

Major Project Report
on
Impact of Work Life Balance on Employee
Turnover using Machine Learning Techniques

Submitted By :

A Sukrithi

2k18/MBA/724

Akruti Rawat

2k18/MBA/739

Under the Guidance of :

Ms. Kusum Lata

Assistant Professor



UNIVERSITY SCHOOL OF MANAGEMENT

& ENTREPRENEURSHIP,

East Delhi Campus,

Delhi Technological University

MAY 2020

CERTIFICATE

This is to certify that project report entitled “**Impact of Work Life Balance on Employee Turnover using Machine Learning Techniques**” which is submitted by **Akruti Rawat and A Sukrithi** in the partial fulfilment for the award of Masters of Business Administration degree of University School Of Management & Enterpreneurship, East Delhi Campus, Delhi Technological University, Delhi is the record of the candidates own work carried out by them. The matter embodied in this thesis is original and has not been submitted for the award of any degree.

Date: May 2020

DECLARATION

We, Akruti Rawat and A Sukrithi, hereby declare that the project work on the topic “**Impact of Work Life Balance on Employee Turnover using Machine Learning Techniques**“ completed at Delhi Technological University and submitted under the guidance and supervision of Ms. Kusum Lata, Assistant Professor, University School of Management and Entrepreneurship, East Delhi Campus, Delhi Technological University, is original work carried out by us during the period Jan 2020 to May 2020. The preparation of our project report is based on our personal findings, interaction with the employees of different organizations and secondary sources.

Date.....

A Sukrithi
2K18/MBA/724

Akruti Rawat
2K18/MBA/739

ACKNOWLEDGEMENT

We wish to express our gratitude to all those who have been instrumental in preparation of our report. Our work would not have been completed without their support and guidance. First of all we would like to express our gratitude toward Delhi Technological University for allowing us to undertake this research work. This research work allowed us to explore the some of the HR issues organizations are facing and how machine learning can be used to solve these issues.

We are deeply indebted to Ms. Kusum Lata, Assistant Professor, University School of Management and Entrepreneurship, East Delhi Campus, Delhi Technological University, for her valuable and enlightened guidance. We would have not been able to execute this project without her constant support, advice and encouragement.

Also we owe special thanks to all the respondents whom we approached with our questionnaire and who took time off their busy schedules to answer them patiently.

A Sukrithi

Akruti Rawat

ABSTRACT

Employee turnover is one of the most significant problems. The solution of this problem is that organizations are heavily investing their time and cost into machine learning techniques in order to predict the employee's turnover. This study was conducted to establish an impact of employee's work life balance on employee turnover. With the advancement in technology, several machine learning algorithms are there along with the respective platform tools available in the market. We used the WEKA Machine Learning Tool for our project.

In this project we are understanding which variables are most influential in predicting employee turnover. Job satisfaction is our dependent variable and several independent variables are also chosen. The sample of our study was 124 in total. Most of the respondents were from the IT industry. Primary data was collected through the structural questionnaires. The research findings say that most of the IT industries have adopted several work life balance practices in their organizations. These practices include flexible working hours, childcare facilities, casual leave facility , rotational shifts, work from home facility . It was revealed that a strong relationship occurs between work life balance practices relating to Work from home facility and turnover intent respectively.

TABLE OF CONTENTS

S.NO.	Topics	Page
1	Chapter 1- Introduction	1
2	Chapter 2- Literature Review	3
3	Chapter 3- Research Methodology	10
4	Chapter 4- Results	26
5	Chapter 5- Findings and Recommendations	29
6	Chapter 6- Limitations of the study	30
7	Bibliography/References	31
8	Annexure	32
9	Plagiarism	38

CHAPTER 1 – INTRODUCTION

Organizations today invest a huge amount in their human resources. Human resources are treated as a true asset to an organization. This involves huge investment in hiring the right candidate, their induction, continuous training and development programs to enhance their skill and knowledge, so when an employee leaves an organization it is a big loss to them. Organizations have realised that to achieve the organizational goals it is not only important to source the right candidate, it is equally important to retain these employees. The reason is that HR professionals have heavily invested both their time and money in training and grooming these candidates so that they can work effectively and efficiently. So if these employees leave it can lead to bigger losses to the organization. The organization along with all the upper management staff need to be concerned with the fact that what will be the return on investment(ROI) of the employees. Hence this makes it important to study about various factors that lead to employee turnover.

Nowadays Machine Learning is a buzzword, especially in the HR industry as gone are the days when HR used to do monotonous and repetitive work all by themselves. With the advent of technological advancements taking place, the HR industry is having great adoption of the machine learning techniques. So, our study has the involvement of several machine learning algorithms which help us in predicting our results effectively. Machine Learning is defined as an application of Artificial Intelligence. It basically allows the system to automatically learn and improve the solutions itself based on experience learning. Here, systems can learn from data, from the patterns and make decisions accordingly with less or minimal involvement of the human beings.

1.1 Industry Profile

Most of our respondents were from the IT industry out of the total number of respondents. IT industry has been one of the top most contributors in India for the economy growth. It has played a key role in putting forward the best of India globally. The IT industry has contributed up to much extent in India's GDP growth. It provides several opportunities for(FDI) Foreign Direct Investment. The IT industry is undergoing rapid changes or evolution which is shaping the world's business standards. The factors which heavily influence the global market. Also, the employment opportunities are created for the youth of the country with the growth of the computers

and the IT industry. It will give a boost to the people who are unemployed. The top IT companies in India are Infosys, Wipro, Genpact, TCS, Google, IBM, Accenture, Capgemini, HCL Technologies etc. The services provided by them are software development, telecommunication, e-commerce, BPO, KPO, mobile commerce which helps in the development of the economy. IT hubs in India are Bangalore often called Silicon Valley Of India, Chennai, Pune, Hyderabad, NCR Delhi Noida Gurgaon. The future of IT industry is very bright due to the fact that with the advancement in the upcoming technologies like Artificial Intelligence, Cloud computing, Machine Learning, Search Engine Optimization, Blockchain, Internet of Things (IOT), big data, it will be on peak. Various government schemes like Make in India are helping the country to create more jobs. All the other industries like the hotel industry, airline industry, pharmaceutical industry etc. start making use of the Information industry to automate and speed up the growth of their sector. Maximum number of responses were from the IT industry in our research project.

1.2 Objectives of the study

1. To study the impact of employee's work life balance on employee turnover using machine learning.
2. To find various independent and dependent variables.
3. To evaluate which independent factors have the most influence on our dependent variable.
4. To build employee turnover predicting models using the most influencing variables.
5. To compare and evaluate each of the model developed and based on this evaluation select the best model suited for predicting employee turnover.

CHAPTER 2 - LITERATURE REVIEW

2.1 Employee Turnover

Employee turnover can be defined as the number of employees leaving an organization and this vacant position being replaced by new employees. This parting of employees from the organization can happen due to retirement, resignation, transfers, termination, death. Turnover can be defined as rotation of staff around the job market; between various organizations and occupations. It can also be defined as the ratio of the number of employees who have resigned/have been terminated during a certain period considered to the total number of employees in that firm during that period. If an organization has a high turnover rate, the average tenure of an employee is really low. High turnover rate can impact the organizations negatively since organizations will lose their skilled employees and will be left with a more inexperienced workforce. This can lead to lower productivity, lower outputs which will lead to lower client/customer satisfaction. Hence will reduce overall revenues of the organization.

2.2 Types of Employee Turnover

There are mainly two types of employee turnover :

a. Voluntary

An employee voluntarily exits an organization due to personal reasons, better pay packages, better job roles, if they are unsatisfied with their current job in terms of their workload, reward and recognition, work life balance, future growth prospects.

Organizations today have started conducting exit interviews, surveys to know why their employees are leaving. This information is being used to properly analyse the needs of the employees and give them a more healthy and happy environment. A happy employee is said to stay with an organization which can in turn be beneficial to the organization.

b. Involuntary

Involuntary turnover happens when the employer himself discharges employees due to poor performance of the employee, lack of enough revenue so he resorts to downsizing to help the organization function for some other reasons.

Most of the organizations try to avoid downsizing unless absolutely necessary. Also organizations today are heavily investing in training programmes to give their employees a chance of improvement or giving upgrading their skill to the latest technology in the market. When employees feel that they are given enough opportunities for growth they tend to stick with the organization for long. Also this helps in avoiding to discharge employees whose performance has been improved due to relevant training.

Employees are the most valuable assets in running any business.

2.3 Employee turnover factors

There can be a lot of reasons for employee turnover like their age, marital status, nature of job, number of years spent in an organization, reward and recognition, incentives, work environment, annual pay package so on. One of the important factors that helps in predicting the turnover rate is employee's work life balance. If an employee is satisfied with his job, there are less chances that he will leave the organization. On-going through various research papers and by talking to some of the employees we decided on some independent variable to conduct our research. Some factors like whether the employee has unrealistic deadlines, heavy workload, long working hours can make the employee exit the organization.

Hence organizations must focus on their employee wellbeing, positive work atmosphere, their needs, should suitably reward them whenever they perform good, conduct trainings so that they feel they are having personal growth along with professional growth, look after their families well beings, give them sufficient time off from their duties so that they have a good work life balance. All these positive changes can in turn help organizations retain their most valuable employees for long.

2.4 Effects of Employee turnover

High turnover can impact an organization negatively. It can also lead to a negative impact on the employees of the organization. When there is a high turnover rate, organizations will have to focus on hiring and training new employees. This will result in organizations not being able to achieve their goals. When employees stay with an organization for long, it gives the organization more efficient workforce which in turn benefits the organization.

a. Revenue

When an employee leaves an organization, the organization again has to rehire a new employee as a replacement. This process involves advertising cost, travel expenses, interview scheduling cost, induction cost, training costs. Even if an external agency is appointed to hire a new employee, it becomes expensive for an organization to replace a candidate. All this money that is being spent on replacement will impact the overall revenue of the organization. There is less profitability if the employee turnover is high.

b. Employee Morale

High turnover rate can result in increased workload on the employees. There will be high pressure on the new recruits to learn quickly and adapt to the job role. High level of output will be expected from the employees. All these issues will lead to a negative work environment which will make it difficult to keep employees motivated and retain them.

c. Output

High turnover will also result in poor quality of product or service. Organization will be left with more inexperienced employees. In some organizations the nature of jobs is repetitive, for example the manufacturing industries. If an employee leaves such manufacturing units it will affect the smooth functioning of their daily operations and result in poor quality of products. Also service industries like airlines, hotel invest a lot in their training programs to provide better services to their customers than their competitors. Employee turnover can affect the kind of services being offered by these

industries, which will result in low customer satisfaction and hence loss in revenue and profits. Losing customers due to low product or service quality will increase the cost one-time customer since the company will have lower return customers. This will decrease the overall return on investment of the organization.

2.5 Relationship between Work Life Balance and Employee Turnover

Work Life balance is defined as the possibility of the integration of the job related tasks along with the family life of the employee within the organization wherever he/she is working. On the other hand, turnover intent can be described as the number of employees who themselves are leaving the organization or it may be possible that the organization themselves are removing the employees from the organization. Turnover Intent can be either voluntary or non-voluntary in nature. Work Life balance is considered as the phenomenon where both the parties i.e. employees as well as their organization get benefit in work related tasks as well as their own life events.

The research shows that support and motivation provided by the managers as well as the supervisors of the company is having an inverse relationship over turnover intent . Work Life balance also affects the productivity of the employee. Work Life Balance practices if correctly and accurately implemented within the organization give a positive impact on the overall productivity of the employee, they will definitely give better and improved results as they are working with a fresh mind.

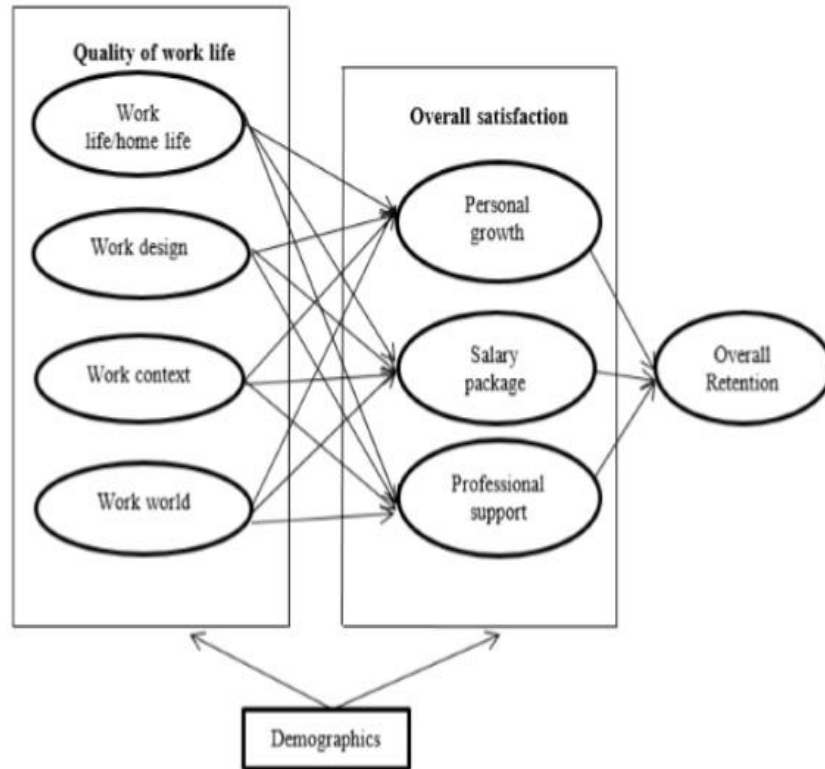


Figure 2.1: Diagrammatic representation of work life balance, job satisfaction and employee retention.

2.6 Strategies to minimize Employee Attrition

Employee turnover can be defined as the number of people leaving the organization. This major problem is something that every business faces with workers' experiences. Whenever an employee leaves it is obviously very costly and a lot of investment in time is there, to find the replacement of that employee within the organization. Hence, it is best for the companies to reduce their turnover rate as much as possible.

First of all, we should hire the right kind of people within the organization who seems to be the right fit , job fit and culture fit for the organization. We should hire the employees who have strong competencies related to the position for which they have applied for.

Secondly, the company should offer competitive pay and benefits to its employees as employees want to be compensated well.

Thirdly, our employees need encouragement and motivation for their hard work . Motivation can be either monetary or non-monetary. When employees do something good, and right they should be appreciated well for their efforts. This would boost their

confidence and they will feel that they are a valuable asset of the company. It will definitely increase their productivity as well as the turnover intent will be less as well. The goal here should be to create a positive and motivating work environment.

2.7 Work Life Balance and Job Satisfaction

Work Life balance has become one of the important aspects for employee wellbeing. Organizations work for maximising the profit and achieve their organizational goals. For the maximization of the profit , the companies need to focus on the efficient and effective management of the overall system organization. Human Resources play a very important role in this aspect. Job satisfaction and employee's work life balance are one of the two factors on which the growth of the organization depends. Due to the increase in the complexity of the work, asking for more efficient and effective work tasks, and demands for more specialisation, there is increased popularity of work life balance within the companies. Work Life balance is all about prioritizing our work tasks given to us and our life events or lifestyle.

2.8 Impact of Machine Learning

Nowadays the HR department is completely moving towards analytics. With the help of machine learning we can get accurate results , workflow automation will be there, the HR professionals will be able to attract the top talent, moreover employee engagement can be measured as well followed by the attrition detection. It is data driven. Machine learning is entirely defined as objective in nature but it definitely depends upon the quality of data which we are using for our research.

There are supervised machine learning algorithms through which the computers can learn from the historical , and labelled datasets. With the advancement in the field of information technology, researchers have researched and studied various machine learning approaches to improve the outcomes of human resources.

Employees today spend the maximum of their time in the workplace for completing their tasks related to work than ever before. Being overloaded by such hectic schedule can lead to great amount of stress and can cause an unbalance situation for them and their families. Having unbalanced life can cause poor performance and ineffective performance in life both at the workplace as well as their home. So , many of the

organizations now follow and value the ideation of work life balance and consider it to be one of the most critical criteria for the success of their organization.

2.9 Consequences of Poor Work life balance

If there will be no work life balance followed by the organizations then obviously there will be conflicts, stress, less productivity, inefficient work. It includes

- a.** Increased level of stress
- b.** Low life satisfaction
- c.** Growth in the rate of problems related to supervision of the children

If this continues to happen then there will be high consequences of work life balance including:

- a.** High absenteeism rate of employees
- b.** Lower productivity
- c.** Low job satisfaction

CHAPTER 3 - RESEARCH METHODOLOGY

3.1 Data Collection

The data for the research has been collected from various employees working in the IT sector in India. We have used a number of variables that could have an impact on the employee turnover. Based on these variables we defined our dependent and independent variables. The independent variables used are age, gender, Flexible working hours, casual leaves facility, rotational shift, childcare facilities, sabbatical leave, flexible working hours, unachievable deadlines, heavy workload, work from home. The interpretation of each independent variable is as follows:

Table 3.1: Independent variables used in the study

S No.	Variables	Data Description
1	Age	To check what is the age group that is most likely to leave the organisation soon
2	Gender	To check whether males/ females are most likely to leave the organisation
3	Education Qualification	To check if the education qualification have any impact on employee turnover
4	Current Job Duration	To check if the number of years spent in current job has any significant impact on the turnover
5	Changed Job in 3 years	To check if the individual has switched his job in the last 3 years and is likely to switch now or not?
6	Flexible Working Hours	To check the impact of giving the employees flexibility to work as per their convenient time on employee turnover

7	Teleworking Allowed	To check if the employee is allowed to work from a location other than the office will have an impact over the turnover rate or not.
8	Rewards and Recognition	To see if employee rewards and recognition helps in making the employee stay with the company for long
9	Casual Leave Facility	To check the impact whether casual leaves have an impact on employee turnover
10	Rotational Shift	To check the impact of shift timings on the turnover rate
11	Compensation Leave	To check how being granted compensation off leave to employees have any impact on their turnover
12	Sabbatical Leave	To check if being granted sabbatical leave by the employer has an impact on employee turnover
13	Child Care Facilities	To evaluate if the special facilities provided by an organization like creche within the premises would make the employee stay with the organization
14	Work from Home	To check if the employee is given full time work from home, will he stay with the organization or not?
15	Insufficient Breaks	To see if the employee is getting enough breaks at work and its impact on their turnover
16	Supportive Feedback	To Check if employees are getting a supportive environment and feedbacks from their managers and its effect on their turnover
17	Unachievable Deadlines	To check if the deadlines for completion of task given to an employee have a direct effect on employee turnover

18	Heavy Workload	To check the impact of heavy workload on employee turnover
----	----------------	--

The dependent variable that we took was whether the job satisfaction of that individual is high or low. If the Job satisfaction is low it is likely that the employee will leave the organization.

Once the variables were defined, a questionnaire was prepared to collect the data from the population. We also interviewed a few respondents to have a better understanding about the subject.

For sampling the data we have used the **Convenience sampling technique**. We chose convenience sampling because it allowed us to take samples from that part of the population that was easily approachable.

3.2 Data Processing

Before developing the employee turnover prediction model we pre-processed the data. The main purpose of pre-processing the data is to improve the quality of the data. By improving the quality of data we get better results since it reduces any noise, inconsistency, redundancy, incomplete data.

We pre-processed the data using the following steps:

- a. Removed null values.
- b. Removed the columns that were unnecessary.
- c. Arranged columns as per the weightage of each of the column.
- d. Saved this data as CSV file format.

3.3 Developing the Model

In Machine Learning , classification has two meanings , firstly when the set of observations are given With the aim of establishing the existence of clusters in data and hence is called unsupervised learning or the second type be when there is a need to accomplish a task by providing training and the input patterns and the output patterns that is called as supervised learning.

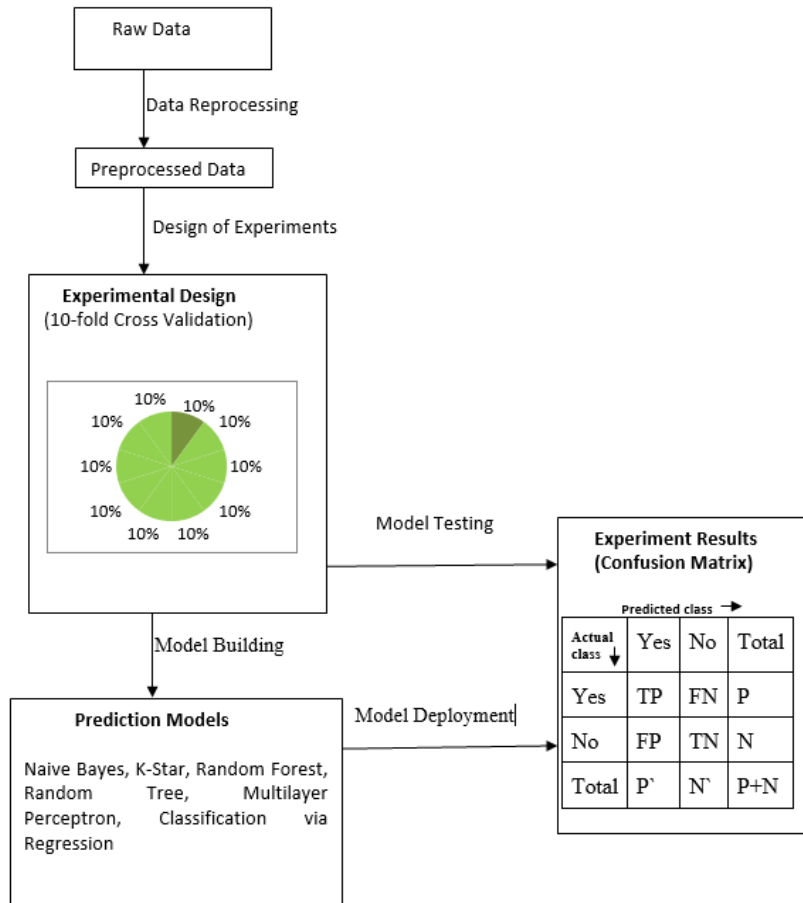


Fig 3.1: Process of building model using Machine learning Technique

Accordingly, this study consists of six popular classification methods (i.e., **Random Forest, Random Tree, K star, classification via regression and multilayer perceptron, Naive Bayes**) along with three ensemble techniques (i.e., bagging, busting and information fusion) are built and compared to each other using their predictive accuracy on the hold-out samples. A huge number of researches compare data mining techniques in different ways. So, the following machine learning techniques are used in our research methodology.

a. Naive Bayes

Naive Bayes is known for its simplicity which the algorithm holds in itself as compared to others. It is Bayes Classification methods. Bayesian classifiers are the statistical classifiers. It is based on probabilities. Here, X = data tuple/ evidence, then H = Hypothesis , $P(H|X) = P(H \text{ belongs to class } X)$, so

The Bayes Theorem is

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)}$$

b. Random Forest

The Random Forest algorithm is one of the popular tree based algorithms. It falls under the category of supervised learning algorithm. Basically, as the name suggests it randomly creates and then merges multiple (decision) trees in one forest. The type of 'ensembling' used in the Random Forest algorithm is known as bagging. In bagging, successive trees do not depend on the earlier trees as each is constructed independent of each other by using a different bootstrap sample of their data set. The last step includes that , majority vote is taken for the final result or prediction.

c. Random Tree

Random Tree algorithm falls under the category of supervised learning. A group of regression and classification trees are used.

d. K Star

It is an algorithm of instance based classifiers. It is different as compared to other types of algorithms as it uses entropy based distance function. It is used to find the K shortest paths between the given set of directed weighted graphs and the desired pair of vertices. It's advantage over other algorithms is that it does not require the graph to be stored in the main memory. Here the class of the test instance is based on the class of those training instances similar to it, as determined by some similarity function.

e. Classification Via Regression

Class is used for doing classification using the regression methods. Binarized class is used and one regression model is built for each class value.

f. Multilayer Perceptron

Backpropagation is used in such kinds of classifiers to learn the multi-layer perceptron for instances to classify.

The network can be built in two ways i.e. either by hand or by setting up or using a (simple) heuristic. Modification as well as the networks parameters are to be monitored during the training time. The nodes in this network are all sigmoid functions.

3.1 Evaluating each model in comparison to each other and Performance Measures

10 fold cross validation - Normally , in K fold cross validation technique, the data set is divided into the k parts wherein the k parts are used for training the model and k-1 parts are being used for the validation purpose. Then , the same method is iterated for k number of times.

So, in the Weka tool(Machine learning platform), we used 10 fold cross validation. Cross validation is a method to find the models predicted. This is done by dividing the sample originally collected into a set of data used to train the model and hence further evaluation is done. It is also called as rotation estimation. So, here we used $K=10$, hence it is 10 fold cross validation.

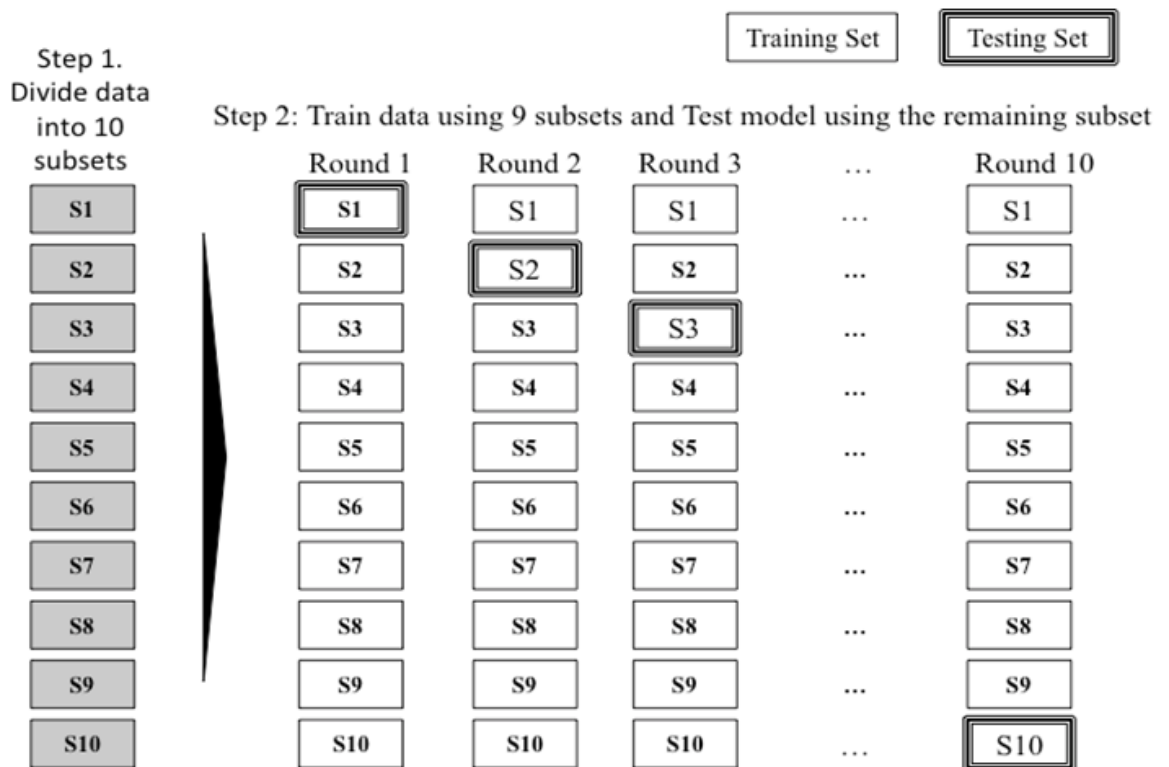


Fig 3.2: Ten - cross fold validation

Once the independent and dependent variables were decided we had to build the model and calculate accuracy, sensitivity, specificity and precision for each model. Using these four parameters we had to compare each of these models and select the model with most accurate results.

The formulas for the given performance measures are :

Table 3.2: Confusion matrix

		Predicted Class		
		Yes	No	Total
Actual Class	Yes	TP	FN	P
	No	FP	TN	N
	Total	P'	N'	P+N

a. Confusion Matrix

Consider P as number of positive occurrences and N as number of negative occurrences available to us under certain conditions , so the outcomes are depicted in a 2X2 matrix. It is called as contingency matrix. This matrix is helpful in evaluation of the performances of the models predicted respectively. This table which consists of 2 rows and 2 columns. The rows are called actual classes and the columns are called a predicted class as shown in the diagram above.

- b. True Positive(TP): It can be defined as the number of positive instances that can be predicted correctly.
- c. True Negative(TN): It can be defined as the number of positive instances that could not be predicted correctly.
- d. False Positive(FP): It can be defined as the number of negative instances that have been predicted incorrectly.
- e. False Negative(FN): It can be defined as the number of negative instances that have been predicted correctly.

- f. Accuracy: Accuracy is measured as the ratio of the number of accurately classified classes to the total given number of classes.

$$\text{Accuracy} = \left(\frac{TP+TN}{P+N} \right) \times 100$$

- g. Sensitivity: Sensitivity can be measured as the ratio of positive instances that have been correctly classified to the total number of actual positive instances.

$$\text{Sensitivity} = \left(\frac{TP}{P} \right) \times 100$$

- h. Specificity: Specificity can be defined as the ratio of negative instances that have been correctly classified to the total number of actual negative instances.

$$\text{Specificity} = \left(\frac{TN}{N} \right) \times 100$$

- i. Precision: Precision can be defined as the ratio of actual positives instances which are correctly predicted to the total number of instances that are predicted positive.

$$\text{Precision} = \left(\frac{TP}{TP+FP} \right) \times 100$$

Using the above defined variables and the four parameters we obtained the following results:

- a. Naive Bayes

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
39 11 |  a = Low  
18 56 |  b = High
```

Fig 3.3 : Naive Bayes confusion matrix

Table 3.3: Naive Bayes performance parameters

Accuracy:	0.767
Sensitivity:	0.78
Specificity:	0.757
Precision:	0.684

b. Random Forest

=== Confusion Matrix ===

```
a  b  <-- classified as
38 12 | a = Low
 7 67 | b = High
```

Fig 3.4: Random Forest confusion matrix

Table 3.4: Random Forest performance parameters

Accuracy:	0.871
Sensitivity:	0.80
Specificity:	0.919
Precision:	0.869

c. Random Tree

```
=== Confusion Matrix ===
```

```
  a  b  <-- classified as
34 16 |  a = Low
15 59 |  b = High
```

Fig 3.5: Random Tree confusion matrix

Table 3.5: Random Tree performance parameters

Accuracy:	0.653
Sensitivity:	0.62
Specificity:	0.676
Precision:	0.564

d. K Star

```
=== Confusion Matrix ===
```

```
  a  b  <-- classified as
38 12 |  a = Low
11 63 |  b = High
```

Fig 3.6: K star confusion matrix

Table 3.6: K star performance parameters

Accuracy:	0.815
Sensitivity:	0.76
Specificity:	0.851
Precision:	0.775

e. Classification via Regression

=== Confusion Matrix ===

```

a  b  <-- classified as
34 16 |  a = Low
12 62 |  b = High

```

Fig 3.7: Classification via Regression confusion matrix

Table 3.7: Classification via Regression performance parameters

Accuracy	0.774
Sensitivity	0.70
Specificity	0.838
Precision	0.739

f. Multilayer Perceptron

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
36 14 |  a = Low  
18 56 |  b = High
```

Fig 3.8: Multilayer Perceptron confusion matrix

Table 3.8: Multilayer Perceptron performance parameters

Accuracy	0.742
Sensitivity	0.72
Specificity	0.757
Precision	0.667

Further, we used the Correlation feature Selection (CFS) Subset Evaluator and the search method used was BestFirst.

Correlation Feature Selection: is one of the most popular filter algorithms which is used for the feature selection which is based on the correlation function. WEKA is the tool and is the open source software through which it can be freely implemented.

Some examples of attribute evaluation methods are:

CfsSubsetEval: “Evaluates the worth of a subset of attributes by considering the individual predictive ability of each feature along with the degree of redundancy between them.”

ClassifierSubsetEval: “Evaluates attribute subsets on training data or a separate hold out testing set. Uses a classifier to estimate the 'merit' of a set of attributes.”

After running the algorithm , the selected attributes whose value were:

- a. FullTimeWFH
- b. Unachievable Deadlines
- c. Heavy Workload

The dependent class was job satisfaction which could take two values either high or low.

Once these most influencing variables were selected, we used these variables to build all the models. Accuracy, sensitivity, specificity and precision for each model was calculated. These four parameters will further be used to compare each model against the other to find the most accurate model.

- a. Naive Bayes

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
34 16 |  a = Low  
 8 66 |  b = High
```

Fig 3.9: Naive Bayes confusion matrix

Table 3.9: Naive Bayes performance parameter

Accuracy:	0.806
Sensitivity:	0.68
Specificity:	0.892
Precision:	0.809

b. Random Forest

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
35 15 |  a = Low  
10 64 |  b = High
```

Fig 3.10: Random Forest confusion matrix

Table 3.10: Random Forest performance parameters

Accuracy:	0.798
Sensitivity:	0.7
Specificity:	0.865
Precision:	0.778

c. Random Tree

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
36 14 |  a = Low  
11 63 |  b = High
```

Fig 3.11: Random Tree confusion matrix

Table 3.11: Random Tree performance parameters

Accuracy:	0.798
Sensitivity:	0.72
Specificity:	0.851
Precision:	0.766

d. K Star

=== Confusion Matrix ===

```

a  b  <-- classified as
34 16 | a = Low
 7 67 | b = High

```

Fig 3.12: K Star confusion matrix

Table 3.12: K star performance parameters

Accuracy:	0.815
Sensitivity:	0.68
Specificity:	0.905
Precision:	0.829

e. Classification via Regression

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
35 15 |  a = Low  
10 64 |  b = High
```

Fig 3.13: Classification via Regression confusion matrix

Table 3.13: Classification via Regression performance parameters

Accuracy	0.798
Sensitivity	0.7
Specificity	0.865
Precision	0.778

f. Multilayer Perceptron

```
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
31 19 |  a = Low  
 8 66 |  b = High
```

Fig 3.14: Multilayer Perceptron confusion matrix

Table 3.14: Multilayer Perceptron performance parameters

Accuracy	0.782
Sensitivity	0.62
Specificity	0.892
Precision	0.795

CHAPTER 4 - RESULTS

Our original dataset consisted of 124 responses, in which further we applied 10- fold cross validation. A tabular representation of the values obtained for accuracy, sensitivity, specificity, precision is as follows:

Table 4.1: Performance measures before applying Correlation selection feature

Parameters	Naive Bayes	Random Forest	Random Tree	K Star	Classification via Regression	Multilayer Perceptron
Accuracy	0.766	0.871	0.653	0.815	0.774	0.742
Sensitivity	0.78	0.80	0.62	0.76	0.70	0.72
Specificity	0.757	0.919	0.677	0.851	0.838	0.757
Precision	0.684	0.869	0.563	0.775	0.739	0.667

Random Forest was the best model as it was giving highest accuracy of 87.1% followed by K Star algorithm with an accuracy of 81.5% then was classification via regression with an accuracy of 77.4% followed by Naive Bayes having an accuracy of 76.6% then came, an accuracy of 74.2% of multilayer perceptron and lastly , it was random tree algorithm having 65.3% of accuracy respectively.

Further , we applied Correlation feature selection along with the Best Fit search method. Correlation feature selection was used basically to determine the influence of the specific features which affect the employee turnover the most. So, the findings were that FullTimeWFH, Unachievable Deadlines, Heavy Workload were the attributes that came out to be mostly affecting the dependent variable i.e. job satisfaction . Further we prepared different models, based on correlation feature selection method and calculated accuracy, precision, sensitivity and specificity and then compared the six models.

Table 4.2: Performance measures after applying Correlation selection feature

Parameters	Naive Bayes	Random Forest	Random Tree	K Star	Classification via Regression	Multilayer Perceptron
Accuracy	0.806	0.798	0.798	0.814	0.798	0.782
Sensitivity	0.68	0.7	0.72	0.68	0.7	0.62
Specificity	0.892	0.865	0.851	0.905	0.865	0.892
Precision	0.809	0.778	0.766	0.829	0.778	0.795

K star had the highest accuracy with a percentage of 81.4% followed by Naive Bayes algorithm with an accuracy rate of 80.6%, Random forest, Random tree and Classification via Regression had an accuracy of 79.8 %.The algorithm with least accuracy among these models was Multilayer Perceptron with a percentage of 78.2%.

Best model's parameter selection by using Correlation based feature selection:

1. Full time work from home
2. Unachievable deadlines
3. Heavy workloads

4.1 Comparison of Results before and after using Correlation feature selection(CFS)

Table 4.3: Comparison of results using CFS

Parameters	Before CFS	After CFS
Accuracy	Random Forest (87.1%)	K-Star (81.4%)
Sensitivity	Random Forest (80%)	Random Tree (72%)
Specificity	Random Forest (91.9%)	K-Star (90.5%)
Precision	Random Forest (86.9%)	K-Star (82.9%)

Even Though K star algorithm after applying the correlation feature selection has less accuracy than Random Forest which does not use this feature, this algorithm will give more accurate results since this model is based on most significant parameters in determining the employee turnover.

4.1 Conclusion

Employee turnover is one of the major concerns of the organizations. It can lead to negative impacts on the overall organizational growth. In this research, we used six machine learning algorithms that were evaluated on various HR datasets. Further, one of the data mining techniques was also introduced in this research like cross validation. To improve the accuracy of results, we also used correlation feature selection technique. Among the developed models, K Star can give the most accurate results with an accuracy of 81.4%.

CHAPTER 5 - FINDINGS AND RECOMMENDATIONS

5.1 Findings

Results that were collected from the survey helped us to understand reasons for employee turnover.

1. We were able to build a model for predicting employee turnover with a proximity of 80%.
2. By using proper variables, sufficient data and right data mining techniques it will be easy for organizations to identify and resolve employee concerns.
3. Most of the employees had high job satisfaction so their chances of leaving the organization will be low.
4. Low job satisfaction among employees was mainly due to three factors namely unachievable deadlines, heavy workload and work from home facility.

5.2 Recommendations

Based on the analysis and suggestions of some of the respondents we would like to recommend the following to organizations.

1. Many employees felt that they were being burdened by heavy workload. Due to this they are not being able to spend enough time with their families. Organizations should try and find ways to reduce workload.
2. Some employees felt the need to have more work from home options. One of the reasons could be to reduce the travel time from home to office since Offices are located quite far away from home. Organizations should consider ways to implement more work from home culture to make it more convenient for their employees and at the same time the effectiveness of work also is increased.
3. Also the task expectations and deadlines should be well discussed with the employee so that he doesn't feel that his work has unrealistic and unachievable deadlines.
4. Also it was observed that the employees who had low job satisfaction were not being rewarded or recognized for their good work. Rewarding employees boosts the morale of employees which in turn leads to higher productivity and low turnover.

CHAPTER 6 - LIMITATIONS OF THE STUDY

1. The study undertaken has been mostly restricted to the IT industry as we got a maximum number of responses from this industry only, so we are not able to study the impact of work life balance on employee turnover of other industries. It excludes hotel industry, airlines industry etc.
2. This study has been evaluated in small HR datasets. The HR dataset should be more in terms of number of respondents. Small HR datasets may contain high variance and randomness.
3. Most of the responses were recorded online, so there might be a possibility that the respondents did not give accurate, honest answers. It was difficult to get in contact with all the 124 respondents.

BIBLIOGRAPHY/REFERENCES

- Ongori, Henry. (2007). A review of the literature on employee turnover. African Journal of Business Man. 1. 49-54.
- Abassi SM, Hollman KW (2000). "Turnover: the real bottom line", Public Personnel Management, 2 (3) :333-342.
- Price, J.L (1977). The study of turnover, 1st edition, Iowa state university press, IA pp10-25.
- Musrrat Parveen, K. M. N. M. K., 2016. QUALITY OF WORK LIFE: THE DETERMINANTS OF JOB SATISFACTION AND JOB RETENTION AMONG RNs AND OHPs. International Journal for Quality Research, 11(1)(ISSN 1800-6450), p. 173– 194.
https://businessperspectives.org/images/pdf/applications/publishing/templates/article/assets/8193/PPM_2017_01_Mamun.pdf
<https://machinelearningmastery.com/bagging-and-random-forest-ensemble-algorithms-for-machine-learning/>
- John G. Cleary, Leonard E. Trigg: K*: An Instance-based Learner Using an Entropic Distance Measure. In: 12th International Conference on Machine Learning, 108-114, 1995
- E. Frank, Y. Wang, S. Inglis, G. Holmes, I.H. Witten (1998). Using model trees for classification. Machine Learning. 32(1):63-76.
- Zhao, Yue & Hryniewicki, Maciej & Cheng, Francesca & Fu, Boyang & Zhu, Xiaoyu. (2018). Employee Turnover Prediction with Machine Learning: A Reliable Approach. 10.1007/978-3-030-01057-7.
<https://www.gnapartners.com/article/what-is-employee-turnover/>
<https://www.indiaeducation.net/computers-it/the-future-scope-of-the-it-industry-in-india.html>
<https://smallbusiness.chron.com/negative-impacts-high-turnover-rate-20269.html>
https://shodhganga.inflibnet.ac.in/bitstream/10603/37171/18/18_appendix.pdf
- Malhotra, Ruchika & Lata, Kusum. (2020). Using Ensembles for Class Imbalance Problem to Predict Maintainability of Open Source Software. International Journal of Reliability, Quality and Safety Engineering. 10.1142/S0218539320400112.

ANNEXURE

Research Questionnaire:

Section A: Demographic Information

1. Email ID _____
2. Name _____
3. Age
 - a) Below 25 years
 - b) 25-35 years
 - c) 35-45 years
 - d) Above 45 years
4. Gender
 - a) Male
 - b) Female
 - c) Others
5. Educational Qualification
 - a) Undergraduate
 - b) Postgraduate
 - c) Doctorate
6. Experience
 - a) 0-2 years
 - b) 2-5 years
 - c) 5-10 years
 - d) More than 10 years
7. Industry Part Of
 - a) IT
 - b) Education
 - c) Pharmaceutical
 - d) Banking and insurance

- e)Automobile
- f)Manufacturing
- g)Healthcare
- h)Financial Services
- i)Others_____

8. Have you changed jobs in the last 3 years?

- a)Yes
- b)No

9. Nationality_____

Section B: Work Life Balance Practices Relating to Flexible Working Schedules

Kindly select the option that is most suitable to your organization in regards to work life balance practices implemented.

Use the following key: Use the key: 1= strongly disagree; 2= disagree; 3=Not sure; 4=agree; 5=strongly agree

S No.	Questions related to Flexible Working Schedules	1	2	3	4	5
1	I have the flexibility of reporting to work any time but have to ensure observance to working hours					
2	I am allowed to work from home					
3	I am given compensation leave for working overtime					
4	I can avail casual leaves as per my convenience					
5	We have rotational shifts					

Section C: Work Life Balance Practices Relating to Employee benefits

Kindly select the option that is most suitable to your organization in regards to work life balance practices implemented.

Use the following key: Use the key: 1= strongly disagree; 2= disagree; 3=Not sure; 4=agree; 5=strongly agree

S No.	Questions Related to employee benefits	1	2	3	4	5
1	Paid maternity leave is granted to me whenever required					
2	Paid paternity leave is granted to me whenever required					
3	My organization has sabbatical leave policy					
4	Organization has creche services to take care of employees children.					

Section D: Work Life Balance Practices Relating to Employee Assistance Programs

Kindly select the option that is most suitable to your organization in regards to work life balance practices implemented.

Use the following key: Use the key: 1= strongly disagree; 2= disagree; 3=Not sure; 4=agree; 5=strongly agree

S No.	Questions related to employee assistance	1	2	3	4	5
1	I have the freedom to work full time from anywhere other than office					
2	Teleworking hours are specified to me per each working week					
3	My company has given teleworking sources to me					
4	It is possible for me to contact others even if I am teleworking					

Section E: Work Life Balance Practices related to stress

Kindly select the option that is most suitable to your organization in regards to work life balance practices implemented.

Use the following key: Use the key: 1= strongly disagree; 2= disagree; 3=Not sure; 4=agree; 5=strongly agree

S No.	Questions related to stress	1	2	3	4	5
1	I have a lot of pressure so unable to take sufficient breaks					
2	My managers appreciate me for the good work I do					
3	My supervisors are reachable for any doubts/clarification					
4	Work pressure does not allow me to spend time with family and friends					
5	I am given unrealistic deadlines					

Section F: Work Life Balance Practices Relating to Job satisfaction

Kindly select the option that is most suitable to your organization in regards to work life balance practices implemented.

Use the following key: Use the key: 1= strongly disagree; 2= disagree; 3=Not sure; 4=agree; 5=strongly agree

S No.	Questions related to job satisfaction	1	2	3	4	5
1	I have a heavy and unrealistic workload					
2	I might change my job few years down the line					

1	Please indicate your level of job satisfaction	High	Low
---	--	------	-----

Impact of Work Life Balance on Employee

ORIGINALITY REPORT

14%

SIMILARITY INDEX

9%

INTERNET SOURCES

6%

PUBLICATIONS

11%

STUDENT PAPERS

PRIMARY SOURCES

1	pdfs.semanticscholar.org Internet Source	4%
2	upstate.illiad.oclc.org Internet Source	1%
3	Submitted to St. Petersburg College Student Paper	1%
4	www.stratebi.es Internet Source	<1%
5	www.irjet.net Internet Source	<1%
6	www.managementparadise.com Internet Source	<1%
7	people.cs.ksu.edu Internet Source	<1%
8	Submitted to Central Queensland University Student Paper	<1%
9	Ahmeda, Rana Alaa El-Deen, M. Elemam Shehaba, Shereen Morsya, and Nermeen	<1%

Mekawiea. "Performance Study of Classification Algorithms for Consumer Online Shopping Attitudes and Behavior Using Data Mining", 2015 Fifth International Conference on Communication Systems and Network Technologies, 2015.

Publication

10

Submitted to Ealing, Hammersmith and West London College

Student Paper

<1%

11

Submitted to University of Lincoln

Student Paper

<1%

12

Submitted to University of Central Lancashire

Student Paper

<1%

13

"Artificial Intelligence: Theories, Models and Applications", Springer Science and Business Media LLC, 2008

Publication

<1%

14

Submitted to National College of Ireland

Student Paper

<1%

15

Ankit Agrawal, Reda Al-Bahrani, Mark J. Russo, Jaishankar Raman, Alok Choudhary. "Lung transplant outcome prediction using UNOS data", 2013 IEEE International Conference on Big Data, 2013

Publication

<1%

16

www.coursehero.com

Internet Source

<1%

17

Submitted to Anglia Ruskin University

Student Paper

<1%

18

Submitted to University of Greenwich

Student Paper

<1%

19

Submitted to University of Melbourne

Student Paper

<1%

20

Advances in Intelligent Systems and Computing,
2016.

Publication

<1%

21

Submitted to Manchester Metropolitan
University

Student Paper

<1%

22

Submitted to The Robert Gordon University

Student Paper

<1%

23

Submitted to iGroup

Student Paper

<1%

24

jms.nonolympictimes.org

Internet Source

<1%

25

gsblibrary.uct.ac.za

Internet Source

<1%

26

docplayer.net

Internet Source

<1%

27	Submitted to University of Hertfordshire Student Paper	<1%
28	Submitted to University of Pretoria Student Paper	<1%
29	Submitted to Auckland University of Technology Student Paper	<1%
30	Submitted to Glion Institute for Higher Education Student Paper	<1%
31	Adel M. Aladwani. "The 6As model of social content management", International Journal of Information Management, 2014 Publication	<1%
32	pt.scribd.com Internet Source	<1%
33	Submitted to Swinburne University of Technology Student Paper	<1%
34	Submitted to American University of Beirut Student Paper	<1%
35	Submitted to Jawaharlal Nehru University (JNU) Student Paper	<1%
36	Submitted to VIT University Student Paper	<1%
37	repository.unika.ac.id Internet Source	<1%

38 D Antai, A Oke, P Braithwaite, DS Anthony. "A 'Balanced' Life: Work-Life Balance and Sickness Absence in Four Nordic Countries", The International Journal of Occupational and Environmental Medicine, 2015 <1%

Publication

39 Kristina Palm. "chapter 18 A Case of Phased Retirement in Sweden", IGI Global, 2017 <1%

Publication

40 Submitted to Eiffel Corporation <1%

Student Paper

41 www.cheap-dissertation-writing.co.uk <1%

Internet Source

42 Submitted to University College Birmingham <1%

Student Paper

43 Submitted to University of York <1%

Student Paper

44 Submitted to Segi University College <1%

Student Paper

45 Submitted to Asia Pacific University College of Technology and Innovation (UCTI) <1%

Student Paper

46 Submitted to University of Sheffield <1%

Student Paper

47 Submitted to Visvesvaraya Technological University <1%

Student Paper

48 Carolynne Hultquist, Gang Chen, Kaiguang Zhao. "A comparison of Gaussian process regression, random forests and support vector regression for burn severity assessment in diseased forests", Remote Sensing Letters, 2014 <1%

Publication

49 "Advanced Computing Technologies and Applications", Springer Science and Business Media LLC, 2020 <1%

Publication

50 Submitted to University of Liverpool <1%

Student Paper

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off