and the factors that influence the turnover intention of the employees are discussed in the following.

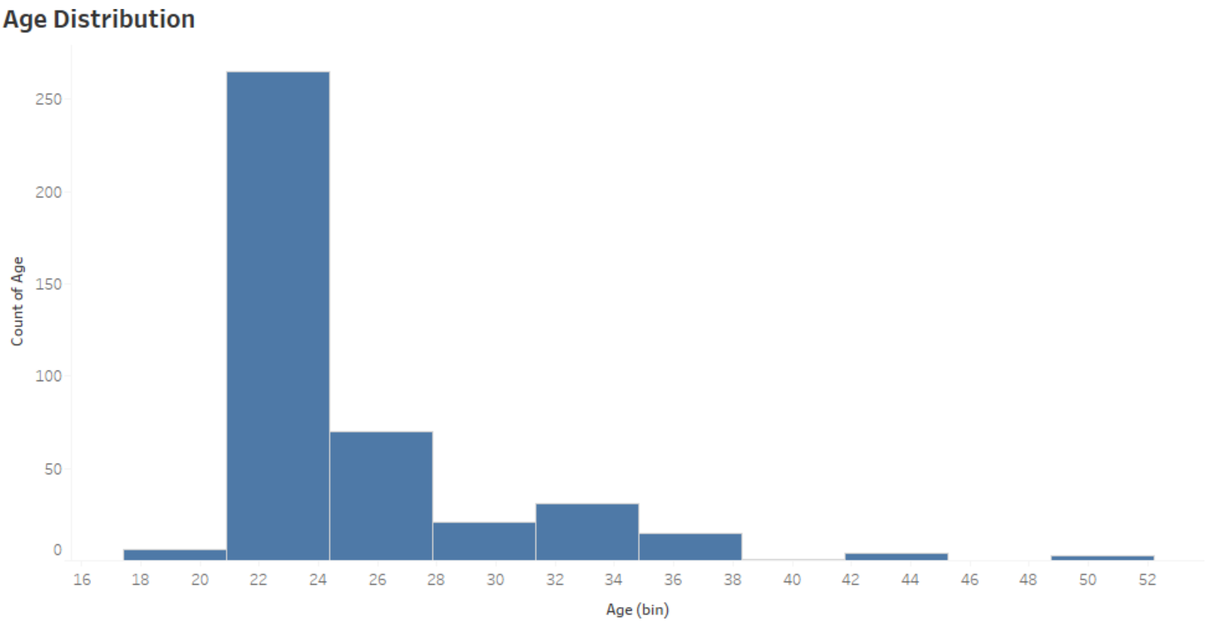

Fig 5.1 Age distribution

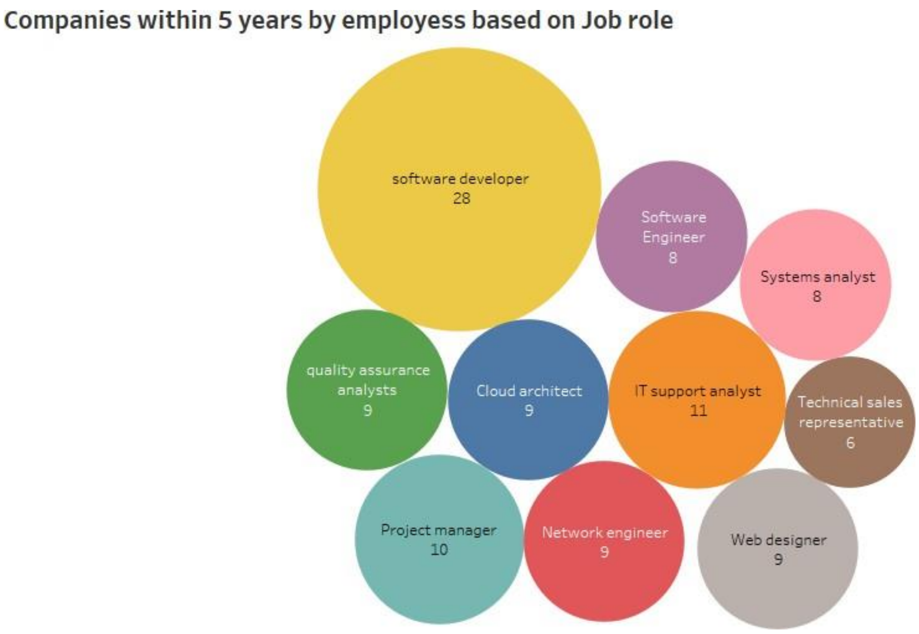

Fig 5.2: Switching companies based on job role

From the fig 5.2 more software developers switched job within 5 years followed by that project manager hopped job within 5 years.

5.1 LOGISTIC REGRESSION

Logistic Regression is used to predict the probability of occurrence of a dependent variable using independent variables. This algorithm has been used to determine the factors that affect the intention of the employee to leave the company. The analysis was performed in SPSS software.

Forward method:

Table 5.2: Model Summary

Step	-2Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	539.711[a]	.085	.113
2	466.512[b]	.232	.310
3	452.986[b]	.257	.343
4	445.520[b]	.270	.360
5	440.047[b]	.280	.373
6	433.554[b]	.291	.388
7	427.620[b]	.301	.401
8	422.109[b]	.310	.414
9	416.919[b]	.319	.425
10	411.389[b]	.328	.437
11	403.197[b]	.341	.454

This table contains the Cox & Snell R Square and Nagelkerke R Square values, which are both methods of calculating the explained variation. These values are sometimes referred to as pseudo R^2 values (and will have lower values than in multiple regression). Therefore, the explained variation in the dependent variable based on our model ranges from 34.0% to 45.0%, depending on whether in reference to the Cox & Snell R^2 or Nagelkerke R^2 methods, respectively

Table:5.3: Confusion matrix of forward logistic regression:

		Predicted class	
		0	1
Actual class	0	147	57
	1	53	159

Table 5.4: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 11[k]	Gender	.654	.270	5.849	1	.016	1.923
	Education	1.097	.400	7.537	1	.006	2.996
	Depended family members	-.396	.098	16.194	1	.000	.673
	Marital Status	-1.354	.400	11.450	1	.001	.258
	Job role	-.039	.014	7.344	1	.007	.962
	Willing to relocate from the work place	.731	.261	7.843	1	.005	2.076
	Job satisfaction	-1.044	.215	23.503	1	.000	.352
	Alternative job opportunity	1.695	.256	43.825	1	.000	5.449
	Job stress	.493	.209	5.572	1	.018	1.636
	Organizational commitment	-.671	.189	12.561	1	.000	.511
	Covid	.497	.167	8.851	1	.003	1.643
	Constant	-3.424	1.284	7.112	1	.008	.033

The last table is the most important one for our logistic regression analysis. It shows the regression function and we can predict the model that affect the turnover intention of the employees.

Log (p/1-p) = -3.424+0.654*gender+1.097*education-0.396*depended family members -1.354*Marital status-0.039*job role+0.731*willing to relocate from work place-1.044*job satisfaction+1.695*alternative job opportunity+0.493*job stress-0.671*organizational commitment+0.497*covid

The factors that influence the turnover intention of the employees has been predicted by using logistic regression. The factors are education, depended family members, marital status, job role, willing to relocate form work place, job satisfaction, alternative job opportunity, job stress, organizational commitment and Attitude towards covid.

5.2 LOGISTIC REGRESSION (Backward method):

Table 5.5: Model Summary

Step	-2Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	385.502[a]	.368	.491
2	385.502[a]	.368	.491
3	385.635[a]	.368	.491
4	385.842[a]	.368	.490
5	386.071[a]	.367	.490
6	386.289[a]	.367	.489
7	386.568[a]	.367	.489
8	387.101[a]	.366	.488
9	387.941[a]	.365	.486
10	388.563[a]	.364	.485
11	389.601[a]	.362	.483
12	390.637[a]	.360	.481
13	394.029[a]	.355	.474
14	398.407[a]	.348	.464
15	399.697[a]	.346	.462
16	403.197[a]	.341	.454

This table contains the Cox & Snell R Square and Nagelkerke R Square values, which are both methods of calculating the explained variation. These values are sometimes referred to as pseudo R^2 values (and will have lower values than in multiple regression). Therefore, the explained variation in the dependent variable based on our model ranges from 34.0% to 45.0%, depending on whether in reference to the Cox & Snell R^2 or Nagelkerke R^2 methods, respectively

Table:5.6: Confusion matrix of backward logistic regression:

		Predicted class	
		0	1
Actual class	0	147	57
	1	53	159

Table 5.7: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 16[a]	Gender	.654	.270	5.849	1	.016	1.923
	Education	1.097	.400	7.537	1	.006	2.996
	Depended family members	-.396	.098	16.194	1	.000	.673
	Marital Status	-1.354	.400	11.450	1	.001	.258
	Job role	-.039	.014	7.344	1	.007	.962
	Willing to relocate from the work place	.731	.261	7.843	1	.005	2.076
	Job satisfaction	-1.044	.215	23.503	1	.000	.352
	Alternative job opportunity	1.695	.256	43.825	1	.000	5.449
	Job stress	.493	.209	5.572	1	.018	1.636
	Organizational commitment	-.671	.189	12.561	1	.000	.511
	Covid	.497	.167	8.851	1	.003	1.643
	Constant	-3.424	1.284	7.112	1	.008	.033

The last table is the most important one for our logistic regression analysis. It shows the regression function and we can predict the model that affect the turnover intention of the employees.

Log (p/1-p) = (-3.424) + 0.654*gender + 1.097*education – 0.396*depended family members – 1.354*marital status – 0.039*job role + 0.731*willing to relocate from workplace – 1.044*job satisfaction + 1.695*alternative job opportunity + 0.493*job stress – 0.671*organizational commitment + 0.497* Attitude towards covid

The factors that influence the turnover intention of the employees has been predicted by using logistic regression. The factors are education, depended family members, marital status, job role, willing to relocate form work place, job satisfaction, alternative job opportunity, job stress, organizational commitment and Attitude towards covid.

Classification algorithm comparison:

The following algorithms Naïve bayes, K nearest neighbour, kernel SVM, linear SVM, random forest and XG Boost has been performed in python with the model selection of K fold cross validation of 10 folds. Jupyter notebook has been used for analysing the data where scikit library has been enabled. Standard scaler has been used for standardizing the feature mean and then scaling the unit variance to 0. It has been used after splitting of the data. The precision, recall and f score has been detected with ROC curve.

Positive class: Employee have intention to leave the company mentioned as 1

Negative class: Employee doesn't have intention to leave the company mentioned as 0

Precision: Precision defines the ratio of correct prediction of employees having intention to switchover to the total prediction of employees having intention to switchover. It is defined as ratio of correct positive predictions to the total predicted positives.

Precision = TP / (TP+FP)

Recall: Recall defines the ratio of correct prediction of employees having intention to switchover to the total employees having an intention to switchover. It is the probability of detecting true positive rate. This is defined as ratio of correct positive predictions to the total positives.

Recall = TP / (TP+FN)

F score: F1 Score is defined as the weighted average of both Precision and Recall

F1 Score = 2*(Recall * Precision) / (Recall + Precision)

ROC curve: It tells how much model is capable of distinguishing between classes. The ROC curve is plotted with true positive rate against the false positive rate where true positive rate is on y-axis and false positive rate is on the x-axis

5.3 NAÏVE BAYES:

Naïve Bayes is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. From the performance of the algorithm the confusion matrix are as follows,

Table:5.8: Confusion matrix of Naïve bayes:

		Predicted class	
		1	0
Actual class	1	105	61
	0	150	100

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.6

test precision: 0.6

test recall: 0.71

test f1 score:0.65

The naïve bayes algorithm gives 0.6 which means the model shows overall accuracy of 60% of predicting the intention of employee to leave the company and to stay within the company. Precision value 60% shows the percentage of correct positive predictions to the total predicted positives. Recall value 71% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.65. The area under curve is about 0.55 which shows that 55% accurate of performance of the model at distingushing between positive and negative classes.

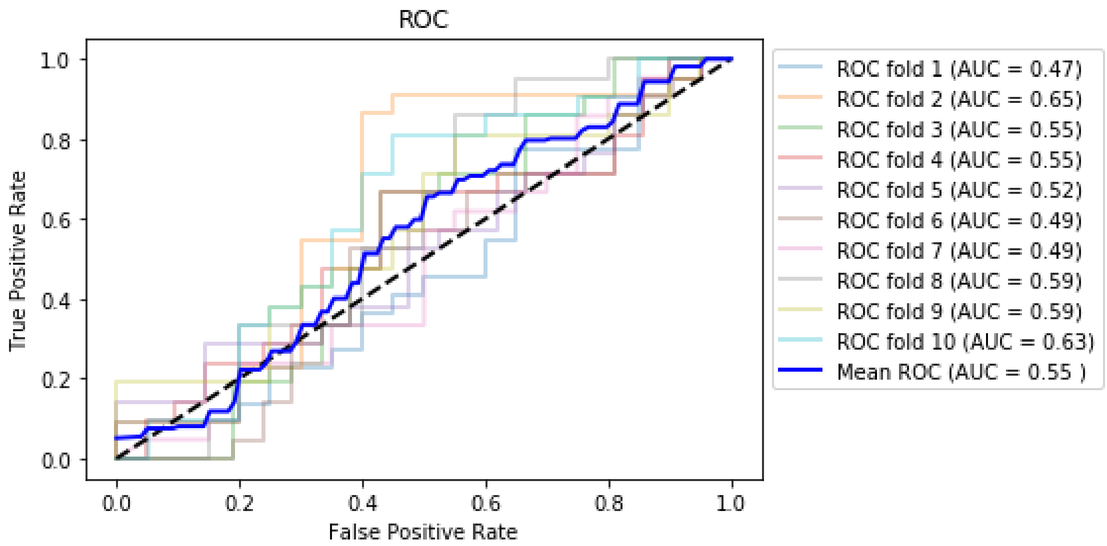

Fig 5.5 ROC curve of naïve bayes

5.4 K NEAREST NEIGHBOUR:

K nearest neighbour algorithm is one of the simplest forms of classification algorithms and it is one of the most used learning algorithms. Its purpose is to use a database in which the data points are separated into several classes to predict the classification of a new sample point. The k nearest will be found by using set k=sqrt(n). The total variables in the study that has been defined was 27. So, for defining the K value of the model the square root of n=27 has been taken as 5. And k value is defined as 5. From the performance of the algorithm the confusion matrix are as follows,

Table:5.9: Confusion matrix of K nearest neighbour:

		Predicted class	
		1	0
Actual class	1	173	44
	0	33	166

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.81

test precision: 0.83

test recall: 0.78

test f1 score:0.81

The K nearest algorithm gives 0.81 which means the model shows overall accuracy of 81% of predicting the intention of employee to leave the company and to stay within the company. Precision value 83% shows the percentage of correct positive predictions to the total predicted positives. Recall value 78% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.81. The area under curve is about 0.73 which shows that 73% accurate of performance of the model at distingushing between positive and negative classes.

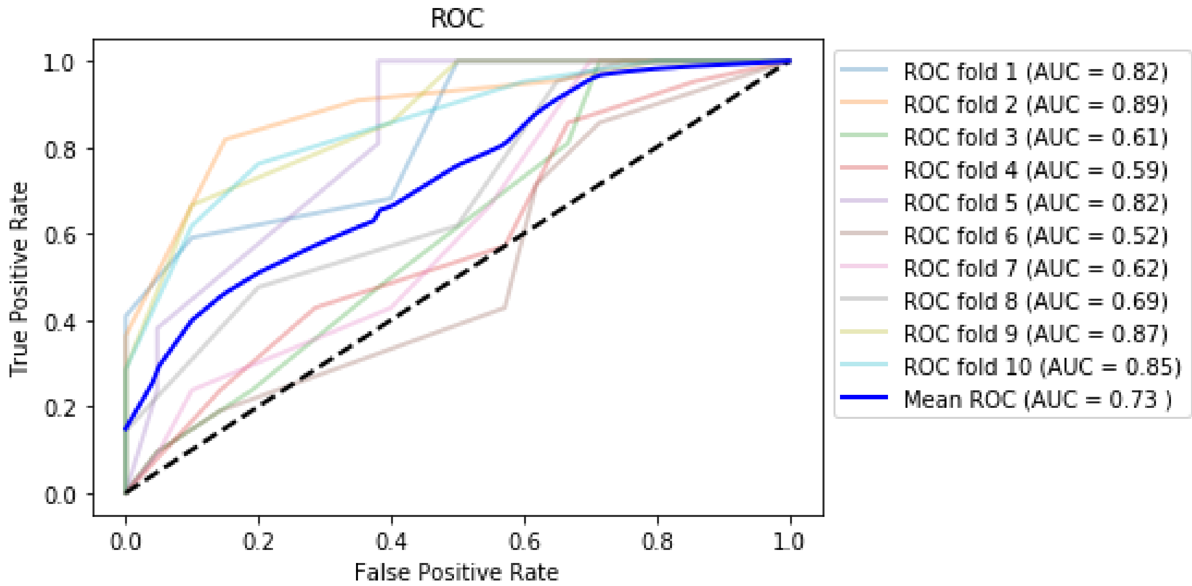

Fig 5.3: ROC curve of K Nearest Neighbor

5.5 KERNEL SVM:

The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types. For example, linear, nonlinear, polynomial, radial basis function (RBF), and sigmoid. In this study radial basis function is being used for kernelizing the function. From the performance of the algorithm the confusion matrix are as follows,

Table:5.10: Confusion matrix of Kernel SVM:

		Predicted class	
		1	0
Actual class	1	183	8
	0	22	183

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.88

test precision: 0.89

test recall: 0.86

test f1 score:0.88

The kernel SVM algorithm gives 0.88 which means the model shows overall accuracy of 88% of predicting the intention of employee to leave the company and to stay within the company. Precision value 89% shows the percentage of correct positive predictions to the total predicted positives. Recall value 86% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.88. The area under curve is about 0.65 which shows that 65% accurate of performance of the model at distingushing between positive and negative classes.

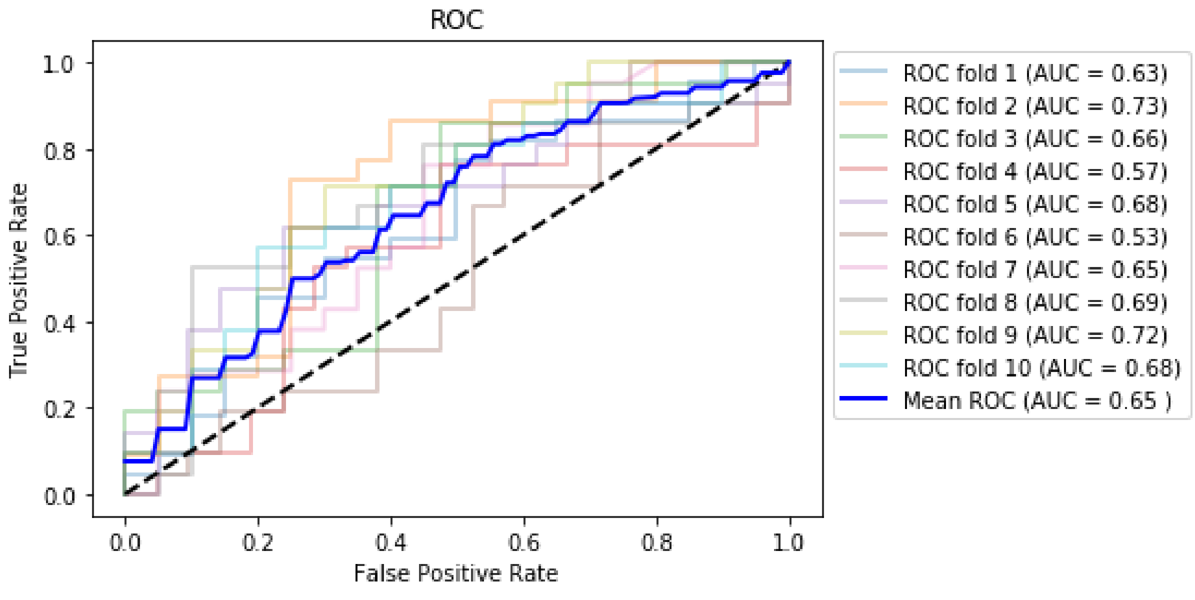

Fig 5.4 ROC curve of kernel SVM

5.6 RANDOM FOREST:

The random forest is a classification algorithm which consist of many decisions' trees. It uses bagging (ensemble) and feature randomness for predicting when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. A random forest reduces the variance of a single decision tree which leads to better predictions and understanding on new data. Entropy criterion has been used for the prediction. From the performance of the algorithm the confusion matrix are as follows,

Table:5.11: Confusion matrix of Random forest:

		Predicted class	
		1	0
Actual class	1	183	28
	0	22	183

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.88

test precision: 0.89

test recall: 0.86

test f1 score:0.88

The random forest algorithm gives 0.88 which means the model shows overall accuracy of 88% of predicting the intention of employee to leave the company and to stay within the company. Precision value 89% shows the percentage of correct positive predictions to the total predicted positives. Recall value 86% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.88. The area under curve is about 0.96 which shows that 96% accurate of performance of the model at distingushing between positive and negative classes.

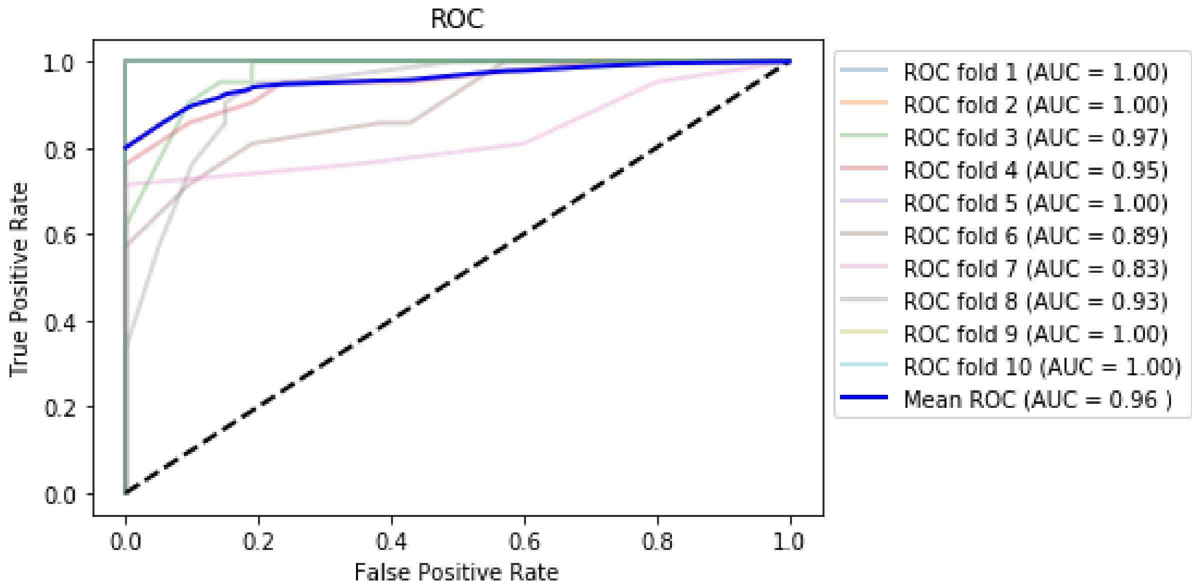

Fig 5.5 ROC curve of random forest

5.7 XG BOOST:

The XG Boost library implements the gradient boosting decision tree algorithm. Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction. It has proven to push the limits of computing power for boosted trees algorithms as it was built and developed for the sole purpose of model performance and computational speed. From the performance of the algorithm the confusion matrix are as follows,

Table:5.12: Confusion matrix of XG Boost:

		Predicted class	
		1	0
Actual class	1	188	44
	0	17	205

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.94

test precision: 0.92

test recall: 0.97

test f1 score:0.94

The XG boost algorithm gives 0.94 which means the model shows overall accuracy of 94% of predicting the intention of employee to leave the company and to stay within the company. Precision value 92% shows the percentage of correct positive predictions to the total predicted positives. Recall value 97% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.94. The area under curve is about 0.97 which shows that 97% accurate of performance of the model at distingushing between positive and negative classes.

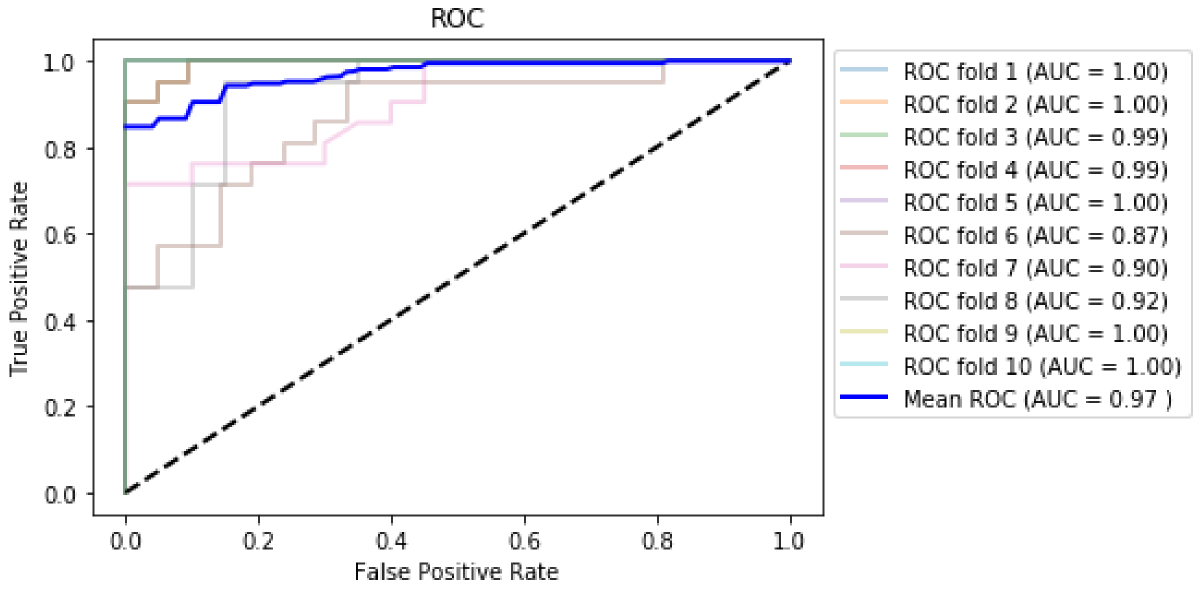

Fig 5.6 ROC curve of XG Boost

5.8 ARTIFICIAL NEURAL NETWORK

An artificial neural network (ANN) is the piece of a computing system designed to simulate the way the human brain analyses and processes information. ANN have self-learning capabilities that enable them to produce better results as more data becomes available. Artificial Neural network is typically organized in layers. Layers are being made up of many interconnected ‘nodes’ which contain an ‘activation function’. The activation function used in this network is sigmoid and rectifier is used. The rectifier activation function is used in the hidden layer and the sigmoid activation function is used in the output layer. Neural network may contain the following 3 layers:

Input layer: The purpose of the input layer is to receive as input the values of the explanatory attributes for each observation.

Hidden layer: The Hidden layers apply given transformations to the input values inside the network. Rectifier activation function is used

Output layer: The hidden layers then link to an 'output layer '. Output layer receives connections from hidden layers or from input layer. It returns an output value that corresponds to the prediction of the response variable. Sigmoid activation function is used.

The ability of the neural network to provide useful data manipulation lies in the proper selection of the weights. This is different from conventional information processing.

From the performance of the algorithm the confusion matrix are as follows,

Table:5.13: Confusion matrix of Artificial neural network:

		Predicted class	
		1	0
Actual class	1	139	54
	0	25	198

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.83

test precision: 0.80

test recall: 0.82

test f1 score:0.80

The artificial neural network gives 0.88 which means the model shows overall accuracy of 83% of predicting the intention of employee to leave the company and to stay within the company. Precision value 80% shows the percentage of correct positive predictions to the total predicted positives. Recall value 82% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.80. The area under curve is about 0.80 which shows that 80% accurate of performance of the model at distingushing between positive and negative classes.

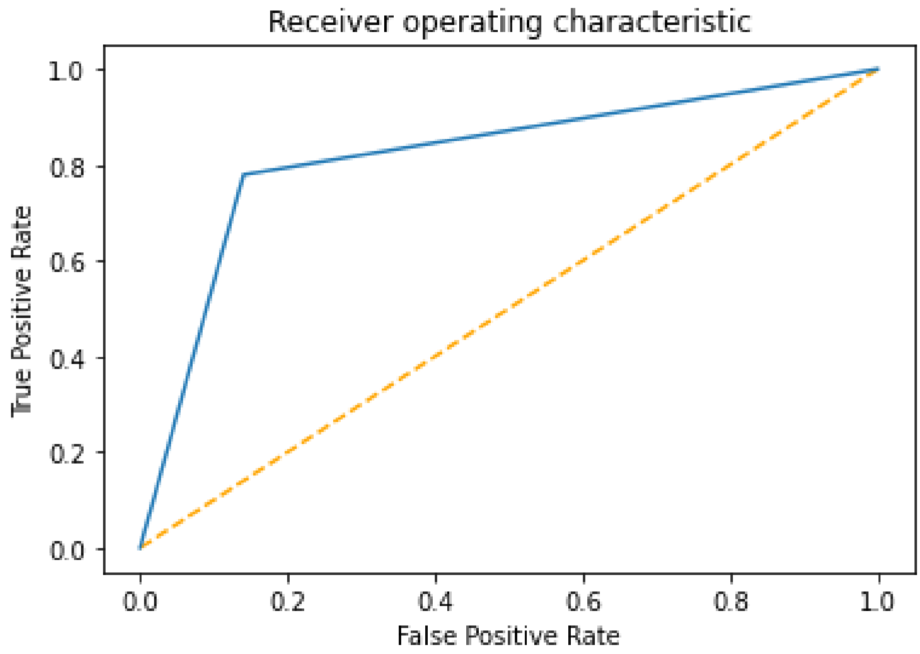

Fig 5.7 ROC curve of artificial neural network

The following is the neural network structure which uses 7 and 7 hidden layer and he output is being detected.

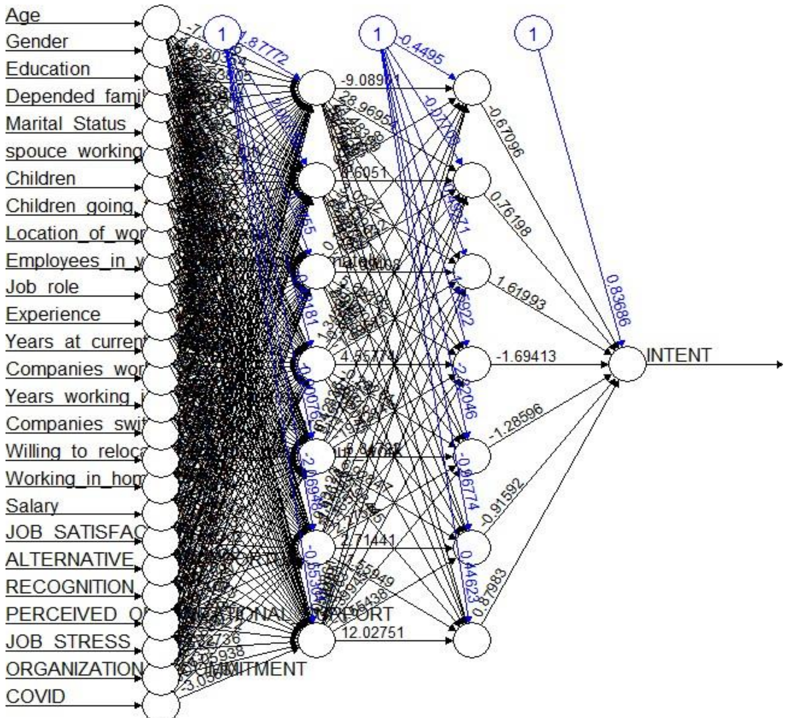

Fig 5.11: neural network diagram

5.9 DECISION TREE

Decision tree builds classification models in the form of a tree structure. It breaks down a data set into smaller and smaller subsets while at the same time an associated decision tree is incrementally developed. The final result is a tree with decision nodes and leaf nodes. The decision tree was performed in Minitab software. The nose splitting has been performed with the Gini index. Gini index or Gini impurity measures the degree or probability of a particular variable being wrongly classified when it is randomly chosen. Kfold cross validation has been used for the model validation.

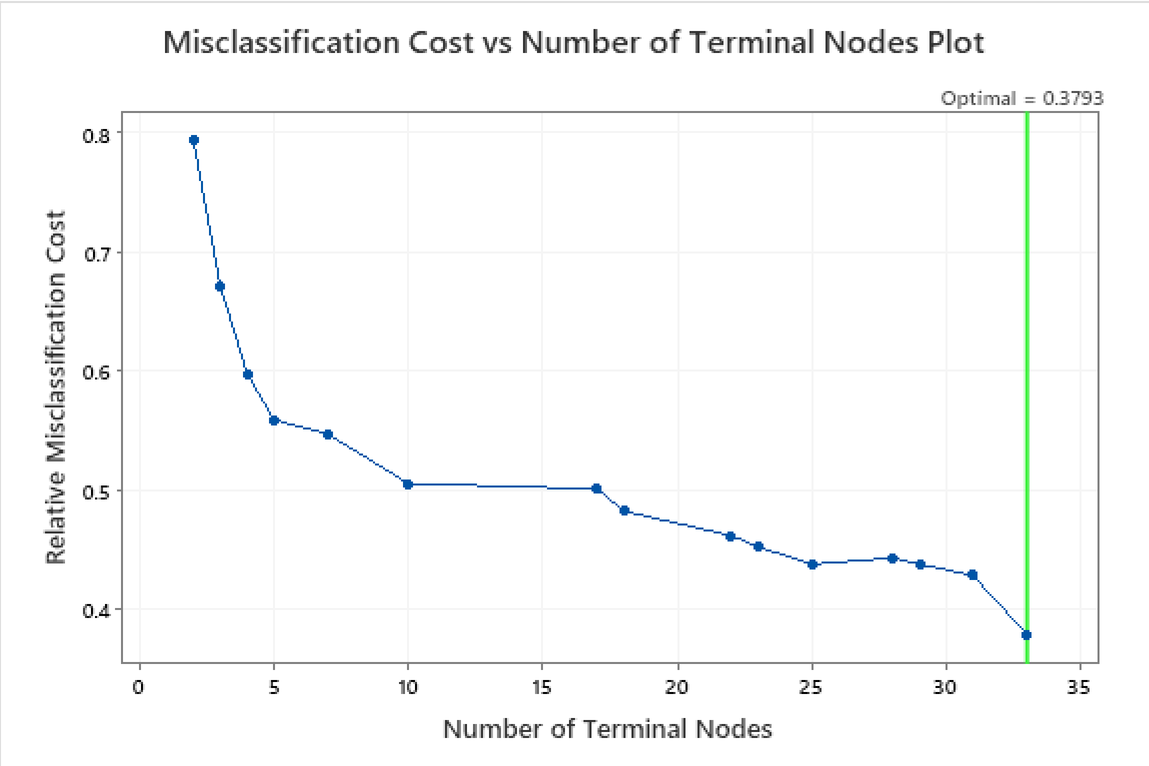

Fig 5.8 optimal score

In this plot, the pattern where the misclassification cost decreases continue after the 30-node tree. The optimal node value as 0.3.

Optimal Tree Diagram

Decision tree model

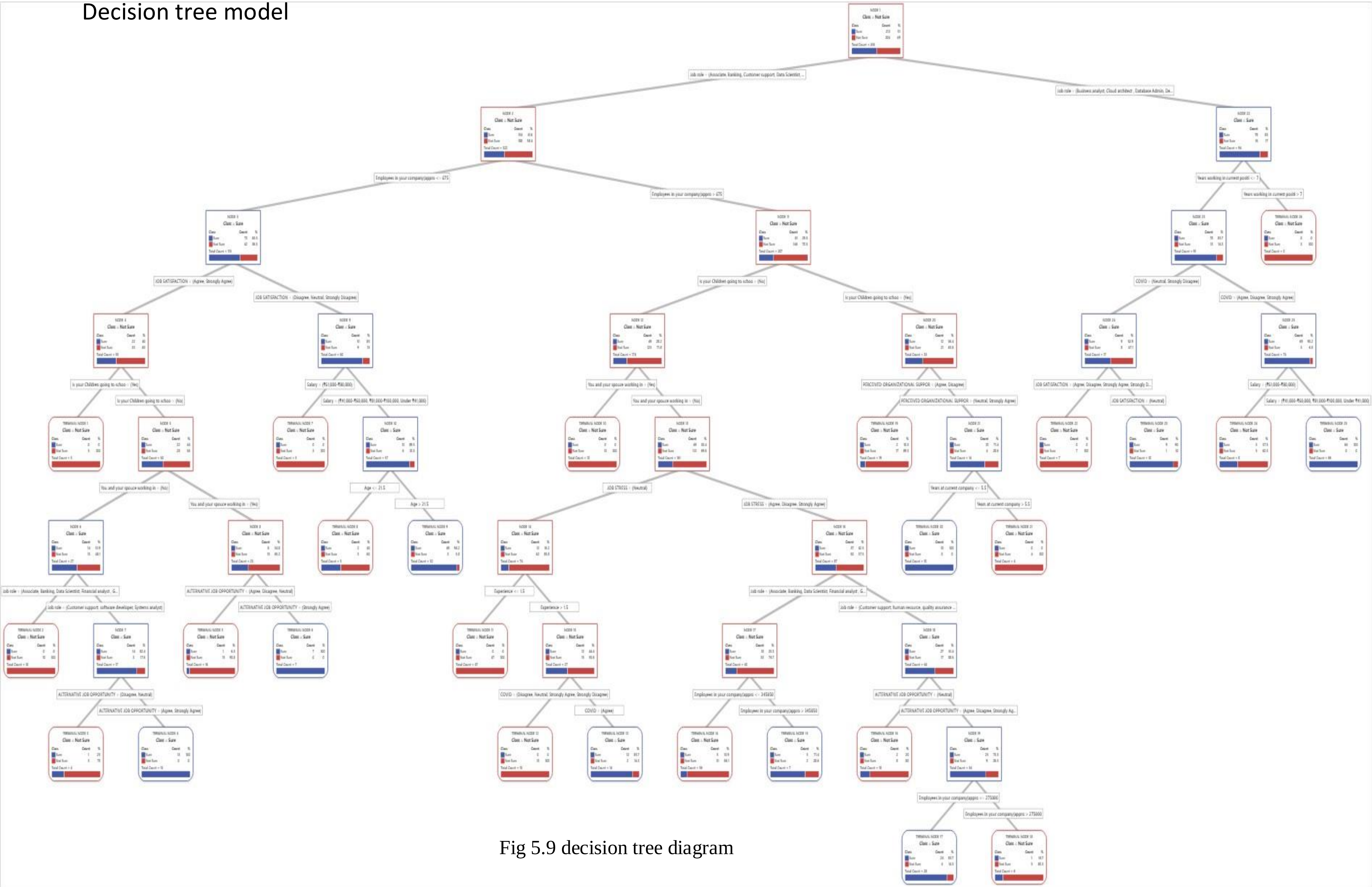

Fig 5.9 decision tree diagram

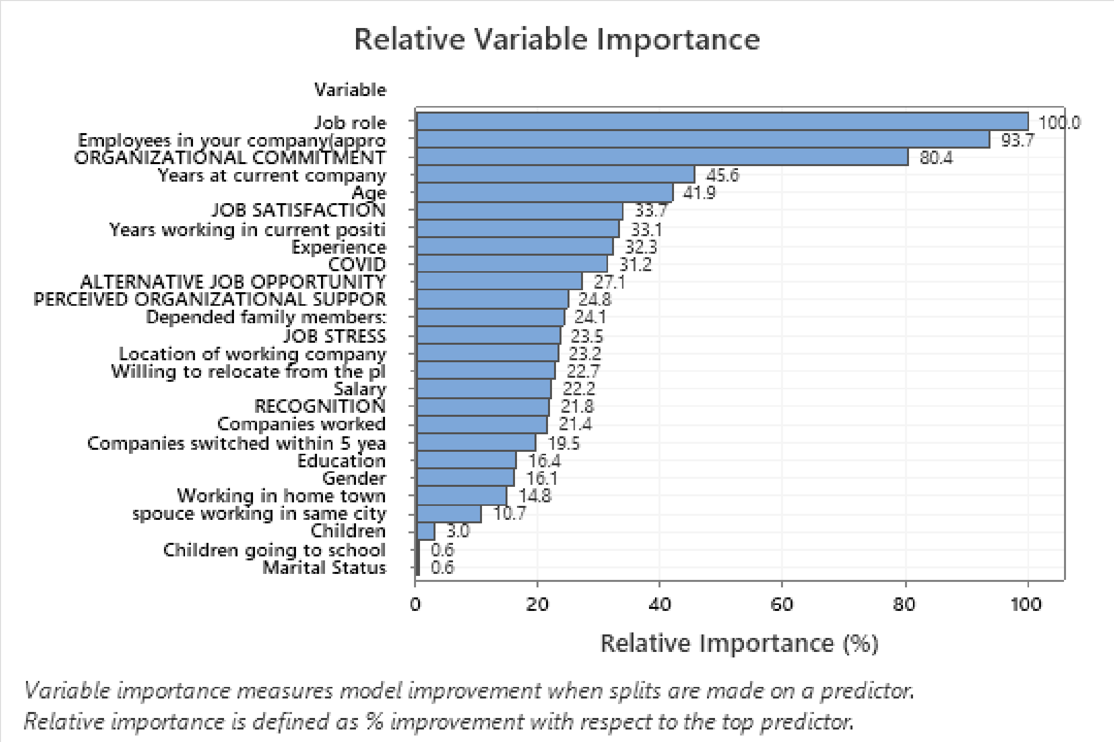

Fig 5.10 relatively important variables

The table shows the relatively important variables. Job role was the relatively highest predictor followed by employees in the company and organizational commitment. Years at current company, age, job satisfaction, years in current position, alternative job opportunity, perceived organizational support, depended family members, job stress, location of working company, willing to relocate, salary, recognition and other variables are in the same improvement.

5.14 Confusion Matrix

		Predicted Class (Training)			**Predicted Class (Test)**		
Actual Class	**Count**	**1**	**0**	**%Correct**	**1**	**0**	**%Correct**
1 (Event)	212	198	14	93.4	173	39	81.6
0	204	5	199	97.5	39	165	80.9
All	416	203	213	95.4	212	204	81.3

Statistics	**Training (%)**	**Test (%)**
True positive rate (sensitivity or power)	93.4	81.6
False positive rate (type I error)	2.5	19.1
False negative rate (type II error)	6.6	18.4
True negative rate (specificity)	97.5	80.9

From the above confusion matrix accuracy, precision, recall and f score has been calculated.

test accuracy: 0.80

test precision: 0.81

test recall: 0.81

test f1 score:0.81

The decision tree algorithm gives 0.80 which means the model shows overall accuracy of 80% of predicting the intention of employee to leave the company and to stay within the company. Precision value 81% shows the percentage of correct positive predictions to the total predicted positives. Recall value 81% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.81. The area under curve is about 0.85 which shows that 85% accurate of performance of the model at distingushing between positive and negative classes.

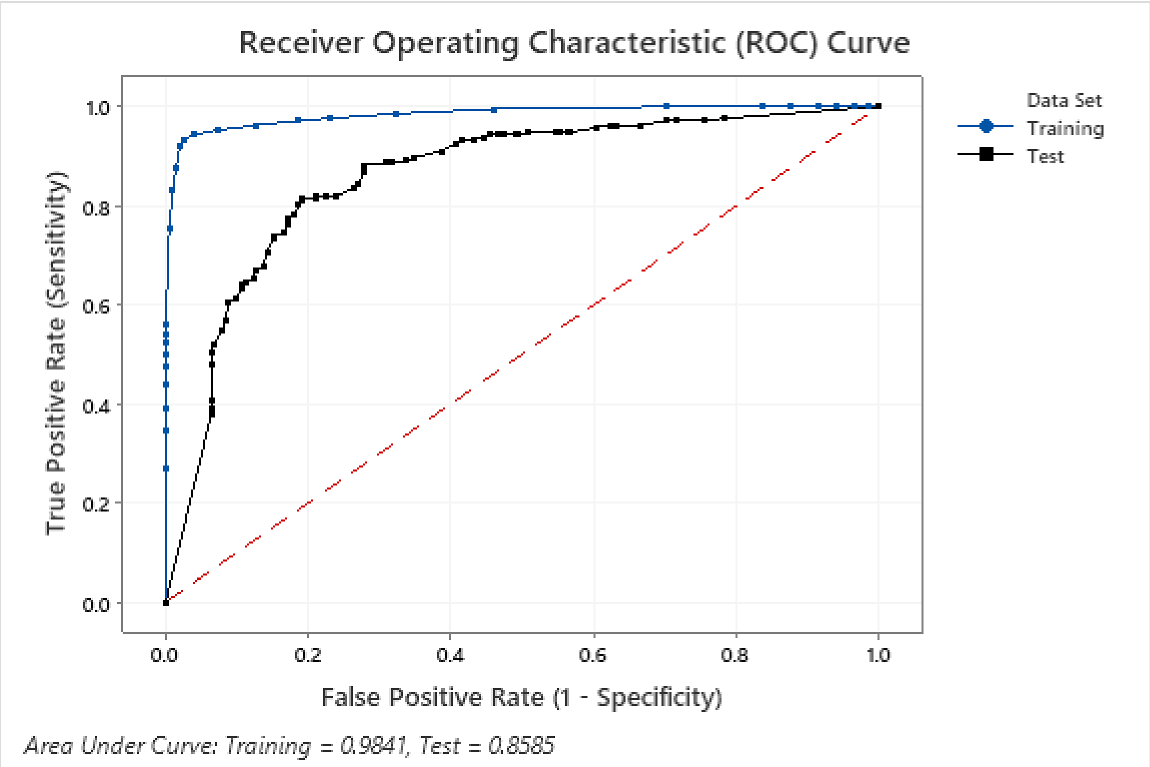

Fig 5.11 ROC curve of decision tree

A gain and lift chart is a visual way to evaluate different the effectiveness of different models. As well as helping to evaluate how good predictive model might be, it can also show visually how the response rate of a targeted group might differ from that of a randomly selected group.

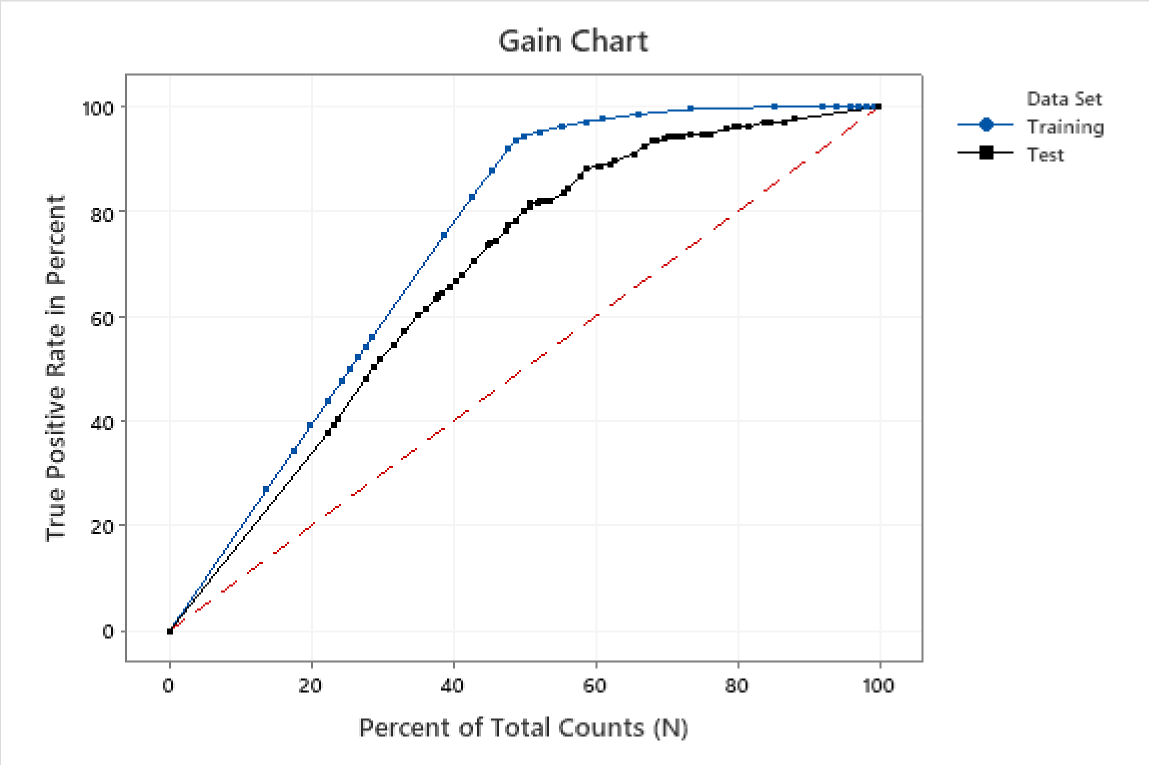

Fig 5.12 Gain chart

It shows the percentage of targets reached when considering a certain percentage of the population with the highest probability to be target according to the model. The above table shows the results obtained from the training and the test results of the model.

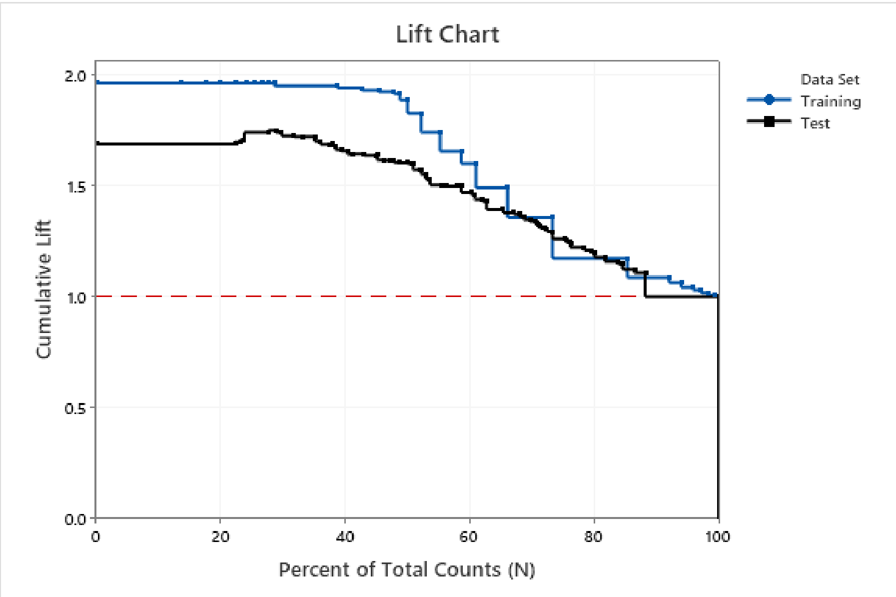

Fig 5.13 Lift chart

Lift is a measure of the effectiveness of a predictive model calculated as the ratio between the results obtained with and without the predictive model. This indicates the results obtained from the lift of training and test set.

Rules in decision tree:

Class = Sure (Event)

Terminal Node	Criterion
25	Years working in current position <= 7, Salary = {₹41,000-₹60,000, ₹81,000-₹100,000, Under ₹41,000}, COVID = {Agree, Disagree, Strongly Agree}, Job role = {Business analyst, Cloud architect , Database Admin, DevOps Engineer, IT consultant, IT support analyst, IT trainee, Quality Testing Officer, Sales and marketing, Senior Project Engineer, Software Testing Officer, Team Leader, Telemarketing, Test engineer}
4	Employees in your company(appro <= 675, You and your spouse working in = {No}, Is your Children going to school = {No}, ALTERNATIVE JOB OPPORTUNITY = {Agree, Strongly Agree}, JOB SATISFACTION = {Agree, Strongly Agree}, Job role = {Customer support, software developer, Systems analyst}

Class = Not Sure

Terminal Node	Criterion
11	Employees in your company(appro > 675, Experience <= 1.5, spouse working in same city= {No}, Children going to school = {No}, JOB STRESS = {Neutral}, Job role = {Associate, Banking, Customer support, Data Scientist, Financial analyst , General manager, human resource, Network engineer, Process executive, Programmer analyst, Project manager, quality assurance analysts, Quality control, Research and development, Researcher, Senior officer in banking process , software developer, Software Engineer, Systems analyst, Technical publication and training, Technical sales representative, Web designer}
12	Employees in your company(appro > 675, Experience > 1.5, spouse working in same city = {No}, Children going to school = {No}, JOB STRESS = {Neutral}, COVID = {Disagree, Neutral, Strongly Agree, Strongly Disagree}, Job role = {Associate, Banking, Customer support, Data Scientist, Financial analyst , General manager, human resource, Network engineer, Process executive, Programmer analyst, Project manager, quality assurance analysts, Quality control, Research and development, Researcher, Senior officer in banking process , software developer, Software Engineer, Systems analyst, Technical publication and training, Technical sales representative, Web designer}

The above-mentioned criterion are the rules that was derived from the decision tree algorithm.

5.15 Comparison of classification algorithm results:

model	accuracy	precision	recall	F1 score
Logistic regression(forward)	0.72	0.73	0.72	0.73
Logistic regression(backward)	0.72	0.73	0.72	0.73
Naïve bayes	0.6	0.6	0.71	0.55
K nearest neighbour	0.81	0.83	0.78	0.81
Kernel SVM	0.88	0.89	0.86	0.88
Random forest	0.88	0.89	0.86	8.88
Artificial neural network	0.83	0.80	0.82	0.80
Decision tree	0.80	0.81	0.81	0.81
XG Boost	**0.94**	**0.92**	**0.97**	**0.94**

From the comparison of the classification algorithms the XG Boost gives 0.94 which means the model shows overall accuracy of 94%. Precision value 92% shows the percentage of correct positive predictions to the total predicted positives. Recall value 97% shows the percentage of correct positive predictions to the total positives and f1 score is the harmonic mean of precision and recall shows the accuracy of predicting the model with low false positive and false negative rate which is to be 0.65. The area under curve is about 0.97 which shows that 97% accurate of performance of the model at distingushing between positive and negative classes. Thus XG Boost gives better results compared to other models.

5.11 TEXT MINING

Text mining methods allow us to highlight the most frequently used keywords in a paragraph of texts. One can create a word cloud, also referred as text cloud or tag cloud, which is a visual representation of text data. The text mining package (tm) and the word cloud generator package (word cloud) are available in R for helping us to analyse texts and to quickly visualize the keywords as a word cloud. The most frequent of 0 words has been detected.

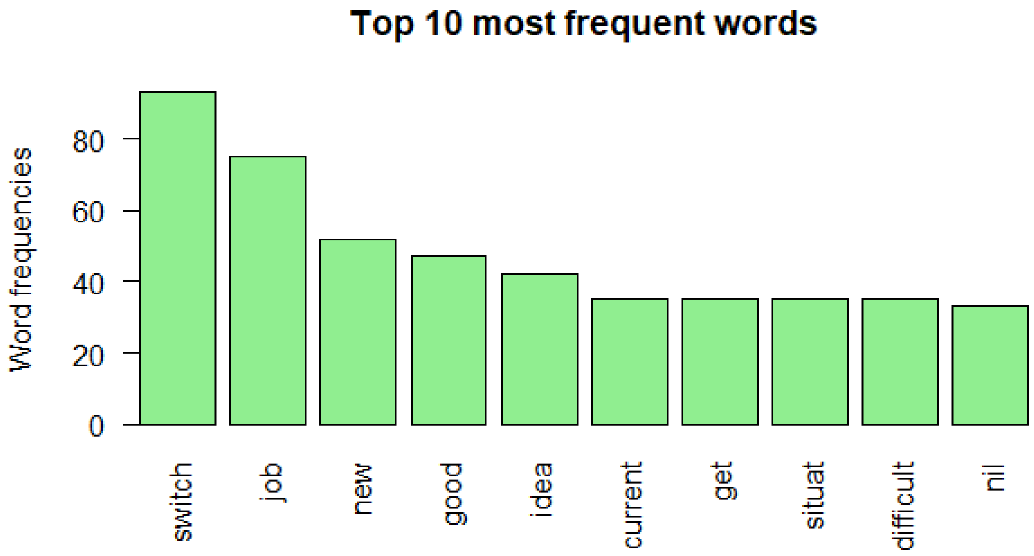

Fig 5.14 10 frequently used words

From the above picture the most frequently used words have been detected. Switch is the most frequently used word by the respondents. 80% of the people use the word switch and followed by the word switch job is frequently used for about 75%. Then the words such as new, good, idea are the words used within range of 40-50%. The word get, situation, difficult and nil are used in same frequency.

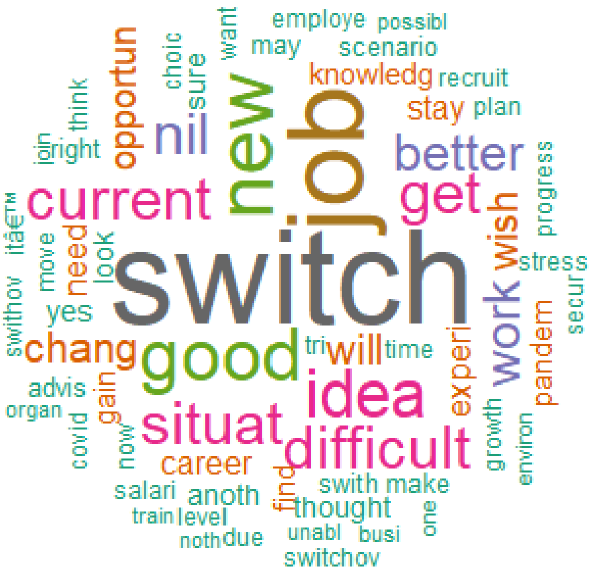

Fig 5.15: word cloud

The above image is the word cloud of the respondents. The word cloud shows additional words that occur frequently and could be of interest for further analysis. Words like “opportunity”, “situation”, “pandemic” etc. could provide more context around the most frequently occurring words and help to gain a better understanding of the main themes.

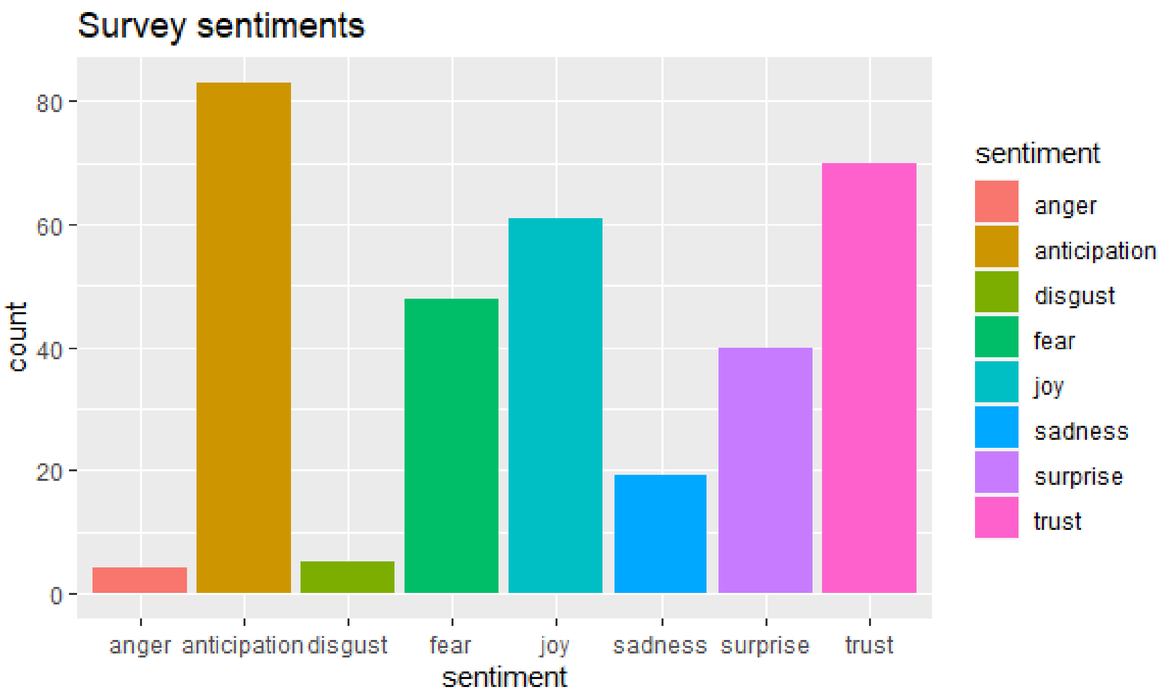

Fig 5.16 survey sentiments

The above table shows the sentimental analysis of the respondents over the intention to switch over to another company or intention to stay in the same company. From the analysis the chart says that anticipation of the employees was high about 85 employees from 417 says that they feel anticipated about the intention to switchover followed by anticipation trust plays a major role here the employees trust their company and joy was expressed that they enjoy the work they do in the present company at the same time they also have fear that they will get fired out of the company and also they if they try to switch to another company they are not sure that they will get a better job due to the pandemic.

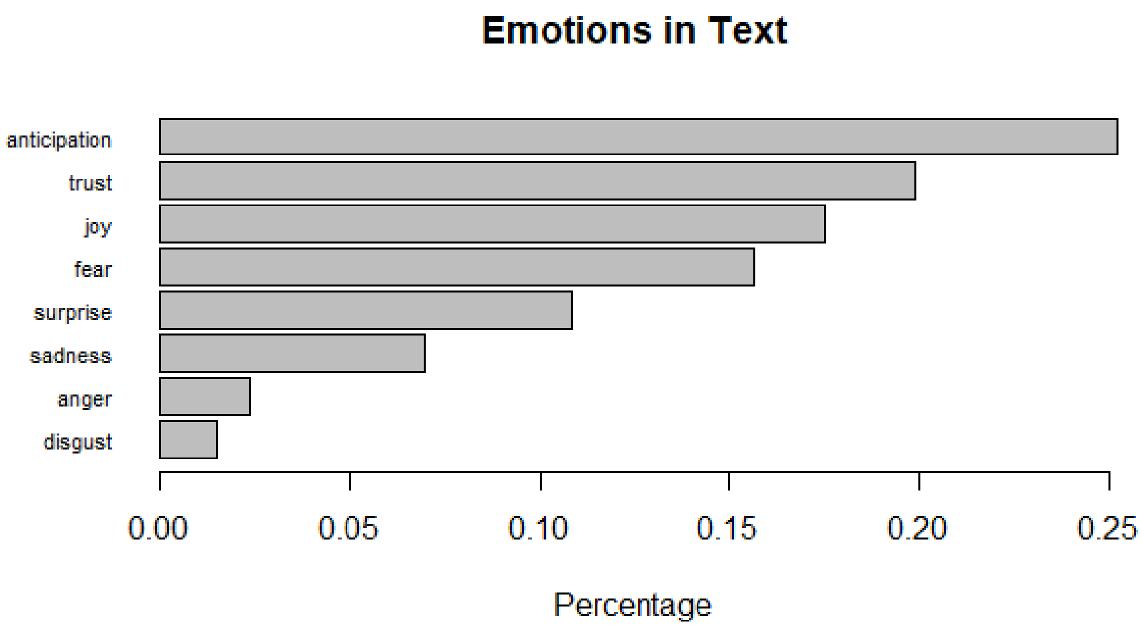

Fig 5.17 emotions in text

This bar plot allows for a quick and easy comparison of the proportion of words associated with each emotion in the text. The emotion “anticipation” has the longest bar and shows that words associated with this positive emotion constitute just over 25% of all the meaningful words in this text. On the other hand, the emotion of “disgust” has the shortest bar and shows that words associated with this negative emotion constitute less than 2% of all the meaningful words in this text. Overall, words associated with the positive emotions of “anticipation” and “trust” account for almost 45% of the meaningful words in the text, which can be interpreted as a good sign of team health

CHAPTER 6

CONCLUSION FINDINGS AND SUGGESTION

CONCLUSION

As a result of analysis logistic regression gives the important variables that affect the intention of the turnover from that alternative job opportunity has a major impact for the intention of the employee to switch to another company. In the prediction of the accuracy of the result the XG Boost gives the highest accuracy of about 0.94. where ensemble learning of boosting technique that helps to predict without overfitting, bias and variance of the data. The text mining results shows that the word “switch” was the most widely used by the respondents. “new”, “job”, “current”, “situation”, “difficult” are the most used words. From this the respondents are willing to switch to new job if they get alternative job opportunity but the current pandemic make them to retain in the same company. From the sentimental analysis the anticipation is the major emotion that was analysed. Based on their age and experience their intention to switchover tends to change and the satisfaction on the job also based on the experience.

FINDINGS

- The factors that influence the turnover intention of the employees has been predicted by using logistic regression. The factors are education, depended family members, marital status, job role, willing to relocate form work place, job satisfaction, alternative job opportunity, job stress, organizational commitment and attitude towards covid. These factors have significant influence over the turnover intention.
- The XG boost algorithm tuned out to be the best classification algorithm for predicting turnover intention of the employees. Where the ensemble technique has been adopted in order to avoid overfitting.
- The decision tree algorithm shows the relatively important variables. job role, age, years at current company, number of employees in the company and organizational commitment has a major influence. Rules have been derived in decision tree based on the positive and negative class.
- From the text mining, the sentiment that has been derived was anticipation and the most frequently used words has been identified. The words are switch, new, job, good, idea, current, situation and difficult. It gives the insights that, the employees have an idea to

switchover but the current pandemic situation make them to retain in the current company because finding better job is difficult. Form the word cloud the additional words like new, experience and knowledge has been highlighted were they wish to switch to gain new knowledge and experience.

- The word anticipation highlights the most frequently used words of the employees. The employees feel anticipated about the pandemic situation to switchover.

SUGGESTION

The factors like willing to relocate from work place, alternative job opportunity, job stress and attitude towards covid has huge impact on the intention of the employee to switchover.

The employee will be willing to relocate from the work place as they as far from their hometown or not adopted to the new place. Their intention relocate from the company can be decreased by recognition his\her thoughts and views and providing better compensation. The company should ensure the employee that they have the right talent ant the right place in the right time.

The company should show the trust over the employee by giving them responsibilities and allow them to grow. They should be respected, make them feel committed to the job and appreciated for the work they do. So, the company can retain the employee form seeking other job opportunity.

To reduce the job stress and make the employee to be retained, the company should add personal touch with the employee and should give importance to their personal feelings. During this pandemic situation the employee should be provided with flexible work timings and should give enough time to complete the work. So, the employee can be retained from switchover.

In decision tree algorithm, job role, age of the employee, number of years working in the current company and organization commitment has been identified as the important variables. based on the employee's job role, age and number of years working in the current company the commitment towards the job and their adherence to the company changes. The main issue is that more employees work in IT sectors are millennial (between 22-25) and tend to hop the job more when compared to other generations. To retain the millennials the company should have them opportunity to grow and better workplace culture that attract the millennials.

REFERENCES

1. Tzeng, H-M., Hsieh, J. G., & Lin, Y. L. (2004). Predicting nurses' intention to quit with a Support Vector Machine: A new approach to set up an early warning mechanism in human resource management. *CIN - Computers Informatics Nursing*, *22*(4), 232-242. https://doi.org/10.1097/00024665-200407000-00012

2. Campbell, J. W., Im, T., & Jeong, J. (2014). Internal Efficiency and Turnover Intention: Evidence from Local Government in South Korea. *Public Personnel Management*, *43*(2), 259–282. https://doi.org/10.1177/0091026014524540

3. Esmaieeli Sikaroudi, A., Ghousi, R., Sikaroudi, A. (2015). A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing). *Journal of Industrial and Systems Engineering*, 8(4), 106-121.

4. Jaiswal, Rakesh & Dash, Satya & Sharma, J. & Mishra, Abhishek & Kar, Suryatapa. (2015). Antecedents of Turnover Intentions of Officers in the Indian Military: A Conceptual Framework. Vikalpa. 40. 145-164. 10.1177/0256090915590335.

5. Ramachandran, P & Baranidharan, M & RAMESH, VAMANAN & Prabhakaran, S. (2011). An Interactive Mining Approach to find the job Satisfaction and Staff Turnover Intentions 1. IJCSET. March 2011.Vol 1, Issue 2,88-9

6. Khera, S. N. (2019). Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques. Vision, 23(1), 12–21. https://doi.org/10.1177/0972262918821221

7. Sudhakaran, P., & Senthilkumar, D. G. (2019). Understanding the Relationship Between Work Variables and Voluntary Turnover Intentions of Software Professionals in India. *Restaurant Business*, *118*(5), 143-148. Retrieved from https://journals.eduindex.org/index.php/rb/article/view/7999

8. Saranya R (2015), impact of perceived organisation support and organisation commitment on turnover intention of women employees in IT industry, JATIT & LLS, Journal of Theoretical and Applied Information Technology Vol.75. No.2

9. Iyer, Vidya V. (2011). Understanding turnover intentions and behavior of Indian information systems professionals: A study of organizational justice, job satisfaction and social norms. Dissertations. 389.

10. Bhatt, Ranna. (2020). Job Satisfaction Dimensions and Organizational Commitment: Tools to Understand Employee Turnover Intention of IT/ITES Industry of the Gujarat State with a Focus on BPO Segment. 10.35940/ijrte. D7133.118419.

11. Kenneth. (1999). Organizational Behavior and Human Decision Processes, science direct. Science direct. 77(1). 3-21

12. Sharma, A., & Nambudiri, R. (2015). Job-Leisure Conflict, Turnover Intention and the Role of Job Satisfaction as a Mediator: An Empirical Study of Indian IT Professionals [Dagger]. South Asian journal of management. 22. 7.

13. Mary C. Lacity & Vidya V. Iyer & Prasad S. Rudramuniyaiah. (2008.) Turnover intentions of Indian IS professionals. Information Systems Frontiers. Springer. 10(2). 225-241

14. Albalawi, A., Naugton, S., Elayan, M., & Sleimi, M. (2019). Perceived Organizational Support, Alternative Job Opportunity, Organizational Commitment, Job Satisfaction and Turnover Intention: A Moderated-mediated Model, Organizacija, 52(4), 310-324. doi: https://doi.org/10.2478/orga-2019-0019

15. Li et al (2019), Empirical Analysis of Factors Impacting Turnover Intention among Manufacturing Workers, International Journal of Business and Management, Vol. 14, No. 4. doi: https://doi.org/ 10.5539/ijbm. v14n4p1

16. Antonio Frian, Fransiska Mulyani. (2018). Millennials employee turnover intention in indonesia. Innovative Issues and Approaches in Social Sciences, vol.11, no.3:90-111, DOI: http://dx.doi.org/10.12959/issn.1855-0541.IIASS-2018-no3-art5

17. Jacobs, E., & Roodt, G. (2008). Organisational culture of hospitals to predict turnover intentions of professional nurses. Health SA Gesondheid, 13(1), 63-78. doi: https://doi.org/10.4102/hsag.v13i1.258

18. Fasanmi Samuel Sunday. (2106). Organizational Citizenship Behavior and Turnover Intent: A Path Analysis of Nigeria Bankers' Behavioural Variables. American Journal of Applied Psychology. Vol. 5, No. 6, 2016, pp. 51-59. doi: 10.11648/j.ajap.20160506.13

19. Arnoux-Nicolas C, Sovet L, Lhotellier L, Di Fabio A and Bernaud J-L. (2016) Perceived Work Conditions and Turnover Intentions: The Mediating Role of Meaning of Work. Front. Psychol. 7:704. doi: 10.3389/fpsyg.2016.00704

20. Hemdi, Mohamad Abdullah & Nasurdin, Aizzat. (2006). Predicting Turnover Intentions of Hotel Employees: The Influence of Employee Development Human Resource Management Practices and Trust in Organization. Gadjah Mada International Journal of Business. 8. 21-42. 10.22146/gamaijb.5625.

21. Hussain, S., & Xian, S.H. (2019). Factors Affecting Employees' Turnover Intention in Construction Companies in Klang, Selangor. KnE Social Sciences, 108–131-108–131.

22. Ugural, Mehmet & Giritli, Heyecan & Urbański, Mariusz. (2020). Determinants of the Turnover Intention of Construction Professionals: A Mediation Analysis. Sustainability. 12. 954. 10.3390/su12030954.

23. Wen B, Zhou X, Hu Y and Zhang X (2020) Role Stress and Turnover Intention of Front-Line Hotel Employees: The Roles of Burnout and Service Climate. Front. Psychol. 11:36. doi: 10.3389/fpsyg.2020.00036

24. Jacobs, Everd & Roodt, Gert. (2007). The development of a knowledge sharing construct to predict turnover intentions. Aslib Proceedings. 59. 229-248. 10.1108/00012530710752034.

25. Uludag, Orhan & Khan, S & Güden, N. (2011). The effects of job satisfaction, organizational commitment, organizational citizenship behavior on turnover intentions. FIU Hospitality Review. 29. 1-21.

26. Rathakrishnan, T. & Ng, Siew & Kok, T.K. (2016). Turnover intentions of lecturers in private universities in Malaysia. Pertanika Journal of Social Sciences and Humanities. 24. 129-146.

27. Kaur, B., Mohindru, & Pankaj (2013). Antecedents of Turnover Intentions: A Literature Review.3. 1219-1230

28. Singh, A., & Singh, G. (2017). Antecedents of turnover intention: testing a conceptual model in the context of professionals in India.2. 23-52

29. lvia Santoni, Muhammad Nusjirwan Harahap. (2019). The model of turnover intentions of employees.Vol 8. No 6 (2018).
DOI: https://doi.org/10.32479/irmm.7284

30. Belete UK (2018). Turnover Intention Influencing Factors of Employees an Empirical Work Review. J Entrepren Organiz Manag 2018. 7:3.

APPENDIX 1:

REVIEW OF LITERATURE

AUTHOR NAME	PROBLEM STUDIED	VARIABLES USED	SAMPLE SIZE	OBJECTIVES	FINDINGS	ALGORITHMS USED
Huey-Ming Tzeng,Rn jer-Guang hsieh,Yih-lon Lin	Predicting Nurses' Intention to Quit With a Support Vector Machine	Intention to quit, using working motivation, job satisfaction, and stress levels	389	To train a learning machine (SVM) for predicting nurses' intention to quit by using the values of a group of predefined predictors	Predictions with 89.2% accuracy. This Support Vector Machine can predict nurses' intention to quit, without asking these nurses whether they have an intention to quit.	Support vector machine
Jesse W. Campbell, Tobin Im, and Jisu Jeong	Internal efficiency and turnover intention: evidence from local government from South Korea	Efficiency emphasis (EE), Public Service Motivation (PSM), Procedural justice, 5. Innovation climate, Merit-based promotions	1407	empirically investigated turnover intentions and behaviors of Indian IS professionals using a theoretical framework that was most relevant in the Indian context.	the relationship of job satisfaction and organizational justice with turnover intentions was strongly supported	apriori

Amir Mohammad Esmaieeli Sikaroudi, Rouzbeh Ghousi, Ali Esmaieeli Sikaroudi	A data mining approach to employee turnover prediction (case study: Arak automotive parts manufacturing)	previous job changes, knowledge about the working conditions, perseverance and interest to work, compatibility of body with job		prediction of quit among staff with an empirical research on a manufacturing company staff.	By consideration of accuracy, time and user friendliness, decision trees generally had the best performance.	Multilayer perceptron, Probabilistic neural network, support vector machine, Classification and regression tree, k nearest, apriori, Naive Bayes
Tanya Attri	Why an Employee Leaves: Predicting using Data Mining Techniques	JobLevel, OverTime, Stock Option Level, Environment Satisfaction, Age, Monthly Income, Job Level	1471	To find the factors affect most for the turnover intention of the employees	work pressure, job security, job previews, which they mentioned as the leading factors for employee attrition	Random Forest, Logistic Regression, Support Vector Machine, Gradient Boosting Machine
P.ramachandran, m.baranidharan, v.ramesh, s.prabhakaran	An interactive mining approach to find the job satisfaction and staff turnover intentions	Job satisfaction, job stress		to find the job Satisfaction and Staff Turnover Intentions	looking deeply into individual’s con’s can help the IT firms to maintain steady growth and development.	The software used ASP.NET MYSQL

Shikha N. Khera, Divya	Predictive Modelling of Employee Turnover in Indian IT Industry Using Machine Learning Techniques	Age, gender, marital status, job level, job profile, job role, travelling	1650	to develop a prediction model based on the employee data in order to tackle the problem of employee turnover of the Indian IT sector.	The accuracy of the model was found to be 0.85 (or 85%), which is a significantly accuracy level. This indicates that the trained SVM model classifies	Support vector machine
Priyada Sudhakaran , Dr. G Senthilkumar	Understanding the relationship between work variables and voluntary turnover intentions of software professionals in India	Total experience, current experience, number of jobs changed	949	To analyse the relationship between the work variables and the voluntary turnover intentions of software professionals in India.	The success of retention plans devised by the Human Resources practitioners.	one-way ANOVA
Dr Saranya R , Dr Muthumani S	Impact of perceived organisation support and organisation commitment on turnover intention of women employees in it industry	Perceived organizational support, organizational commitment	598	to examine if a relationship exists between Perceived organization support and	there exists powerful relationship between perceived organization support and organization commitment which	Regression

				employee turnover for women employees	decreases the turnover intentions.	
Vidya v. Iyer	Understanding turnover intentions and behaviour of Indian information systems professionals: a study of organizational justice, job satisfaction and social norms	Job satisfaction, organizational justice, organizational alternatives, social norms	75	To find the strongest variable that affect turnover intention.	the relationship of job satisfaction and organizational justice with turnover intentions was strongly supported	Discriminant validity
Dalke purva	Workers turnover intention in it sector in Indore forecasting and planning	Employee recognition, external career opportunity, job hopping		To determine the motives underlying turnover intentions	Employee recognition, external career opportunity, job hopping has a strong influence over the turnover intention	Dummy variable regression
J m a jayasundera j a s, k ayakody a k l, Jayawardana a	Structural Equation Modeling the role of leader-member exchange	Perceived organizational support, organizational commitment, job satisfaction, leader member exchange	238	to identify the effect of POS on TI among Gen Y employees while also examining the impact of LMX on the relationship between POS and TI.	the relationship between POS and TI is mediated by JS and OC. Hence, it was verified that JS and OC, which can be considered as outcomes of POS, also contribute	Structural Equation Modeling

					in reducing turnover intention.	
Anupama sharma and ranjeet nambudiri	Job-Leisure Conflict, Turnover Intention and the Role of Job Satisfaction as a Mediator: An Empirical Study of Indian IT Professionals	Job-leisure conflict, job satisfaction	173	Finding out all the factors which contribute to the turnover intention among employees is employers' high priority	a significant positive relationship between job-leisure conflict and turnover intention and significant negative relationship between job satisfaction and turnover intention.	reliability analysis, Baron and Kenny Mediation Test, correlation
Mary c. Lacity, vid ya v. Iyer & pras ad s. Rudramuni yaiah	Turnover intentions of Indian is professionals	Job satisfaction, organization commitment		To identify that the Job Satisfaction affects Turnover Intentions among Ind ian IS professionals.	Job Satisfaction, Organizational Satisfaction, and Social Norms as the main determinants of Turnover Intentions among Indian IS Professionals.	
Abdulmaje ed saad albalawi, shahnaz naugton,	Perceived organizational support, alternative job opportunity, organizational commitment, job satisfaction and turnover	Perceived organizational support, organizational commitment, job satisfaction, perceives	270	examines the mediating role of organizational commitment on the link between	organizational commitment mediates the association between perceived organizational support and turnover	contemporary variance-based structural equation modelling

malek bakheet elayan, mohammad tahseen sleimi	intention: a moderated-mediated model	alternative job opportunity		perceived organizational support, perceived alternative job opportunities, and turnover intention, and the moderating role of job satisfaction on the proposed relationships.	intention, perceived alternative job opportunities and turnover intention	
Yuting li, rapinder sawhney,guilherme luz tortorella	Empirical analysis of factors impacting turnover intention among manufacturing workers	Job satisfaction, job performance, organizational commitment, leadership, work-family conflict	138	to identify the main predictors of manufacturing workers' turnover intention and explore the relationship between turnover intention and these predictors, such as job satisfaction, organizational commitment,	the turnover intention of manufacturing workers was significantly associated with job satisfaction, organizational commitment, and work-family conflict.	SPSS and structural equation modelling

				leadership, job performance, and work-family conflict.		
Antonio frian, fransiska mulyani	Millennials employee turnover intention in Indonesia	Salary and compensation, perceived alternative job employment, employee development system, and employee involvement	200	to find out the other factors that affect millennial employee turnover and able to help companies to face millennial generation	salary and compensation and employee involvement have no significant influence on turnover intention in millennial generation.	multiple regression analysis
Dr e jacobs, prof g roodt	Organisational culture of hospitals to predict turnover intentions of professional nurses	Knowledge sharing, organisational commitment, organisational citizenship behaviour, job satisfaction	530	To determine the relationship between organisational culture and turnover intentions on a bivariate level.	organisational culture, in interaction with the selected variables of knowledge sharing, job satisfaction and OCB's, as well as organisational commitment as independent predictor and various demographic variables, interactively predict turnover intentions.	general linear modelling

Fasanmi samuel sunday	Organizational citizenship behaviour and turnover intent: a path analysis of Nigeria bankers' behavioural variables	Affective commitment, procedural justice, psychological empowerment, turnover intent, organizational citizenship behaviour	885	the influence of affective commitment, procedural justice and psychological empowerment on the negative relationship between citizenship behaviour and turnover intent among survivors of a consolidated bank in Nigeria.	affective commitment, procedural justice and psychological empowerment have direct effects on the negative relationship between citizenship behaviour and turnover intent.	multivariate multiple regression analysis
Caroline arnoux-nicolas, laurent sovet, lin lhotellier, annamaria di fabio, jean-luc bernaud	Perceived work conditions and turnover intentions the mediating role of meaning of work	Work pressure, lack of resources, job insecurity, organizational changes, lack of personal development, personal reasons, work climate, public image of the company	336	to examine how the relationships between working conditions and turnover intentions are mediated by meaning of work among a sample of French workers	job characteristics influence critical psychological states, with in return have significant and various impacts on employee's work outcomes	multiple regression

Mohamad abdullah hemdi aizzat mohd. Nasurdin	Predicting turnover intentions of hotel employees: the influence of employee development human resource management practices and trust in organization	performance appraisal, training and development, career advancement, trust in organization	380	to investigate the influence of human development in HRM practices on trust in organization and on turnover intentions, and to examine whether trust in organization serves to mediate the relationship between perceptions of HRM practices and turnover intentions.	when employees perceive that their organizations show greater concern for their personal growth and welfare via the provision of adequate and continuous training and development, fair and formal performance appraisal and feedback system, and adequate career advancement opportunities, they will experience a positive emotional state	multiple regression, Principal component factor analyses
Suhaidah hussain and see huei xian	Factors affecting employees' turnover intention in construction companies in Klang, Selangor	Colleague relations, organizational commitment, organizational justice, organizational reputation,	160	to find out the factors that may influence the employees' turnover intention in construction companies through identify the factors	all the independent and dependent variables in this research area acceptable with high internal consistency and reliability.	Multiple Regression Analysis

		communication, organizational politics		and determine the relationship between the factors with the employees' turnover intention.		
Mehmet nurettin ugural, heyecan giritli and mariusz urbanski	Determinants of the turnover intention of construction professionals: a mediation analysis	organizational identification, perceived external prestige, and turnover intention	525	to determine the key factors that contribute to the voluntary turnover intentions of qualified construction professionals	individual differences in self-orientation may be related to turnover intention indirectly through perceptions of organizational prestige	Confirmatory Factor Analysis (CFA), mediation analysis
Biyan wen, xiaoman zhou, yaou hu and xiao zhang	Role stress and turnover intention of front-line hotel employees: the roles of burnout and service climate	Role stress, burnout, organizational service climate,	454	the moderating effect of service climate on the underlying mechanism that links role stress with turnover intention	hotels can improve the organizational service climate through communication, incentive schemes, service quality and corporate responsibility plans	exploratory factor analysis

Everd jacobs, gert roodt	The development of a knowledge sharing construct to predict turnover intentions	Organisational culture, organisational commitment, organisational citizenship behaviour and job satisfaction	530	to discuss the development of a knowledge sharing questionnaire and the role of knowledge sharing in predicting turnover intentions of registered professional nurses.	A significant negative relationship was found between knowledge sharing behaviour and turnover intentions. Furthermore, knowledge sharing interacted with organisational culture in a final model where all the selected mediating and demographic variables were simultaneously entered into the equation to predict turnover intentions.	general linear modelling
Orhan ulndag, sonia khan, nafiya guden	The effects of job satisfaction, organizational commitment, organizational citizenship behaviour on turnover intentions	job satisfaction, organizational commitment, organizational citizenship behaviour	116	the effects of job satisfaction and affective organizational commitment on organizational citizenship behaviour	, the effect of organizational citizenship behaviour and satisfaction on turnover intentions was found to be significant	Multiple regression

				and turnover intentions.		
Thanuja rathakrishnan, ng siew imm and tee keng kok	Turnover intentions of lecturers in private universities in Malaysia	Job security, supervisor support, compensation satisfaction, job autonomy, KPI achievability, job satisfaction	253	To examine the factors determining the turnover intention of lecturers in private universities in Malaysia	compensation satisfaction, job autonomy, KPI achievability, and job satisfaction explained turnover intention.	Multiple regression
Bandhanpreet kaur , mohindru and dr. Pankaj	Antecedents of turnover intentions: a literature review	Job stress, job satisfaction, quality of work life, organizational justice		To measure the extent to which the old employees leave and new employees enter the organization in a given period.	quality of work life, job stress, job satisfaction and organizational justice have an impact on the turnover intentions	
Archana singh, feza tabassum azmi, ganesh singh	Antecedents of turnover intention: testing a conceptual model in the context of professionals in india	Job person fit, job stability, perceived organization support, psychological contract, pay satisfaction, job satisfaction, organization	303	To identify Antecedents of turnover intention	satisfaction with pay, and promotion has a stronger relationship with Organization Commitment and Job Embeddedness	Discriminant validity, factor analysis

		commitment and job embeddedness.			compared to Job Satisfaction	
Alvia santoni, muhammad nusjirwan harahap	The model of turnover intentions of employees	Leadership, work environment, compensation, job satisfaction	260	to find out the influence of leadership, work environment, compensation, partial evaluation and jointly against job satisfaction and know the influence of leadership, work environment, compensation and job satisfaction as partial and jointly over turnover intentions of employees in the plastic industry of household appliances in the special capital region of Jakarta.	the turnover intention employees, especially the intention to move but for fear not getting better job can be lowered if employees feel satisfied with the work itself that is reinforced by work environment extern/internal factors that in the form of selfish abstinence.	SEM

Belete ak	Turnover intention influencing factors of employees: an empirical work review	Job satisfaction, job stress, organizational culture, organizational commitment, salary, organizational justice, promotional opportunity, demographic variables, leadership styles, and organizational climate		Factors influencing turnover intention among technical employees	The result shows that job commitment is an important variable toward job satisfaction that leads to turnover intention	

APPENDIX 2

Code snippet(python):

Importing the libraries:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

importing the data set:

```
dataset = pd.read_csv('C:\\Users\\HP\\Desktop\\DATA.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

import stratified Kfold for training the data:

```
from sklearn.model_selection import StratifiedKFold
skf = StratifiedKFold(n_splits=10)
skf.get_n_splits(x, y)
print(skf)
for train_index, test_index in skf.split(x, y):
   print("TRAIN:", train_index, "TEST:", test_index)
   X_train, X_test = x[train_index], x[test_index]
```

```
    y_train, y_test = y[train_index], y[test_index]
```

Feature scaling:

```
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

Training the model on the training set

```
from xgboost import XGBClassifier
classifier = XGBClassifier()
classifier.fit(X_train, y_train)
```

confusion matrix:

```
from sklearn.metrics import confusion_matrix
def tp(y_train, y_pred): return confusion_matrix(y_train, y_pred)[0, 0]
def tn(y_train, y_pred): return confusion_matrix(y_train, y_pred)[1, 1]
def fp(y_train, y_pred): return confusion_matrix(y_train, y_pred)[1, 0]
def fn(y_train, y_pred): return confusion_matrix(y_train, y_pred)[0, 1]
scoring = {'tp' : make_scorer(tp), 'tn' : make_scorer(tn),
           'fp' : make_scorer(fp), 'fn' : make_scorer(fn)}
cv_results = cross_validate(classifier.fit(X_train, y_train), X_train, y_train, scoring=scoring)
```

Evaluate the score by cross validation

```
from sklearn.model_selection import cross_validate
cross_validate(estimator=classifier, X=X_train, y=y_train, scoring=scoring)
```

ROC curve

```
import matplotlib.patches as patches
from sklearn.metrics import roc_curve,auc
from numpy import interp
tprs = []
aucs = []
mean_fpr = np.linspace(0,1,100)
i = 1
for train,test in skf.split(x,y):
    prediction = classifier.fit(x[train],y[train]).predict_proba(x[test])
    fpr, tpr, t = roc_curve(y[test], prediction[:, 1])
    tprs.append(interp(mean_fpr, fpr, tpr))
    roc_auc = auc(fpr, tpr)
    aucs.append(roc_auc)
    plt.plot(fpr, tpr, lw=2, alpha=0.3, label='ROC fold %d (AUC = %0.2f)' % (i, roc_auc))
    i= i+1
```

```
plt.plot([0,1],[0,1],linestyle = '--',lw = 2,color = 'black')
mean_tpr = np.mean(tprs, axis=0)
mean_auc = auc(mean_fpr, mean_tpr)
plt.plot(mean_fpr, mean_tpr, color='blue',
        label=r'Mean ROC (AUC = %0.2f )' % (mean_auc),lw=2, alpha=1)

plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('ROC')
plt.legend(bbox_to_anchor=(1,1), loc="upper left")
plt.show()
```