



Predicting the impact of post-COVID distance working on IT employees through a Neural Network

A dissertation submitted in partial fulfilment of the requirements for the degree of
Master of Science (MSc)

The qualification for which the dissertation is a part
(Master's degree in data science)

Author : Vyshnavi Muthumula
Name of School : Cardiff School of Technology
Name of university : Cardiff Metropolitan University
Supervisor : Abdelrahman Abuarqoub

Declaration

I hereby declare that this dissertation entitled “**Predicting the impact of post-COVID distance working on IT employees through a Neural Network**” is entirely my own work and has not been submitted nor is being submitted for any other degree.

Student: Vyshnavi Muthumula

Signature: Vyshnavi Muthumula

Date: 30-03-2023

Supervisor: Abdelrahman Abuarqoub

Signature:

Date:

Abstract

All the world is suddenly faced many problems because of COVID-19 pandemic, and this changed entire world activities. Government made decisions which no one ever thought that would happen. All the borders are shut down and everyone stays at home because of lock down. Only in the emergency, people are moving around otherwise everyone has to stay at home. Important things like hospitals stay open. This research focuses on Industrial technology employee emotions about WFH which played important role during this pandemic. The growth of the organization is depending on the productivity of the employees and the productivity of the employees depends on their feelings. If employee is happy with work environment and workload, they can make the maximum productivity which leads best outcomes for the organization. To predict the impact of work from home on employees due to pandemic is discussed in this paper. The models used for this research is Exploratory data analysis, Machine learning algorithms and Deep learning techniques to find the future predictions. Exploratory data analysis is used for better visualizations and to give statistical analysis. Under machine learning techniques the performed models for this multi classification is Support vector machine, Logistic regression, KNN, Decision tree and Neural network. Neural networks have the highest accuracy rates (71.6%), whereas logistic regression has the lowest accuracy rates (43%). Age 35-44 people are more positively responds to WFH activities than other age group people and IT professionals working actively during COVID-19 which results high performance.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor, Abdelrahman Abuarqoub, for his invaluable guidance and support throughout this research project.

I am grateful to Cardiff Metropolitan University for providing me with the opportunity to pursue a master's degree. It has been an incredible journey.

I would also like to express my sincere appreciation to God, my family, friends, and the university staff who have been a great source of motivation and support during this challenging period.

Table of Contents

Declaration	2
Abstract	3
Acknowledgements	4
List of Figures	7
List of Abbreviations	8
Chapter 1	9
1.1 Research Title	9
1.2 Introduction	9
1.3 Research Rationale	10
1.4 Research Contribution	10
1.5 Research Aim	11
1.6 Research Objectives	11
1.7 Research Questions	11
1.8 Dissertation Outline	12
1.9 Summary	12
Chapter 2	13
Literature review	13
2.1 Importance of Employee satisfaction in an organization	13
2.2 Remote Work and Mental health	14
2.3 Work–Life Balance, Work Productivity and Work Satisfaction	15
2.4 Gender differences	16
2.5 Stress levels for employees due to WFH	17
2.6 What is meant by machine learning and how it could be beneficial towards identifying employment satisfaction?	18
2.7 Literature Gap	22
Chapter 3	22
Research Methodology	22
3.1 Philosophy	24
3.2 Approach	25
3.3 Strategy	27

3.4 Methods	27
3.5 Time Horizon.....	27
3.6 Techniques and Procedures Sampling, Data collection and Tools	28
3.7 Ethical Principles.....	29
3.8 Performance Evaluation	29
3.9 Research Limitations.....	29
Chapter 4.....	30
Data Analysis and Results	30
4.1 Introduction	30
4.2 Rationale Behind Choosing Models.....	31
4.2.1 Support Vector Algorithm	31
4.2.2 Logistic Regression	32
4.2.3 Decision Tree.....	32
4.2.4 K-NN Algorithm	33
4.2.5 Neural Networks	33
4.3 Implementation.....	33
4.3.1 Importing the libraries.....	34
4.3.2 Data processing	34
4.3.3 Methods Effectiveness	34
4.4 Challenges and Lessons Learnt	51
Chapter 5	52
Conclusion and Recommendations	52
5.1 Conclusion.....	52
5.2 Limitations	53
5.3 Recommendations and Future work.....	53
References.....	54
Appendix.....	60

List of Figures

Figure 1: Characteristics of work from home	14
Figure 2: Stress levels	18
Figure 3: Machine learning model structure	19
Figure 4: Neural network structure	20
Figure 5: Onion Research diagram	24
Figure 6: Research methodology	26
Figure 7: Age count.....	36
Figure 8: Age analysis.....	37
Figure 9: Designation analysis.....	38
Figure 10: Overall Satisfaction report.....	39
Figure 11: Resource usage	41
Figure 12: Preference of work by employees before COVID-19	42
Figure 13: Issues faced by employees by cyber threads	43
Figure 14: Security measures.....	44
Figure 15: Assets.....	45
Figure 16: Future preference about remote work by employee	46
Figure 17: Awareness created by employer	47
Figure 18: Best Model	47
Figure 19: Neural network output.....	48
Figure 20: Confusion matrix for Decision tree	49
Figure 21: Confusion matrix for Neural network	50
Figure 22: Classification report for Neural network.....	50
Figure 23: Scrapped data	60
Figure 24: Cleaned data	61
Figure 25: Word count	62

List of Abbreviations

Abbreviation	Definition
IT	Information Technology
WFH	Work From Home
AI	Artificial Intelligence
ML	Machine Learning
KNN	K-Nearest Neighbour
ANN	Artificial Neural Networks
SVM	Support Vector Machine
LR	Logistic Regression
PLS-SEM	Partial least squares structural equation modelling
RBF	Exploratory data analysis
CMIE	Centre for Monitoring Indian Economy

Chapter 1

1.1 Research Title

Predicting the impact of post-COVID distance working on IT employees through a Neural Network

1.2 Introduction

Due to the COVID-19 virus, the year 2020 will likely rank among the most severe year in recent history. Significant losses, including human deaths, economic losses and societal unrest, were brought on by this pandemic. The COVID-19 virus has had an effect on 213 nations and territories worldwide. The government imposed a lockdown where no one was allowed to leave their house in order to stop the spread of COVID-19. People are losing their jobs globally as a result of the closure of several organizations. The Centre for Monitoring Indian Economy (CMIE) estimates that 27 million employees between the ages of 20 and 30 lost their employment as a result of this closure (Sharma, 2020). These individuals are new to the market and should receive training before beginning the project; however, virtual trainings have little impact. This work from home (WFH) was more widely available because of the COVID-19 pandemic restrictions (Bao et al., 2022). As a result, WFH is frequently used by software companies; however, this WFH has a number of advantages and disadvantages that the author will examine in this study. This new technique (WFH) enhanced productivity with 13% more than working from office during COVID-19 (O'Hara, 2020), and this transformation will have a significant impact in the future, according to Murugesh, the chief executive of business process management (Financial Express, 2020). The idea of WFH is not new because several organizations both in India and throughout the world have embraced it.

The workspace environment was entirely changed due to many regulations brought by the COVID-19 pandemic. Because of this, businesses that allowed remote employees to continue operating were able to succeed at some level while those that required personnel presence, such as hospitals and mechanical, electronic, or electrical enterprises, failed. Next, each company carried out multiple staff interviews and surveys to gather feedback and ascertain the impact of WFH. Nowadays it's difficult to evaluate work efficiency, well-being, and employee participation in

activities (Russo et al., 2021). Twitter recently stated that it plans to have its employees join businesses permanently working remotely like Facebook, Microsoft, Amazon, and Google in extending its WFH policy (Ghosh, 2020). As a consequence, 90% of individuals working remotely, with 65% of them hailing from cities and the other 35% working from smaller villages (Mitta, 2020).

While grey-hat and black-hat hackers considered the pandemic as a greater opportunity to compromise the networks, 81% of organisations worldwide experienced cyber-attacks. Due to cyber security threats during the busiest season, 79% of firms experienced downtime (BusinessWire, 2021). Researchers conclude that working remotely is not always challenging for companies or employees. The purpose of this research is to predict the key characteristics of post-COVID remote working IT personnel using a Neural Network with the following factors: employee work-life balance, productivity, cyber issues, and employee satisfaction. Machine learning models are the most used advanced technology to detect the future predictions. The impact of COVID-19 on gender segregation has been the subject of some research articles; however, the impact of WFH on employee age classification and designation has not been the subject of any publications. The purpose of this study is to achieve these goals by using visualisations and ML models to determine employee satisfaction with WFH.

1.3 Research Rationale

The unanticipated changes brought by the outbreak of the Corona virus, the world's financial system is almost completely helpless. Scientists should be ready to decrease harmful viral impacts on society even if they are unable to stop the generation of deadly viruses (Donthu and Gustafsson, 2020). Thus, studying the effects of COVID-19 on software firms to determine employee happiness is increasingly crucial to maintaining employee work-life balance and the expansion of the business.

1.4 Research Contribution

This study contributes to the development of a machine learning model for NN that predicts the key characteristics of post-COVID remote IT employees, allowing firms to forecast the outcomes and sentiments of such workers. According to such findings, organizations can forecast what will

be needed in the eventuality of another epidemic. It uses a variety of techniques, including machine learning (ML) algorithms for regression, classification, and image processing, to produce efficient and reliable results (Kushwaha et al., 2020). Due to the outbreak, organizations learn the crucial role that remote working has to play and realize that this new function will be the only one in the future that will enable organizations to function in a regular manner (Gartner, 2020). In order to develop society and respect employee preferences, it is important to know the features and procedures that enable the workplace to change from a cruel one into a pleasant one (Michaelson et al., 2014). The purpose of analyzing employees based on their designation is important because it will help to find out who want WFH and who does not. These findings enable businesses to structurally separate the role responsibilities.

1.5 Research Aim

The aim of this research is to develop an advanced machine learning model that will assist in determining employee satisfaction as well as organizational scope changes, work-life balance, perspectives of employees based on their age and designation through visualizations.

1.6 Research Objectives

The objective of the research is given below.

1. Evaluating how the pandemic has affected business priorities and scope changes.
2. Analysing the impact of work from home (Remote working) in IT organizations using machine learning models and neural networks.
3. Classifying the workforce according to job titles and keeping track of which departments permit work from home and which do not.
4. Exploring the Age based employee work-life balance through visualizations.

1.7 Research Questions

- How business agendas changed in IT organizations during COVID-19?
- What are the scope changes in organizations after the pandemic?

- What is the importance of machine learning models and how it will be helpful for this study?
- Does WFH increased productivity?
- What factors affect the cyber issues?
- How to increase the efficiency of employees during WFH?
- Is employee's work depending on geographic locations?

1.8 Dissertation Outline

The dissertation is organised as follows:

Chapter 1: The introduction discusses the context of the research project and the rationale behind the topic selection.

Chapter 2: This chapter is about literature review, and it is a demonstration on current research/theories that are already available about the impact of COVID-19 on employees.

Chapter 3: All the research methodology such as research approach, research design, strategies, time horizon, ethical principles and performance evaluation details are explained in this chapter.

Chapter 4: This chapter explains the results with detail explanation. Data collection, Data preprocessing, data cleaning, exploratory data analysis, applying machine learning models, finding best model and concluding the impact of COVID-19 on IT employees are discussed.

Chapter 5: Summarized the findings and recommendations from the research objectives. It has been discussed how machine learning can be used to improve employee satisfaction in the future.

1.9 Summary

This chapter provides an overview of the research before going into greater detail about the importance of machine learning algorithms to predict the impact of post-COVID on distance working IT employees. This can be done through a machine learning models along with Neural Network by considering these parameters such as employee work life balance, productivity of the

project, employee satisfaction and client satisfaction. The introduction provides detailed information on how COVID-19 altered global ecosystem and what new technology was adopted into organizations. It also illustrates how employee satisfaction affects the productivity and why it is important to analyse. The chapter came up with couple of research objectives and research questions.

Chapter 2

This chapter discusses literature reviews and provides examples of recent studies and hypotheses that have been published about the effects of COVID-19 on employees.

Literature review

In order to do a literature review, one must look at previous research in their subject and use that as the basis for their own data collection. This section is entirely about how other research papers are done and what are the new changes added to this research.

2.1 Importance of Employee satisfaction in an organization

Employee happiness is a key factor for determining organisational effectiveness, particularly in the service industry. Employee satisfaction increases long-term team performance and maintains valuable clients, as it is crucial for better business operations. Using the Partial least squares structural equation modelling (PLS-SEM) technique, it was discovered that employee happiness and customer or client satisfaction were causally related and that identifying the employees' role in this relationship was significant (Kurdi, Alshurideh and Alnaser, 2020). Any firm must prioritize employee satisfaction if it wants to succeed. Lower production rates are directly correlated with high levels of employee satisfaction.

Several factors, such as excessive stress levels, a lack of communication within the organization, a lack of recognition, or few opportunities for professional advancement, might cause employees to lose interest in their work and resign (Sullivan, 2012). In situations like this, the employee would look for another job that provides the necessary financial security and job requirements that challenge them suitably, increasing the turnover rate at the initial organization

(Koslowsky&Krausz, 2002). Using a qualitative comparative analysis with fuzzy sets Jun Yu and Yihong (2021) investigated the five job characteristics that affected how happy employees were to do WFH. These include time, a suitable work environment, job autonomy, virtual networking support, and monitoring tools.

The below picture illustrates more about the characteristics of WFH.

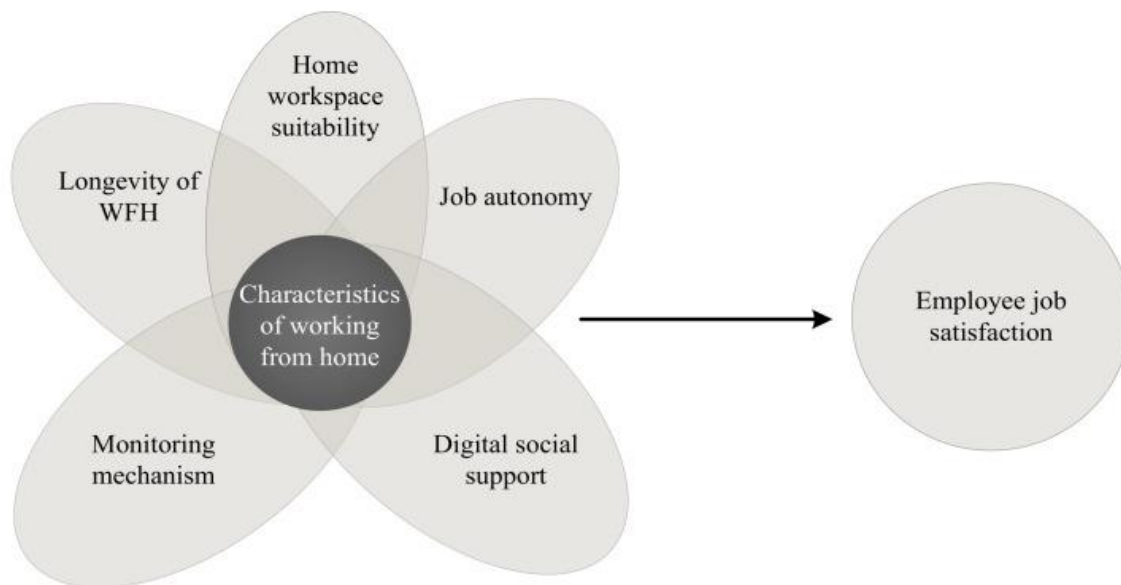


Figure 1: Characteristics of work from home

(Source: Yu and Wu, 2021)

2.2 Remote Work and Mental health

Before the pandemic, a lot of businesses offered remote work chances, but they were only available for a short time, not for very long, thus organizations lacked a comprehensive analysis of WFH. Also, numerous organizations stated that job autonomy can help to minimize stress (Song and Gao, 2020). Another researcher claimed that employing tele work would raise stress levels. As a result, employees could find it challenging to maintain a positive work-life balance (Sullivan, 2012).

Nonetheless, there are various ways in which WFH during the crisis is different from prior distant work. Such as

- i) It is not voluntary for both Employer and Employee.

- ii) The organizations don't have any chance to consider whether the employee is able to work remotely or not.
- iii) Irrespective of the employee skills the employer brought this system to everyone and if the employee is not skilled then learning team will train them remotely.

Stress is a major component that affects how well employees balance their professional and personal lives. The following less evident behaviours may indicate a worker's disengagement from the workplace: doing personal business while at work, playing games, conversing about things unrelated to work, spending time on social media, and lowering performance standards. When exhibited as signs of unhappiness, these withdrawal behaviours could result in an employee quitting their job; "the bayesian theory suggests that thinking about quitting is the most likely consequence of job discontent" (Koslowsky&Krausz, 2002). Not just adults, but also younger generations of people, have experienced enormous mental issues as a result of the continuing issues and after consequences of this outbreak (Yasmin, 2020).

A key benefit of these procedures has been the mathematical and statistical results, which have allowed for the most accurate detection of stress levels over extended periods of time using a predetermined criterion established at the beginning of training (Chaudhary, Gaur and Chakrabarti, 2022). Businesses these days are constantly developing creative and effective strategies to maintain employee motivation in these challenging circumstances. Because of the employee engagement activities they offer, several companies are developing professionally and learning new skills. Employees are still enthusiastic and feel a sense of connection to the organisation despite the difficult COVID-19 pandemic (Chanana and Sangeeta, 2020).

2.3 Work–Life Balance, Work Productivity and Work Satisfaction

Organizations won't keep track of an employee's location or method of work while they are working remotely. They simply take the task's outcome into account. This raises the possibility of work-life balance (Chung and van der Horst, 2017). Employees that have flexibility in their schedules boost the likelihood that they will put in more hours at the office. There are some people who are so enthusiastic about work and never stop doing, which has a negative impact on the balance between work and life (Bjärntoft et al., 2020). In conclusion, employees who work from

home will experience more stress since additional family issues will disrupt their ability to focus on their work (Sandoval-Reyes, Idrovo-Carlier and Duque-Oliva, 2021). There are some reports saying the productivity levels are increased unexpectedly because of remote work. The reason for this is employee can work in their productive time (Gajendran, Harrison and Delaney-Klinger, 2014).

The distractions from colleagues are less (Martin and MacDonnell, 2012) which results increase in productivity (Nakrošienė, Bučiūnienė and Goštautaitė, 2019). Even with beneficial outcomes, there are some drawbacks as well, such as the potential decline in productivity associated with insecure WIFI connections and unsafe work environments (Gorlick, 2020). Employees feel less engaged and inspired as a result of WFH's reduction in interactions with colleagues and weakening of relationships among them (Allen, Golden and Shockley, 2015). To feel satisfied at work, one has to feel related to other (Brunelle and Fortin, 2021) and given that remote work became the norm (over than 15.1 hours per week) throughout the pandemic and was used by the majority of co-workers and managers.

Overall, the effectiveness of the results is based on how well employees are doing their jobs. According to studies, working remotely is associated with higher levels of employee satisfaction (Bae and Kim, 2016). Several reports claim that even with remote labour, the rate of profit growth stays the same (Delanoeije and Verbruggen, 2020). Sardeshmukh (2021) discovered the negative correlation between remote work, job engagement and the mediation of these interactions by job demands with resources. Organizations recognized that providing security guidelines and safe guidance is important. Numerous businesses have set up flexible work hours and provided necessary equipment like a laptop or desktop, a keyboard, mouse, and shared internet etc.

2.4 Gender differences

Another aspect affecting productivity is gender. Compared to men, women are more responsible for taking care of the family. Women have to work more to fulfil both career and family obligations (Allen, Golden and Shockley, 2015). The fact that they don't have to rush to the office in the morning after finishing all housework is one of the benefits of WFH for women. Everyone in the household is at home, including the kids, and this puts a double or triple burden on women who must simultaneously care for the household, the kids' care and education, and their jobs

(Nakrošienė, Bučiūnienė and Goštautaitė, 2019). In addition, businesses could use prevention techniques like telemedicine (psychology) and unofficial support groups to give staff members the psychological assistance they require (Ng et al., 2020). Governments and companies should also create regulations that, in light of the new situation, protect the physical and emotional health of employees (Sandoval-Reyes, Idrovo-Carlier and Duque-Oliva, 2021).

2.5 Stress levels for employees due to WFH

Stress is a complex notion with no clear cultural definitions, and it is hampered by both a lack of comprehension and an excess of knowledge (Selye, 1980). In recent years, academics have come to the conclusion that stress has a significant impact on both employee health and organizational productivity. Due to social, cultural, and lifestyle changes, there will be a disruption in work-life balance, employee performance, and workforce efficiency, which will result in lower production (Ijcms.in, 2022). If people can control their work-related stress, organizational tension then it won't be a problem, but there are instances when workplace stress is intolerable (Ivancevich & Matteson, 1993). Age, education, work history, position, income, and marital status are all factors that affect a person's likelihood of experiencing job stress. An analysis of two industries information technology and agricultural research found that both had medium levels of stress, with women employees in both having high levels of stress as a result of job overload and role ambiguity (Prasad, et al., 2016). Task prioritising reports, as well as the solutions, are required for remote working, in addition to the accessibility of tools like team viewer, splash top, Microsoft remote desktop, zoom, Microsoft teams, as well as some communication tools (Melanie, 2020).

Due to a lack of managerial expertise and organisational experience which needed to create and implement strategies for WFH policies and procedures, according to Hickman (2019), workplace isolation will negatively affect remote workers (Ijcms.in, 2022). The workplace management team provides safe furniture, a clean environment, and some energizing parties, so the home setting is not always safe (Pillastrini et al., 2010). In addition, they frequently displayed increased negativity and irritation, which were linked to their disconnectedness and inability to discuss their problems at work and come up with solutions with co-workers (www.thecdi.net, n.d.). Yet because the employee won't have that kind of furniture at home, working in the same position all the time would result in back and neck ache (Côté et al., 2008). Compared to employee, employer has more

advantage because of WFH like the productivity increase, reduced office rent, lower infrastructure, maintenance expenses, and more employee involvement all result in improved cost savings (www.proquest.com, n.d.).

The below picture is the results of remote workers stress levels during COVID-19.



Figure 2: Stress levels

(Source: Armstrong, 2021)

2.6 What is meant by machine learning and how it could be beneficial towards identifying employment satisfaction?

The globe is interconnected, and analysing the large amounts of data using conventional methods is challenging. Millions of data points are produced every second by computers and other advanced technology. New technologies are developed to analyse these enormous data sets for example Artificial Intelligence, Big data and Machine Learning, etc. For extensive data analytics projects,

the term "big data" has gained popularity. As a result, data is a crucial aspect of organizational success. A wide range of analysis and storage techniques are needed for this type of data due to its size, transformation, and quick rate of change. To predict future results from a huge data set more accurately, one of the techniques is known as machine learning (ML) (Choi and Choi, 2022). There are different types of ML and in these supervised ML and unsupervised ML. Supervised ML is used for predicting the output based on input and the input is labelled data. Unsupervised machine learning algorithms are trained on unlabelled data and predict the future output based on that data. Algorithms for unsupervised machine learning do not need labelled data because they examine unlabeled data in order to find patterns that can be utilised to divide it into subsets.

The structure of Machine learning models is shown below:

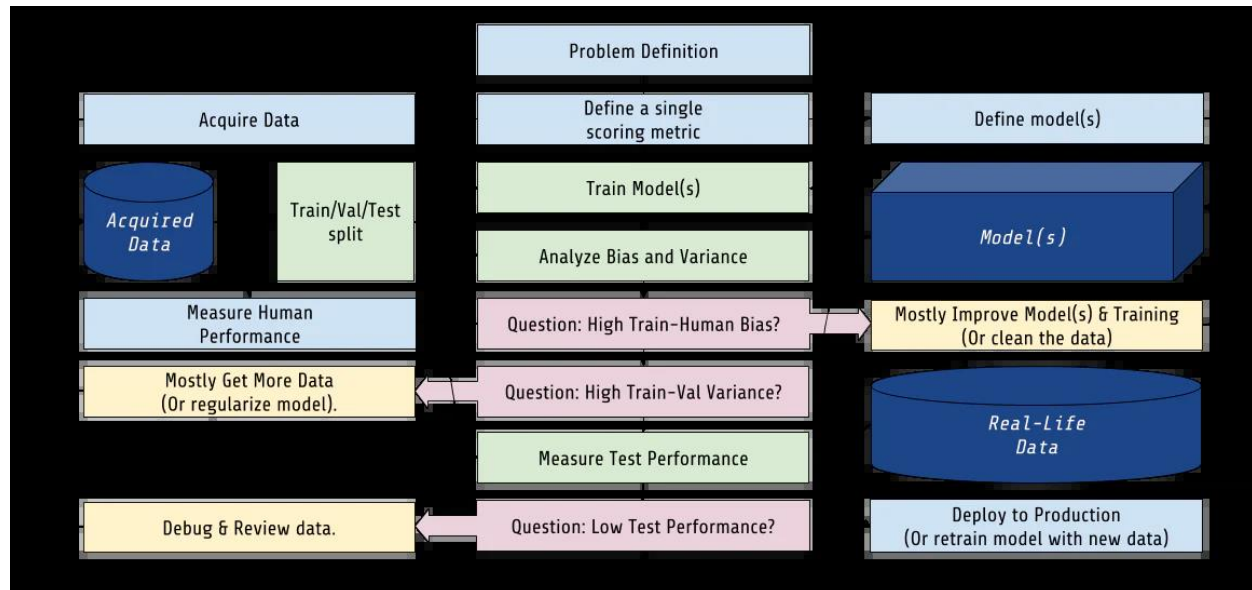


Figure 3: Machine learning model structure

(Source: Neuraxio, n.d.)

Machine learning methods are the key to business operations, and AI is becoming more useful in the real world. Researchers investigated a variety of ML techniques to enhance HR (human resources) management performance in response to changes in IT organizations (Li et al., 2011; LaFayette et al., 2019; Hughes et al., 2019). Support vector algorithm is another technique for solving classification tasks and this algorithm is also used for regression, density estimation and for other applications as well. The SVM algorithm's primary goal is to locate a subset of features

in an N-dimensional space that effectively labels the input points. So, by using SVM technique predicting accuracy for the employee satisfaction is one of the good approaches (Liu, 2021; Mutalib et al., 2021). Neural networks are the advanced deep learning technique which is also called as brain model architecture which connects the input layer, hidden layer with coefficients for finding the output. There are different types of neural networks are present and these are mostly used for face generations (Mutalib et al., 2021). The input layers are formed with input variables.

The default structure of Neural network is shown below:

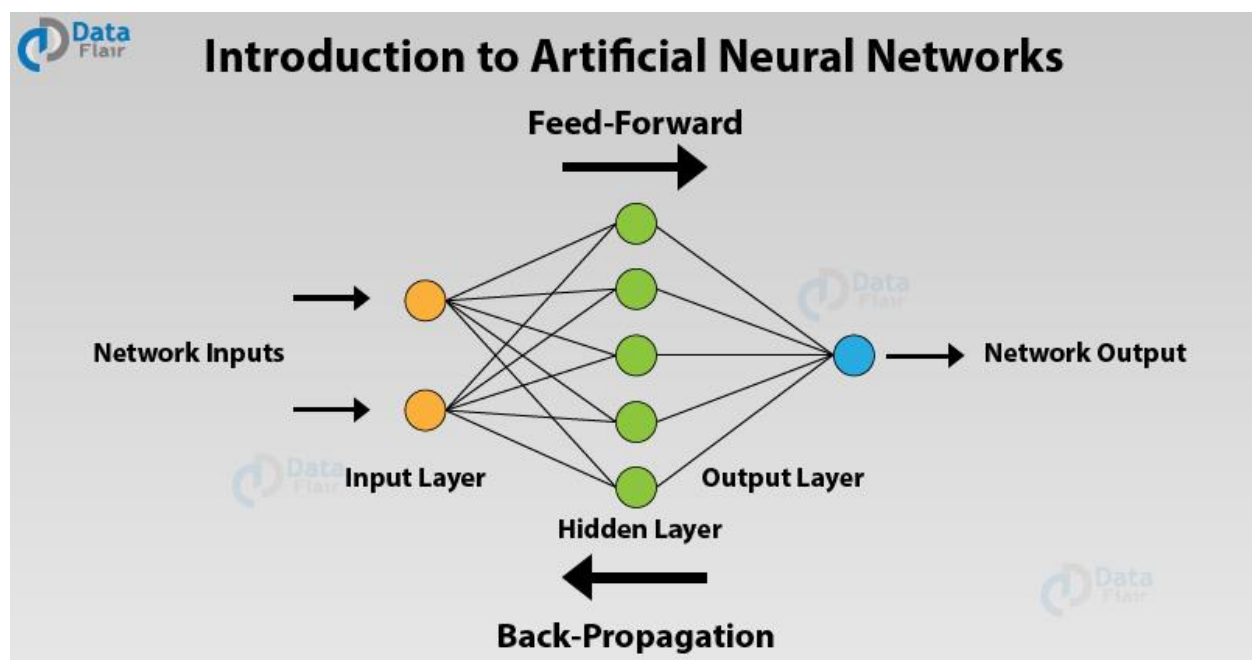


Figure 4: Neural network structure

(Source: DataFlair, 2017)

For analysing employee satisfaction machine learning are the best technique to predict the response which results employer to think about future invention remote work. There are many ML algorithms, out of those the author chose to perform supervised ML classification as the output variable has 5 classes. The author performs ML classification and advanced neural network which is one of the lead techniques in the market. Later the author compared both the models to predict which is giving best accuracy. The results are lead to get the most featured output which effects employee satisfaction. In recent years, one of the fundamental methods for detecting stress has

been machine learning. Supervised algorithms are used to solve mental health problems (Mutalib et al., 2021). With the use of probability, the basic approach for resolving classification issues is logistic regression, which makes use of the more complex sigmoid function as a cost function. The input variables are combined linearly with coefficient to detect the output variable (Mutalib et al., 2021). The coefficients are nothing but weights of the elements.

The present dataset contains higher dimension features spaces, so the chance of over fitting is more. To solve this risk the author performed data processing and then selected the best model for prediction which is NN as this as kernel which will help to extract the higher features.

The processes below are carried out to determine the impact of COVID-19 on employees.

Step 1: Input WFH Dataset: In the first phase, collecting data either from primary or secondary. In this research the author collected the data from data Mendeley which is the hub for data.

Step 2: Pre-Processing Data: In the second stage, the data is examined for any irrelevant fields or characteristics that are not crucial to employee happiness. Additionally, the data is examined for systematic, duplicate, and undefined errors. To avoid making any illogical errors when computing the dataset for prediction, do this. Data visualizations also carried out to find the features which effects employee satisfaction.

Step 3: Train-Test Data: The dataset is split into training and testing data in the third phase. Typically, 80% of the data will be used for training and the remaining 20% for testing. With training data, the model will learn to differentiate between various levels of employee happiness and determine threshold values for particular circumstances. The trained model will be used with testing data to look for issues with over fitting or under fitting.

Step 4: Implementation of the Model: The chosen methods will be applied in the fourth phase in order to determine employee satisfaction using the variables or features fed into the algorithms.

Step 5: Performance Evaluation: In the final stage, the model's efficiency will be evaluated in terms of many metrics, including accuracy, precision, F1-score, etc., to ensure that the method was applied correctly.

2.7 Literature Gap

After reviewing a number of research publications, the author concluded that many people used machine learning models to forecast future sample results. Sentiment analysis was used by several researchers to study employee emotions; however, they were unable to identify the optimal model or the feature that would be most closely associated with the input. Analysing the impact of COVID-19 on employees using Twitter data is one of famous use case but the chances for accurate predictions is less in this scenario. For effective data analysis the data should be large with more features so that performance will be more. To increase the performance the author should take the higher quantity of data. During the literature review it is found that there is a lot of awareness needs to be created for the employers about the employee's happiness. More review should need to be done. Organizations should maintain their profile more updated and available to the employee where they can add their feedback.

Chapter 3

This chapter covers all the methodological approaches in depth, including research method, research design, strategies, time horizon, ethical considerations, and data collection with data analysis.

Research Methodology

Research is a combination of two Words're' and 'search' which search again or finding the new facts or updating the old reviews with new ones. Research methodology is a strategy for methodically solving the research problem. Research approaches go through various stages. Research philosophies, research techniques, research design, research tactics, time frame, data collection, and data analysis are some of them (RESEARCH METHODOLOGY a step-by-step guide for beginners, n.d.).

Importance of Conducting Research

There are numerous benefits for conducting research, some of which are described below:

It enhances decision-making: Organizations are more capable of making decisions as a result of research's inculcation of the practice of logical and inductive reasoning (Gupta and Gupta, 2022). After carefully examining the current circumstances, all commercial groups and governments make their decisions. When making judgments about company policies and day-to-day operations, they actively use research techniques.

It aids in identifying trends: Making specific future predictions is a must for having the fortitude to confront it. These forecasts are not wild guesses; rather, they are based on an analysis of historical trends (Gupta and Gupta, 2022). For instance, the growth forecasts issued by the education minister are based on a detailed trend analysis of the educational system. Techniques including growth, peak, trough, and recovery, among others, are used to anticipate business cycles.

Helps to improve the productivity: After learning about their lows and highs as a consequence of their study, the organizations will be in better position. They will develop new ways to boost the productivity of the firms once they have identified their deficiencies (Gupta and Gupta, 2022).

The onion research diagram shown below:

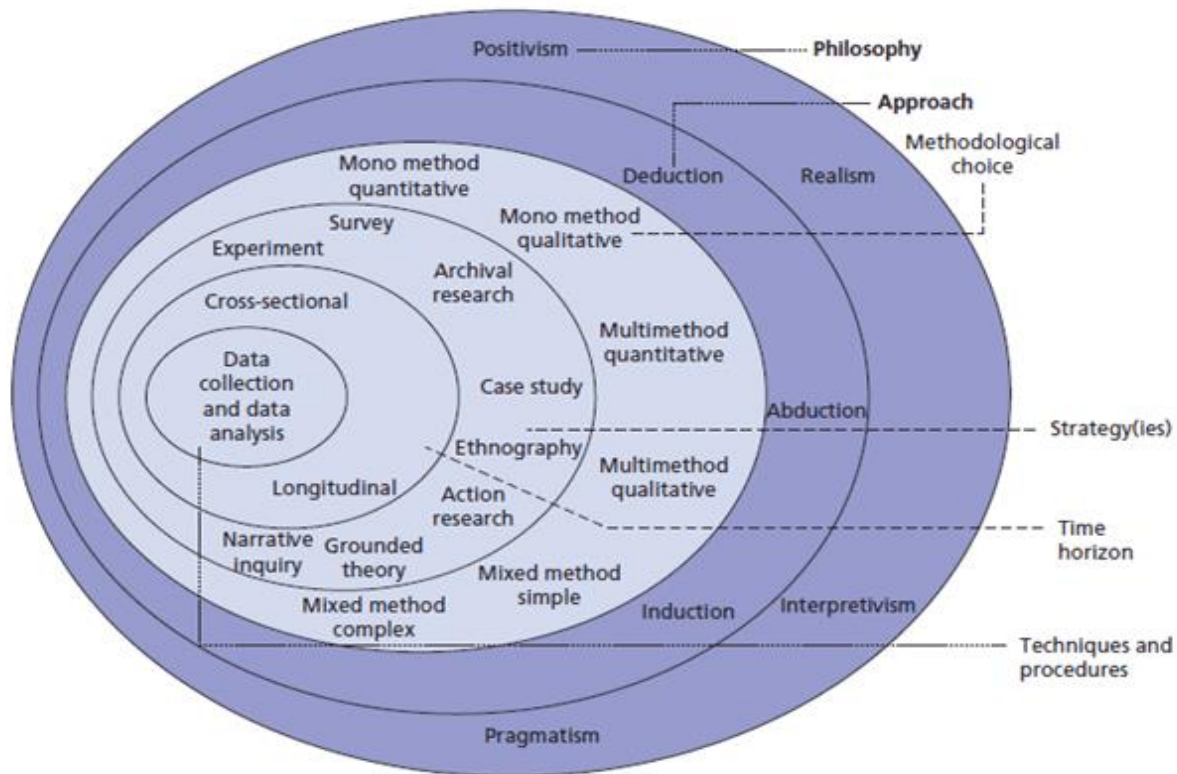


Figure 5: Onion Research diagram

(Source: Alamgeer, 2022)

Exploratory research is a form of research that investigates all sides of a topic and undertakes the following before coming to a previously undetermined conclusion. This survey is being conducted in an objective manner to ascertain each employee's viewpoint on WFH in order for organisations to understand the impact of WFH for future prediction and to compare the findings with the current models.

3.1 Philosophy

Several alternative research philosophies exist, and they are positivism, critical realism, interpretivism, postmodernism, and pragmatism (Saunders & Townsend, 2016, p836). The data collected for this study is secondary from primary data which was done by a survey. The primary data philosophy is interpretivism because the information for the earlier study was gathered through questionnaires given to employees who participated in surveys. But for this research the author taken the existing survey results so this suite both interpretivism and positivism philosophy.

Employee job satisfaction can be determined by the outcomes of a subsequent review using machine learning algorithms. The subjective meaning and experiences of people are highlighted by the study theory known as interpretivism. In this sense, using a 5 output fields (such as 'agree', 'strongly agree', 'disagree', 'strongly disagree', 'undecided') to predict impact of COVID-19 on employees can be considered as a method of gathering qualitative data that might reveal insights into the personal perspectives and experiences of employees.

3.2 Approach

The methodology includes several distinct kinds of research approaches. The deductive method is appropriate for this investigation. Data analysis follows a deductive process that starts with theories. The theory will be created first, and then it will be examined using the data gathered throughout the research. Using the ideas discovered through the literature study, questionnaires were found from secondary data for employees then the researcher applied exploratory data analysis and machine learning techniques to reveal employee opinions about WFH. Because it deals with unknowns, this deductive method is effective in producing the accurate answer. The research approach results are effective to predict the future values with the help of machine learning algorithms with advanced neural networks (Subasi *et al.*, 2021).

The below picture illustrates the steps for research methodologies:

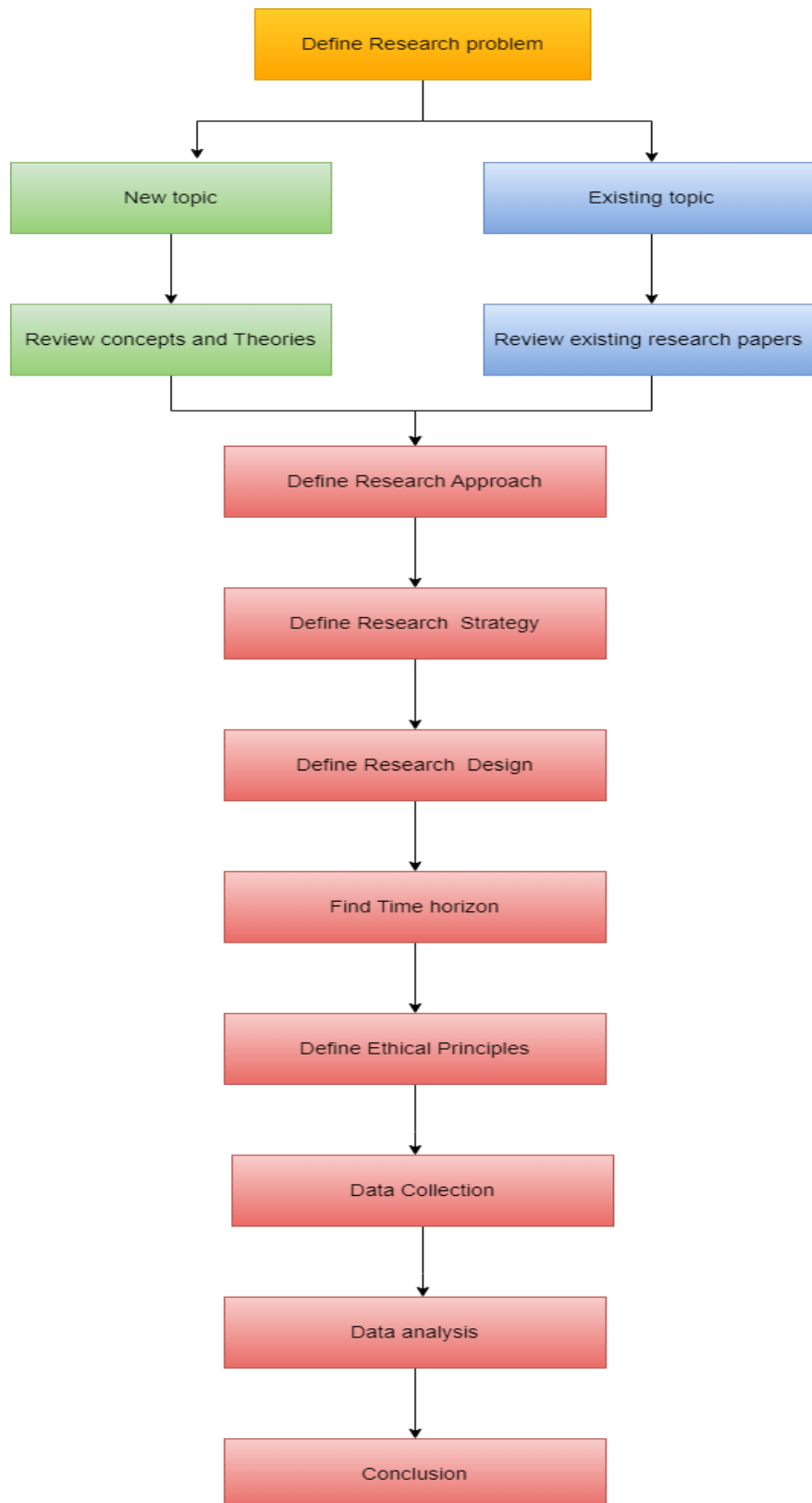


Figure 6: Research methodology

3.3 Strategy

It is merely a strategy for how one may attempt to reach their study objective. Many research methods are available, including experimentation, case studies, surveys, archival research, grounded theory, and others. The quantitative techniques category includes an experiment, a survey, and a case study. The author can select any methodology because this study falls under the quantitative research category, but the author felt that an advanced computer research case survey would be more suitable for this research because it integrates well with the exploratory research's nature, deductive research methodology, and quantitative research design.

3.4 Methods

Quantitative design and Qualitative design are the two methods in research. Through statistical analysis (which works with digits), quantitative analysis enables the researcher to identify the connection between ongoing and forthcoming projects. Quantitative research frequently aims to examine ideas deductively by seeking evidence to either confirm or refute them. Due to the fact that deductive research methodology progresses from theories to data, it functions extremely well with deductive research methodology but not with inductive methodology. Also, these hypotheses were created by scientists using inductive reasoning. Even though the research approach is deductive, the author chooses quantitative research for this study. When conducting case studies, organisations should take into account previous findings and continue their investigation (Soiferman, 2010).

3.5 Time Horizon

It is a timeline that the author specifies based on the goals and scope of the research. Currently, there are two different time frames present. They are longitudinal time horizon and cross-sectional time horizon. Longitudinal time horizon means performing the survey for data collection in longer period for better results. Cross-sectional time horizon carrying out the study within a predetermined period of time. The researcher chose **cross-sectional time horizon** because as this research has less time to complete the study. Because it would take too long to continuously collect the employee feedback across multiple conditions in several geographies using COVID-19,

secondary data was chosen instead. Participants from a range of ages were studied concurrently in this investigation. Hence, as opposed to choosing a longitudinal time horizon, the author chose a cross-sectional one.

3.6 Techniques and Procedures Sampling, Data collection and Tools

There are different samplings techniques present and, in this research, the primary research considered each employee opinion so the best suited technique is **unbiased (probability) simple random sampling**. The research is based on the individual perspectives of employees about the pros and cons of the remote working arrangement. The primary data defined as collecting live data from surveys, questionnaires and interviews. The term "secondary data" refers to information gathered from already-existing sources, such as official government websites. Data world, twitter, Kaggle, GitHub and IEEE are one the data repositories. The author chooses **secondary data** for this research and this secondary data found from **data Mendeley** which is the repository for providing data sets. Researcher reviewed numerous papers on Google Scholar, IEEE, SCOPUS, and Research Gate to understand more about the application of ML models in COVID-19 pandemic-related applications.

The author performed some statistics, visualizations and machine learning models to fill answer the research questions. The programming language used for this analysis is Python which is a open source and it provides many in-built libraries for machine learning models. Exploratory data analysis is performed along with machine learning models to find the statistics and represent the remaining objectives through visualizations. The main purpose of choosing ML models is due to its advance techniques for finding future unknown samples. It is very important for any industry to predict the future outcomes so that the industries can have an idea to proceed further. One of the most widely utilised methods for determining the subjective opinions of employees, or other individuals is sentiment analysis. For finding the impact of COVID-19 on employees many authors performed sentiment analysis which discussed in literature review. All the programming is represented in Jupiter notebook with a lucid manner comments.

3.7 Ethical Principles

Prior to the research, the author carefully reviewed the questions to determine whether they would involve sensitive personal data or not. The researcher must declare that the information gathered will be utilised purely for analysis and forecasting the key characteristics of remote workers as a result of WFH. Employee details from the gathered secondary dataset, including age, title, working hours, preferences, satisfaction with WFH, feedback, and other variables, will be retained in a secure environment. There are no ethical concerns with this secondary data because it was obtained from a secure source.

3.8 Performance Evaluation

The selected Load is operating a deep learning model NN to determine how the COVID-19 affects employees who work remotely. Accuracy is used in this study as a metric because it is a performance indicator for machine learning. The objective is to use descriptive statistics to examine and compare the accuracy of the existing model's neural network to the outcomes of earlier models.

3.9 Research Limitations

In order to distinguish the crucial features of post-COVID remote working IT workers, the researcher will take the essential steps to use a Neural Network as opposed to other multi categorization machine learning models. After reviewing the necessary literature reviews with numerous search engines, the author explained the research methodologies. As mentioned in the research methodology chapter, the researcher used a variety of data sources to collect the information. The most significant prominent factor was an advanced computer case study using secondary data with deductive technique. Cross-sectional time horizon is selected as this analysis has less time to finish. For developing new strategies more time is needed and it leads to find the new leads. As this research is conducted with secondary data, real facts are needed and gathering primary data is time consuming. After analysing several research papers many ML algorithms performed on the impact of COVID-19 on IT employees and this research is more advanced than

the present findings. In future more advanced techniques also can be done on this research using Convolutional neural networks.

Chapter 4

This chapter explains the results with detail explanation. Data collection, Data preprocessing, data cleaning, exploratory data analysis, applying machine learning models, finding best model and concluding the impact of COVID-19 on IT employees are discussed.

Data Analysis and Results

4.1 Introduction

All the world is suddenly faced many problems because of COVID-19 pandemic, and this changed entire world activities. Government made decisions which no one ever thought that would happen. All the borders are shut down and everyone stays at home because of lock down. Only in the emergency, people are moving around otherwise everyone has to stay at home. Important things like hospitals stay open. This research focuses on Industrial technology employee emotions about WFH which played important role during this pandemic. The growth of the organization is depending on the productivity of the employees and the productivity of the employees depends on their feelings. If employee is happy with work environment and work load, they can make the maximum productivity which leads best outcomes for the organization. That is the reason the author decided to find the employee satisfaction based on the few features. The models used for this research is Exploratory data analysis, Machine learning algorithms and Deep learning techniques to find the future predictions. Exploratory data analysis is used for better visualizations and to give statistical analysis. The purpose of visualizations is because of the view or understanding the problem with pictorial representation will be more effective than mathematical derivations.

4.2 Rationale Behind Choosing Models

The goal of this study is to determine employee satisfaction utilising descriptive analysis and machine learning model predictions. The elements influencing employee satisfaction using EDA approaches were defined by the researcher using pie charts, count plots, and correlogram plots. Along with descriptive analysis ML models such as KNN, Decision tree, Logistic regression and Support vector machine learning (SVM) also performed to predict the impact. Also, the author used neural networks, a more advanced deep learning that is superior to machine learning. The researcher used 4 fundamental ML models and sophisticated neural network comparisons to obtain the optimal accuracy model for the output. In ML, Decision tree provided best accuracy, but Neural network is the advanced AI with more accuracy than Decision tree. ML models are employed more frequently today because of their improved performance and ability to forecast future untested data. SVM is considered to be among the best-supervised machine learning methods, while KNN is a slow algorithm (Lim et al., 2020).

4.2.1 Support Vector Algorithm

SVM is an efficient supervised machine learning technique since they can handle both linear and non-linear regression and classification tasks (Zeebaree, n.d.). A difficult task in SVM is selecting hyperparameter values and author performed the research with different hypers parameters to check the variations for better accuracy. The first chosen hypermeters are kernel = rbf, C= 0.1 and gamma = 0.5. The accuracy with these hyper parameters is 45.10 % and F1 score is 28.03 %. The second chosen hyper parameters are kernel = poly, C= 1 and degree = 3. The accuracy with these hyper parameters is 47.06 % and F1 score is 32.16 %. To eliminate overfitting or underfitting from the noise and outliers, hyperparameters are applied. Reason for using the kernel as the radial basis function (rbf) is that it lowers the problems with space complexity and is similar to the K-NN technique. Gamma indicates the impact of a single training data set. Gamma and performance are inversely proportional to each other. As gamma increases performance decreases. A regularisation or optimisation parameter called "C" is utilised to manage the trade-off point (Friedrichs and Igel, 2005). Since there is no linear way to separate the data, the polynomial kernel is frequently utilized in SVM classification issues. In certain cases, the polynomial kernel can locate a class label that divides the classes by mapping the data into a higher-dimensional space. As this research output

variable contains multi class labels the author has to perform multi label classification. The polynomial kernel is the best kernel to this analysis, and it's proven with better accuracy than rbf kernel.

4.2.2 Logistic Regression

The likelihood of the target values from the data set is determined by using logistic regression (Stoltzfus, 2011). The hyper parameters for this logistic regression are multiclass equals to 'multinomial' and solver equals to 'lbfgs'. Obviously, reason behind choosing the multinomial class because this operation is executed on 5 multiple classes such as 'Agree', 'Strongly agree', 'Disagree', 'Strongly disagree', and 'Undecided'. By executing repeated approximation reductions along with coordinate directions or coordinate hyper planes, the **Coordinate Descent (CD) approach** used by the solver helps to address optimization issues. The decision tree approach is a classification tree method that collects bootstrap data and randomizes predictions in order to achieve a high level of projected accuracy (Rigatti, 2017).

The L-BFGS stands for 'Limited-memory Broyden–Fletcher–Goldfarb–Shanno' functions approximately the fastest when the dataset is small compared to other methods, especially since it saves a lot of memory. The phrase "Limited memory" merely indicates that it only maintains a small number of units that implicitly reflect the prediction. Nevertheless, there are some "severe" limitations, such as the possibility that if it is not properly protected, it will not converge to anything. The accuracy for this algorithm is 38.4 % which is very low than other algorithms.

4.2.3 Decision Tree

Decision trees can estimate the future unknown samples using straightforward decision rules and the predictions of the past (Myles et al., 2004). They have the ability to accommodate complicated data sets and let the user see how the decision was made. The hyper parameter criteria which control the method used to measure the impurity of a split. Entropy as the criteria can be used as a metric for impurity instead of "gini," which is the default value. Max depth defines the most depth they should let the tree reach. The model will become more sophisticated with the more depth they permit. It is simple to predict what will occur in the case of training error. Training error will constantly decrease as max depth increases. The hyper parameters for this algorithm are criterion

as ‘entropy’, max depth as 3 and random state is 101. The accuracy for this algorithm is 51 % which is the highest than other algorithms.

4.2.4 K-NN Algorithm

Analogy based learning is the foundation of nearest neighbor classifiers. The training samples are building with n-dimensional samples (Cunningham and Delany, 2021). Each point represents signal point in one dimensional space. As to conclude all the n-dimensional space patterns are filled with each sample. In future if any new unknown sample occurs then k-nearest neighbour’s searches for the patterns which match the unknown sample. Nearest neighbour classifiers are instance-based, or a lazy approach since they maintain all of the training samples and wait to build a classifier until a single (unlabelled) sample needs to be classed (Satyanarayana, Ramalingaswamy and Ramadevi, 2014). For this analysis KNN provided accuracy is 49 % and which is lower than decision tree.

4.2.5 Neural Networks

It is a deep learning technique that trains the software to analyse input data, and the name was chosen based on the neurons found in the human brain. Neural networks are used in many applications recently as it has advanced artificial intelligence (Ripley, 1994).

It has three segments in the process, and they are input layer, hidden layer and output layer. The training data is delivered to the input unit, where it is processed by the hidden layer, and then the output layer stores the outcomes. Neural networks are best for classification algorithms and image processing. There are multiple layers exist in the hidden layers which handle the process. For this analysis the author selected ‘relu’ to activate the layers (Lam et al., 2014).

4.3 Implementation

In programming everything goes in step by step. The following steps define the entire process for finding the employee satisfaction.

4.3.1 Importing the libraries

In python there are many in-built libraries are present which will help to visualize the results or to perform statistics or to predict the accuracy. They all are very simple and understandable.

As per the required operations user should need to download the libraries. The below all libraries are used for reading the data, processing it and then finally to visualize.

Pandas: This library is used to manipulate the data and it is built from Numby library.

Numpy: This library is used to perform mathematical operations on arrays.

Matplotlib: The name itself saying that it is used for plotting the graphs.

stats: This is used to perform statistics on the dataset.

seaborn: It is used to draw the statistic plots python.

4.3.2 Data processing

Finding pure data is impossible; the data got for this analysis contains very less information and contains some Nan values, other language titles and unwanted characters. The author renamed columns by using rename function on the long columns to make them shorter and make understandable. The columns which contain Nan values are dropped and by using replace function some of the string values are replaced. Like 'Yes' is replaced with 1 and 'No' is replaced with 0 for better predictions then converted into float to check the correlations. In the cleaning process some of the unwanted rows are deleted with drop function. The deleted rows are ID, Start Date, Completion data and Comments. There is no use in keeping these rows for the analysis.

4.3.3 Methods Effectiveness

The techniques to make a decision on the question will come from an analysis of the EDA, followed by the implementation of accuracy analyses for the five models that were selected in order to determine the most appropriate algorithm. Based on the accuracy the best model will be developed and then by using that model accuracy the performance of the model can be defined.

By using confusion matrix, the author can identify the performance rate with recall and precision values. F1-score will be very useful for when the data is uneven, and it is the average of recall and precision. All these parameters are calculated for the best accuracy model i.e., Neural network. The detailed explanation of all the programming part is explained below.

Exploratory data analysis (EDA)

EDA is used to analysis the data sets using visualizations and statistics. This EDA techniques used here is to

1. Organize the personnel into categories based on their job titles and noting which departments support working from home and which do not.
2. Exploring the Age based on employee work-life balance through visualizations.
3. For finding the resource details.
4. Checking whether any cyber issues are present or not.
5. What devices employee prefers to use during WFH.

Age analysis

The pie chart below displays the overall participation rates for each age group.

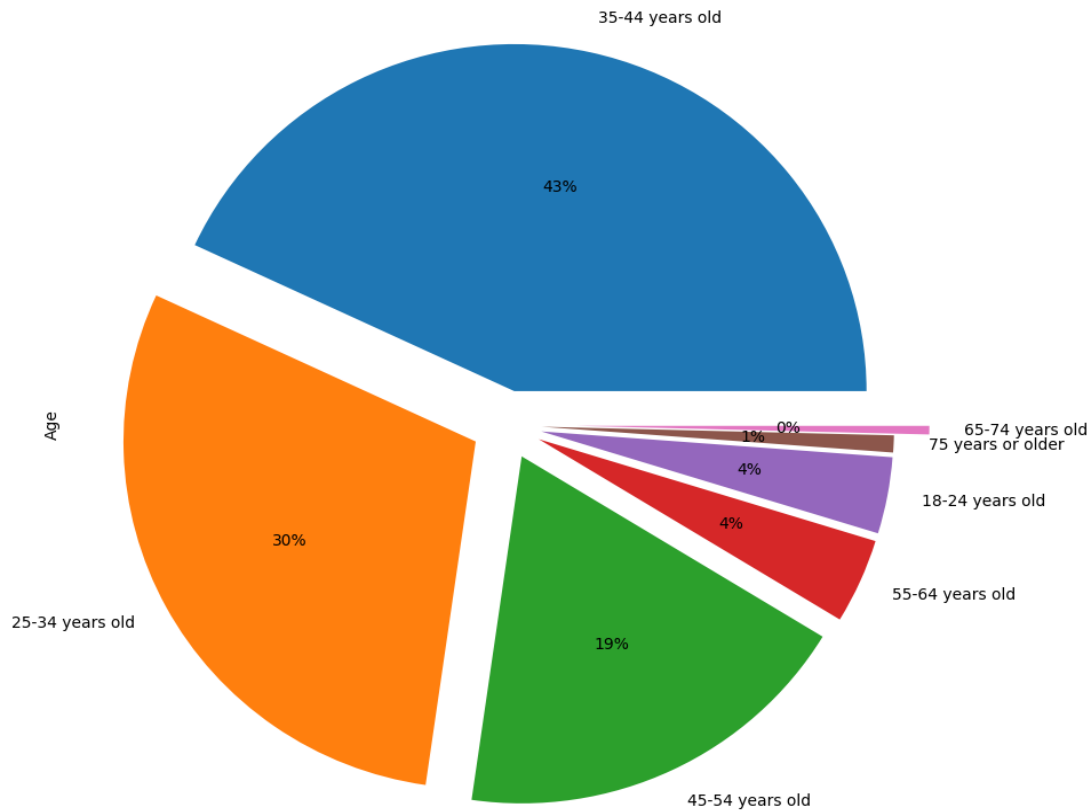


Figure 7: Age count

Based on the above pie chart age 35-44 years old people give more response than other age group people.

The below graph illustrates which age group people want work from home.

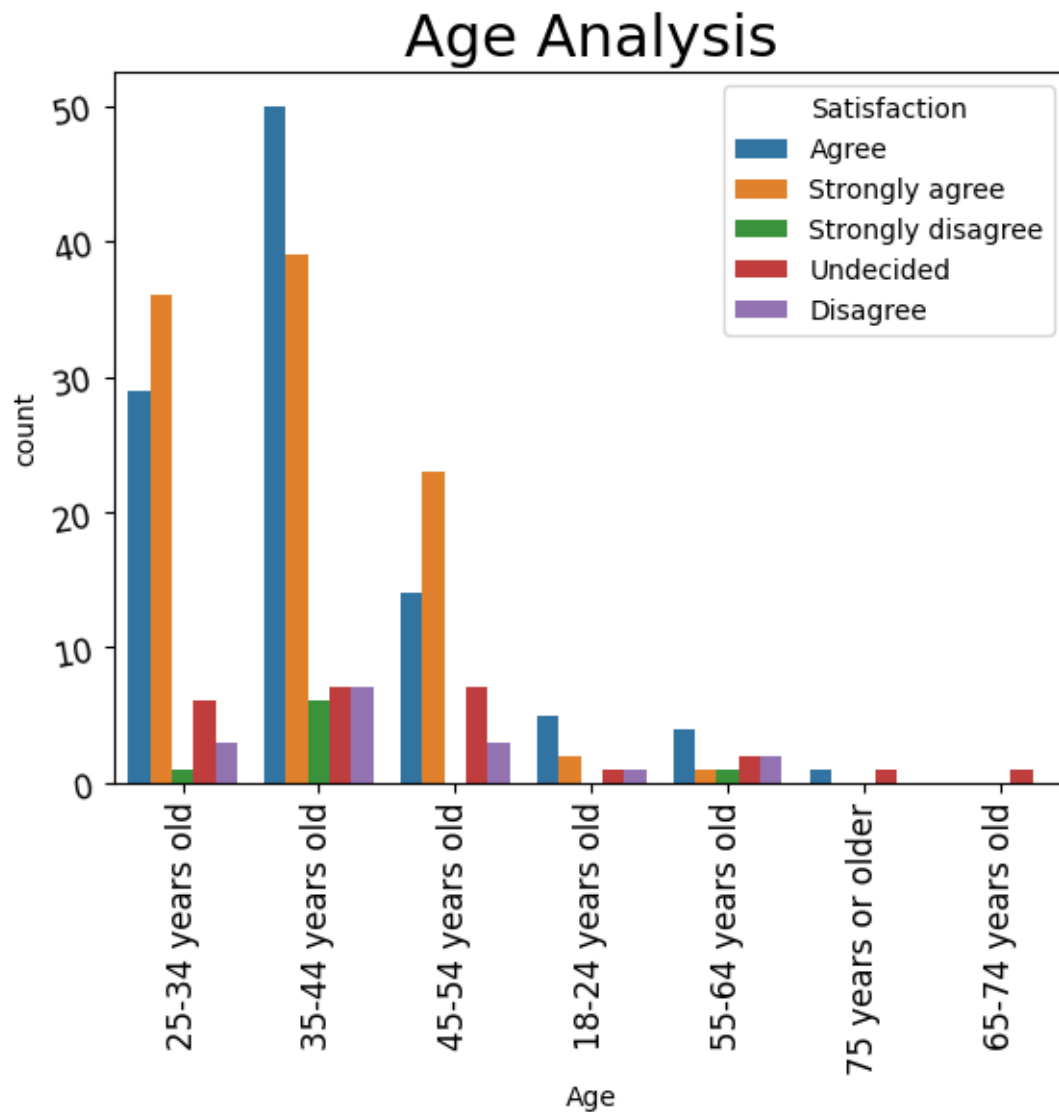


Figure 8: Age analysis

Age 35-44 years people are more satisfied with remote work features. Above age 65 people are not to decide whether they are comfort with remote or not.

Designation analysis

Every employee workload is based on their designation and security levels of work is also depends on their role. Signalizing the employee satisfaction based on their designation is also important. The below graph is drawn to check which role employees want to work remotely and who are not.

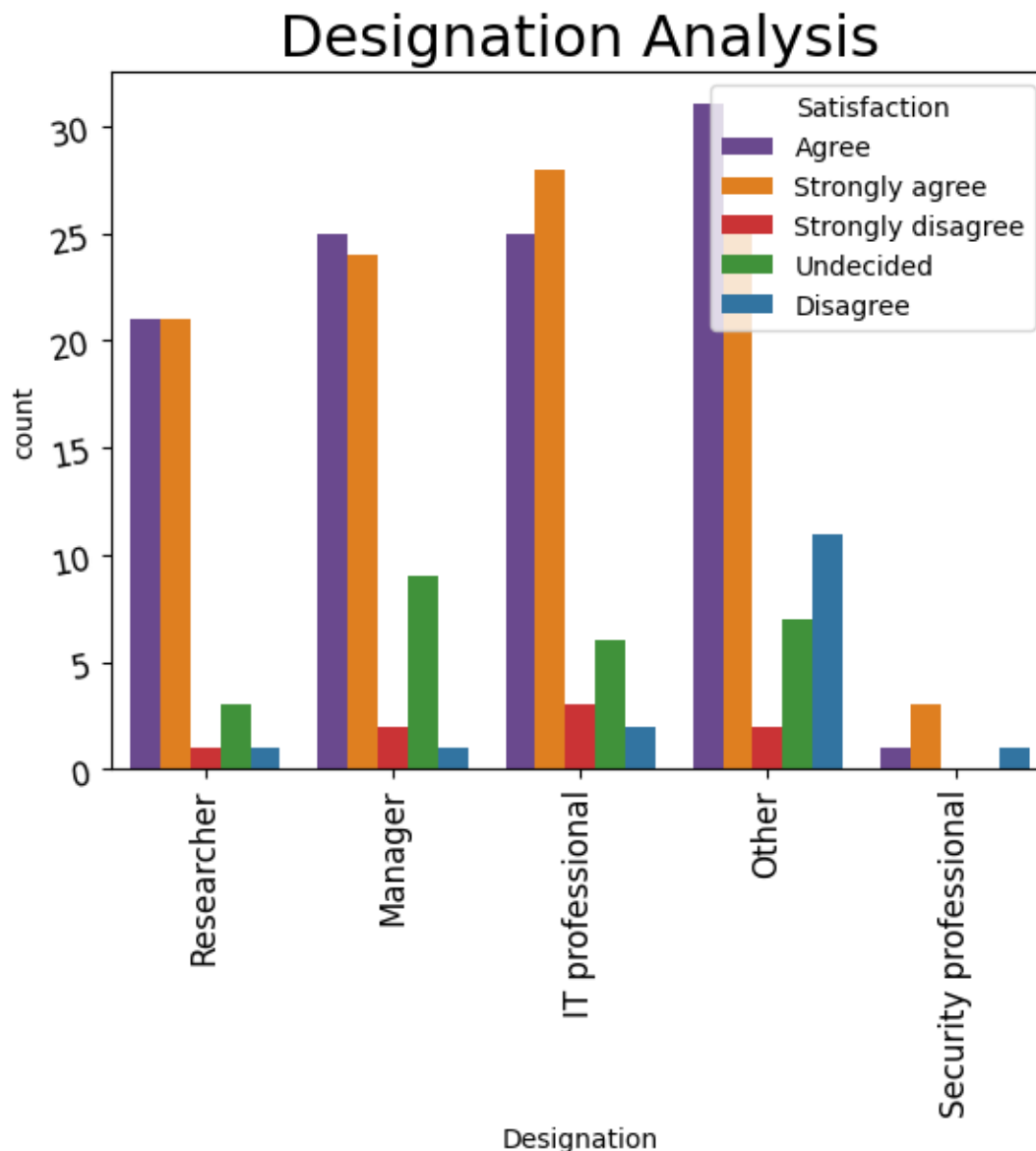


Figure 9: Designation analysis

As per the results IT professionals are interested in WFH than other roles. The rationale for this might be that these people are developers, and they don't necessarily need to manage clients in person because they can manage them virtually without any problems. Other means financial team and other support teams these are also giving positive response towards WFH. Security professionals are giving negative response about remote work because if employees work in remote the cyber issues will be more. It is because employee may not be able to follow all the security guidelines when they are at home.

Overall Satisfaction report

This visualization is specially meant for analysing the overall satisfaction of the employees based on the features which are shown below:

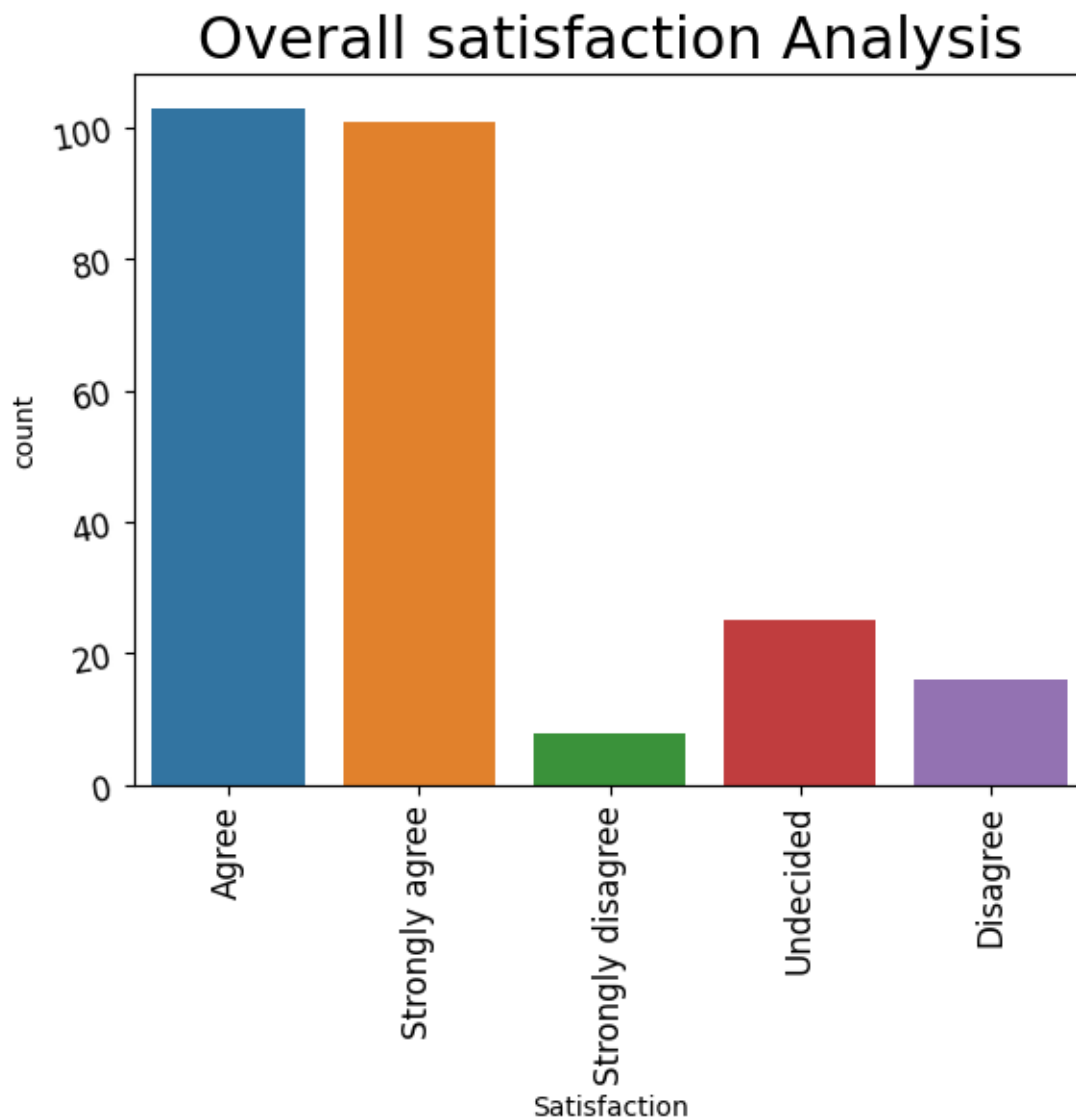


Figure 10: Overall Satisfaction report

This report is out of 264 employee responses, 90 people are strongly agreed about WFH environment. In remaining people 100 are merely agreed then this can conclude as a positive sign for remote work future.

Resource usage

The employer should know the resource details like how many resources need to be provided and what resources employee preferred. This analysis will help organizations to find the resource requirements.

The below will explain more:

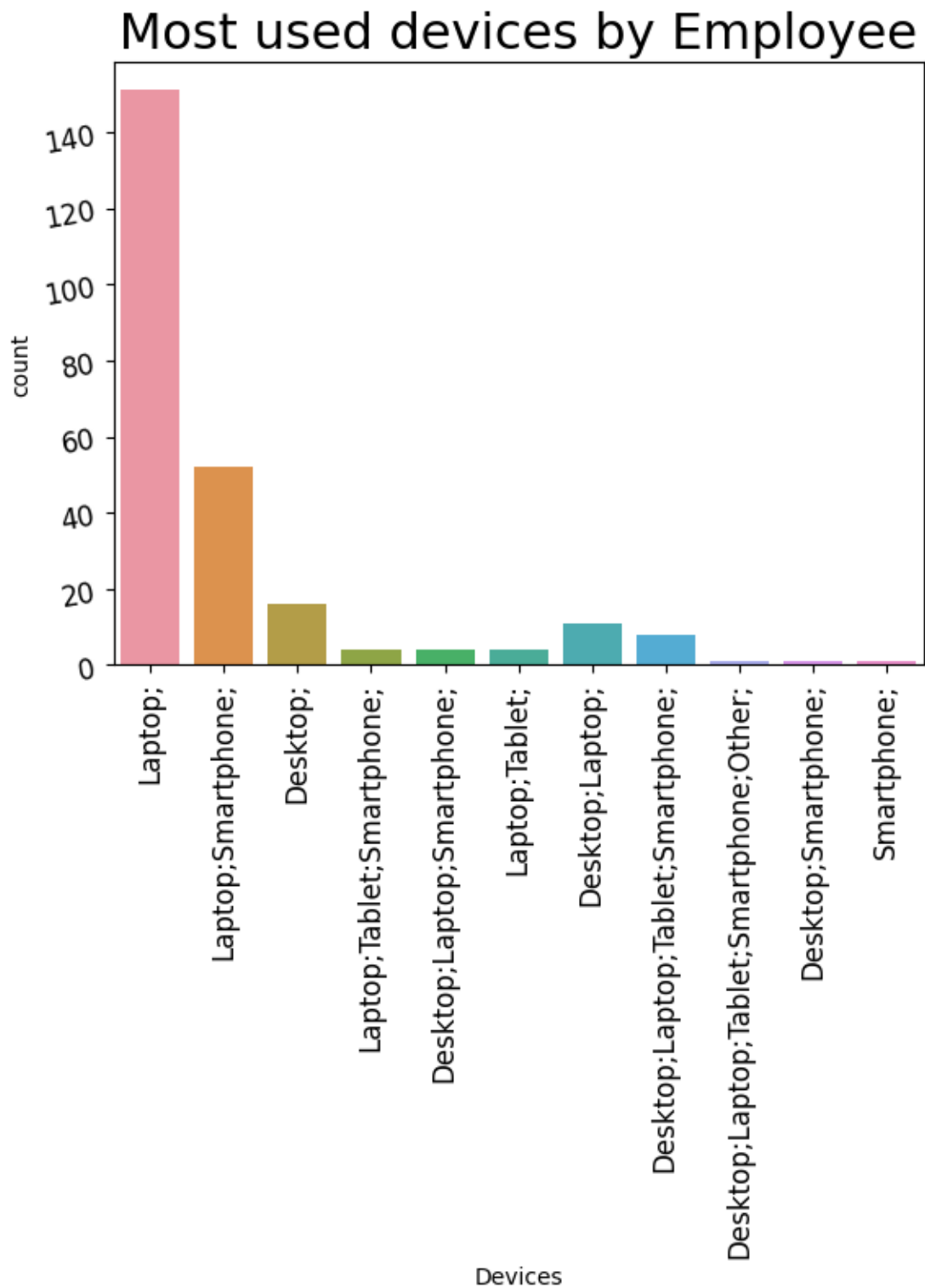


Figure 11: Resource usage

When it comes to work, maintaining security is very important along with employee comfort. The study results mostly employees prefer to use Laptop than other devices because latest laptops are very easy to carry and having good security features with more advanced technologies.

Preference of work by employees before COVID-19

The author explored how people emotions on remote work before COVID-19.

The results are explained through the below graph:

Preference of work from home by Employee before covid

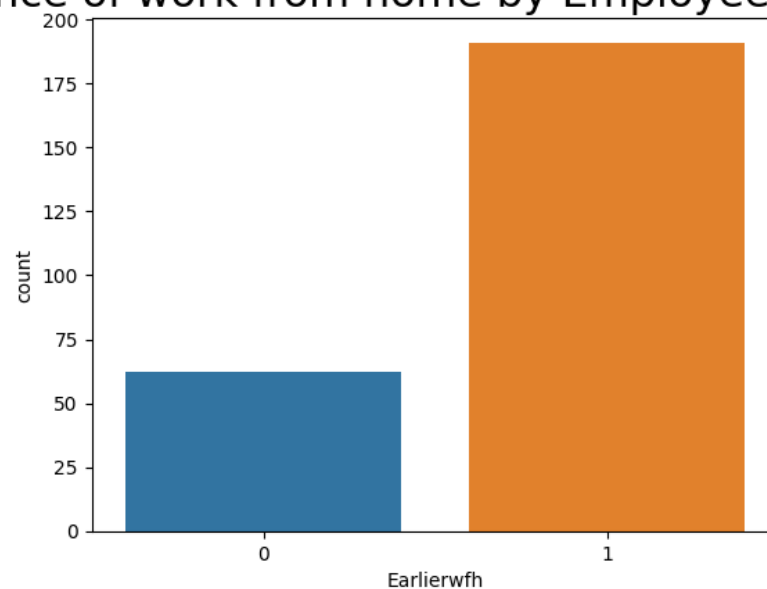


Figure 12: Preference of work by employees before COVID-19

Even before the epidemic, IT professionals were interested in remote work since it reduced travel time and allowed them to spend more time with their families at home.

Issues faced by employees by cyber threads.

Every new technique will have advantages and disadvantages. One of the disadvantages for this remote work is cyber issues because when employee working remotely, they may not be able to use secured environment all time. Let's see how this survey reports about cyber issues which given below:

Issues faced by employees by cyber threads

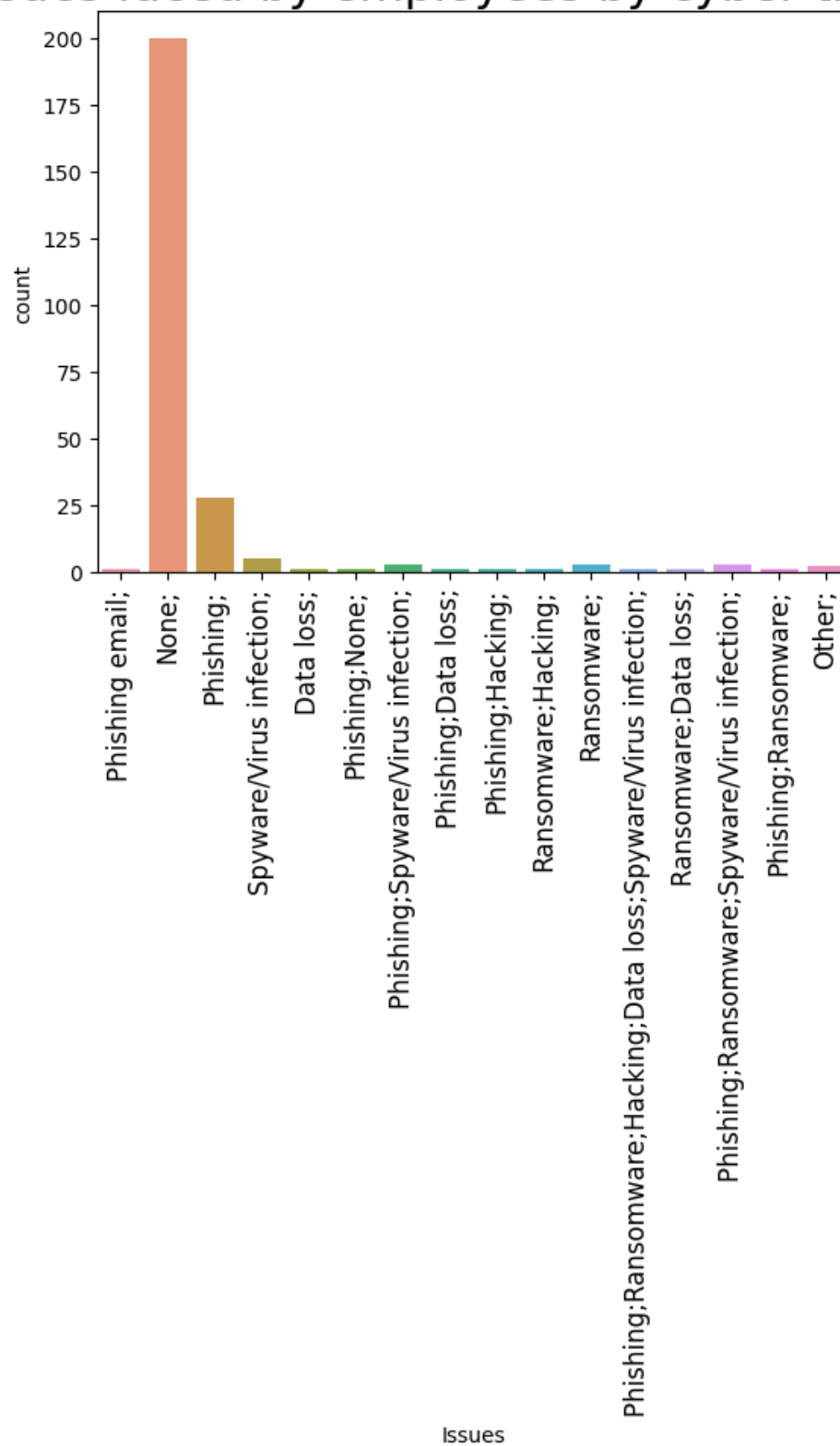


Figure 13: Issues faced by employees by cyber threads

Most of the employees said there are cyber issues, and their work progress is going well. Few employees reported that they are facing problem with suspicious mails which is called phishing.

By this fraud mails cyber attackers are trying to grab the employee personal information like credentials. Along with phishing, ransom ware attacks are also more which means injecting the malware software to the system which will block the computer.

Security measures

To get rid of the cyber issues the organizations should follow some safe guidelines. Let's see what the third-party environments are used by IT organizations by using below picture.

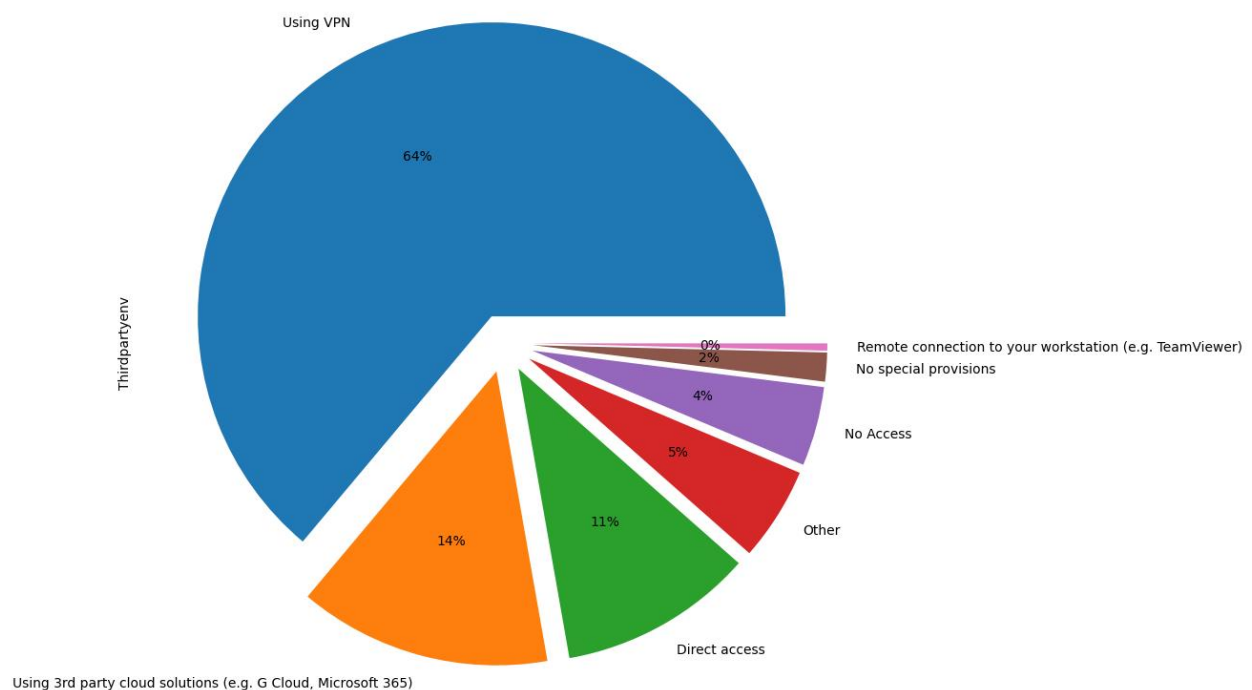


Figure 14: Security measures

As per the above pie chart most of the people are using VPN which is meant by 'virtual private network' used to protect the sensitive data from the network providers by providing protective connection and this also hides the internet protocol IP address. Simply most of the clients are providing VPN as the virtual network will reduce risks. No one is using Team Viewer as per this research.

Assets used by employees.

Employee preference matters to the employer for the organization growth. So, this analysis will be very useful for the organization to know more about their requirements. The most frequently utilized assets by workers are shown in the graph below.

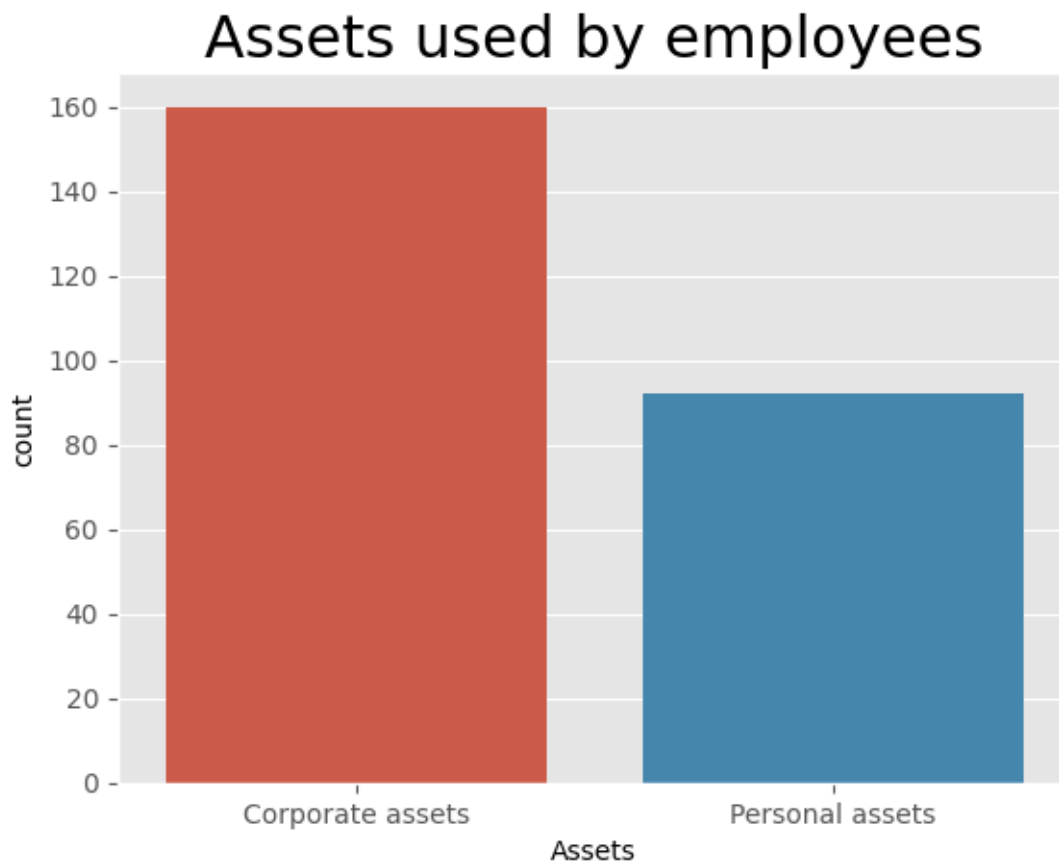


Figure 15: Assets

More than half employees are referencing corporate assets over personal assets because office people will provide necessary software with more secure operating systems but in employee personal assets the employee must install all the required software's for the project without IT support.

Future preference about remote work by employee

Finally, the overall concept for this entire analysis is to predict the future of the remote work. The below graph will explain more details:

Future preference about remote work by employee

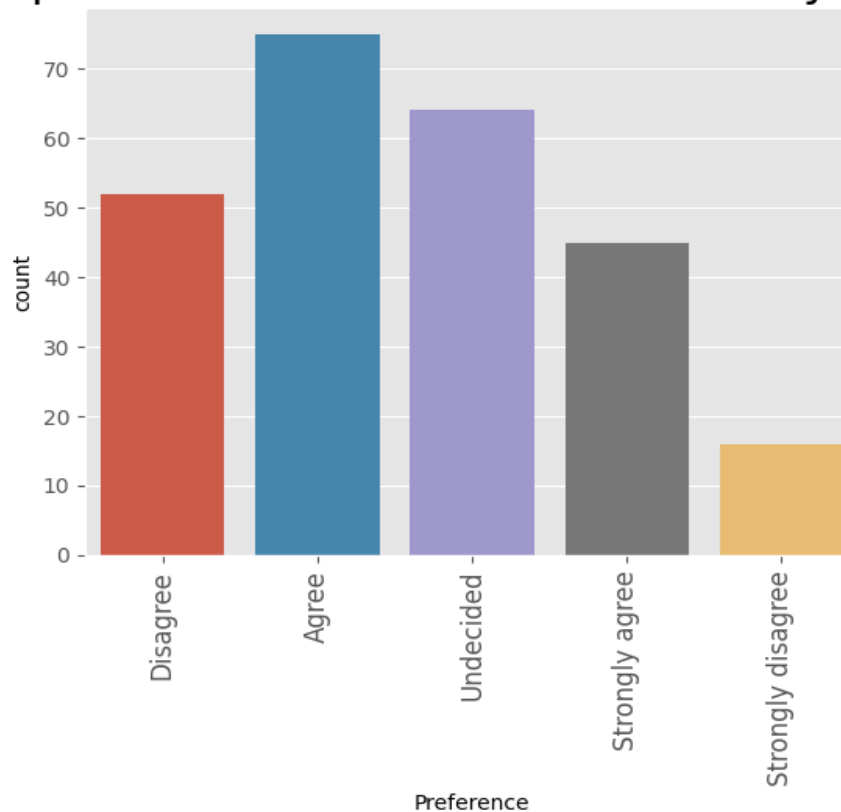


Figure 16: Future preference about remote work by employee

Compared to negative reviews positive reviews are more for remote work that's why some people requesting permanent work from home and 40% of the employees don't want work from home as they are not able to balance their mental health and work pressure. 50% people are not able to decide what to choose.

Awareness created by employer.

When any new technology introduced then the employer should give training to the employee for better performance. Creating awareness to the employees is more important for productivity.

A bit glimpse of the results shown below:

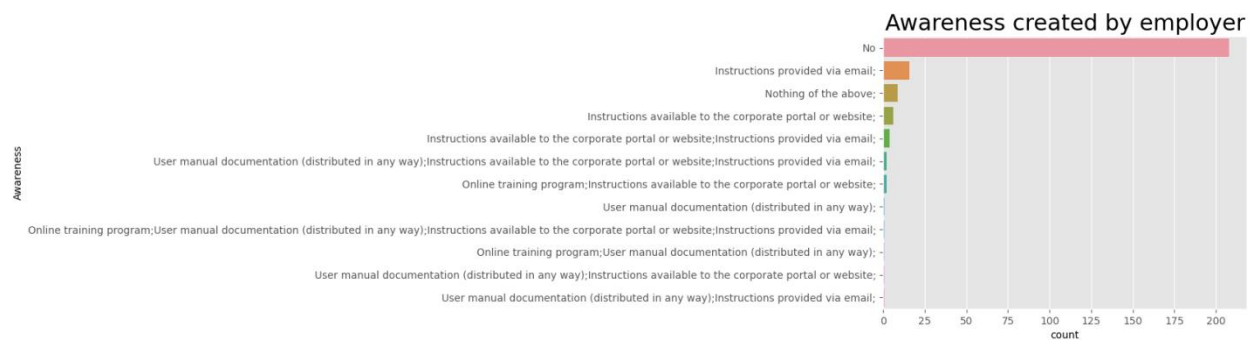


Figure 17: Awareness created by employer

Many employees are saying that they didn't get proper awareness of the work, cyber issues, how to use the asset properly and further notifications. The majority of the information is provided via email, the company website or portal, online training, and user manual documents.

Machine learning outcomes

Machine learning models are used to predict the future data based on some trained data. By calculating accuracy or error with the help of classification or regression tasks researcher can identify how accurately the future data can be predicted. The primary objective of the ML algorithms used in this research is to estimate employee satisfaction using a few input variables including 'Awareness', 'Preference', 'Designation', 'devices' and 'guidelines given'.

Best Model

The author tried to check which model given best accuracy and results are represented in a data frame shown below.

Finding the best model

```
In [200]: #Creating a Two-dimensional array which contains above 5 machine learning models
best_acc = [['Logistic Regression', LR_accuracy], ['Decision Tree', DT_accuracy], ['KNN', employee_knn_acc], ['SVM', poly_accuracy]]
```

```
In [201]: best_model = pd.DataFrame(best_acc, columns = ["Model", "Accuracy"])
best_model
```

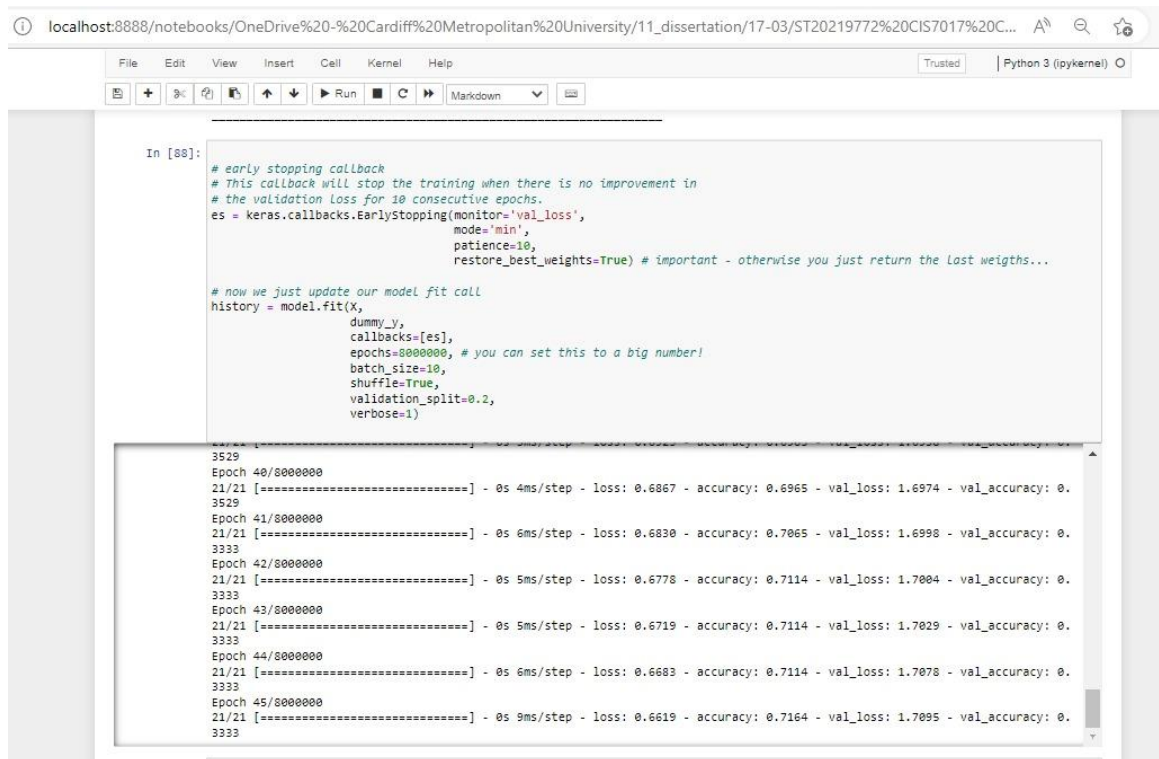
```
Out[201]:
```

	Model	Accuracy
0	Logistic Regression	38.369231
1	Decision Tree	50.980392
2	KNN	49.019608
3	SVM	47.058824

From the above results Decision tree is giving the highest accuracy with 51 compared to other algorithms and least provided accurat is by Logistic regression classifier.

Figure 18: Best Model

Neural network output is shown below:



```

In [88]:
# early stopping callback
# This callback will stop the training when there is no improvement in
# the validation loss for 10 consecutive epochs.
es = keras.callbacks.EarlyStopping(monitor='val_loss',
                                   mode='min',
                                   patience=10,
                                   restore_best_weights=True) # important - otherwise you just return the last weights...

# now we just update our model fit call
history = model.fit(X,
                    dummy_y,
                    callbacks=[es],
                    epochs=8000000, # you can set this to a big number!
                    batch_size=10,
                    shuffle=True,
                    validation_split=0.2,
                    verbose=1)

```

```

3529 Epoch 40/8000000
21/21 [=====] - 0s 4ms/step - loss: 0.6867 - accuracy: 0.6965 - val_loss: 1.6974 - val_accuracy: 0.
3529 Epoch 41/8000000
21/21 [=====] - 0s 6ms/step - loss: 0.6830 - accuracy: 0.7065 - val_loss: 1.6998 - val_accuracy: 0.
3333 Epoch 42/8000000
21/21 [=====] - 0s 5ms/step - loss: 0.6778 - accuracy: 0.7114 - val_loss: 1.7004 - val_accuracy: 0.
3333 Epoch 43/8000000
21/21 [=====] - 0s 5ms/step - loss: 0.6719 - accuracy: 0.7114 - val_loss: 1.7029 - val_accuracy: 0.
3333 Epoch 44/8000000
21/21 [=====] - 0s 6ms/step - loss: 0.6683 - accuracy: 0.7114 - val_loss: 1.7078 - val_accuracy: 0.
3333 Epoch 45/8000000
21/21 [=====] - 0s 9ms/step - loss: 0.6619 - accuracy: 0.7164 - val_loss: 1.7095 - val_accuracy: 0.
3333

```

Figure 19: Neural network output

Based on the table Decision tree is the best algorithm for this research compared to others. But neural networks are better than decision tree because NN provided highest accuracy than machine learning models with 71.6%. The accuracy for Logistic regression is 38.3%, for Decision tree 51%, for KNN 49% and for SVM 47%. To increase the accuracy for all the algorithms the chosen dataset must be large and cleaned.

Confusion matrix for the best model

It is used to compare the actual values with predicted values so that the performance of the ML can be analysed. It is used for the optimization of ML. As this is multi classification with 5 classes the confusion matrix size is 5 X 5. For this analysis decision tree matrix given maximum accuracy in ML models so the confusion matrix for selection tree is proven underneath.


```
In [167]: #Confusion matrix
DT_con_matrix = plt.axes()
sns.heatmap(metrics.confusion_matrix(y_test, y_pred),annot=True, cmap="Blues")
#Setting title
DT_con_matrix.set_title(('Decision Tree Accuracy score: ', DT_accuracy), fontsize=20)
#Setting x label
DT_con_matrix.set_xlabel("Predicted values")
#Setting ylabel
DT_con_matrix.set_ylabel("Actual Output Classification")
plt.show()
```

('Decision Tree Accuracy score: ', 50.98039215686274)

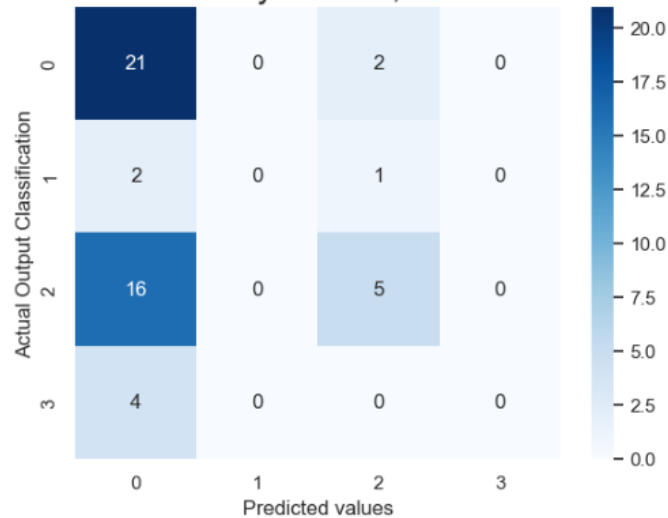


Figure 20: Confusion matrix for Decision tree

Confusion matrix for Neural network is shown below.

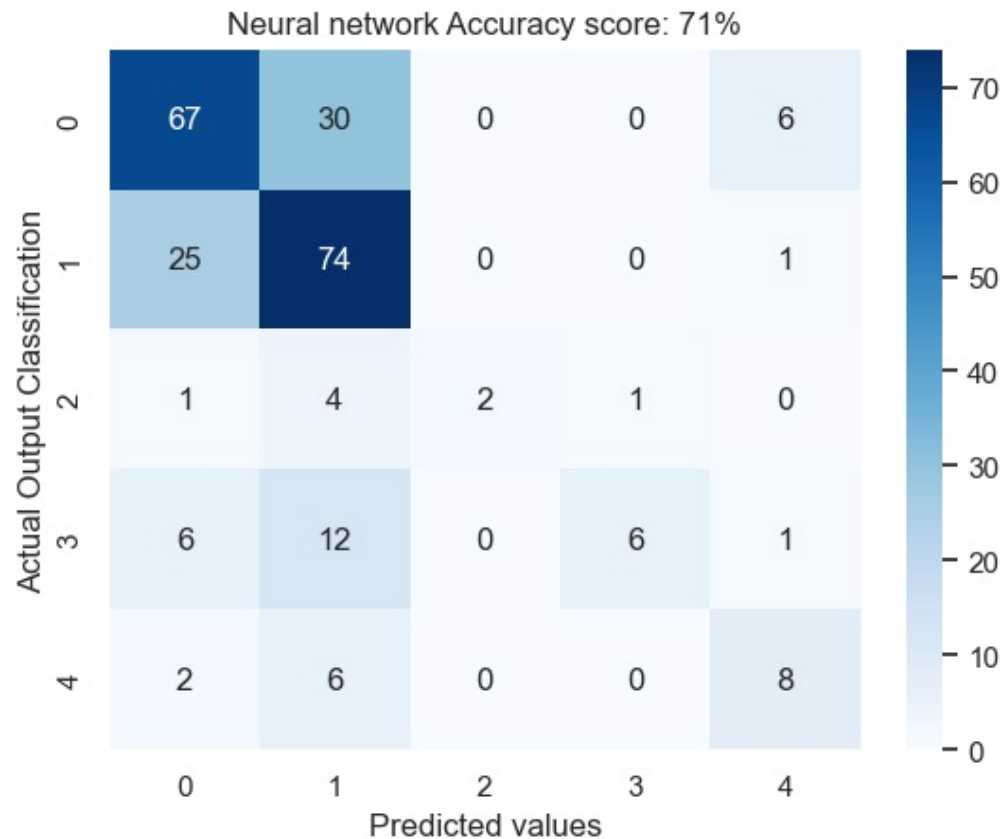


Figure 21: Confusion matrix for Neural network

Along with matrix confusion matrix is also used to find the precision and recall. Precision tells us how proper the version is at predicting a selected elegance. Recall tells you about the way in many instances the model was capable of come across a specific magnificence.

The classification for neural network is shown below:

```
In [154]: # more detail on how well things were predicted
print(classification_report(dummy_y.argmax(axis=1), preds.argmax(axis=1)))
```

	precision	recall	f1-score	support
0	0.57	0.64	0.60	103
1	0.63	0.67	0.65	100
2	0.75	0.38	0.50	8
3	0.50	0.20	0.29	25
4	0.47	0.44	0.45	16
accuracy			0.59	252
macro avg	0.58	0.46	0.50	252
weighted avg	0.58	0.59	0.58	252

Figure 22: Classification report for Neural network

4.4 Challenges and Lessons Learnt

The dataset has very limited amount of data which results less training for the model. By using visualizations, the author predicted the output, but visualizations are not productive for future predictions so, the author selected few best supervised machine learning models for better performance. Inappropriate columns are removed because some of them are not helpful for answering the research objectives (Example: ID). Although choosing a kernel for SVM is challenging, the researcher chose a polynomial basis function over a radial basis function as this is a non-linear problem (Norkin and Keyzer, 2009). Effectiveness of ML models is influenced by hyperparameters. Gamma and C values are key factors in SVM. The gamma value affects the ML's performance, which shifts from exponential to linear (Al-Mejibli, Alwan and Abd, 2020). There are no specific guidelines for selecting the values of gamma and C, because these are optimal parameters. Hard tests should be used to determine the most appropriate values for greater accuracy (Duan, K.B. and Keerthi, S.S., 2005). During several trials, it is discovered that the gamma value of 0.5 and the "C" value of 0.1 offer more accuracy. It has been established through numerous experiments that polynomial, sigmoid, and RBF are the best-suited kernels, with a 47% accuracy rate.

The decision tree algorithm also given headache; the author tried to plot the tree leaf's but as the input variable has dummy values which given mismatch to the classifier. So, the author performed only accuracy test along with confusion matrix. The classification report is used to assemble the results which got from the multi classification algorithms. Decision tree given highest accuracy which is 51% than other ML techniques. In neural network while selecting the number of hidden layers the author faced problems, after multiple runs it found that two hidden layers provided the good accuracy than the machine learning models with 'relu' activation and 32 echoes. The accuracy for this NN is 71.6%. As per the study most of the employees are satisfied with WFH environment and they are having good work-life-balance. The cyber issues are little more because of this security team wants to return the office for delivering the secured products to client.

Chapter 5

Summarized the findings and recommendations from the research objectives are discussed in this chapter. It has been discussed how machine learning's can be used to improve employee satisfaction in the future.

Conclusion and Recommendations

5.1 Conclusion

The goal of doing this research is to find the impact of COVID-19 on IT employees and what are most important factors that effects employee satisfaction. The analysis is conducted through EDA techniques and machine learning models. As per the visualizations results age 35-44 group people are responded positively compared to other age groups. According to CMIE the younger generations lost their job due to the pandemic regulations which is WFH. The author did very interesting study on employee designation analysis. That means the author perspective is to find which type of employees are enjoying the work structure during pandemic and which are struggling to work. The results are pretty reasonable i.e. IT professionals are the second highest count people whose response is more positive. The first positive highest count response is others and they are Human Resource team, testers, business analyst and support developers. Employers are providing more corporate assets than personal assets because of security issues. The records are saying that Laptop is the most used device by employees for work. There are some other devices also present like desktop, tap and smart phone. But these devices are not so in popular use than laptop. The cyber issues are less as they are getting proper awareness about WFH through documentation and by using VPN. Along with these devices organizations are also providing mouse, keyboard and WIFI connection.

While EDA never provides accurate future forecasts, ML algorithms are used to determine the relevance of the feature. By contrasting the accuracy between the five ML algorithms, it is demonstrated that the neural network algorithm is the best algorithm. Support vector machine, decision tree, KNN, and logistic regression are the machine learning models that have been used. For better predictions, the author additionally used an advanced neural network in conjunction with these. The decision tree model's accuracy is 51%, which is a respectable level of accuracy.

With this accuracy, we may conclude that future forecasts will be effective (Gold and Sollich, 2003). Comparing various multi categorization techniques, neural networks have the highest accuracy (71.6%), which is impressive. With the aid of a confusion matrix, the author measured the AUC-ROC curve, recall, precision, and accuracy of the four ML models.

5.2 Limitations

One of the limitations of the study is the minimum amount of secondary data that was obtained. Considering all parameters in the future may produce better outcomes than evaluating only a few columns for training purposes. Because there is not enough time to complete the study, this research uses a cross-sectional time horizon. However, careful consideration of predictions made before and after COVID-19 will produce useful results. Hence, it will be recommended to fulfil expectations in the long term. This study only focused on a few features related to IT workers; further research in other sectors is required. As every restart of the program the train set will takes new reputation of values the accuracy will changes for ML techniques.

5.3 Recommendations and Future work

As per the research, it is found that finding employee satisfaction is very important for the growth of the organization. The entire research is done with secondary data and the dataset got from the third-party repository called data Mendeley. The data set don't have much data and for taking primary data the time for the dissertation is not sufficient. But if the research has time, then going for primary data is very effective. Predicting future outputs need more data for more accurate results. It's simple if the training data is more than predicting the unknown data properties will be effective. The comments field doesn't have much feedback about WFH from employees which lead author to not perform sentimental analysis, but the author successfully performed multi label classification with the health of satisfaction column. Converting string object to float data type given trouble to the author for converting it the author performed pre-processing to get effective models. The entire key to solve the challenges is to take the good data set where all the columns' data values are numerical or convertible strings. Along with machine learning models performing sentimental analysis will also give accurate results when the data is textual comments. By giving more awareness about cyber issues can be useful for avoiding cyber threats.

References

1. Alamgeer, Z. (2022). *Understanding Research Onion for Research Methodology*. [online] The Innovidea. Available at: <https://theinnovidea.com/understanding-research-onion-for-research-methodology/>.
2. Allen, T.D., Golden, T.D. and Shockley, K.M. (2015). How Effective Is Telecommuting? Assessing the Status of Our Scientific Findings. *Psychological Science in the Public Interest*, 16(2), pp.40–68. doi:<https://doi.org/10.1177/1529100615593273>.
3. Armstrong, M. (2021). *2020 was a record year for feeling stressed at work*. [online] World Economic Forum. Available at: <https://www.weforum.org/agenda/2021/12/employees-stress-mental-health-workplace-environment/>.
4. Bae, K.B. and Kim, D. (2016). The Impact of Decoupling of Telework on Job Satisfaction in U.S. Federal Agencies. *The American Review of Public Administration*, 46(3), pp.356–371. doi:<https://doi.org/10.1177/0275074016637183>.
5. Bao, L., Li, T., Xia, X., Zhu, K., Li, H. and Yang, X. (2022). How does working from home affect developer productivity? — A case study of Baidu during the COVID-19 pandemic. *Science China Information Sciences*, 65(4). doi:10.1007/s11432-020-3278-4.
6. Bjärntoft, S., Hallman, D.M., Mathiassen, S.E., Larsson, J. and Jahncke, H. (2020). Occupational and Individual Determinants of Work-life Balance among Office Workers with Flexible Work Arrangements. *International Journal of Environmental Research and Public Health*, [online] 17(4). doi:<https://doi.org/10.3390/ijerph17041418>.
7. Brunelle, E. and Fortin, J.-A. (2021). Distance Makes the Heart Grow Fonder: An Examination of Teleworkers' and Office Workers' Job Satisfaction Through the Lens of Self-Determination Theory. *SAGE Open*, 11(1), p.215824402098551. doi:<https://doi.org/10.1177/2158244020985516>.
8. BusinessWire (2021). *Cyber Threats Have Increased 81% Since Global Pandemic*. [online] www.businesswire.com. Available at: <https://www.businesswire.com/news/home/20211108005775/en/Cyber-Threats-Have-Increased-81-Since-Global-Pandemic>.
9. Chanana, N. and Sangeeta (2020). Employee engagement practices during COVID-19 lockdown. *Journal of Public Affairs*, [online] 21(4). doi:<https://doi.org/10.1002/pa.2508>.

10. Choi, Y. and Choi, J. (2022). Job Satisfaction Prediction and Machine Learning Technique. [online] doi:<https://doi.org/10.21203/rs.3.rs-1683972/v1>.
11. Chung, H. and van der Horst, M. (2017). Women's employment patterns after childbirth and the perceived access to and use of flexitime and teleworking. *Human Relations*, 71(1), pp.47–72. doi:<https://doi.org/10.1177/0018726717713828>.
12. Côté, P., van der Velde, G., Cassidy, J.D., Carroll, L.J., Hogg-Johnson, S., Holm, L.W., Carragee, E.J., Haldeman, S., Nordin, M., Hurwitz, E.L., Guzman, J. and Peloso, P.M. (2008). The Burden and Determinants of Neck Pain in Workers. *Spine*, 33(Supplement), pp.S60–S74. doi:<https://doi.org/10.1097/brs.0b013e3181643ee4>.
13. Cunningham, P. and Delany, S.J. (2021). k-Nearest Neighbour Classifiers - A Tutorial. *ACM Computing Surveys*, 54(6), pp.1–25.
14. DataFlair. (2017). *Artificial Neural Networks for Machine Learning - Every aspect you need to know about*. [online] Available at: <https://data-flair.training/blogs/artificial-neural-networks-for-machine-learning/>.
15. de Lucas Ancillo, A., del Val Núñez, M.T. and Gavrilă, S.G. (2020). Workplace Change within the COVID-19 context: a Grounded Theory Approach. *Economic Research-Ekonomska Istraživanja*, [online] 34(1), pp.1–20. doi:10.1080/1331677x.2020.1862689.
16. Delanoeije, J. and Verbruggen, M. (2020). Between-person and within-person effects of telework: a quasi-field experiment. *European Journal of Work and Organizational Psychology*. [online] Available at: <https://www.semanticscholar.org/paper/Between-person-and-within-person-effects-of-a-Delanoeije-Verbruggen/59b6fbff83dffaadf174d4812ec4038823895c02> [Accessed 15 Mar. 2023].
17. Donthu, N. and Gustafsson, A. (2020). Effects of COVID-19 on Business and Research. *Journal of Business Research*, 117(1), pp.284–289. doi:<https://doi.org/10.1016/j.jbusres.2020.06.008>.
18. ELAC (2022). *Digital technologies for a new future*. [online] Available at: https://www.cepal.org/sites/default/files/publication/files/46817/S2000960_en.pdf
19. Financial Express (2020), “COVID-19 impact: more staff in services sector companies to work from home in future as well”, available at: www.financialexpress.com/industry/covid-19-impact-more-staff-in-services-sector-companies-to-work-from-home-in-future-as-well/1940050/ (accessed 23 July 2020).

20. Gajendran, R.S., Harrison, D.A. and Delaney-Klinger, K. (2014). Are Telecommuters Remotely Good Citizens? Unpacking Telecommuting's Effects on Performance Via I-Deals and Job Resources. *Personnel Psychology*, 68(2), pp.353–393. doi:<https://doi.org/10.1111/peps.12082>.
21. Ghosh, S. (2020), "Working from home: monitoring employee productivity in the times of lockdown", Financial express, available at: www.financialexpress.com/jobs/work-from-home-monitoring-employee-productivity-in-the-times-of-lockdown/1967197/ (accessed 24 July 2020).
22. Gorlick, A. (2020). *The productivity pitfalls of working from home in the age of COVID-19*. [online] Stanford News. Available at: <https://news.stanford.edu/2020/03/30/productivity-pitfalls-working-home-age-covid-19/>.
23. Ijcms.in. (2022). *View of Effect of Occupational Stress and Remote Working on Psychological Well-Being of Employees: An Empirical Analysis during Covid- 19 Pandemic Concerning Information Technology Industry in Hyderabad*. [online] Available at: <https://www.ijcms.in/index.php/ijcms/article/view/304/287>.
24. Koslowsky, M. and Krausz, M. eds., 2002. *Voluntary employee withdrawal and inattendance: A current perspective*. Springer Science & Business Media.
25. Kurdi, B., Alshurideh, M. and Alnaser, A. (2020). The impact of employee satisfaction on customer satisfaction: Theoretical and empirical underpinning. *Management Science Letters*, [online] 10(15), pp.3561–3570. Available at: <http://m.growingscience.com/beta/msl/4068-the-impact-of-employee-satisfaction-%20on-customer-satisfaction-theoretical-and-empirical-underpinning.html>.
26. Lam, H.K., Ekong, U., Liu, H., Xiao, B., Araujo, H., Ling, S.H. and Chan, K.Y. (2014). A study of neural-network-based classifiers for material classification. *Neurocomputing*, [online] 144, pp.367–377. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0925231214006274> [Accessed 21 Mar. 2023].
27. Martin, B.H. and MacDonnell, R. (2012). Is telework effective for organizations? : a meta-analysis of empirical research on perceptions of telework and organizational outcomes. *Management research review*, [online] 35(7). Available at: <https://www.econbiz.de/Record/telework-effective-for-organizations-meta-analysis->

[empirical-research-perceptions-telework-and-organizational-outcomes-martin-brittany-harker/10009573311](https://www.researchgate.net/publication/3545496_Impacts_of_Home-Based_Telework_on_Quality_of_Life_for_Employees_and_Their_Partners_Quantitative_and_Qualitative_Results_From_a_European_Survey).

28. Mitta, S. (2020), “*Work from home has been successful during COVID-19 lockdown. What next?*”, available at: <https://economictimes.indiatimes.com/magazines/panache/work-from-home-has-been-successful-during-covid-19-lockdown-what-next/articleshow/75470580.cms> (accessed 23 July 2020).
29. Nakrošienė, A., Bučiūnienė, I. and Goštautaitė, B. (2019). Working from home: characteristics and outcomes of telework. *International Journal of Manpower*, [online] 40(1), pp.87–101. doi:<https://doi.org/10.1108/ijm-07-2017-0172>.
30. Nakrošienė, A., Bučiūnienė, I. and Goštautaitė, B. (2019). Working from home: characteristics and outcomes of telework. *International Journal of Manpower*, [online] 40(1), pp.87–101. doi:<https://doi.org/10.1108/ijm-07-2017-0172>.
31. Neuraxio. (n.d.). *The Business Process of Machine Learning, with AutoML*. [online] Available at: <https://www.neuraxio.com/blogs/news/the-business-process-of-machine-learning-with-automl> [Accessed 27 Mar. 2023].
32. Ng, Q.X., Chee, K.T., De Deyn, M.L.Z.Q. and Chua, Z. (2020). Staying connected during the COVID-19 pandemic. *International Journal of Social Psychiatry*, 66(5), pp.519–520. doi:<https://doi.org/10.1177/0020764020926562>.
33. O’Hara, C. (2020), “*5 Ways to work from home more effectively*”, Harvard business review, available at: <https://hbr.org/2014/10/5-ways-to-work-from-home-more-effectively> (accessed 23 July 2020).
34. Pillastrini, P., Mugnai, R., Bertozzi, L., Costi, S., Curti, S., Guccione, A., Mattioli, S. and Violante, F.S. (2010). Effectiveness of an ergonomic intervention on work-related posture and low back pain in video display terminal operators: A 3 year cross-over trial. *Applied Ergonomics*, 41(3), pp.436–443. doi:<https://doi.org/10.1016/j.apergo.2009.09.008>.
35. ResearchGate. (n.d.). (PDF) *Impacts of Home-Based Telework on Quality of Life for Employees and Their Partners. Quantitative and Qualitative Results From a European Survey*. [online] Available at: https://www.researchgate.net/publication/23545496_Impacts_of_Home-Based_Telework_on_Quality_of_Life_for_Employees_and_Their_Partners_Quantitative_and_Qualitative_Results_From_a_European_Survey.

36. Ripley, B.D. (1994). Neural Networks and Related Methods for Classification. *Journal of the Royal Statistical Society: Series B (Methodological)*, 56(3), pp.409–437.
37. Russo, D., Hanel, P.H.P., Altnickel, S. and van Berkel, N. (2021). The Daily Life of Software Engineers During the COVID-19 Pandemic. *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. doi:10.1109/icse-seip52600.2021.00048.
38. Russo, D., Hanel, P.H.P., Altnickel, S. and van Berkel, N. (2021). The Daily Life of Software Engineers During the COVID-19 Pandemic. *2021 IEEE/ACM 43rd International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. doi:10.1109/icse-seip52600.2021.00048.
39. Sandoval-Reyes, J., Idrovo-Carlier, S. and Duque-Oliva, E.J. (2021). Remote Work, Work Stress, and Work–Life during Pandemic Times: A Latin America Situation. *International Journal of Environmental Research and Public Health*, [online] 18(13), p.7069. doi:<https://doi.org/10.3390/ijerph18137069>.
40. Sandoval-Reyes, J., Idrovo-Carlier, S. and Duque-Oliva, E.J. (2021). Remote Work, Work Stress, and Work–Life during Pandemic Times: A Latin America Situation. *International Journal of Environmental Research and Public Health*, [online] 18(13), p.7069. doi:<https://doi.org/10.3390/ijerph18137069>.
41. Sandoval-Reyes, J., Idrovo-Carlier, S. and Duque-Oliva, E.J. (2021). Remote Work, Work Stress, and Work–Life during Pandemic Times: A Latin America Situation. *International Journal of Environmental Research and Public Health*, [online] 18(13), p.7069. doi:<https://doi.org/10.3390/ijerph18137069>.
42. Sandoval-Reyes, J., Idrovo-Carlier, S. and Duque-Oliva, E.J. (2021). Remote Work, Work Stress, and Work–Life during Pandemic Times: A Latin America Situation. *International Journal of Environmental Research and Public Health*, [online] 18(13), p.7069. doi:<https://doi.org/10.3390/ijerph18137069>.
43. Satyanarayana, N., Ramalingaswamy, C. and Ramadevi, Y. (2014). Survey of Classification Techniques in Data Mining. *IJISSET -International Journal of Innovative Science, Engineering & Technology*, [online] 1(9). Available at: https://www.ijiset.com/v1s9/IJISSET_V1_I9_42.pdf [Accessed 27 Apr. 2022].

44. Shamir, B. and Salomon, I. (1985), “Work-at-home and the quality of working life”, *Academy of Management Review*, Vol. 10 No. 3, pp. 455-464.
45. Sharma, Y.S. (2020), “27 Million youth in age group of 20-30 years lost jobs in April: CMIE”, available at: <https://economictimes.indiatimes.com/news/economy/indicators/unemployment-rate-dips-to-23-97-data-from-cmie-shows/articleshow/75689370.cms?from=mdr> (accessed 23 July 2020).
46. Song, Y. and Gao, J. (2020). Does Telework Stress Employees Out? A Study on Working at Home and Subjective Well-Being for Wage/Salary Workers. *Journal of Happiness Studies*, [online] 21(7), pp.2649–2668. Available at: https://ideas.repec.org/a/spr/jhappi/v21y2020i7d10.1007_s10902-019-00196-6.html [Accessed 12 Mar. 2023].
47. Sullivan, C. (2012). Remote Working and Work-Life Balance. *Work and Quality of Life*, [online] pp.275–290. doi:https://doi.org/10.1007/978-94-007-4059-4_15.
48. www.ilo.org. (2021). *Analysis of the activities of employer and business membership organizations in the COVID-19 pandemic and what comes next*. [online] Available at: https://www.ilo.org/actemp/publications/WCMS_766658/lang--en/index.htm.
49. www.proquest.com. (n.d.). *The Relationship Between Remote Work and Job Satisfaction: The Mediating Roles of Perceived Autonomy, Work-family Conflict, and Telecommuting Intensity* - ProQuest. [online] Available at: <https://www.proquest.com/openview/2d31cf88ac3d494d612b66c7fa12e0e2/1?pq-origsite=gscholar&cbl=18750&diss=y>.
50. www.thecdi.net. (n.d.). *River*. [online] Available at: https://www.thecdi.net/write/Documents/Wales%20Forum/Sandi_Mann_Holdworth.pdf.
51. Yu, J. and Wu, Y. (2021). The Impact of Enforced Working from Home on Employee Job Satisfaction during COVID-19: An Event System Perspective. *International Journal of Environmental Research and Public Health*, [online] 18(24), p.13207. doi:<https://doi.org/10.3390/ijerph182413207>.

Appendix

GitHub link for all the required files link given below

<https://github.com/vyshnavimuthumula/Dissertation>

Original data shown below

ID	Ώρα έναρξης	Ώρα ολοκλήρωσης	Prior to the COVID	Did you receive any	Please describe the	What kind of device	Are these devices a	These devices are:	Are these
5	4-7-20 16:53:59	4-7-20 16:56:10	Yes	No		Laptop;	No	Corporate assets	Yes, partly
6	4-7-20 17:37:30	4-7-20 17:41:50	Yes	No		Laptop;	No	Corporate assets	No
7	4-8-20 14:16:24	4-8-20 14:19:31	Yes	No		Laptop;Smartphone;	No	Personal assets	
8	4-8-20 14:17:51	4-8-20 14:19:55	Yes	Yes		Laptop;	No	Corporate assets	No
9	4-8-20 14:16:45	4-8-20 14:20:02	Yes	Yes	Use of VPN	Laptop;Smartphone;	No	Corporate assets	No
10	4-8-20 14:18:12	4-8-20 14:20:52	Yes	No		Laptop;	No	Corporate assets	Yes, fully
11	4-8-20 14:20:15	4-8-20 14:22:51	Yes	No		Laptop;	Yes	Personal assets	
12	4-8-20 14:30:52	4-8-20 14:33:59	Yes	No		Laptop;Smartphone;	No	Corporate assets	Yes, partly
13	4-8-20 14:31:03	4-8-20 14:35:01	Yes	No		Laptop;Smartphone;	Yes	Corporate assets	No
14	4-8-20 14:45:53	4-8-20 14:49:32	Yes	No		Laptop;	No	Personal assets	
15	4-8-20 14:46:29	4-8-20 14:49:59	Yes	No		Laptop;	No	Corporate assets	Yes, partly
16	4-8-20 14:49:40	4-8-20 14:53:46	Yes	No		Laptop;	No	Personal assets	
17	4-8-20 14:47:51	4-8-20 14:53:52	Yes	No		Laptop;Smartphone;	Yes	Personal assets	
18	4-8-20 14:57:49	4-8-20 15:00:28	Yes	No		Laptop;Smartphone;	No	Corporate assets	Yes, partly
19	4-8-20 15:04:23	4-8-20 15:06:47	Yes	No		Laptop;	No	Personal assets	
20	4-8-20 14:55:12	4-8-20 15:11:04	Yes	Yes		Laptop;	No	Corporate assets	Yes, partly
21	4-8-20 14:46:32	4-8-20 15:11:28	Yes	No		Laptop;	No	Personal assets	
22	4-8-20 15:16:36	4-8-20 15:19:12	No	No		Laptop;	No	Corporate assets	Yes, partly
23	4-8-20 15:17:27	4-8-20 15:22:30	No	No		Laptop;	No	Corporate assets	Yes, fully
24	4-8-20 15:21:58	4-8-20 15:23:41	No	No		Desktop;	Yes	Corporate assets	No
25	4-8-20 15:23:47	4-8-20 15:26:13	Yes	Yes	use passwords on com	Laptop;	Yes	Personal assets	
26	4-8-20 15:27:03	4-8-20 15:32:12	Yes	No		Laptop;	No	Personal assets	

Figure 23: Scrapped data

Printing the first 5 rows of the data after data cleaning shown below

In [74]:

```
# # Printing first 5 rows by using head funtion
WFH_COVID_crisis_df.head()
```

Out[74]:

	ID	Start time	Completion time	Earlierwfh	Isguidelines	Guidelinesgiven	Devices	Useofdevicesbyothers	Assets	Devicesmanaged	...	Satisfaction
0	5	2020-07-04 16:53:59	2020-07-04 16:56:10	Yes	No	No	Laptop;	No	Corporate assets	Yes, partly	...	Agree
1	6	2020-07-04 17:37:30	2020-07-04 17:41:50	Yes	No	No	Laptop;	No	Corporate assets	No	...	Agree
2	7	2020-08-04 14:16:24	2020-08-04 14:19:31	Yes	No	No	Laptop;Smartphone;	No	Personal assets	No	...	Strongly agree
3	8	2020-08-04 14:17:51	2020-08-04 14:19:55	Yes	Yes	No	Laptop;	No	Corporate assets	No	...	Agree
4	9	2020-08-04 14:16:45	2020-08-04 14:20:02	Yes	Yes	Use of VPN	Laptop;Smartphone;	No	Corporate assets	No	...	Strongly agree

5 rows × 32 columns

Figure 24: Cleaned data

Based on few comments the author performed word count which shown below

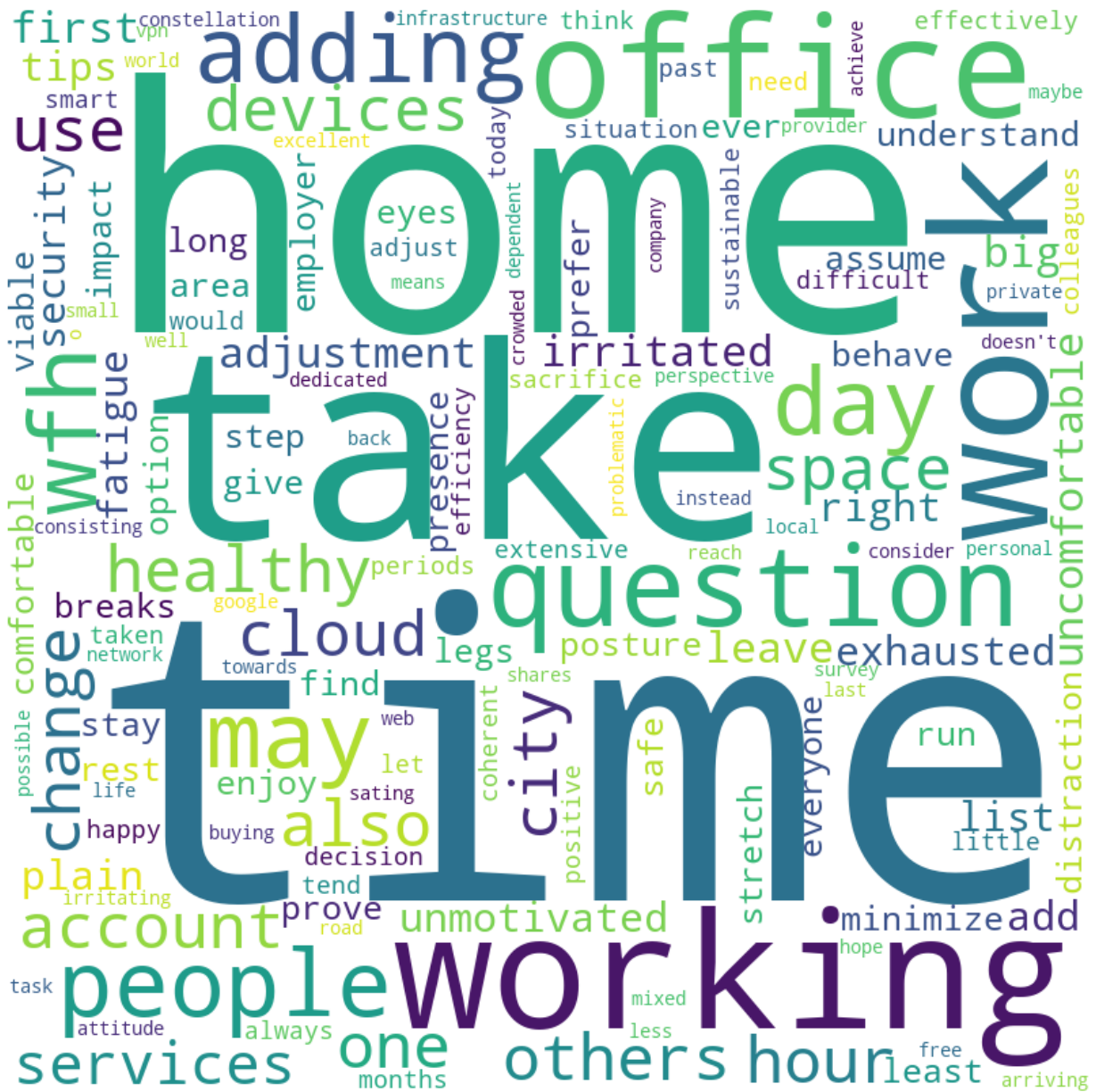


Figure 25: Word count