Bachelor of Science in Computer Science & Engineering

Developing an Assistance System for Women's Work Life Balance using Machine Learning

by

Urshi Barua

ID: 1704058

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

May, 2023

Developing an Assistance System for Women's Work Life Balance using Machine Learning

Submitted in partial fulfilment of the requirements for
Degree of Bachelor of Science
in Computer Science & Engineering

by

Urshi Barua

ID: 1704058

Supervised by

Dr. Mahfuzuhoq Chowdhury

Associate Professor

Department of Computer Science & Engineering

Chittagong University of Engineering & Technology (CUET)

Chattogram-4349, Bangladesh.

The thesis titled '**Developing an Assistance System for Women's Work Life Balance using Machine Learning'** submitted by ID: 1704058, Session 2017-2018 has been accepted as satisfactory in fulfilment of the requirement for the degree of Bachelor of Science in Computer Science & Engineering to be awarded by the Chittagong University of Engineering & Technology (CUET).

Board of Examiners

Chairman

Dr. Mahfuzuhoq Chowdhury
Associate Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Member (Ex-Officio)

Dr. Abu Hasnat Mohammad Ashfak Habib
Professor & Head
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Member (External)

Dr. Asaduzzaman
Professor
Department of Computer Science & Engineering
Chittagong University of Engineering & Technology (CUET)

Declaration of Originality

This is to certify that I am the sole author of this thesis and that neither any part of this thesis nor the whole of the thesis has been submitted for a degree to any other institution.

I certify that, to the best of my knowledge, my thesis does not infringe upon anyone's copyright nor violate any proprietary rights and that any ideas, techniques, quotations, or any other material from the work of other people included in my thesis, published or otherwise, are fully acknowledged in accordance with the standard referencing practices. I am also aware that if any infringement of anyone's copyright is found, whether intentional or otherwise, I may be subject to legal and disciplinary action determined by Dept. of CSE, CUET.

I hereby assign every rights in the copyright of this thesis work to Dept. of CSE, CUET, who shall be the owner of the copyright of this work and any reproduction or use in any form or by any means whatsoever is prohibited without the consent of Dept. of CSE, CUET.

Signature of the candidate

Date:

Acknowledgements

I would like to acknowledge and show tremendous appreciation to all individuals without whom, it would be impossible to finish this study. I offer my heartfelt gratitude to my supervisor Dr. Mahfuzulhoq Chowdhury, Associate Professor, Department of Computer Science and Engineering. I am indebted to him for his motivation, proper guidance, constructive criticism and immense support towards the progress of this study. All these things have helped me to grow as a machine learning practitioner.

I also want to express my gratitude to every person who made this study possible, including Dr. Md. Shafiul Hasan , MBBS, DPM (Psychiatric Medicine), Institute of Applied Health Science, USTC, Chittagong who helped to validate the dataset.

I would like to express gratitude to my respective parents and family members for the unceasing encouragement, support, and attention throughout my whole life.

Abstract

Many researchers currently hold out hope that artificial intelligence, and machine learning in particular, can enhance a number of areas of daily life for people, communities, and entire nations. For instance, it has been suggested that so-called machine learning may soon relieve workers of some responsibilities, which may improve procedures or assist in determining the most efficient ways to do jobs. In the long run, it would improve employees' ability to combine their work and personal lives. As a result, the general quality of life for workers would also increase. The current research aims to find out work life balance status of working women. We also tried to build some additional features that can help a working woman to live a balanced work life. Different kinds of ML (Machine Learning) classification algorithms were used for this purpose. The study also prepares a set of 22 attributes for appropriately classifying work life balance status. The data is examined. Several machine learning methods are tested using the dataset. The research has found and suggests K-Nearest Neighbour to be the most effective, with a classification accuracy of 98%, despite the fact that previous studies have identified Support Vector Machine to be the most efficient model in these kind of classification tasks.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ANN Artificial Neural Network. 11

API Application Programming Interface. 6, 18, 20

BMI Body Mass Index. 16

GNU GNU's Not Unix. 12

HR Human Resource. 7, 13

IT Information Technology. 12, 13

KNN K-Nearest Neighbour. 3, 23, 37, 48, 49, 55, 56

ML Machine Learning. vi, 13, 16

SDK Software Development Kit. 26

SVC Support Vector Classifier. viii, 38, 39

SVM Support Vector Machine. ix, 4, 12, 13, 37, 48, 50, 51, 55

UI User Interface. viii, 18, 24, 25

WLB Work Life Balance. 12–14

XML Extensible Markup Language. 18, 21

Chapter 1

Introduction

1.1 Introduction

Due to society's broad expectations and the mounting burden of personal and professional obligations, modern women today are under tremendous stress. The majority of people experience stress as a physical and emotional reaction to life's events. Although stress is perfectly normal, it can be prolonged by persistent pressure and unreasonable expectations, which can result in a number of health problems. When under stress, the body secretes the hormone cortisol, which temporarily boosts energy production at the expense of other crucial bodily functions including digestion and the immune system. If this reaction persists, it may cause women's health issues. So, maintaining a work life balance is very important.

A common term for work-life balance is the ratio of time spent working to time spent with family, friends, or engaging in personal interests and hobbies. Many people strive to strike a better balance between their personal and professional lives so that work does not monopolize their time. The phrase itself, though, might be deceptive because our personal and professional lives aren't usually kept apart. In today's environment, the notion "work-life balance" is considered to be one of the primary challenges that play an essential part in organizational effectiveness.

The overall well-being, encompassing the physical, emotional, and mental health, may be enhanced by improving the balance between work and personal life. Working long hours has been associated to major health problems like "impaired sleep, depressive disorders, excessive drinking, diabetes, diminished memory, and heart

disease," according to studies [1]. Because it affects both personal and professional lives in today's competitive economy, researchers and academics have concentrated on the issue of work-life balance.Evidence indicates that while a healthy work-life balance generates harmony between the professional and personal worlds and an unhealthy work-life balance can negatively impact an employee's personal life, which results in workplace discontent and damages the company's reputation and productivity [2].

So, here we can understand the importance of balancing among work life and personal life, as this issue can affect her mental health and can affect the economy also. To overcome this problem, we want to create a model that can predict whether a woman is having a balanced work life or not. The method of machine learning called prediction will be utilized for analyzing datasets obtained from women. Yet again, it is insufficient to determine whether or not a woman has a balanced work-life schedule because this is a worrying issue. We want to develop an assistance system (an android platform based mobile application) that can also provide a woman mental healthcare who is suffering from stress, depression for having an imbalanced life. They can handle their issues and get assistance when they need it with the help of the mobile application.

1.1.1 Machine learning

Machine learning (ML) is a technique that helps computers how to handle data with greater effectiveness. Even after examining the data, there are times when we are unable to interpret or extrapolate the results. In that case, we apply machine learning. The demand for machine learning has expanded as a result of the abundance of datasets [3]. Machine learning is a specialized area of artificial intelligence (AI) and computer science that focuses on using data and algorithms to simulate how humans learn, gradually increasing the accuracy of the system. It is a crucial element in the developing discipline of data science. Algorithms are trained using statistical techniques to produce classifications or predictions and to find significant findings in data mining projects.

1.1.2 Work Life Balance Prediction

We can determine if one person's work life is balanced or not by using the prediction of a work life balance system based on a machine learning technique. According to speculation, referred to as machine learning could soon relieve workers of some responsibilities, which could improve procedures or lead to the discovery of the most efficient ways to complete jobs [4]. And clustering and classification techniques can be used on same datasets to do different types of analysis [5]. In our work, we have used several machine learning classification algorithms to predict if a woman is having a balanced life or not.

1.1.2.1 K-Nearest Neighbor (KNN) Algorithm

One of the simplest machine learning algorithms, based on the supervised learning method, is K-Nearest Neighbour. It is widely used in intrusion detection, data mining, and pattern recognition. The KNN algorithms select a value for k such that the nearest neighbor's data points are separated. KNN is essential to machine learning. This is also known as lazy learning method [6]. In this classifier algorithm, We must choose the k neighbors and determine their euclidean distance from one another. The Euclidean distance should then be sorted from shortest to longest. Neighbors count the amount of data points in each of these categories. In this study, we've determined the ideal k value for this algorithm.

1.1.2.2 Decision Tree Algorithm

A decision tree is a tree structure that resembles a flowchart, where each internal node represents a feature, branches represent rules, and leaf nodes provide the algorithm's output. It is a flexible supervised machine-learning approach that may be applied to classification and regression issues. One of the most potent algorithms is this one. There are other classification methods that can be found in the literature, however decision trees are the most widely utilized since they are simpler to use and comprehend than other classification algorithms [7].

1.1.2.3 Support vector machine Algorithm

Support vector machines (SVM), also known as support vector networks) are supervised learning models with corresponding learning algorithms that examine data for regression and classification. A study shows, the accuracy of SVM can be higher than other regression models also [8].

1.1.2.4 Linear Regression Algorithm

It is possible to compute the linear relationship between a dependent variable and one or more independent features using a supervised machine learning algorithm called linear regression. Finance, economics, and psychology are just a few of the disciplines that employ linear regression to analyze and forecast the behavior of a given variable.

1.1.2.5 Random Forest Algorithm

A Random Forest is another type of supervised machine learning technique, used for solving classification and regression problems. It generates multiple decision trees at training time to get more accurate and consistent predictions. A classification system known as random forests creates a large number of decision trees, each of which has a higher predictive accuracy than the others.

1.1.2.6 AdaBoost Algorithm

The Boosting technique known as AdaBoost algorithm, sometimes known as Adaptive Boosting, is used as an Ensemble Method in machine learning. The weights are redistributed to each instance, with higher weights being given to instances that were mistakenly identified, hence the name "adaptive boosting." For supervised learning, boosting is used to lower bias and variation.

1.1.2.7 Logistic Regression Algorithm

The principal use of the supervised machine learning algorithm, logistic regression, is classification problems, where the objective is to estimate the likelihood

that a given instance belongs to a particular class. It's a type of statistical technique that examines the correlation between a group of independent variables and a set of binary dependent variables. It is an effective instrument for making decisions.

1.1.3 Matrices used

- **Accuracy**: The accuracy is calculated by dividing the total number of predictions in the dataset by the number of accurate forecasts (ACC). The accuracy levels range from 0.0 to 1.0, with 1.0 being the most accurate.

  $$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**: Precision is obtained by dividing the total number of accurate positive forecasts by the total number of positive predictions (PREC). Another term for it is positive predictive value (PPV). The best and worst precision values are 1.0 and 0.0, respectively.

  $$Precision = \frac{TP}{TP + FP}$$

- **Recall**: Sensitivity is calculated by dividing the total number of correct positive predictions by the total number of positives (SN). It is often referred to as the true positive rate (TPR) or the recall rate (REC) (TPR). The sensitivity ranges from 0.0 to 1.0, with 1.0 being the most.

  $$Recall = \frac{TP}{TP + FN}$$

- **F1-score**: The F-score is a harmonic mean of recall and precision..

  $$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

1.1.4 Visualization

Many insights that data alone cannot offer are made possible by data visualization. Python offers some of the most interesting tools for data visualization. While the simpler plot types are found in many libraries, some are exclusive to a few. We have utilized the Matplotlib and Seaborn visualization tools.

1.2 Work Overview

Data is needed to feed the machine learning model in machine learning systems. Traditional machine learning models are sensitive to the choice of features and the preprocessing of the data. Therefore, choosing features before collecting data is a crucial step. For the purpose of computation, the acquired string data is mapped to its corresponding numerical data. To understand the structure of the numerical data, preprocessing and visualization are used.

The information is then fed into a variety of machine learning models for analysis and selection of the model with the highest accuracy for the system's improved performance. Frontend and backend web technologies are used to create an app. With a focus on excellent usability, the chosen machine learning model is introduced to the application.

1.3 Difficulties

Creating a useful machine learning model is always challenging. The first challenge is finding the proper dataset that can be used to train a machine learning model. Finding the best model is also a difficult one.

The difficulty I faced was preparing the dataset. There is no dataset exact dataset for my model. Secondly, visualizing the data was also difficult for me. The basic process for deploying machine learning on a website entails building the model in any Python IDE, extracting it using a pickle module, and then deploying it as a web app with the use of any web framework, such as flask. As our machine learning model's output will be in JSON format, we have developed a Flask API. We have deployed machine learning models on a website, Railway. Lastly this implementation process was slightly difficult for me.

1.4 Applications

The purpose of this assistance system is to help the working women. The following are a few examples of potential applications:

- The building of personalized mobile applications can use the work-life balance prediction model that has been developed.
- Companies can implement designed policies, practices, and benefits that encourage a healthier work-life balance by integrating the model into their HR systems and gaining insightful information on the work-life difficulties that their workforce faces.
- Companies can integrate work life balance prediction model for employees' wellbeing.

1.5 Motivation

Evidence suggests that while a healthy work-life balance fosters harmony in both the professional and personal spheres, an unhealthy work-life balance can adversely affect an employee's personal life, which culminates in job unhappiness and harms the reputation and productivity of the company. Women in Bangladesh have made a significant contribution to our economy. Despite this, working women still face a number of challenges across both their personal and professional lives. Women strain to strike a work-life balance, which eventually affects their social life. Women experience pressure from both their home and the workplace as they endeavor to strike a balance between work and life. The following list of reasons motivates our work :

1. This work can reduce employee attrition rate as predict imbalanced work life.
2. Women have contributed a significant amount to our economy. So, low attribution rate can ensure our economic stability.
3. Later in life, bad health is caused by an unhealthy work-life balance.

1.6 Contribution of the thesis

Every thesis or piece of study has been done in order to identify a distinctive contribution. It's not anything like integration or information collecting; rather,

it's something like invention, highlighting a cause and effect or suggesting a model that would benefit people. The thesis works' primary contributions are as follows:

1. To gather information that will be useful for future work,
2. To prepare work life balance dataset,
3. To select the ideal machine learning predictive model for the dataset,
4. To develop a mobile application based on machine learning that can predict a person is having a balanced work life or not based on work life balance data,
5. Offer primary health care via mobile app.

1.7 Thesis Organization

Organization of the rest of the thesis are as follows :

- **Chapter 2:** It provides summary of the previous works in the related fields and it's background.
- **Chapter 3:** It represents the details how the proposed system works and the total overview of the system which provides a clear vision of the system.
- **Chapter 4:** It shows all about the evaluation and experimental outcomes of the proposed system.
- **Chapter 5:** In addition, a brief outline of future work has been provided, which makes the proposed system more efficient.

1.8 Conclusion

The work-life balance prediction machine learning model and its significance in resolving the difficulties faced by people in creating a harmonious integration of work and personal life have been thoroughly described in the Introduction chapter. We have discussed the motivation for the work, described its goals and domain, and emphasized the potential significance and uses of the suggested model. We have stressed the growing significance of work-life balance in

contemporary culture throughout this chapter. People are under more pressure than ever to successfully balance their professional and personal duties due to the rapid growth of technology and the changing nature of employment. Achieving a healthy work-life balance is still a challenging endeavor that is influenced by a number of variables, including job obligations, individual circumstances, and social expectations.

Chapter 2

Literature Review

2.1 Introduction

This thesis's literature review portion gives a thorough summary of the body of knowledge regarding work-life balance and machine learning methods for prediction. This part provides a critical examination of the major themes and discoveries in the discipline and serves as the basis for the formulation of the research inquiries and goals. This study seeks to identify imperfections and regions for additional research by synthesizing the available literature and by offering a theoretical foundation for the next empirical research.

Beginning with an overview of the idea of work-life balance, its definitions, and metrics, the literature review is organized to provide a methodical and in-depth analysis of the domain. The discussion of the causes and effects of work-life balance that includes individual, managerial, and societal issues that affect results follows. The literature study also looks at the difficulties in balancing work and life, including how technology plays a part, how work is getting harder, and how work itself is evolving.

The literature review then looks into the potential for predicting work-life balance outcomes using machine learning methodologies. An review of the supervised and unsupervised machine learning methods currently used in prediction algorithms is also included. The paper also discusses the critical aspects of selecting and training machine learning models for work-life balance prediction, such as data quality, feature selection, and assessment metrics.

It is significant to note that the literature evaluation pulls from a wide range of disciplinary viewpoints, including those from computer science, psychology,

sociology, and organizational behavior. This study intends to provide a holistic and multidisciplinary knowledge of work-life balance prediction by fusing these many viewpoints.

2.2 Related Literature Review

In [9], the authors discovered a negligible link between work-life balance and productivity. they conducted structured interviews with 128 working women employees for data collection and for testing hypotheses and creating models, use SMART PLS. Here, The study discovered a strong link between work-life balance and mental health. Employees' mental well-being will also be compromised if they are unable to strike a balance between work and personal life. In this paper, researcher only shows how imbalanced life style can effect an employee's both personal and professional life. Limitation of this work is, they did not use any machine learning algorithm to predict or to give any suggestions if the employee having an imbalanced life.

From the work of [10],the essay examines the causes of women experiencing greater stress in academic settings and separates the project into two categories: personal and professional. They also state that, Women who try hard to manage their professional and personal lives will face difficulties if their jobs are unsatisfying or unreliable. Work-life balance should be balanced between one's personal and professional obligations.

In the work of [4],The major objective was to determine whether a machine learning algorithm could identify any potential relationships between workplace and employee-specific characteristics and employees' perceptions of preserving work-life balance. Here, they used selectKbest algorithm to select the most valuable features to obtain a standard dataset and used artificial neural network (ANN) with 25 and 50 neurons of two hidden layers, ADAM and ReLU was build and trained on 90 percent of the dataset which predicts work life balanced with 81 percentage of accuracy. The limitation of this work is smaller dataset is used for prediction. So the predicted result may not always be accurate.

This article [11] focuses on building a generalized fuzzy logic rule-based forecasting model that will be utilized later to assess the survey data gathered from Hyderabad-based female IT engineers working in the software versus services industry. The work show WLB and its effects on attrition and job unhappiness.

In [12],the researcher used machine learning to to attempt and uncover ways of creating a better work-life balance. The authors of the study employed a machine learning technique to look for features that have an effect on people's subjective feelings of their work-life balance in order to better understand how people perceive it. The results of this study show that supervised machine learning was used to create a prediction model for the level of work-life balance experienced by Information Technology (IT) women employees. 150 women who worked in the information technology industry participated in this poll.

In the work of [13], to find out respondents' opinions on how virtual technologies affect work, study, and social life, a poll was performed online. In order to assess the work-life balance, the data were analyzed using a combination of descriptive methods from PSPP (GNU open source SPSS—Statistical Package for the Social Sciences, Free Software Foundation, Boston, MA, USA). This regression model demonstrates how the incorporation of XR resources in the e-learning process, together with a raised level of culture and living standard, have a favorable impact on the work-life balance.

From this work of [14], we can learn that, the information from the data analysis demonstrates to us clearly that machine learning technologies can be used to assist the Human Resources division in managing the workforce of the organization. This helped turn for predicting employee attrition takes into account, they used a real dataset from the Gamo Gofa Zone Trade and Industry Development Office from Kaggle, which contains a substantial chunk of 12,756 employees under their study's consideration and 23 factors initially under deliberation but later reduced to 16 important factors. They used different machine learning techniques like: Support Vector Machine (SVM), Gaussian Naïve Bayes Classifier, Random Forest Classifier and also incorporating several pre-processing methodologies such as feature selection techniques to predict if one employee is having a balanced life or

not.This work is not only for women, so there must be some features which are not in this dataset they used, but are important reason for having an imbalanced one. Moreover, the main drawback is their accuracy rate for their dataset is 71.52 percent which is pretty low.

In the work of [15], the researcher studies how smart work can enhance the WLB. Both of these works' main limitation is they did not discuss about making any machine that can predict WLB or any solution if anyone is having imbalanced one.

In [16],they state that, retaining employees has become a significant problem for many IT companies. Many businesses are looking for different strategies to keep their employees, such as producing yearly personnel survey reports and identifying employee complaints. The data gathered from a company's alumni as well as current employees is used in this paper to create a model using a machine learning algorithm. The HR staff may utilize this to understand the reasons for attrition that are unique to their company and will provide an early signal and time to retain before it's too late.

In the article [8], supervised machine learning (ML) is used to develop a prediction model for the degree of work-life balance among women in the information technology (IT) industry. 425 women who work in the IT sector in total took part in the study. A questionnaire with 40 attributes was used to collect the data, which is used as an input for supervised machine learning. Regression trees and support vector machines (SVM) are two examples of different ML models that are developed and contrasted. For predicting work-life balance, an optimized SVM model is found to have the maximum accuracy.

This author in [17] proposed the characteristics of the WLB concept by evaluating various ideas, outlining potential WLB implications, and outlining WLB models. Both the positive and negative effects of technology on WLB was discussed. Finally, a Smart Service System was suggested to optimize WLB, and ethical standards and data collecting were discussed. They did not propose any system to detect WLB and did not propose any solution tool that can convert one's imbalanced life to balanced.

2.3 Conclusion

A literature review is a summary of the earlier written works on a certain subject. These previous works mostly discuss about the facts how WLB can effect one employee's both personal and professional life. Also described that, relationship between WLB and attribution rate, job unhappiness which can leads one to leave one's workplace. But, we will be trying to make a prediction model to help the women to find if they are having a imbalanced work life style. And also help them to make their life more balanced by featuring a suggestions module in our application. Moreover, all these previous works very specific sets of data but in our work, one of our goal is to make a large data sets for better result. This work will be beneficial for the working women. Also, if these women can lead a healthy work life by using this mobile application, then the employee attribution rate will be decreased day by day. This is also beneficial to our economy, as women, have been participating and contributing remarkably in our economy.

2.3.1 Implementation Challenges

The preparation of dataset, using machine learning algorithms, visualizing the gathered data, and evaluating the performance of various algorithms are a few implementation problems.

Chapter 3

Methodology

3.1 Introduction

This thesis' methodology section gives a thorough explanation of the methodology used to create a machine learning model for forecasting work-life balance. This section describes the techniques used for data collection and preprocessing, model selection and training, evaluation, and validation. This study guarantees transparency and reproducibility by describing its methods, allowing subsequent studies and practitioners to expand on its findings.

The approach is divided into several important phases. Data preparation strategies are used to assure data quality and relevance after the methods for data gathering are defined. The selection of the best machine learning algorithm for predicting work-life balance is then discussed in detail, taking algorithmic suitability and model training techniques into account. Additionally mentioned are the validation procedures utilized and the assessment criteria used to measure the model's effectiveness.

In order to ensure rigor and reliability in the construction and evaluation of the prediction model, the methodology section is organized to give a clear and systematic account of the actions done. This organized strategy is used in the study in an effort to reduce bias, increase generalizability, and lay the groundwork for further investigation of the topic.

Overall, by describing the research design and methods used to address the research questions and accomplish the goals of this thesis, the methodology section acts as a road map for the next chapters.

3.2 Overview of Framework and Detailed Explanation

The suggested assistance system for women's work life balance using machine learning will work on mainly three steps:

1. Build ML model on work life balance dataset
2. Create Android App
3. Deploy ML Model on Android

3.2.1 Build ML model on work life balance dataset

For predicting if one's work life is balanced or not, we have built an ML model using several classification algorithms. The fig. 3.1 represents the ML model framework. First of all, we collected data from several survey and then created a dataset. The features of this dataset were:

1.How many fruits or vegetables do you eat everyday?

2. How much stress do you typically experience everyday?

3. How many new places do you visit?

4. How many people are very close to you?

5. How many people do you help achieve a better life ?

6. With how many people do you interact with during a typical day?

7. How many remarkable achievements are you proud of?

8. How many times do you donate your time or money to good causes?

9. What is your Body Mass Index(BMI) range?

10. How many works do you complete in a day?

11. How many actions you have to complete in a day?

12. How many steps you walk in a day?

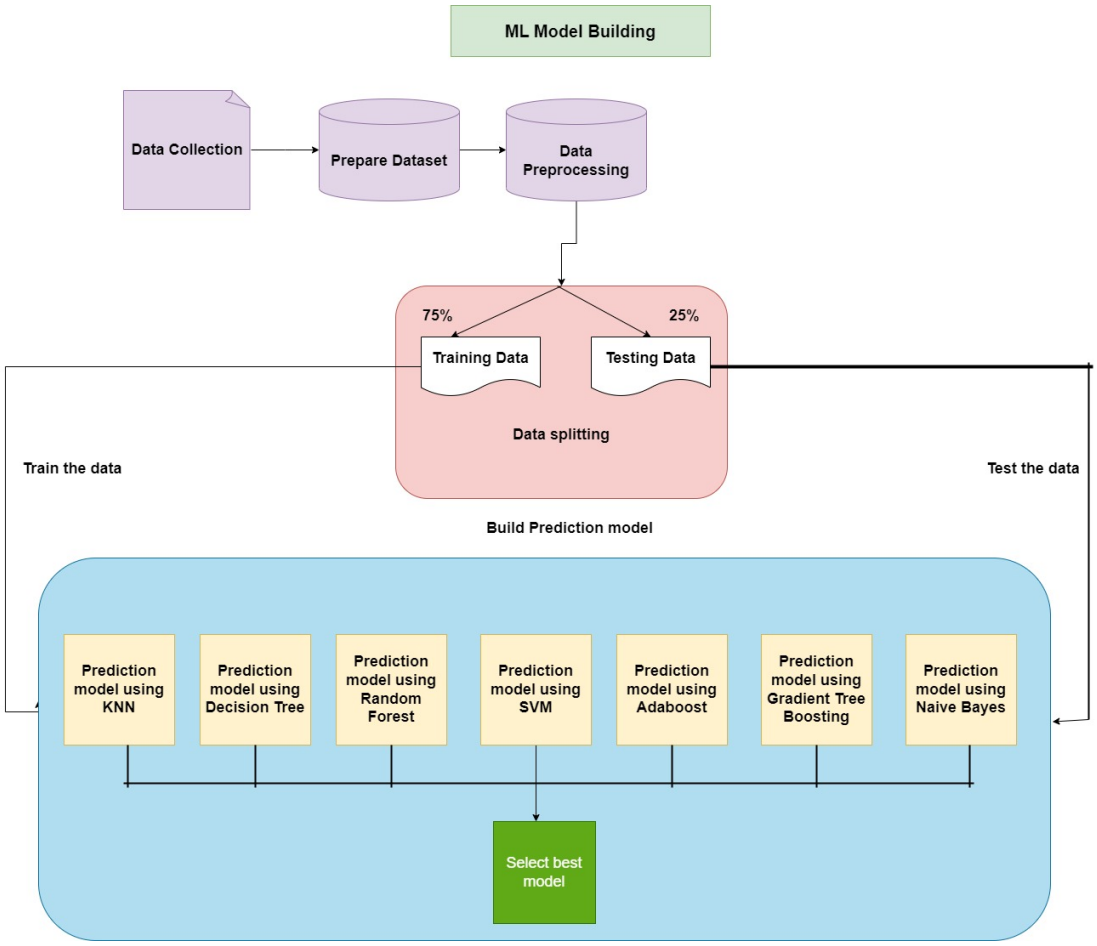


Figure 3.1: ML Model Framework

13. How many hours do you sleep in a day?
14. How many vacation you have lost recently?
15. How much anger do you typically express everyday?
16. what is the range of your income?
17. How many awards or achievements do you have?
18. How many hours you spent in a week for passion?
19. How many hours do you spent in a week for meditation, yoga etc?
20. What is your age?
21. What is your gender?

After preparing the dataset and data processing, we divided the data into training data and testing data. 25% of the data is used for testing, while 75% is used for

training. We have used machine learning algorithms like the KNearest Neighbor (KNN) Algorithm, Decision Tree Algorithm, Support Vector Machine Algorithm, Random Forest Algorithm, Adaboost Algorithm, Gradient Tree Boosting Algorithm and Naïve Bayes Algorithm. After that, we have training data to train and we also have testing data to test the model.

We have chosen the best model for prediction based on accuracy after these models have been trained with training data and tested with testing data. In our work, the model using is the KNearest Neighbor (KNN) algorithm. Finally, we can predict work life balance or not using this model.

3.2.2 Create Android App

We initially required an Android studio in order to work with Android programming. We are aware that the UI (User Interface) is constantly produced in an XML file. After designing our UI, we wrote backend logic to accept data from the frontend in a Java file. Our Android app name is BalanceHER. Firebase is used and it is a backend system. Firebase offers a link between mobile applications and the web to the backend cloud storage and API.

Figure 3.2 represents process flow diagram of our app. First, any wants to use our app, then she has to open our app BalanceHER, and a splash screen will be displayed. If she has already logged in to this app, then she can enter directly into the home fragment. Otherwise, she has to enter the login page. To login to this app, an email and password are required. And if the email is valid, then the home activity interface is displayed. If the user is new, she has to sign up first. To sign up for this app, a user name, email, mobile number, password, and confirm password are required. If the password is matched with the confirmed password, then the home activity interface is displayed. In home activity, four interfaces such as find out, Doctor Info, Suggestion box and compliants will be found. If she wants to find out if she is having a balanced life or not, then she has to click find out. She will enter the prediction interface and she has to fillup the required inputs. Then result will be predicted using machine learning model. There will be other options in home activity. From the doctor info interface an user can

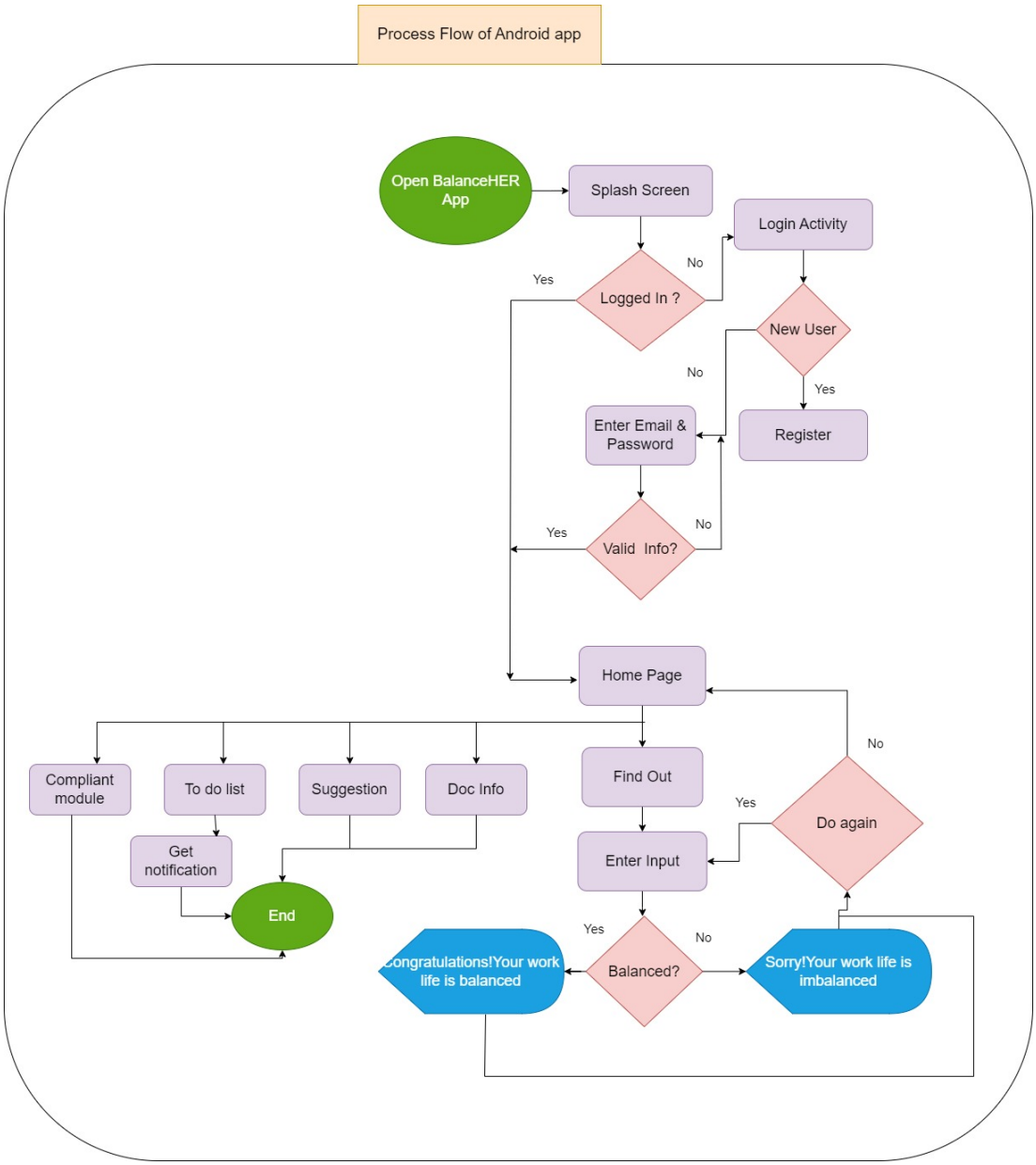


Figure 3.2: Process Flow Diagram of BalanceHER

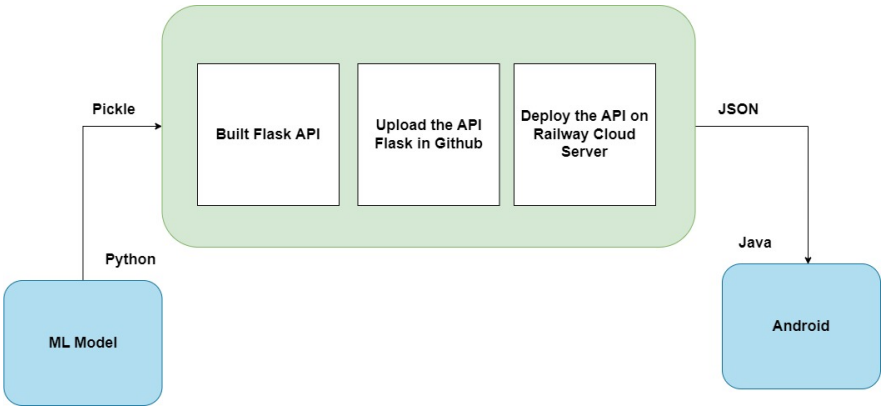


Figure 3.3: Deploy ML Model on Android

get the information about good doctor who can help to get rid of mental illness for having an unbalanced life. There are another interface named suggestion box which is listed some tips to lead more balanced life.

3.2.3 Deploy ML Model on Android

First, we saved our machine learning model as an a.pkl file. Then we have the flask API using Python in Pycharm. In fig. 3.3, it represents the deployment of ML model on Android. The form will be filled out by the user, who will then submit it. The form will then receive a POST request. Additionally, when a user submits a post request, the Flask API will accept their data and send it to the machine learning model so that it can predict the output class. We will send the predicted class as JSON to the Android app. We have used Postman to verify API of our BD Entrepreneur App. An automated and interactive tool called Postman is used to check the project's APIs. One library called Volley is required to hit API.

The logic we have created is that we will use the Android app's inputs to call the API, and then the API's answer will be displayed back in the Android app. So, there was one issue that the Android app was unable to detect because the API we developed was running locally on our system. Therefore, we had to launch our API online, and Railway was used to do so.

3.3 Implementation of the Proposed Method

The implementation section of this thesis presents the details of how the proposed method for predicting work-life balance using machine learning was implemented and executed. This section provides a comprehensive description of the tools, technologies, and procedures employed to develop and deploy the predictive model. By documenting the implementation process, this study aims to ensure reproducibility and facilitate further advancements in the field.

The implementation phase involved translating the research design and methodology into a practical application. It encompassed several key steps, including data collection and preprocessing, model development and training, and performance evaluation. Additionally, considerations such as platform selection, programming languages, and software libraries were taken into account to enable efficient and reliable implementation. The chosen implementation approach focused on creating a user-friendly application that allows users to input their relevant data and obtain work-life balance predictions.

3.3.1 System Requirements

Good hardware is a requirement for completing any computing task effectively and in reasonable amount of time. An proper programming language is required to carry out any project on a computer. Here is a list of the equipment needed to put our system into practice:

- Operating System: Windows 10 with 8 GB RAM,
- Google Colab for machine learning code visualization,
- Android Studio for App development,
- PyCharm IDE for API Flask development,
- Firebase Real time Database used for store data and also for authentication,
- Python is used for model prediction and API development,
- The XML language is used for front-end design,

1	Timestam	FRUITS_VE	DAILY_STF	PLACES_VI	CORE_CIR	SUPPORTI	SOCIAL_N	ACHIEVEN	DONATIO	BMI_RAN	TODO_CO	FLOW	DAILY_STE	LIVE_VISI	SLEEP_HO	LOST_VAC	DAILY_SH	SUFFICIEN	PERSO
2	7/7/2015	3	2	2	5	0	5	2	0	1	6	4	5	0	7	5	5	1	
3	7/7/2015	2	3	4	3	8	10	5	2	2	5	2	5	5	8	2	2	2	
4	7/7/2015	2	3	3	4	4	10	3	2	2	2	2	4	5	8	10	2	2	
5	7/7/2015	3	3	10	3	10	7	2	5	2	3	5	5	0	5	7	5	1	
6	7/7/2015	5	1	3	3	10	4	2	4	2	5	0	5	0	7	0	0	2	
7	7/8/2015	3	2	3	9	10	10	2	3	1	6	1	7	10	8	0	2	2	

1	SLEEP_HO	LOST_VAC	DAILY_SH	SUFFICIEN	PERSONAL	TIME_FOR	WEEKLY_M	AGE	GENDER	RESULT	EXPERT_OPINION
2	7	5	5	1	4	0	5	2	1	1	Balanced
3	8	2	2	2	3	2	6	2	1	1	Balanced
4	8	10	2	2	4	8	3	2	1	1	Balanced
5	5	7	5	1	5	2	0	3	1	1	Balanced
6	7	0	0	2	8	1	5	3	1	1	Balanced
7	8	0	2	2	10	8	3	3	1	1	Balanced

Figure 3.4: Sample data of work life balance dataset

- The Java language is used for controlling the front end,
- An android phone to run our android project.

3.3.2 Implementation Details

The overall system implementation part is divided into several parts. We will discuss all these parts in the below subsection.

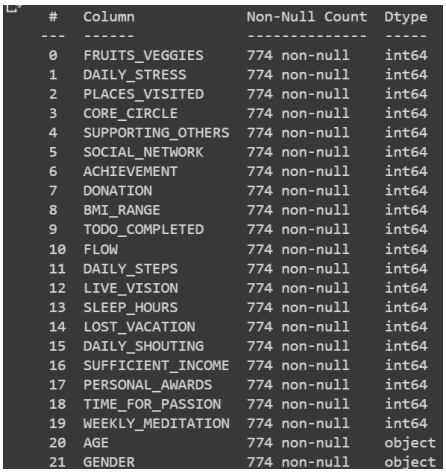
3.3.2.1 Steps for the Work Life Balance Prediction Model

The method of developing the work life balance prediction model will be briefly covered in this subsection. The fig. 3.1 represents the overall steps of prediction model.

Step 1: Firstly, we have collected information on work life balance from several survey. After collecting the data, I have created a work life balance dataset, fig. 3.4 represents the tail data of the work life balance dataset. This dataset contains 770 instances with 23 attributes describing how we live our lives.

Step 2: The fig. 3.5 shows that in this dataset, there was some categorical information. The categorical data cannot be understood by a machine. In order to create numerical data, we convert categorical data using label encoding. Then we determine whether a value is missing or not. Fortunately, our dataset did not contain any missing values. Therefore, handling the missing value wasn't necessary in our dataset.

Step 3: With the use of just pertinent data and the elimination of irrelevant data, feature selection is a technique for lowering the input variable for your model. There are many features which are related to this prediction. We simply drop



#	Column	Non-Null Count	Dtype
0	FRUITS_VEGGIES	774 non-null	int64
1	DAILY_STRESS	774 non-null	int64
2	PLACES_VISITED	774 non-null	int64
3	CORE_CIRCLE	774 non-null	int64
4	SUPPORTING_OTHERS	774 non-null	int64
5	SOCIAL_NETWORK	774 non-null	int64
6	ACHIEVEMENT	774 non-null	int64
7	DONATION	774 non-null	int64
8	BMI_RANGE	774 non-null	int64
9	TODO_COMPLETED	774 non-null	int64
10	FLOW	774 non-null	int64
11	DAILY_STEPS	774 non-null	int64
12	LIVE_VISION	774 non-null	int64
13	SLEEP_HOURS	774 non-null	int64
14	LOST_VACATION	774 non-null	int64
15	DAILY_SHOUTING	774 non-null	int64
16	SUFFICIENT_INCOME	774 non-null	int64
17	PERSONAL_AWARDS	774 non-null	int64
18	TIME_FOR_PASSION	774 non-null	int64
19	WEEKLY_MEDITATION	774 non-null	int64
20	AGE	774 non-null	object
21	GENDER	774 non-null	object

Figure 3.5: Detail information of dataset

one column named timestamp as this feature is not important for our prediction model.

Step 4: In this step, we split our dataset into train and test data. 25% of the data is used for testing, while 75% is used for training. The model is then trained using several machine learning classification algorithms such as KNearest Neighbor (KNN) Algorithm, Decision Tree Algorithm, Support Vector Machine Algorithm, Random Forest Algorithm, Adaboost Algorithm, Gradient Tree Boosting Algorithm and Naïve Bayes Algorithm. The model is trained using each of these machine learning techniques, and then its accuracy is evaluated. The main goal is to choose the most effective classifier for our dataset.

Step 5: It is the last stage of choosing a model. The most effective model has been chosen based on the evaluation matrices. We have now adjusted the settings using the grid search method to get the best performance out of the classification algorithm. In our model, the KNearest Neighbor (KNN) algorithm gives us the best result compared with the other algorithms.

3.3.2.2 Steps for Android Application

After selecting the best machine learning model for work life balance prediction, we have predicted and the result is shown in the mobile application. The fig. 3.2 represents process flow diagram of our app.

Step 1: At first, we need an Android studio to work with Android programming.

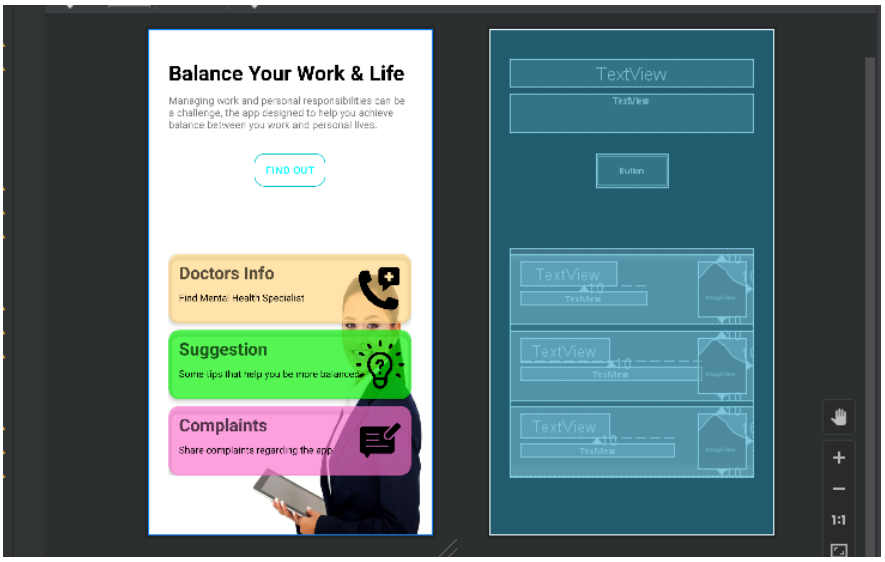


Figure 3.6: UI design of Home Page

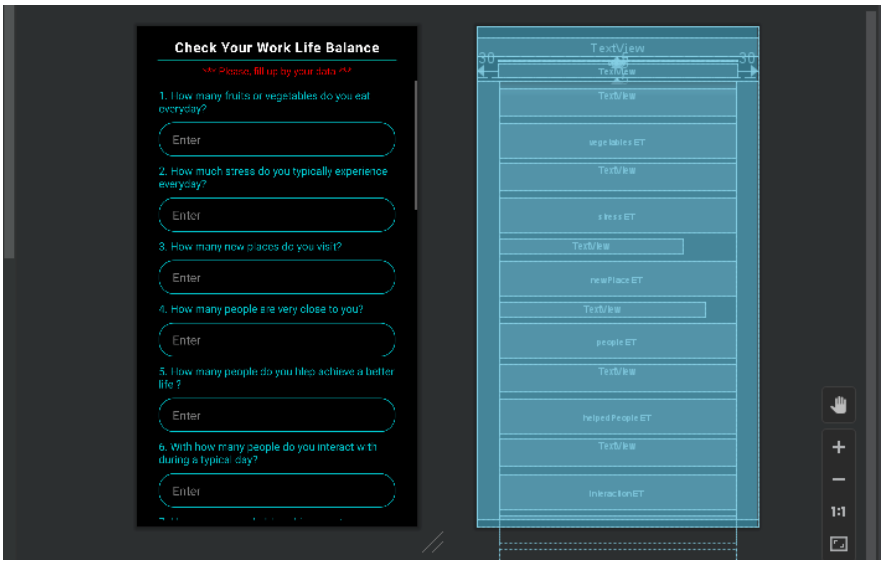


Figure 3.7: UI design of taking user input for prediction

When we launched Android Studio after installation, we created a new project named BalanceHER.

Step 2: After creating the Android project, we designed the user interface. We have completed the front-end UI using the XML language. There were at least eight user interfaces, such as home page, login, signup, user inputs for prediction, showing the result, doctor info interface, suggestion interface , complaint interface. The user interface of our app consists of all the elements that the user can view and use. Some UI design is shown in . fig. 3.6, .fig. 3.7, fig. 3.8.

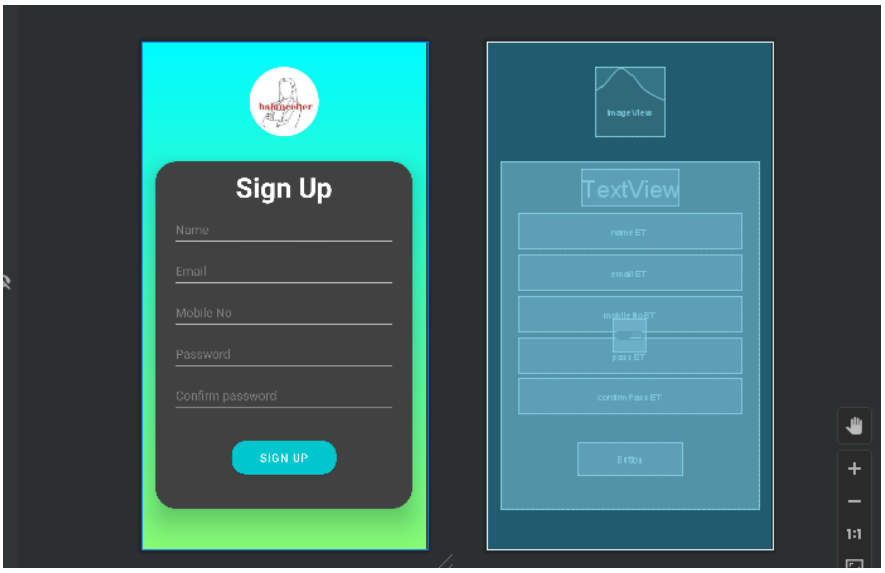


Figure 3.8: UI design of Sign Page

In our home page, there is the tagline and a short introduction of our app BalanceHER. At the center of the design are several cards and one button, the button is for "FIND OUT", if user click on this the user inputs for prediction interface will pop up. The other cards are for respectively doctor info, suggestion and complaints.

In our user input interface, there are two text views in the upper portion which display the text "Check Your Work Life Balance" and another text *** Please, fillup by your data ***. There are three "Autocompletetextviex" and eighteen "Textinputedittext". In the lower portion, there is a predict button. If the user enter all the information and click predict, then result page will pop up but it is also a part of our user input page. There are two option "Go Back" and "Do Again". There are another two interface in our app where doctor info and suggestion to lead more balanced life are listed.

Step 3: In this step, we have created a database for our app. We utilized Firebase Real Time Database to build the database. Stored database in firebase is shown in fig. 3.9, fig. 3.10. Google has also built a platform called Firebase for building mobile and web applications. A cloud-hosted NOSQL database called Firebase Real Time Database allows users to sync and save data in real time. To provide users with notice, we used Firebase Cloud Messaging (FCM) multicast, which can transmit messages to numerous devices. To make developer authentication

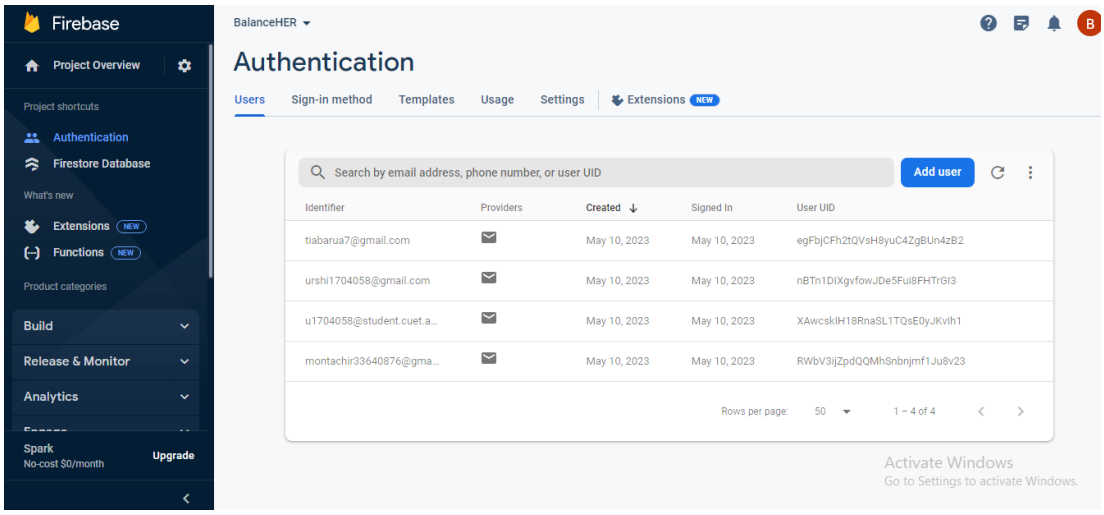


Figure 3.9: Authentication database

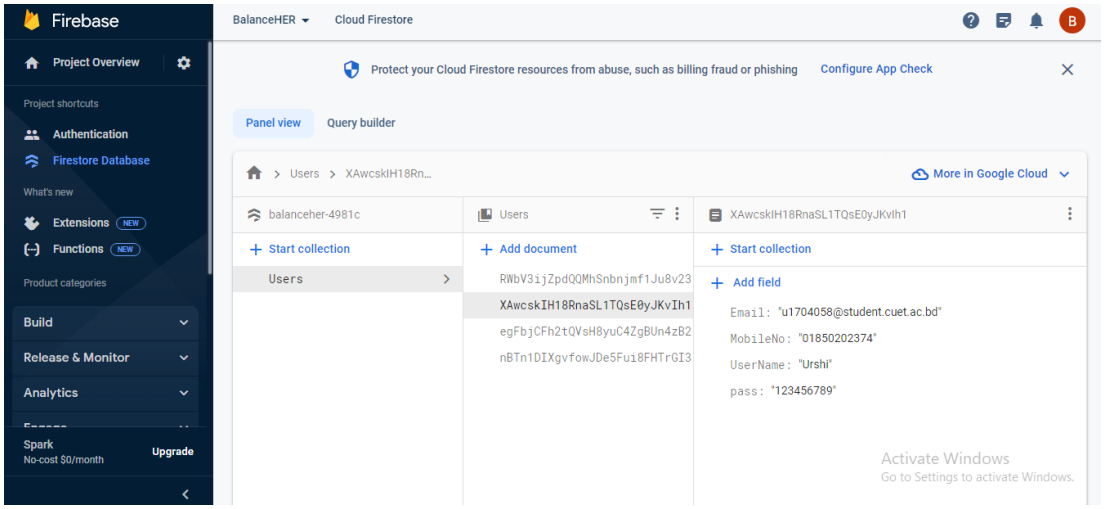


Figure 3.10: Signup database

quick and easy, the Realtime Database interfaces with Firebase Authentication. Our declarative security paradigm can be used to provide access based on user identity or through data pattern matching. In addition, Firebase was used to store the necessary data.

The fig. 3.9 represents the authentication database. We only need to obtain the user's authentication credentials from them and then provide them to the Firebase Authentication SDK in order to authenticate our users. This login information includes an email address, password, mobile number, user name, applicant occupation etc. After receiving the credentials, Firebase will check them, and the user will then receive a response informing the success or failure of the authentication.

To create a user with an email address and a password, use the createUserWithEmailAndPassword() method. The arguments for this function are an email address and a password, which are validated before a new user is created. On the Firebase website, in the Authentication area, then on the Users page, our user will be registered, and we can view the registered user there. Users will encounter an error if they attempt to register several times using the same email address. We have a method called "signInWithEmailAndPassword()" that allows users to sign in with their email and password. This technique takes an email address and a password as arguments, validates them, and then, if the validation is successful, signs the user in to our application.

Step 4: After designing our UI, we wrote backend logic to accept data from the frontend in a Java file.

Step 5: First, we saved our machine learning model as an a.pkl file. Then we have the flask API using Python in Pycharm. Furthermore, when an user submits a post request, the Flask API will receive their data and send it to the machine learning model so that it can predict the output class. We will send the anticipated class as JSON to the Android app.

Step 6: The fig. 3.3 represents the deployment of ML model on Android. The API we developed was running locally on our system, there was a problem. Therefore, we had to launch our API online, and Railway was used to do so.

Step 7: Finally, the user will fill in the required input for the prediction. Then the input will be passed as JSON form through the API flask to the ML model, which has been saved as an a.pkl file. The result will be shown in the mobile app.

3.4 Conclusion

As a summary, this methodology chapter has given a thorough explanation of the process used to create a machine learning model for forecasting work-life balance. To ensure transparency and repeatability of the research, the procedures used for data collection, preprocessing, model selection, training, and evaluation have

been presented in a methodical manner. The dataset was built to include pertinent features linked to work-life balance through thorough data collection. In order to handle missing values, guarantee data quality, and improve the model's prediction ability, data preparation techniques were used. Following thorough model training methods, the right machine learning technique was chosen depending on the dataset's properties and the task at hand.

The usefulness of the model in forecasting work-life balance has been revealed through the performance evaluation of the model using acceptable metrics. The following chapters of this thesis will expand on this methodology by using the constructed model to investigate and examine the variables affecting work-life balance and to make insightful judgments. This study intends to increase our knowledge of work-life balance through the use of ethical research methods and to aid in the creation of interventions and workplace wellness policies.

Chapter 4

Results and Discussions

4.1 Introduction

The findings and analyses from the use and assessment of the work-life balance prediction app are presented in the Results and Discussions portion of this thesis. This section seeks to offer insights into the prediction model's performance, assess the results' importance, and explain their implications in relation to work-life balance. In order to build the app, user data had to be gathered, put through a machine learning model, and predictions had to be made. As indications of a person's work-life equilibrium, these predictions are given as work-life balance scores or qualitative assessments.

We will first give a summary of the application's usage outcomes in this area, including the number of users, the distribution of input variables, and the overall predictions for work-life balance. The discussion will also cover the accuracy and dependability of the forecasts made by the app. In order to determine how well the model predicts work-life balance, we will examine its accuracy, precision, recall, and other performance measures, compare them to the baseline, and evaluate its performance.

The results will be thoroughly analyzed and their importance and consequences will be thoroughly discussed in the following chapters. By doing this, we hope to raise awareness of the variables affecting work-life balance and enhance people's wellbeing and job satisfaction in today's workplace.

4.2 Dataset Description

The dataset includes 774 response from a survey. The survey is called authentic happiness survey. It contains 22 question that can design our lifestyle and help us to achieve a healthy work life. In this work life balance dataset, there are 774 rows and 23 columns. This dataset is made by the responses of women's as our work is basically for women.

4.2.1 Validation of the dataset

Verifying dataset with a psychiatrist is a valuable step in ensuring the accuracy and reliability of a research. We have verified dataset with the help of

- Dr. Md. Shafiul Hasan

  MBBS, DPM (Psychiatric Medicine)

  Mental Diseases, Disorders and Drug Addiction Specialist

  Institute of Applied Health Science, USTC, Chittagong.

This verification process involved consulting with an expert who possesses specialized knowledge and clinical experience in mental health and work-life balance. In this work life balance dataset, there are 774 rows and 23 columns. We add a column is called expert opinion where validated data is stored. If anyone is having balanced work life, then it's value will be balanced and if anyone is having imbalanced work life, then it's value will be imbalanced. Figure 4.1 represents the tail data of the work life balance dataset.

The questionnaires are examined by the expert to gather information about work-life balance. Then he examine the indicators, symptoms, or dimensions of mental health that included and assess their validity and appropriateness in the context of work-life balance. After that, he ensure that the questionnaires used here are accurate and consistent with clinical standards. According to the psychiatrist, the Diagnostic and Statistical Manual of Mental Disorders (DSM) includes diagnostic criteria for mental health conditions that can be influenced by work-life balance issues. So, the reasons behind the imbalanced work life are negative emotion,

1	Timestam	FRUITS_VE	DAILY_STE	PLACES_VI	CORE_CIR	SUPPORTI	SOCIAL_N	ACHIEVEM	DONATIO	BMI_RAN	TODO_CO	FLOW	DAILY_STE	LIVE_VISIC	SLEEP_HO	LOST_VAC	DAILY_SH	SUFFICIEN	PERSO
2	7/7/2015	3	2	2	5	0	5	2	0	1	6	4	5	0	7	5	5	1	
3	7/7/2015	2	3	4	3	8	10	5	2	2	5	2	5	5	8	2	2	2	
4	7/7/2015	2	3	3	4	4	10	3	2	2	2	2	4	5	8	10	2	2	
5	7/7/2015	3	3	10	3	10	7	2	5	2	3	5	5	0	5	7	5	1	
6	7/7/2015	5	1	3	3	10	4	2	4	2	5	0	5	0	7	0	0	2	
7	7/8/2015	3	2	3	9	10	10	2	3	1	6	1	7	10	8	0	2	2	

1	SLEEP_HO	LOST_VAC	DAILY_SH	SUFFICIEN	PERSONAL	TIME_FOR	WEEKLY_M	AGE	GENDER	RESULT	EXPERT_OPINION
2	7	5	5	1	4	0	5	2	1	1	Balanced
3	8	2	2	2	3	2	6	2	1	1	Balanced
4	8	10	2	2	4	8	3	2	1	1	Balanced
5	5	7	5	1	5	2	0	3	1	1	Balanced
6	7	0	0	2	8	1	5	3	1	1	Balanced
7	8	0	2	2	10	8	3	3	1	1	Balanced

Figure 4.1: Sample data of work life balance dataset

stress, workload, lack of social support, proper communication with family. As a result, the features involving these factors are very important for the work life balance. By evaluating these factors he validate our data. If we take an example if anyone experience stress 4 on a 5 scale, expressing negative emotion more than 8 on a 10 scale for our dataset, then the chances of the person is having imbalanced life increased. The most important feature of our dataset is daily shouting, core circle, social network, weekly meditation, sufficient income. The importance of this features are respectively 0.07, 0.05, 0.06, 0.05 and 0.05. The importance of BMI range, sleep hour, places visit, lost vacation are almost similar. They are 0.043, 0.038, 0.045, 0.045. Figure 4.3 represents the comparison of feature importances.

4.3 Impact Analysis

The work-life balance prediction app's potential implications and effects on different stakeholders, including people, organizations, and society at large, are to be evaluated in this thesis's Impact Analysis section. This investigation explores the potential consequences of the app's adoption and use on several facets of work-life balance, both positive and negative. There is a rising understanding of the significance of work-life balance in fostering employee well-being, job satisfaction, and overall productivity as the use of technology in the workplace continues to develop. The creation and use of the work-life balance prediction app offers a cutting-edge strategy to assist people in attaining a harmonious balance between their personal and professional lives.

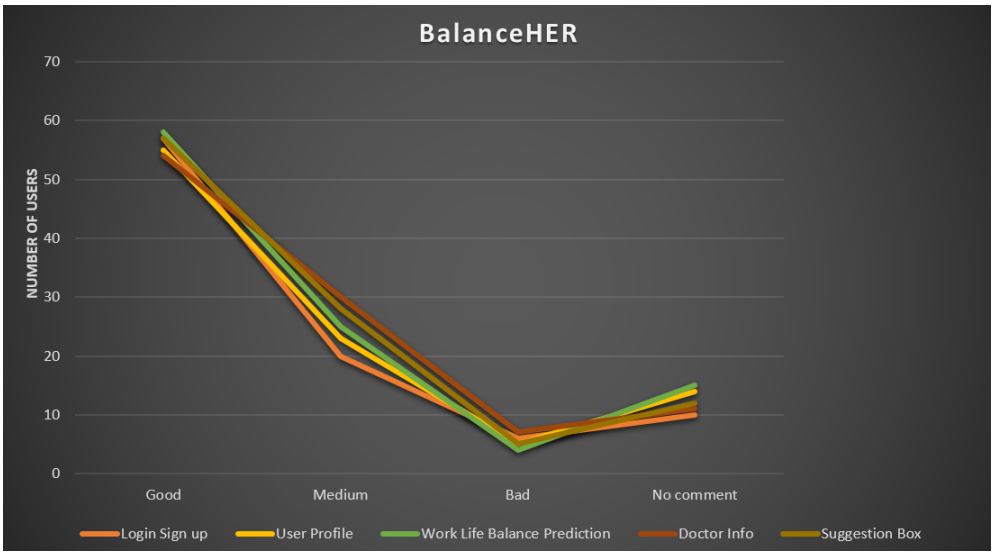


Figure 4.2: User Rating

4.3.1 Social Impact

The software could increase public awareness of the value of work-life balance in society. It can inform users about the elements that lead to a healthy work-life integration by giving them individualized insights and recommendations. The public's awareness of the importance of work-life balance in society could grow as a result of the software. By providing them with tailored insights and recommendations, it may educate users about the factors that contribute to a successful work-life integration.

I have collected some user ratings for this BalanceHER app from my friends. These ratings is shown in fig. 4.2. Fron the user rating graph, we can see that good rating of every features is high. But the prediction feature get the highest good rating. And the number of medium ratings, no comments, and bad ratings descends in that order. The number of bad rating is too low comparatively.

4.4 Evaluation of Framework

The outcomes of the experiment will be presented in this section. The validation results of the project are displayed in this crucial part.

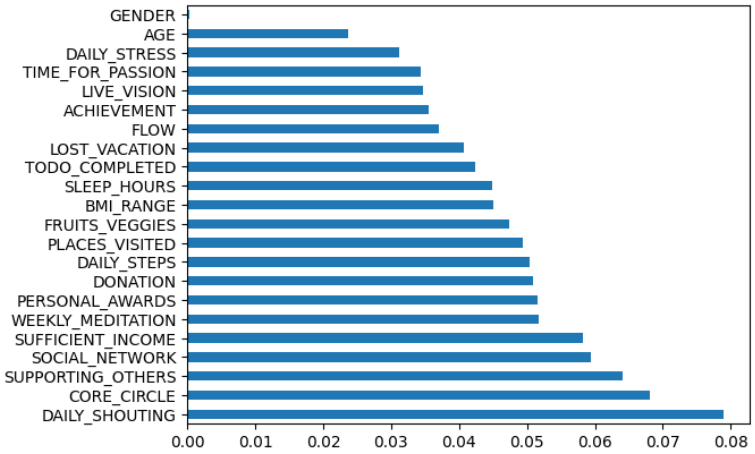


Figure 4.3: Comparison of features in order of importance

4.4.1 Important Feature Selection

We separated the dataset into dependent and independent features after preprocessing it. We used the feature importance technique to identify the key features in our dataset. We've brought in ExtraTreesClassifier, an ensemble approach to applying this function. The relative relevance of each feature in our dataset is shown in fig. 4.3. Here we can see that "DAILY_SHOUTING" is the most important feature. "CORE_CIRCLE", "SUPPORTING_OTHERS" are also important features. The feature "GENDER" is not important for our work life balance dataset.

4.4.2 Data Visualization

The graphic display of data as well as information is known as data visualization. Data visualization tools offer an easy approach to observe and analyze trends, outliers, and patterns in data by utilizing visual elements like charts, graphs, and maps. Additionally, it offers a great tool for staff members or business owners to clearly deliver data to non-technical audiences.

Figure 4.4 represents the possibility of having balanced life loans against age. From the figure we can see that, from the age of 21 to 35 carry the maximum possibility of having balanced life. From the fig. 4.5, we can see the violin plot of time for passion against age. Here we can see that, almost all kind of aged people spent same number of hours for passion.

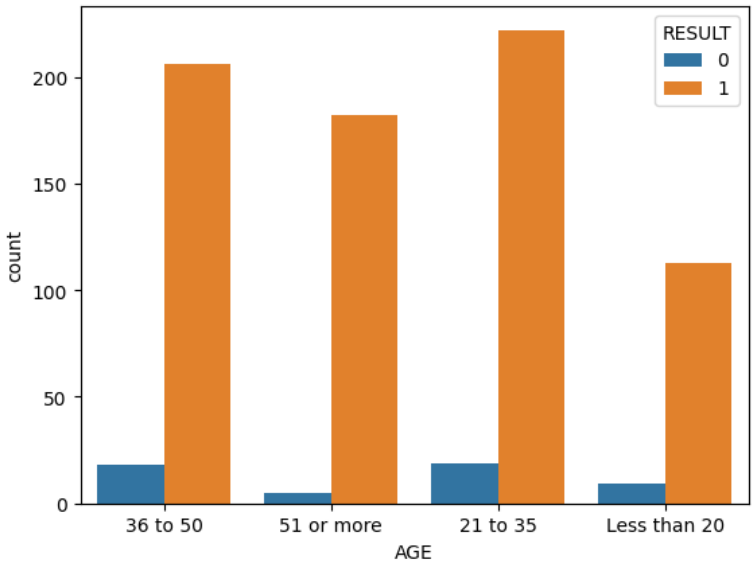


Figure 4.4: Result against age

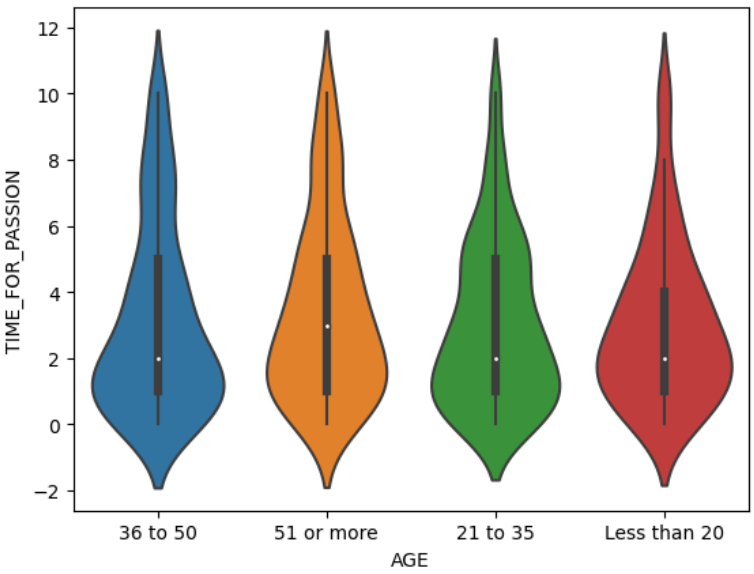


Figure 4.5: Violin plot of time for passion against age

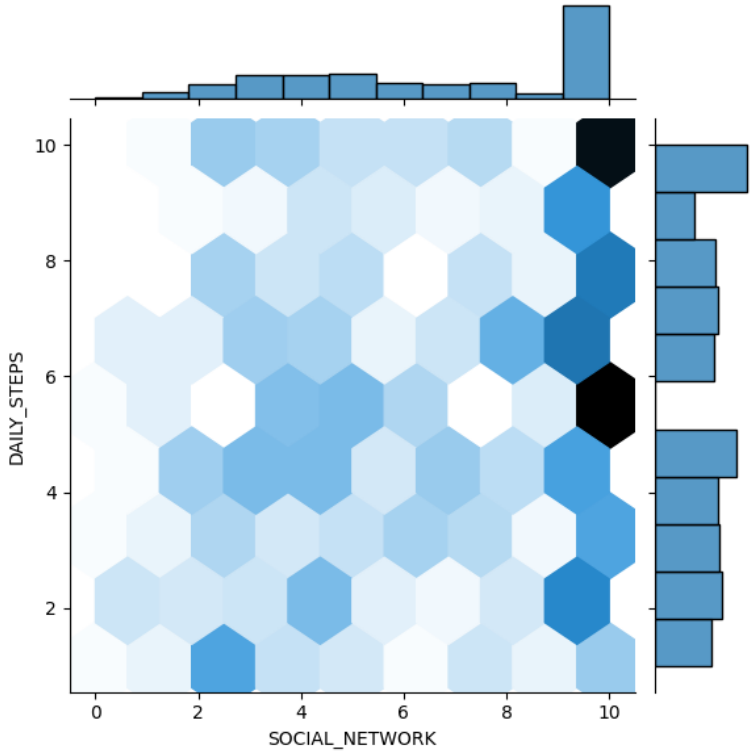


Figure 4.6: Joint plot of social network against daily steps

From the fig. 4.6, we can see the joint plot between daily steps and social networks. This plot also represents the distribution plot of daily step and social network individually. From the fig. 4.7, we can see the box plot of social network against age. Here we can understand that the number of people who have less number of social network are, who are less than 20.

The correlation matrix of the features is shown in figure 4.8. No features are eliminated or changed for the machine learning model because the correlation matrix clearly shows that each feature contributes in some way to the final output. We may create heatmaps using the Seaborn charting software that illustrate the Pearson product-moment correlation coefficient (PPMCC) correlation between features. The bivariate correlation coefficient Pearson quantifies the linear relationship between two features. The strongest correlations are found between Time for Passion with Flow (0.48), Supporting Others with Donation (0.4), Achievement with Flow (0.4).

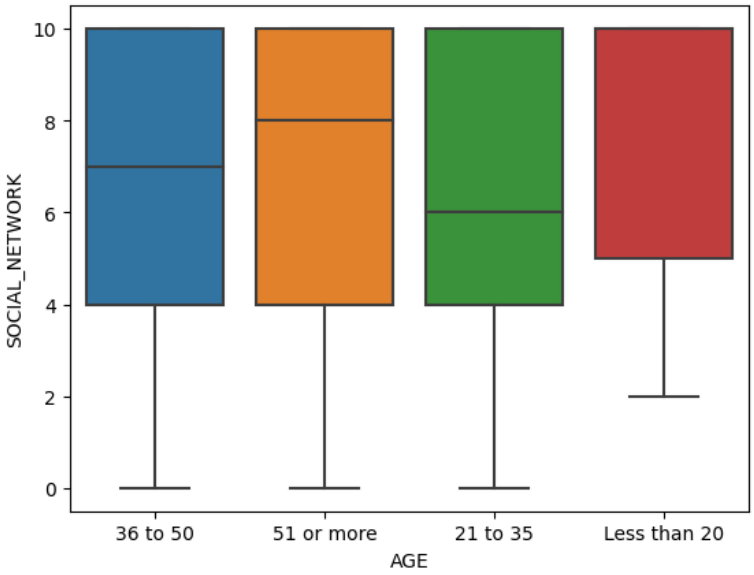


Figure 4.7: Box plot of social network against age

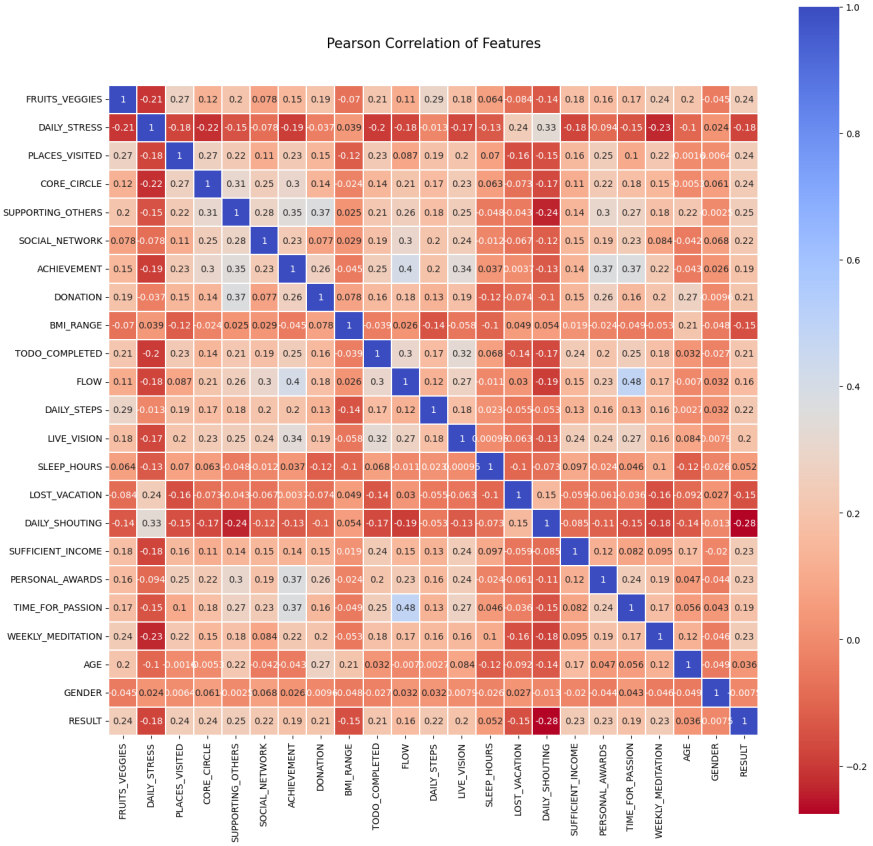


Figure 4.8: Correlation Heatmap

Figure 4.9: Block of code and Output of K-Nearest Neighbor

4.4.3 Building machine learning model

In our research, we used the seven most well-liked machine learning classifiers and trained them to test the performance of machine learning models against our gathered and preprocessed dataset. Such as, K-Nearest Neighbor(KNN), Decision Tree, Support Vector Machine(SVM), Random forest, Adaboost, Naive Bayes, Gradient Tree Boosting. The block code and output of best four algorithms for our research are given below:

4.4.3.1 K-Nearest Neighbor

The KNN classifier categorizes unlabeled observations by placing them in the category of the most comparable labeled samples. We built a model with the Scikit-learn module's KNeighborsClassifier. The model is then trained using a training dataset and evaluated using a testing dataset. The accuracy has been saved in order to further establish the best model. The classification function computed the precision, recall, f1-score displayed in figure 4.9 using metrics from the Scikit-learn module.

4.4.3.2 Decision Tree

Decision tree which is a non-parametric supervised learning technique for classification and regression (DT). To create a model that predicts the value of a target

```
from sklearn.tree import DecisionTreeClassifier

dt = DecisionTreeClassifier().fit(X_train, y_train)

dt_pred = dt.predict(X_test)

dt_met = precision_recall_fscore_support(y_test,dt_pred, average='macro')

print('Accuracy of Decision Tree classifier on test set: {:.2f}'
     .format(dt.score(X_test, y_test)))

dt_acc = dt.score(X_test, y_test)

print(dt_met)
```

```
Accuracy of Decision Tree classifier on test set: 0.95
(0.6867828304292393, 0.8357524828113063, 0.736413043478261, None)
```

Figure 4.10: Block of code and Output of Decision Tree

variable, the goal is to learn simple decision rules generated from the data attributes. It is possible to imagine a piecewise constant approximation of a tree. Using DecisionTreeClassifier from the Scikit-learn module, we created a model. Next, a training dataset is used to train the model, and a testing dataset is used to evaluate it. To further determine the optimal model, the accuracy has been saved. Using metrics from the Scikit-learn library, the classification function calculated the precision, recall, f1-score shown in fig. 4.10.

4.4.3.3 Random Forest

The random forest also can be called decision tree version where a huge number of decision trees are created using training data and then your new data is fit into one of the trees as a random forest. We built a model with the Scikit-learn module's RandomForestClassifier. The model is then trained using a training dataset and evaluated using a testing dataset. Accuracy has been saved in order to further establish the best model. The classification function computed the precision, recall, f1-score displayed in fig. 4.11.

4.4.3.4 Support Vector Machines

The supervised machine learning technique known as SVC, or Support Vector Classifier, is frequently used for classification problems. SVC separates the data into two classes by mapping the data points to a high-dimensional space and then locating the best hyperplane. We built a model with the Scikit-learn module's SVC. The model is then trained using a training dataset and evaluated using

```
from sklearn.ensemble import RandomForestClassifier

rf = RandomForestClassifier(max_depth=8, random_state=0)

rf.fit(X_train, y_train)

rf_pred = rf.predict(X_test)

rf_met = precision_recall_fscore_support(y_test,rf_pred, average='macro')

print('Accuracy of Random Forest on test set: {:.2f}'
     .format(rf.score(X_test, y_test)))

rf_acc = rf.score(X_test, y_test)

print(rf_met)
```

```
Accuracy of Random Forest on test set: 0.98
(0.9921052631578947, 0.7857142857142857, 0.8596575837955148, None)
```

Figure 4.11: Block of code and Output of Random forest

```
from sklearn.svm import SVC
from sklearn.metrics import confusion_matrix

# The default SVC kernel is radial basis function (RBF)
svm= SVC().fit(X_train, y_train)

svm_pred = svm.predict(X_test)

svm_met = precision_recall_fscore_support(y_test,svm_pred, average='macro')

print('Accuracy of Kernalized Support Vector Machine Classifier on test set: {:.2f}'
     .format(svm.score(X_test, y_test)))

svm_acc = svm.score(X_test, y_test)

print(svm_met)
```

```
Accuracy of Kernalized Support Vector Machine Classifier on test set: 0.98
(0.9921052631578947, 0.7857142857142857, 0.8596575837955148, None)
```

Figure 4.12: Block of code and Output of SVC

a testing dataset. Accuracy has been saved in order to further establish the best model. The classification function computed the precision, recall, f1-score displayed in fig. 4.12.

4.4.4 Mobile Application Framework

In this section the interface and feature of our mobile application will be discussed and displayed.

4.4.4.1 User Interface

Our user interface is consists of home page and login sign-up functionality. Users profile can be created in this interface. who will further proceed to share their work life balance information with us. The fig. 4.13 displays the splash screen and the home page of our application. And the fig. 4.14 represents the login page and the sign up page of our application.

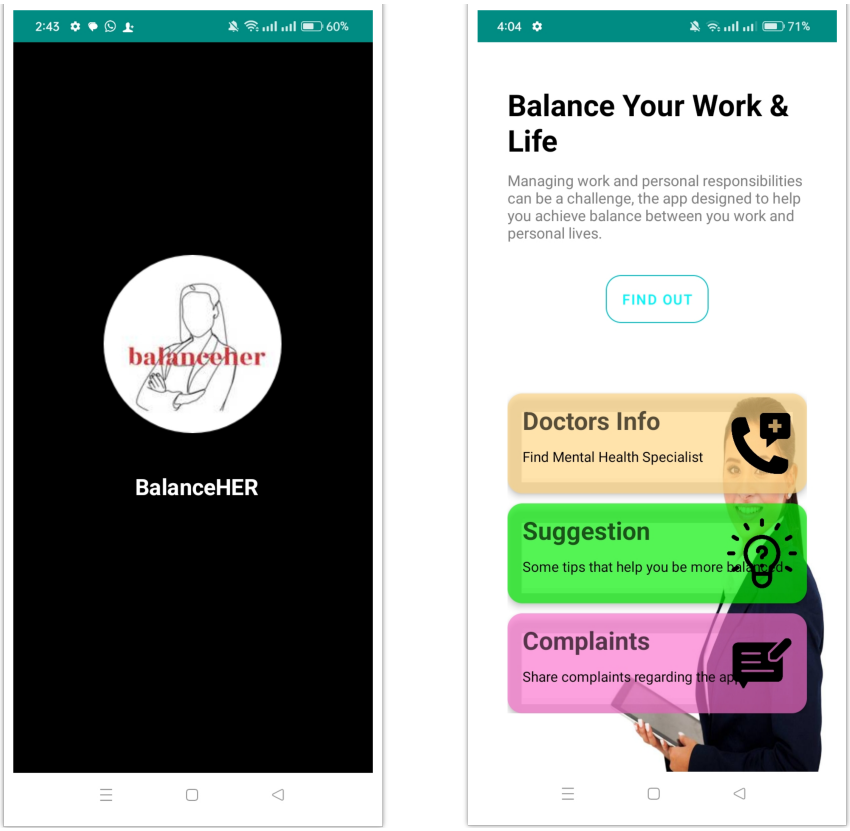


Figure 4.13: User interface(a)

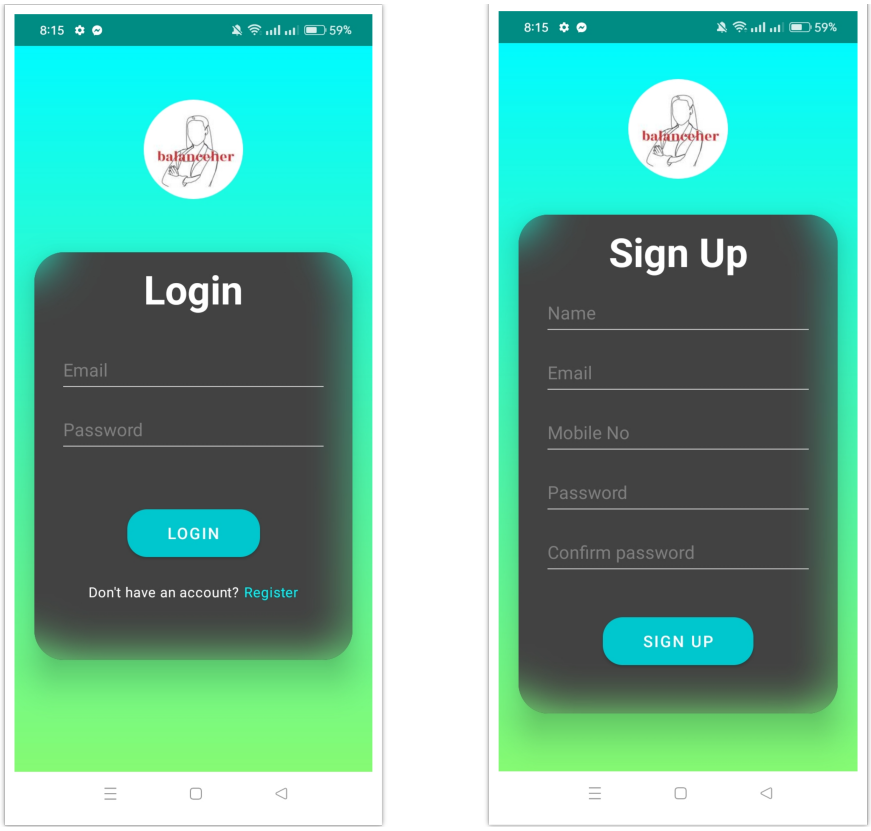


Figure 4.14: User Interface(b)

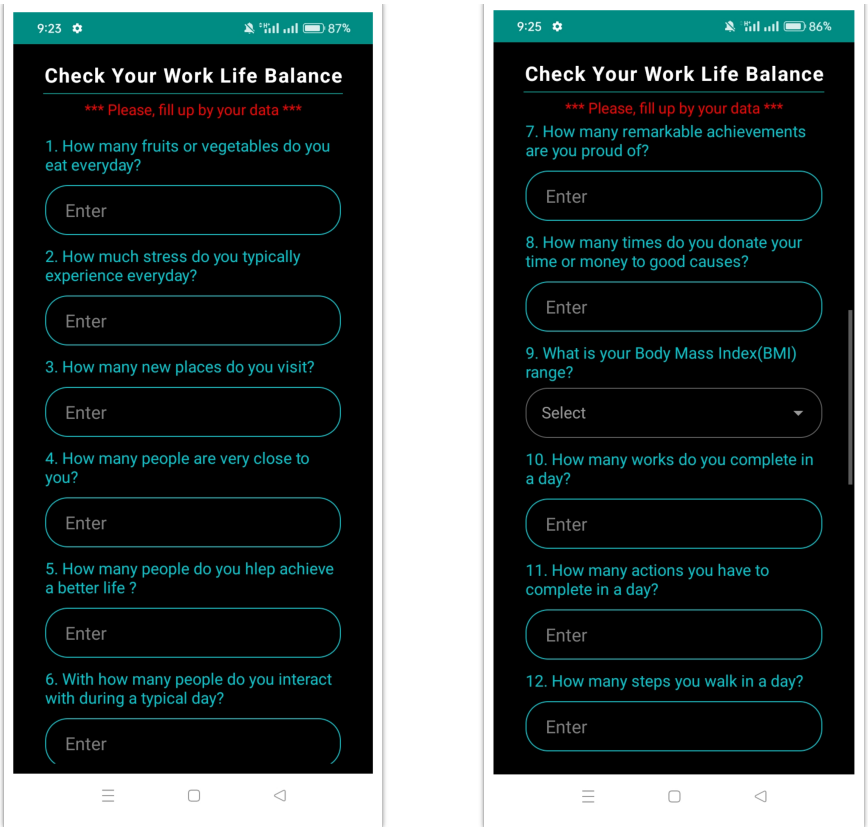


Figure 4.15: Prediction interface(a)

4.4.4.2 Prediction Interface

In this section, where fig. 4.15 and fig. 4.16 shows that, the logged in users will able to share their information that will be used for prediction with our system by filling up the form. And their given information will be taken to the backend where the model is deployed through API. Then the model's result will be shown through the application via another API.

Then the application will show weather the user is having a balanced life or not. If she is having an imbalanced life, and she is suffering from any mental issuses due to this, then she can take help from the doctor information page where she can get the information to meet with any doctor.

4.4.4.3 Doctor Information

Doctor Information system is where nearby doctors' chamber address and contact information will be shared through where they can take medical help from

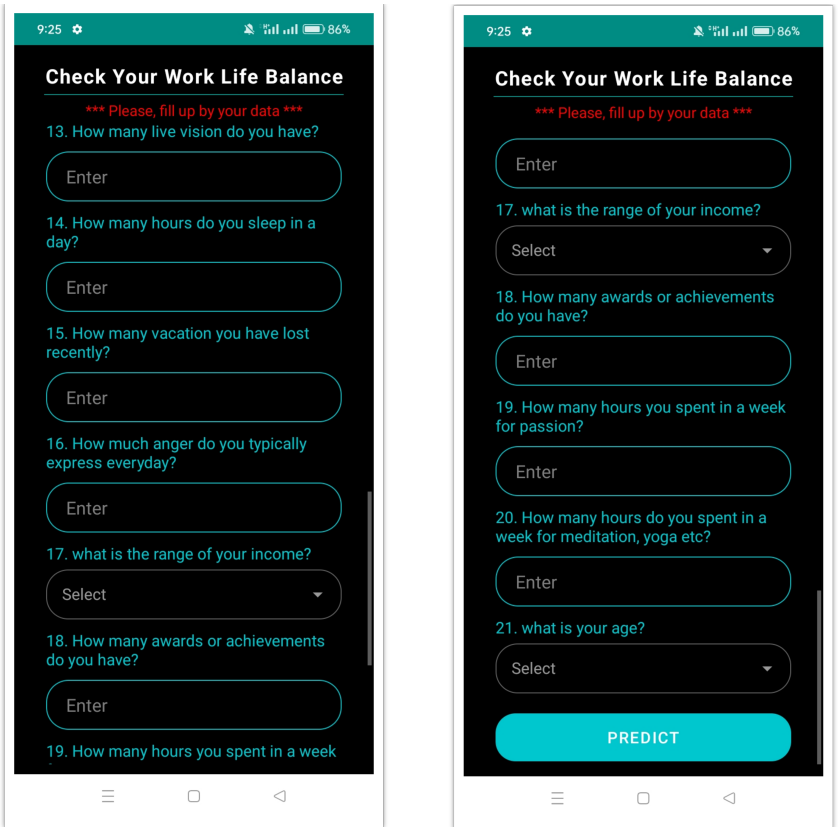


Figure 4.16: Prediction Interface(b)

the doctors in offline method. fig. 4.17 shows the doctors' information which is available in our system.

4.4.4.4 Suggestion Box

There is an interface in our application named suggestion box that can help a woman to make her life more balanced. Different tips are given that can make one's work life and personal life more balanced. There are some information like, how to manage stress level, how to boost self esteem, how to manage finance, etc. The fig. 4.18 represents the suggestion box interface of our app.

4.4.4.5 To Do List Module

Anyone, including women, can manage their duties, obligations, and goals successfully by using a to-do list. In this module, user can add their tasks' description with title and set time for that specific task. Then, user will get notify at the right moment. In fig. 4.19 , we can see that there is a add button in the lower

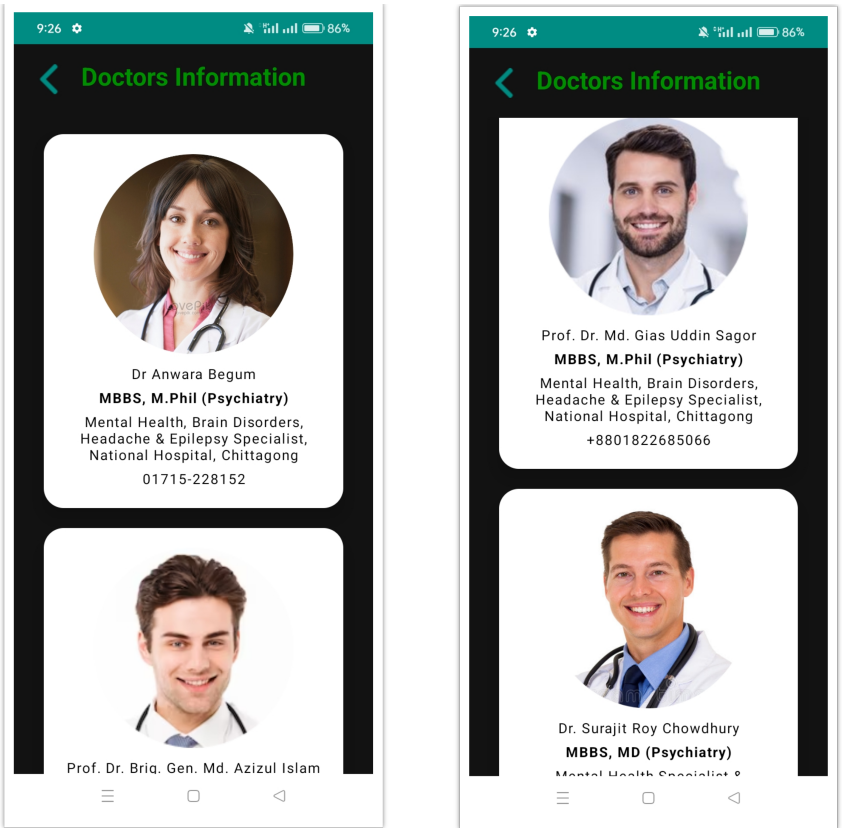


Figure 4.17: Doctor Information Interface

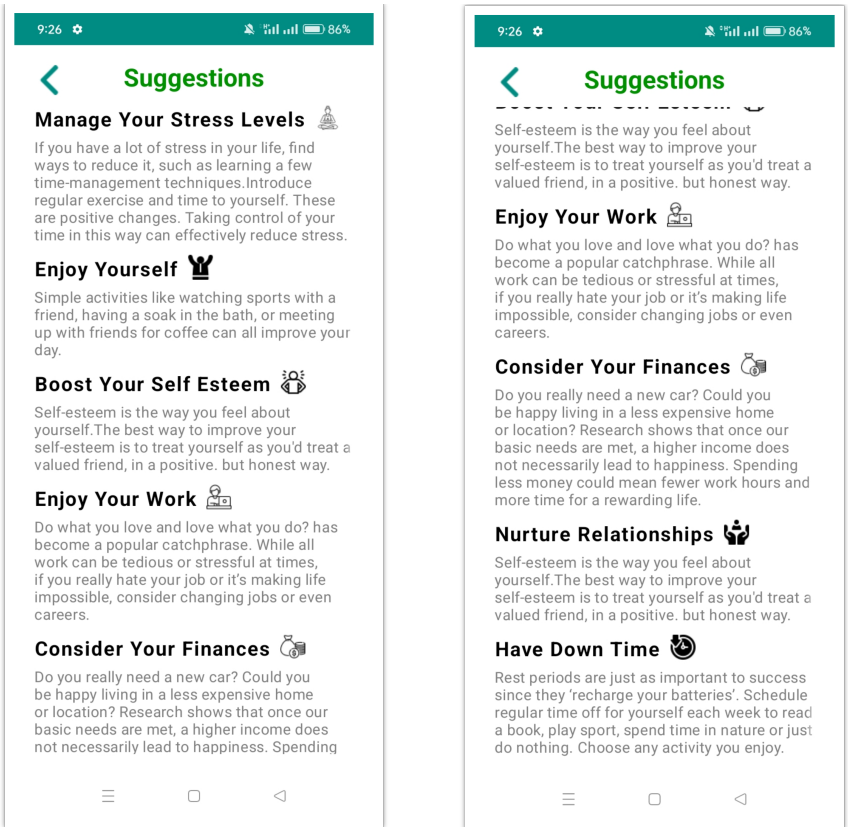


Figure 4.18: Suggestion Box Interface

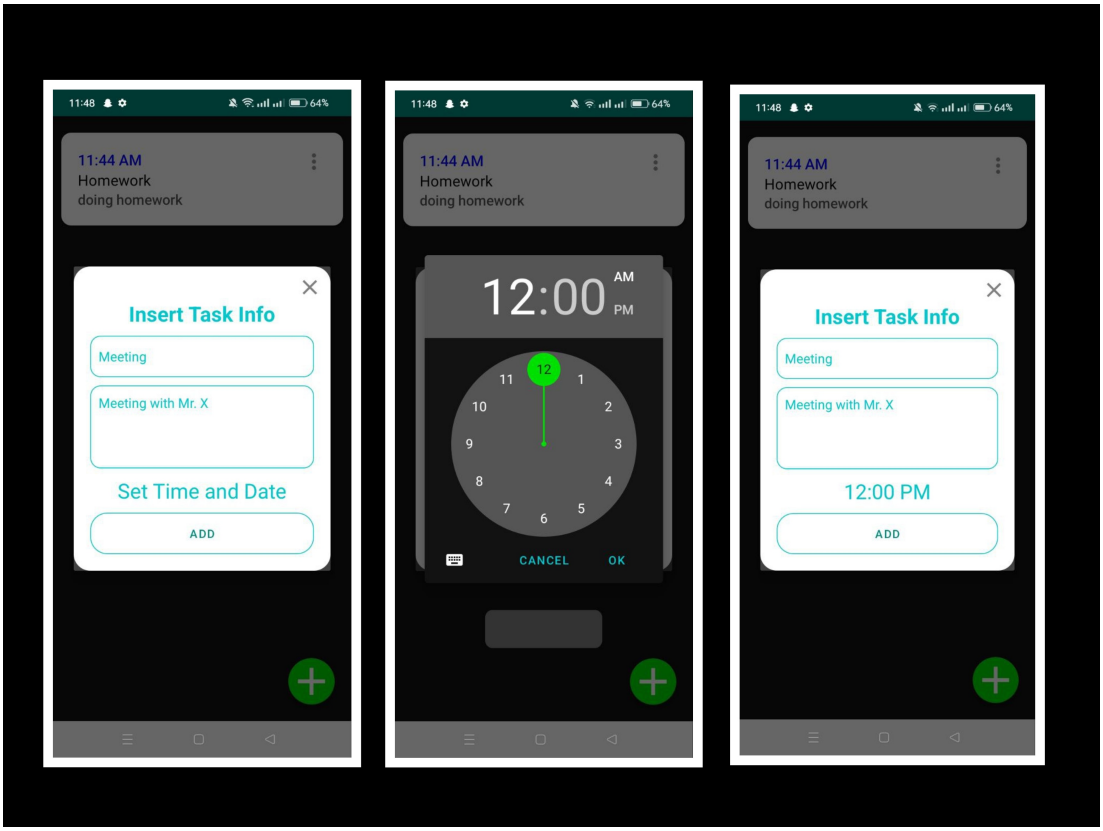


Figure 4.19: To Do List Module(a)

portion of the page. If the user click on that, a insert task info page will pop up. There the user will add the task's title, description and also add time.

After adding the tasks' title, description and time, the user will get notification about that task. The notification will display the text,"You have a scheduled task now.". The user can also edit or delete her tasks. Fig. 4.20 and fig. 4.21, shows the to do list module which is available in our system.

4.4.4.6 Complaints

Our app includes a feature that enables users to send emails directly to the police in case of emergency situations or when they need to report a crime or suspicious activity. This feature seeks to give users a simple and quick way to contact help or report incidents to the proper authorities. Fig. 4.22, shows the complaints module which is available in our system.

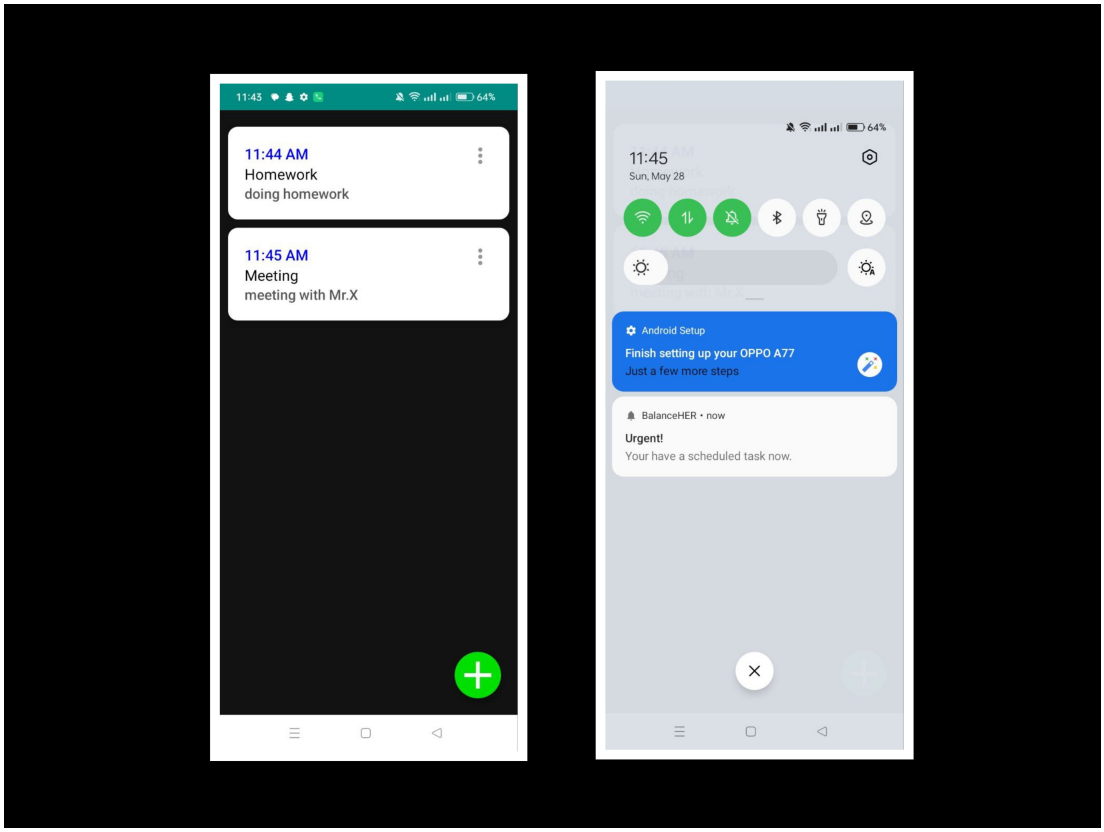


Figure 4.20: To Do List Module(b)

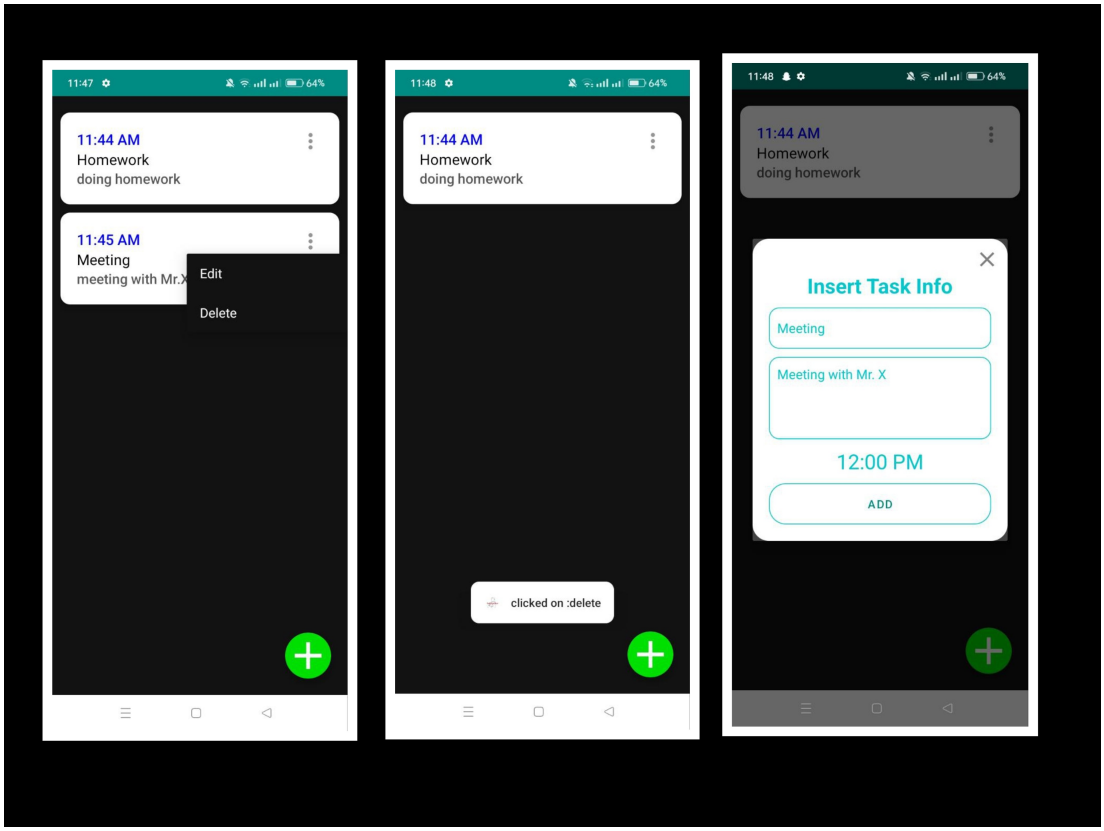


Figure 4.21: To Do List Module(c)

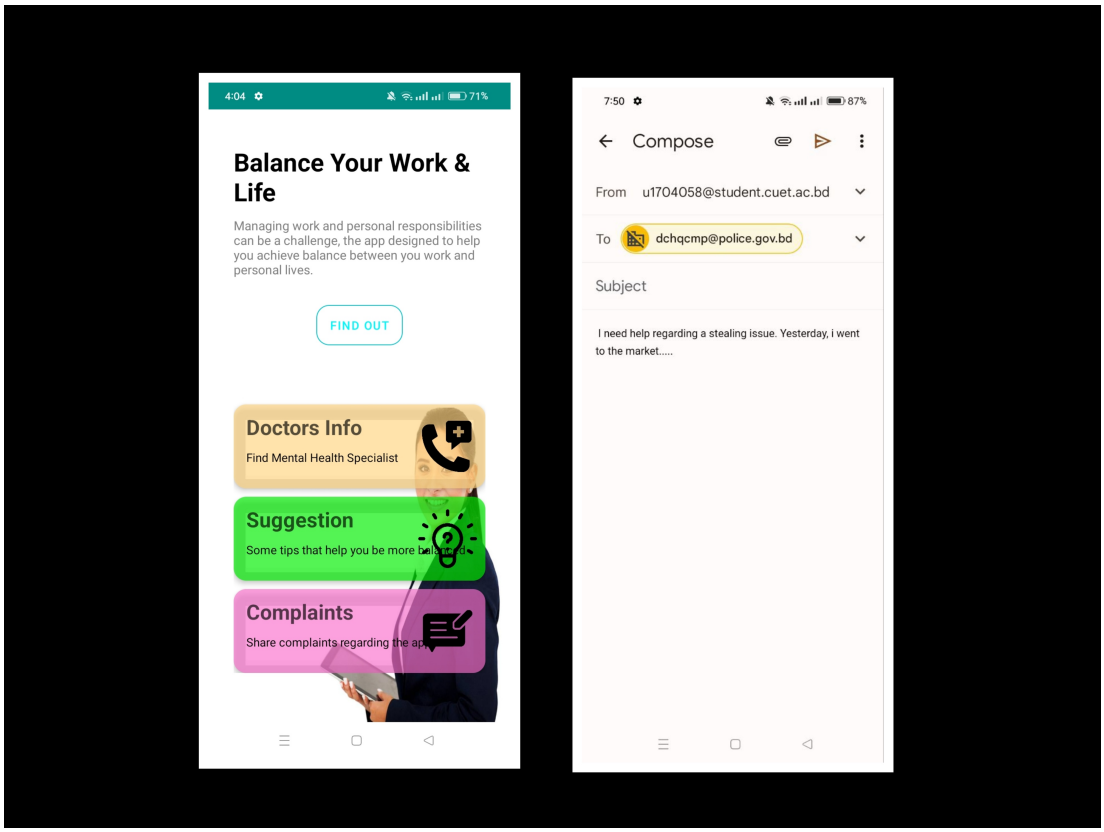


Figure 4.22: Complaints Module

4.5 Evaluation of Performance

One of a prediction model's key tasks is performance evaluation. table 4.1 shows the confusion matrix of Work Life Balance. A classification problem's prediction outcomes are compiled in a confusion matrix. Count values are used to describe the number of accurate and inaccurate predictions for each class. This is the confusion matrix's key. Whether a model is operating correctly or not is certified. Generally speaking, the key evaluation matrices are accuracy, precision, recall and f1-score. The equations are given below:

	predicted as imbalanced life	predicted as balanced life
actual as balanced life	True Negative	False Positive
actual as imbalanced life	False Negative	True Positive

Table 4.1: Confusion matrix of Work Life Balance

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$F1 = \frac{2 * Precision * Recall}{Precision + Recall} = \frac{2 * TP}{2 * TP + FP + FN}$$

- **TP:** TP (True Positive) indicates the number of women who have been correctly identified as having balanced work life.
- **TN:** TN (True Negative) is a measure of the number of women who have been appropriately categorised but are having an imbalanced work life.
- **FP:** FP is a measure of the number of women who were mistakenly labeled as having balanced life when, in fact, they were not. A Type I mistake is often referred to as FP.
- **FN:** FN indicates the amount of women who were mistakenly excluded from having balanced life. A Type II mistake is often referred to as FN.

4.5.1 Performance Evaluation Using Model Accuracy and Confusion Matrix

Performance evaluation is a crucial step in the machine learning process. However, it is a challenging task. Therefore, it needs to be done carefully while implementing machine learning. Any research must include an evaluation of a machine learning algorithm. A machine learning algorithm may yield good results when evaluated using an accuracy score metric, but it may yield substandard results while evaluated using other metrics, such as logarithmic loss or any other metric. Classification accuracy is frequently used to evaluate a learning model's performance, although this is insufficient to evaluate a model completely.

In our research, we used the seven most well-liked machine learning classifiers and trained them to test the performance of machine learning models against our gathered and preprocessed dataset. We have computed the model accuracy, precision, recall, F1 score of each trained model for the evaluation process and to identify the best-performing model. We have divided the dataset into 75% and 25% for training and testing sequentially in order to evaluate performance in a

high-quality manner. Every model was tested with 25% of the data samples after it had been trained with 75% of the shuffled data samples.

For evaluating models and finding optimum result we have used K-Fold Cross Validation. Where value of K is 10. It is found in previous studies, that 10-fold cross validation is better for limited size dataset [18] As in our case we have also dealt with relatively smaller dataset, we have used 10-fold cross validation.

Classifier	**Accuracy**	**Precision**	**Recall**	**F1 Score**
K-Nearest Neighbor	98%	0.85	0.86	0.85
Decision Tree	95%	0.69	0.79	0.72
Support Vector Machine	97%	0.92	0.78	0.85
Random Forest	98%	0.92	0.78	0.85
Adaboost	97%	0.82	0.78	0.80
Gradient Tree Boosting	96%	0.74	0.84	0.78
Naïve Bayes	87%	0.59	0.86	0.62

Table 4.2: Performance Evaluation of ML Algorithms

Classifier	**Accuracy**	**Standard Deviation**
K-Nearest Neighbor	98.28%	1.34%
Decision Tree	95.52%	2.58%
Support Vector Machine	96.90%	1.29%
Random Forest	97.07%	1.10%
Adaboost	93.93%	1.69%
Gradient Tree Boosting	96.21%	2.29%
Naïve Bayes	92.24%	3.01%

Table 4.3: Performance Evaluation after 10-fold cross validation

From analyzing table 4.2 , it is seen that, KNN outperformed other six algorithms with 98% accuracy. It also stands alone with better precision, recall and F1 score. Figure 4.23 shows the performance comparison chart of all seven algorithms. The easiest performance metric to understand is accuracy, which is just the proportion of properly predicted observations to all observations. The model accuracy for K-Nearest Neighbor, Random Forest, Decision Tree, SVM, Adaboost, Gradient Tree boosting, Naive Bayes are respectively 98%, 98%, 95%, 97%, 97%, 96% and 87%. It is seen that the KNN algorithm and the Random Forest algorithm gives the best accuracy among three of these algorithms.

From table 4.3 we can see, after 10-Fold cross validation four algorithms stand out among seven ML algorithms which are KNN, Random forest, SVM and Gradient tree with accuracy of 98.28%, 97.07%, 96.90% and 96.21% respectively. In this case also, Gradient Tree performs slightly better then others. But here we can

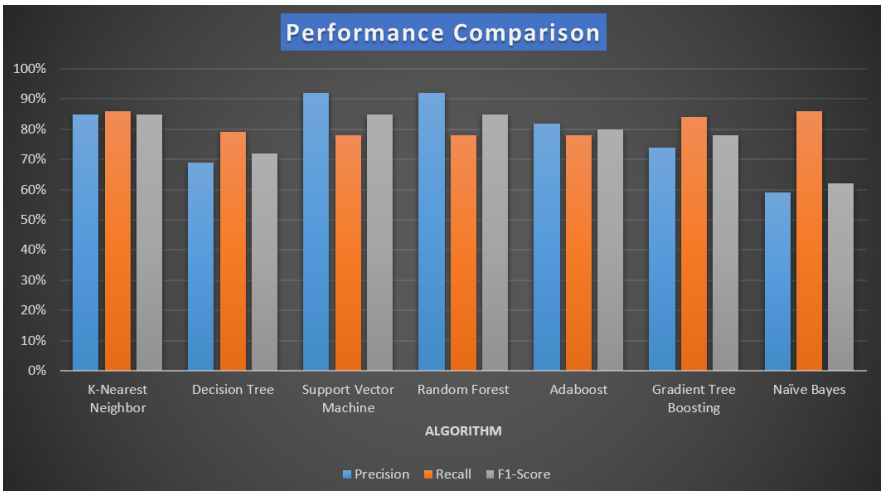


Figure 4.23: Performance Comparison(a)

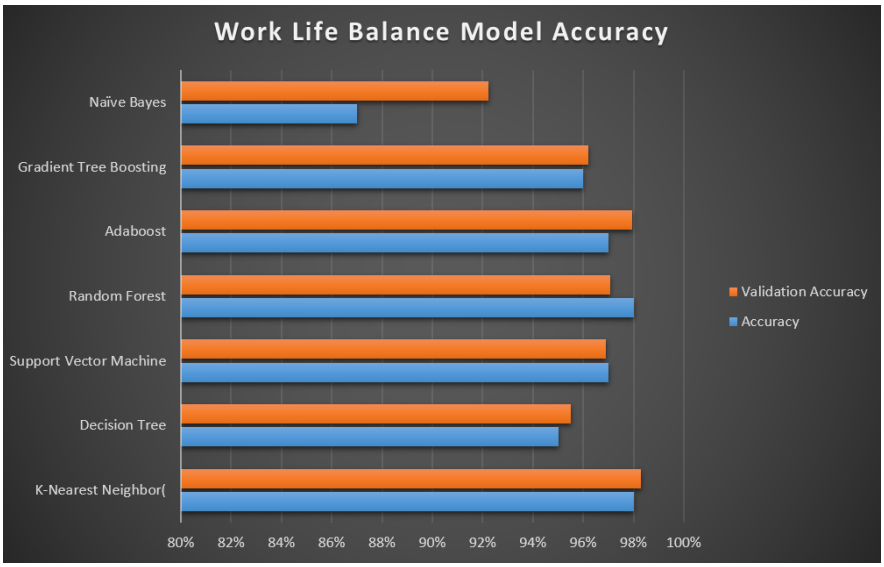


Figure 4.24: Performance Comparison(b)

see Random forest and Adaboost algorithm's accuracy after validation decrease. It can happen due to overfitting probelm. So we can say that K-Nearest Neighbor(KNN) is the best model. So, finally, the KNN classifier is selected as the predictive model of the proposed system. The fig. 4.24 shows performance of training accuracy and validation accuracy.

The fig. 4.25 represents the confusion matrix of K-Nearest Neighbor Algorithm. The prediction model using this algorithm can predict the 185 persons who have been correctly identified as having balanced life and the 5 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 2 women who were mistakenly labeled as having balanced work life

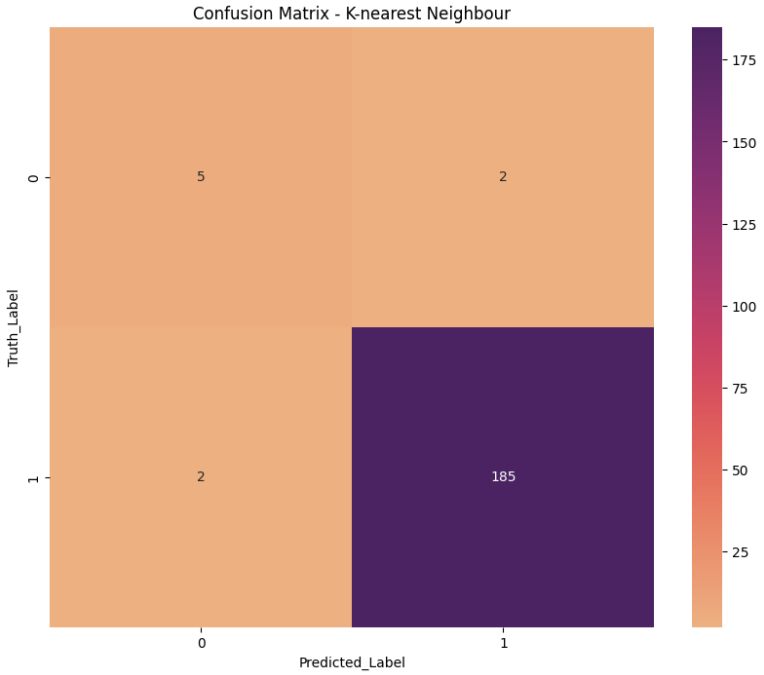


Figure 4.25: Confusion matrix of K-Nearest Neighbor Algorithm

when, in fact, they were not, and also predicted 2 women who was mistakenly excluded from list of having balanced work life.

In fig. 4.26, it represents the confusion matrix of Decision Tree Algorithm. The prediction model using this algorithm can predict the 181 persons who have been correctly identified as having balanced life and the 4 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 3 women who were mistakenly labeled as having balanced work life when, in fact, they were not, and also predicted 6 women who was mistakenly excluded from list of having balanced work life.

In fig. 4.27, it represents the confusion matrix of SVM Algorithm. The prediction model using this algorithm can predict the 187 persons who have been correctly identified as having balanced life and the 4 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 3 women who were mistakenly labeled as having balanced work life when, in fact, they were not, and also predicted 0 women who was mistakenly excluded from list of having balanced work life.

In fig. 4.28, that represents the confusion matrix of Random forest Algorithm. The prediction model using this algorithm can predict the 187 persons who have

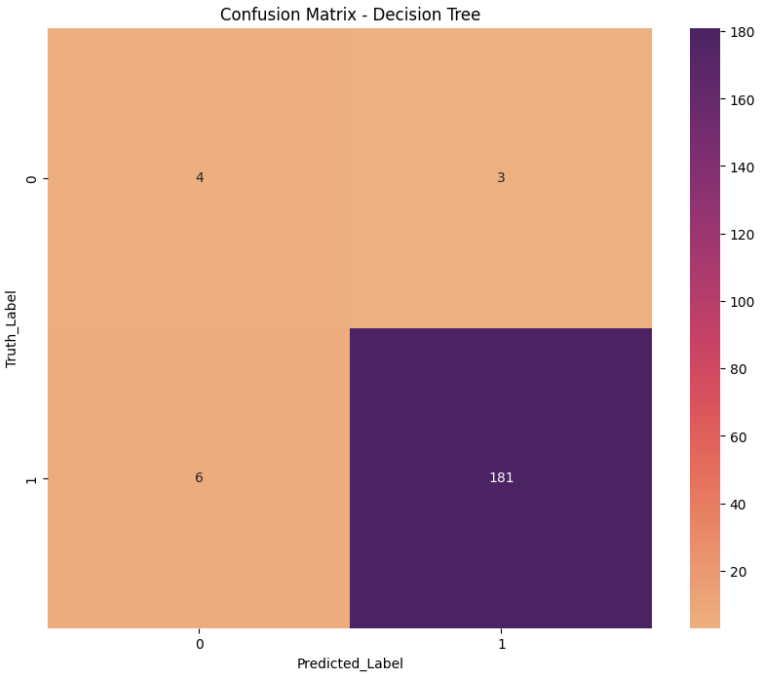


Figure 4.26: Confusion matrix of Decision Tree Algorithm

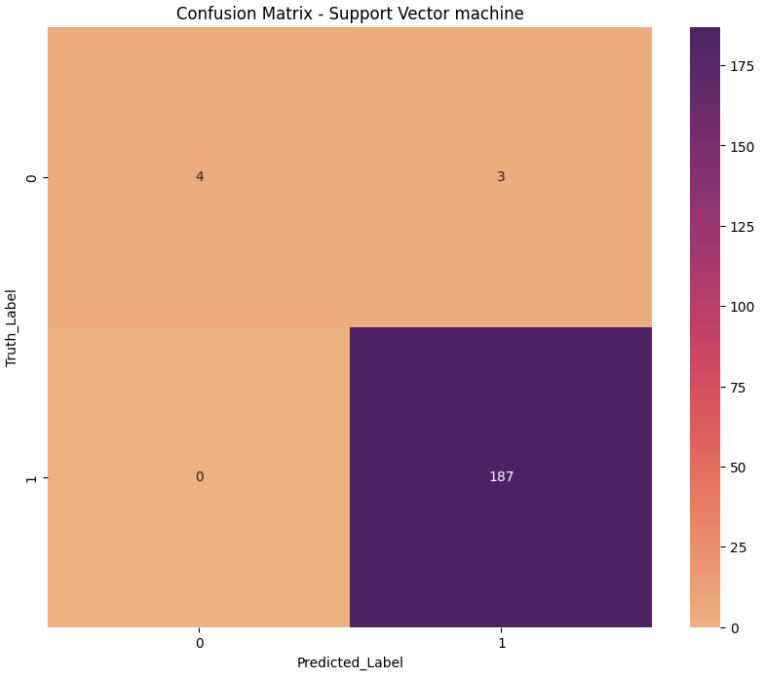


Figure 4.27: Confusion matrix of SVM Algorithm

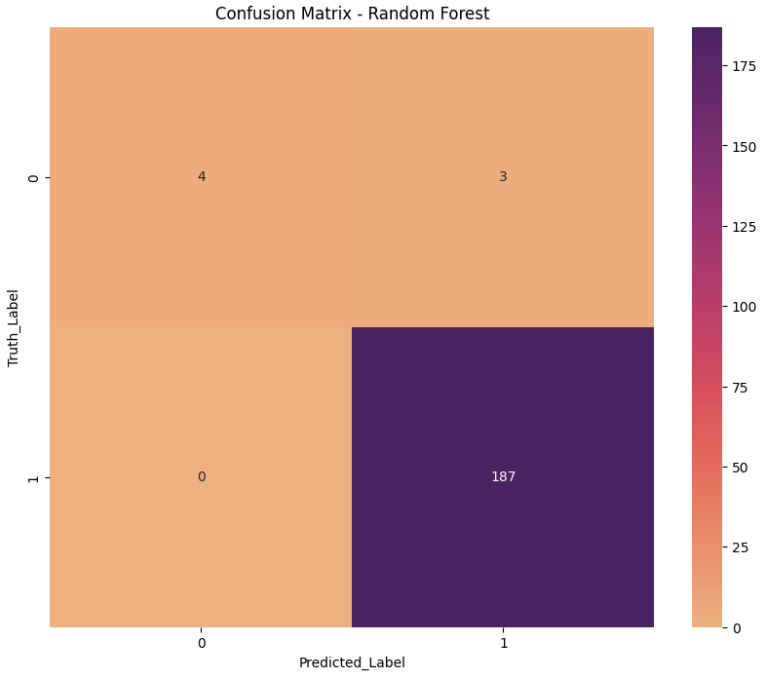


Figure 4.28: Confusion matrix of Random forest Algorithm

been correctly identified as having balanced life and the 4 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 3 women who were mistakenly labeled as having balanced work life when, in fact, they were not, and also predicted no women who was mistakenly excluded from list of having balanced work life.

The fig. 4.29, it represents the confusion matrix of Adaboost Algorithm. The prediction model using this algorithm can predict the 185 persons who have been correctly identified as having balanced life and the 4 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 3 women who were mistakenly labeled as having balanced work life when, in fact, they were not, and also predicted 2 women who was mistakenly excluded from list of having balanced work life.

From the fig. 4.30, we see the confusion matrix of Gradient Tree Boosting Algorithm. The prediction model using this algorithm can predict the 182 persons who have been correctly identified as having balanced life and the 5 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 2 women who were mistakenly labeled as having balanced work life when, in fact, they were not, and also predicted 5 women who was mistakenly

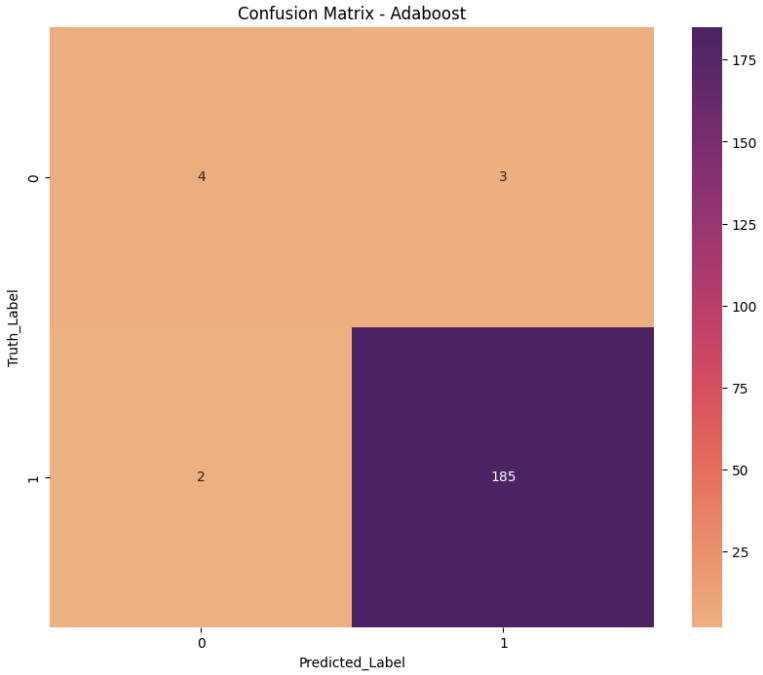


Figure 4.29: Confusion matrix of Adaboost Algorithm

excluded from list of having balanced work life.

The fig. 4.31 represents the confusion matrix of Naive Bayes Algorithm. The prediction model using this algorithm can predict the 163 persons who have been correctly identified as having balanced life and the 6 women who have been appropriately categorized but are having an imbalanced work life. This model predicted 1 women who were mistakenly labeled as having balanced work life when, in fact, they were not, and also predicted 24 women who was mistakenly excluded from list of having balanced work life.

4.5.2 Performance Evaluation Using AUC-ROC Curve

A measurement tool for binary classification issues is the Receiver Operator Characteristic (ROC) curve. In essence, it separates the "signal" from the "noise" by plotting the TPR against the FPR at different threshold values. The capacity of a classifier to differentiate between classes is measured by the Area Under the Curve (AUC), which is used as a summary of the ROC curve. A higher X-axis value on a ROC curve denotes more false positives than true negatives. A larger Y-axis value, however, denotes a greater proportion of true positives compared to false negatives. The capacity to strike a balance between false positives and

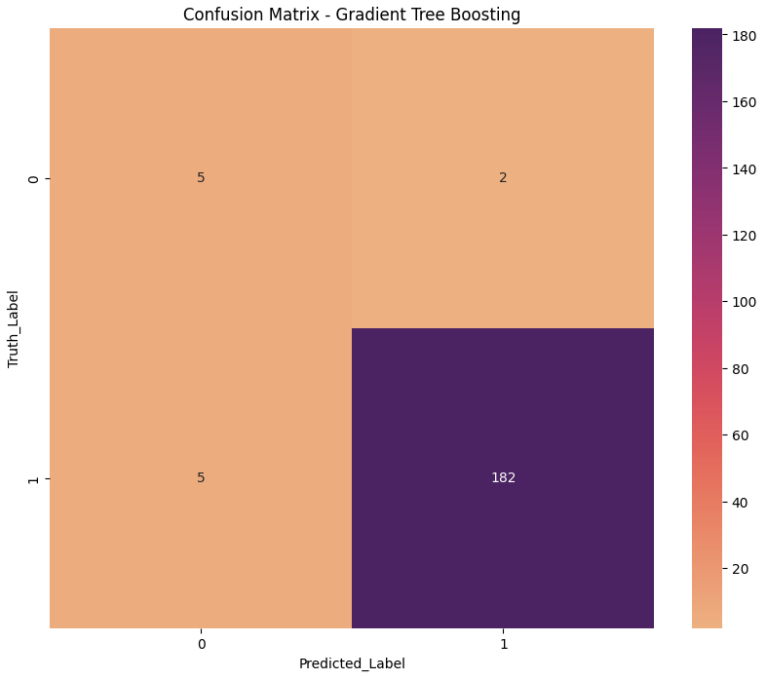


Figure 4.30: Confusion matrix of Gradient Tree Boosting Algorithm

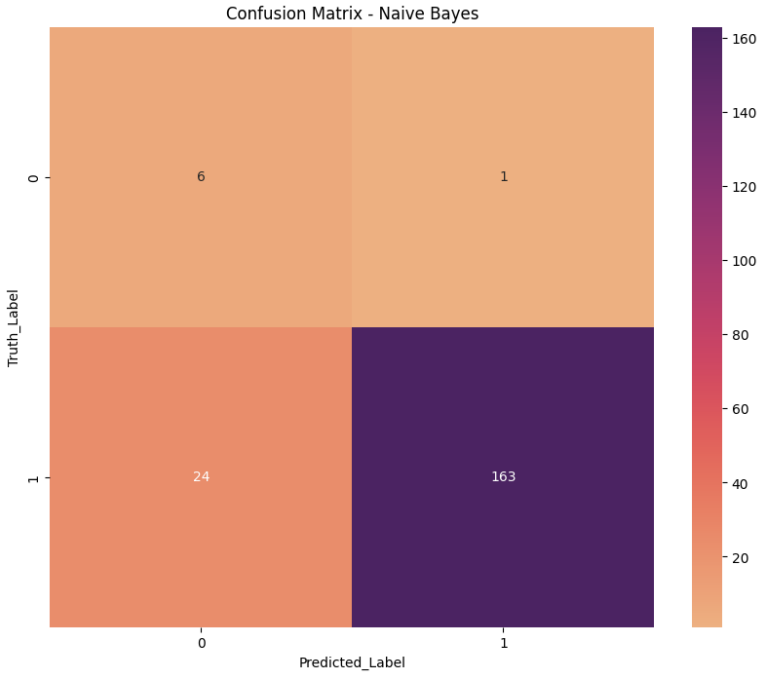


Figure 4.31: Confusion matrix of Naive Bayes Algorithm

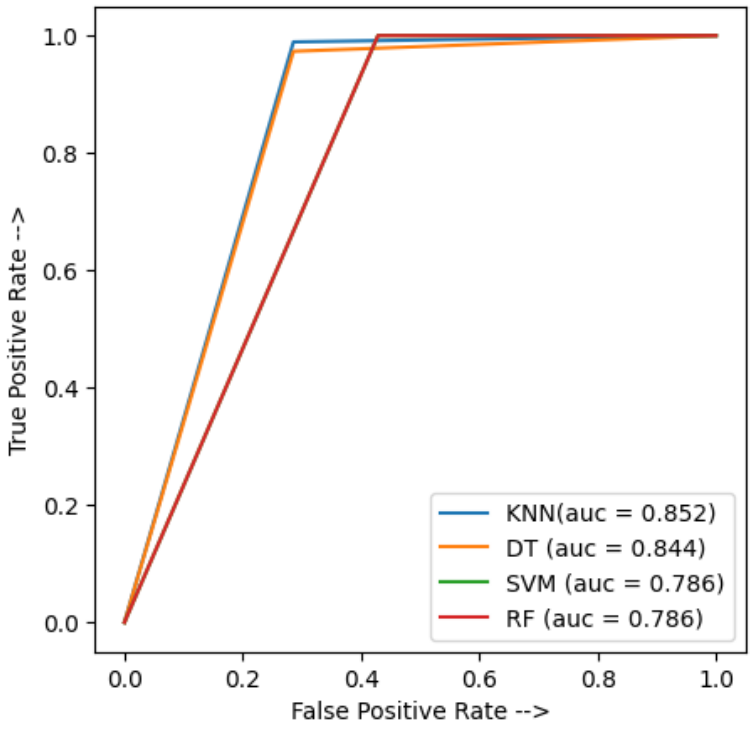


Figure 4.32: ROC-AUC curve of four algorithms model

false negatives will therefore influence the threshold selection. The equation of TPR and FPR is given below:

$$TPR = \frac{TP}{TP + FN}$$

$$FPR = \frac{FP}{TN + FP}$$

Where, TPR denotes True Positive Rate.

FPR denotes False Positive Rate.

In this fig. 4.32, we can see that KNN provided the best performance as it occupies the curve at the far left and top. And also, AUC for K-Nearest Neighbor, Decision Tree, SVM, Random Forest algorithms are respectively, 86%, 84%, 78% and 78%. In a nutshell, the ROC curve is commonly used in the machine learning community to compare the performance of different learning models. It is seen that the KNN algorithm gives the best AUC among four of these algorithms. AUC = 1 represents the best model performance.

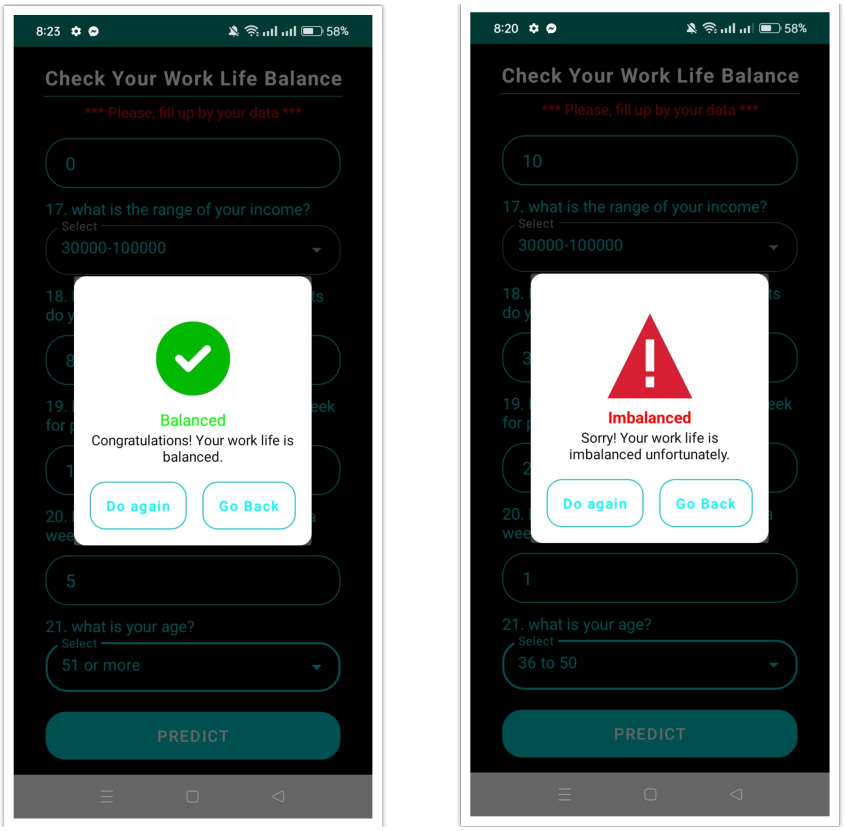


Figure 4.33: Prediction results are displayed in the mobile app

4.5.3 Performance Evaluation for App Prediction

The fig. 4.33 represents the final result displayed in the mobile application of our system. In our model, the KNN algorithm gives us the best result compared with the other algorithms. The model can predict an user is having a balanced work life or not, and the result is displayed in the mobile app. If a particular user is having balanced work life, then the message "Congratulations! Your work life is balanced" will be displayed. Otherwise, a message "Sorry! Your work life is imbalanced unfortunately" will be displayed. If so, then she can visit our other to interface such as doctor info and suggestion box that can help her to get rid of imbalanced work life.

4.6 Conclusion

The findings and analysis from the use and assessment of the work-life balance prediction app have been reported in the Results and Discussions portion of this thesis. We have learned a lot about the app's predictive powers and prospective

effects on work-life balance by looking at its usage, performance data, and implications. Several important conclusions were drawn from the study of the data. First off, according to performance indicators like validation accuracy and precision, the software showed promise in its predictive accuracy. Although there was a little difference in accuracy between training and validation, this is to be expected and can be related to the characteristics of the dataset and the difficulty of the goal of predicting work-life balance. The debate of the findings has also clarified the app's potential effects on people, organizations, and society as a whole. The software has the ability to assist users in better managing their work-life balance by offering individualized insights and recommendations. This can then result in lower stress levels, better health, and greater job satisfaction. Additionally, businesses may use the app's information to create tailored policies and procedures that encourage a positive workplace culture and improve employee engagement and retention.

Chapter 5

Conclusion

5.1 Conclusion

We tend to get less sleep and work more due to the mounting pressure of modern life, despite the damaging consequences it has on our health. An inadequate work-life balance in midlife may have harmful effects decades later, according to a recent study [9]. Numerous studies have demonstrated the detrimental impact of extended work hours on one's physical and mental health. On the other hand, flexible work schedules and hours that provide employees greater control over their time have been demonstrated to improve health and well-being.

Flexible work schedules, remote working possibilities, and telecommuting can all help with work-life balance. Productivity will rise as a result, and expenses will drop. Additionally, a positive work-life balance might result in; a higher level of job satisfaction, employee's attraction for work, a greater sense of commitment to work. For this reason, our intention was to build an assistance system for women that can help them to lead a balanced work and personal life.

The first chapter included a general overview of the study. The reader was introduced to relevant topics in this chapter. Both the issue discovered and the method used to complete this assignment are briefly explained. Under the heading of inspiration, the importance and contribution of the work are emphasised.

The earlier research on this subject is covered in the following chapter. The research's experimental techniques, data collection process, flaws, and algorithms are discussed in relation to the findings. The researchers address a variety of machine learning algorithms that they have presented.

The strategy employed to achieve the work's objective is thoroughly stated and

discussed in chapter 3. The collection of fundamental attributes is where the approach starts. Then the dataset prepared. The acquired data is preprocessed and converted to a numeric values for computing. The final component set is then determined by analyzing the numerical data that has been accumulated. Nothing is altered or removed throughout this process. The attributes and data are then loaded into various machine learning models, that are then tested to see which model performs the best. The idea is implemented into a mobile application and tested for usability.

We described in detail in chapter 4, that how our methodology produced the results we saw. At the start of the chapter, the dataset is discussed. Additionally, the findings of the data gathering and key findings from the data have been analyzed. The results of the model evaluation are then presented, together with our research's conclusions.

5.2 Future Work

This study has a lot of potential to be upgraded in future. They are outlined below:

- **Dataset Quantity:** There are only 774 instances in the dataset. If there had been additional data available for the procedure, the analysis's findings would have been more precise.
- **Dataset Variance:** Not only should the number of data instances be increased, but data should also have been gathered from a variety of sources to enhance diversity.
- **Applying RL:** Implementation of Reinforcement Learning agent instead of conventional machine learning model can be done and discussed.

References

[1] S. G. Carmichael, 'The research is clear: Long hours backfire for people and for companies,' *Harvard Business Review*, vol. 19, pp. 2–4, 2015 (cit. on p. 2).

[2] M. Tasnim, M. Z. Hossain and F. Enam, 'Work-life balance: Reality check for the working women of bangladesh,' *Journal of Human Resource and Sustainability Studies*, vol. 5, no. 1, pp. 75–86, 2017 (cit. on p. 2).

[3] B. Mahesh, 'Machine learning algorithms-a review,' *International Journal of Science and Research (IJSR).[Internet]*, vol. 9, pp. 381–386, 2020 (cit. on p. 2).

[4] A. Pawlicka, M. Pawlicki, R. Tomaszewska, M. Choraś and R. Gerlach, 'Innovative machine learning approach and evaluation campaign for predicting the subjective feeling of work-life balance among employees,' *PloS one*, vol. 15, no. 5, e0232771, 2020 (cit. on pp. 3, 11).

[5] P. J. Grace and N. N. Banu, 'Machine learning on emotional intelligence and work life balance,' *International Journal of Computer Applications*, vol. 116, no. 10, 2015 (cit. on p. 3).

[6] G. Guo, H. Wang, D. Bell, Y. Bi and K. Greer, 'Knn model-based approach in classification,' in *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, Springer, 2003, pp. 986–996 (cit. on p. 3).

[7] A. Priyam, G. Abhijeeta, A. Rathee and S. Srivastava, 'Comparative analysis of decision tree classification algorithms,' *International Journal of current engineering and technology*, vol. 3, no. 2, pp. 334–337, 2013 (cit. on p. 3).

[8] C. Gupta, K. R. Rao and P. Datta, 'Support vector machine based prediction of work-life balance among women in information technology organizations,' *IEEE Engineering Management Review*, vol. 50, no. 2, pp. 147–155, 2022 (cit. on pp. 4, 13).

[9] S. Gaikwad, L. Swaminathan and S. George, 'Impact of work-life balance on job performance - analysis of the mediating role of mental well-being and work engagement on women employees in it sector,' in *2021 International Conference on Decision Aid Sciences and Application (DASA)*, 2021, pp. 204–209. DOI: 10.1109/DASA53625.2021.9681920 (cit. on pp. 11, 58).

[10] M. Sujithra, P. Velvadivu, J. Rathika, R. Priyadharshini and P. Preethi, 'A study on psychological stress of working women in educational institution

using machine learning,' in *2022 13th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, 2022, pp. 1–7. DOI: 10.1109/ICCCNT54827.2022.9984460 (cit. on p. 11).

[11] S. Rajagopalan and L. Rajamani, 'A fuzzy logic rule based forecasting model: Work-life balance in it among software vs. services industry on the view of women software engineer,' in *2013 International Conference on Machine Intelligence and Research Advancement*, IEEE, 2013, pp. 241–246 (cit. on p. 12).

[12] S. D. Paigude and S. Shikalgar, 'Deep learning model for work life balance prediction for working women in it industry,' (cit. on p. 12).

[13] R. Bucea-Manea-Ţoniş, R. Bucea-Manea-Ţoniş, V. E. Simion, D. Ilic, C. Braicu and N. Manea, 'Sustainability in higher education: The relationship between work-life balance and xr e-learning facilities,' *Sustainability*, vol. 12, no. 14, p. 5872, 2020 (cit. on p. 12).

[14] K. Radha and M. Rohith, 'An experimental analysis of work-life balance among the employees using machine learning classifiers,' *arXiv preprint*, 2021 (cit. on p. 12).

[15] W. Sung, 'Improving work life balance through the use of smart work experience,' in *2021 21st ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*, IEEE, 2021, pp. 163–166 (cit. on p. 13).

[16] P. Sadana and D. Munnuru, 'Machine learning model to predict work force attrition,' in *2021 6th International Conference for Convergence in Technology (I2CT)*, 2021, pp. 1–6. DOI: 10.1109/I2CT51068.2021.9418140 (cit. on p. 13).

[17] J. A. Westwood and J. A. Cazier, 'Work-life optimization: Using big data and analytics to facilitate work-life balance,' in *2016 49th Hawaii International Conference on System Sciences (HICSS)*, IEEE, 2016, pp. 1701–1709 (cit. on p. 13).

[18] H. L. Vu, K. T. W. Ng, A. Richter and C. An, 'Analysis of input set characteristics and variances on k-fold cross validation for a recurrent neural network model on waste disposal rate estimation,' *Journal of environmental management*, vol. 311, p. 114 869, 2022 (cit. on p. 48).