B.Sc. Engineering Thesis Defense

Work-Life Balance Prediction For Academic Women In Bangladesh Using Machine Learning

Submitted by

Ireen Ara Haque

MUH1901046F

Under the Supervision of

Dr. Nazia Majadi

Associate Professor

Dept. of CSTE, NSTU

Department of Computer Science and Telecommunication Engineering

Noakhali Science and Technology University

Submission Date:14/03/2024

Work-Life Balance Prediction For Academic Women In Bangladesh Using Machine Learning

RECOMMENDATION

This thesis titled, **"Work-Life Balance Prediction For Academic Women In Bangladesh Using Machine Learning"** submitted by Ireen Ara Haque, Roll No: MUH1901046F, Session: 2018-19, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Telecommunication Engineering as B.Sc. Engg. (CSTE).

Approved By

Thesis and Oral Examination Board

1st Examiner

1. ……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

Department of Computer Science and Telecommunication Engineering, NSTU

2nd Examiner

2. ……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

Department of Computer Science and Telecommunication Engineering, NSTU

3rd Examiner

3. ……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

Department of Computer Science and Telecommunication Engineering, NSTU

External

4. ……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

STATEMENT OF ORIGINALITY

This is to certify that the work presented in this thesis entitled, **“Work-Life Balance Prediction For Academic Women In Bangladesh Using Machine Learning”**, is the outcome of the research carried out by **Ireen Ara Haque** under the supervision of **Dr. Nazia Majadi**, Associate Professor, Department of Computer Engineering and Telecommunication Engineering, Noakhali Science and Technology University (NSTU), Noakhali-3814, Bangladesh.

It is also declared that neither this thesis nor any part thereof has been submitted anywhere else for the award of any degree, diploma, or other qualifications.

Supervisor
Dr. Nazia Majadi
Associate Professor,
Department of Computer Engineering and TelecommunicationEngineering,
Noakhali Science and Technology University

ACKNOWLEDGEMENT

I would like to begin by sincerely thanking Allah Ta'ala for granting me the courage and willpower to start and complete this thesis-
“Work-Life Balance Prediction For Academic Women In Bangladesh Using Machine Learning”.
Next, I want to express my sincere gratitude to my supervisor, **Dr. Nazia Majadi**, for her innovative ideas, invaluable guidance, and unwavering support. Completing this thesis would have been challenging without her assistance. I am truly grateful for her willingness to dedicate her time to help me in such a thoughtful manner.
In addition, I would like to thank all the teachers in the Department of Computer Science and Telecommunication Engineering (CSTE) for their thoughtful advice and assistance with the coursework.
Lastly, I would like to sincerely thank my parents for their love, dreams, and sacrifices throughout my life.

Abstract

Work-life balance is a pressing issue for women employees in the 21st century due to conflicting demands between personal and professional responsibilities. This study focuses on examining the work-life balance of female employees in Bangladesh's educational institutions using machine learning techniques. Data from 205 women employees was collected through a structured questionnaire and preprocessed by label encoding and null value removal. Various machine learning classifiers, including Logistic Regression, Decision Tree, Naive Bayes, Support Vector Machine Algorithm, Random Forest, and K-Nearest Neighbor (KNN), were employed to predict work-life balance. To address the imbalanced dataset, different techniques were applied, prioritizing not only accuracy but also precision, recall, and F1-score. The Naive Bayes classifier yielded the best results with 90% accuracy, 100% precision, and 95% recall on the original dataset.

Keywords: Work-life balance, women employees, educational institutions, conflicting demands, machine learning, Naive Bayes, Logistic Regression.

Table of Content

List of Figures

List of Tables

Chapter 1

Introduction

The introduction is initiated by seeking the background of the problem and setting the stage for the upcoming discussion. Afterward, the problem is explained, and the issues involved are highlighted. Additionally, the aims and objectives of the research are clearly stated, providing a roadmap for the study.

1.1 Background

In the past, women could only handle domestic tasks like cooking, cleaning, laundry, child care, etc. They were confined to homemaking and denied the right to leave home. But things have changed since then. They have a big job to play even outside the home in addition to being the homemaker. Many families became multiple earners due to the rising expense of living on the one hand and the enhanced education and employment options on the other. This led to the husband and wife starting their careers. In this competitive world, women collaborate with men in various life domains while managing domestic responsibilities. These changes in workplace culture have given women employees extra obligations and responsibilities for their families, both in the workplace and in society [1].

Even as women enter the workplace in increasing numbers, they continue to do the majority of unpaid family roles. As a result, it has become harder for female employees to balance work and family obligations, especially when no strategies are accessible. Work and family are important life roles that should ideally be balanced rather than seen as competing for time and resources [2].

Bangladesh is experiencing a problem with women's work-life balance, much like every other nation on earth [2]. Employees who are female will not be able to look forward to career advancement without maintaining a balance between work and personal life, and an organization's productivity will also be hindered.

1.2 Problem and Motivation

Work-life balance is the division of an employee's personal and work life within an institution and is an ongoing endeavor to live a balanced life. It describes the ability of an employee to set priorities in their personal and professional lives. Professional life entails dedication to an institution, career advancement, and fruitful performance. Personal life includes friendships, family, personal growth, and physical fitness and health.

In the competitive era, the organization’s expectations from the employees are increasing day by day. Employees are required to put forth extra effort and concentrate more on their work to meet the employer's demands, which is causing a work-life imbalance [1]. Women prioritize their families over their careers [3]. Work-life balance (WLB) or juggling work, family, and social obligations with self-care, is a significant problem for working women [4]. Work-life balance was seen as the main cause of unhappiness for working women employees as well as a source of distress [1]. Regardless of work schedules, it can be challenging for full-time employees to maintain a healthy work-life balance, especially if they have children. Therefore, organizations must devise plans to support female employees in loving their jobs and living life to the fullest. Hence, if women maintain a good WLB, it not only helps them have successful families but also increases their commitment to the success of their organization [4].

1.3 Aims and Objectives

Due to its effects on both personal and professional life, the issue of work-life balance has attracted the attention of researchers and academics in today's competitive world. According to research, while preserving a healthy work-life balance encourages harmony in all aspects of one's life, doing so may hurt one's personal life, which can cause job discontent and affect an organization's reputation and productivity. The goal of this study is to examine how working women combine their jobs and personal lives in a variety of educational institutions. The study's specific objectives are as follows:

- Explore the current representation of women in various academic disciplines in Bangladesh.
- Assess the challenges and opportunities for women in academia regarding work-life balance.

- To predict if women working in academia in Bangladesh can maintain a work-life balance using machine learning.

In Bangladesh, women have significantly boosted the country's economy. Despite this, there are still many obstacles that working women must overcome in both their personal and professional lives. To create a work-life balance predictor for working women in Bangladesh, this study uses machine learning. This project aims to share knowledge about using ML techniques to predict factors affecting work-life in academia.

1.4 Thesis Organization

The rest of the thesis is organized as follows: Chapter 2 encompasses the literature review of the research. In Chapter 3, the research methodology is addressed, encompassing parameter selection, questionnaire preparation, data collection, data analysis, prediction modeling, and evaluation metrics for predicting outcomes. Furthermore, Chapter 4 examines the expected result. Finally, in Chapter 5, the concluding remarks are provided.

Chapter 2

Literature Review

Numerous researchers have made significant contributions to the prediction of the factors that influence women's personal and professional lives in a variety of ways. Most of them used several machine learning models, such as logistic regression, decision trees, naive bayes, random forests, etc.
The application of machine learning (ML) approaches in OB research for decision-making has been documented in previous literature. Organizational sciences benefit from the smart application of machine learning techniques. Researchers have developed several models using machine learning to find out the factors affecting work and personal life.

Pawlicka, et al. [5] employed machine learning to predict work-life balance feelings among 800 employees in Poland, using an artificial neural network. They achieved 81% accuracy by identifying factors such as working hours, free time, weekend work, self-employment, and financial assessment. This research, one of Poland's largest in this domain, concludes that the innovative methodology and neural network effectively correlate and predict work-life balance. The research notes an imbalance in the dataset, potentially affecting generalizability. It lacks discussion on the limitations of using an artificial neural network for work-life balance analysis. The focus is on correlations, not causation, and external validation is absent. Cultural influences on work-life balance are not explored, warranting further investigation for a comprehensive understanding.

Paigude, et al. [6] used deep learning models to predict work-life balance for women in the IT industry. It explores AI and machine learning's potential impact, analyzes subjective feelings, and identifies influencing factors. Comparing models, the research finds long short-term memory outperforms Multilayer Perceptron in accuracy. The findings contribute to understanding organizational behavior(OB) and enhancing work-life balance in the IT sector. The research's primary focus on women in the IT sector restricts its applicability to other sectors or genders. The sample size of 150 female IT professionals may not represent the entire population. Relying on subjective feelings about work-life balance introduces potential individual biases. They did not include a discussion on specific factors influencing work-life balance, limiting depth. Moreover, the comparison of models was restricted to forecasting accuracy, neglecting other performance metrics or factors.

Sujithra, et al. [7] explored stress factors for women in educational institutions, emphasizing the challenges of balancing personal and professional life, with added responsibilities like caregiving and household work. They investigated stress among working women in India, where 87% experience stress from family responsibilities and the work environment. They employed machine learning models to analyze psychological stress levels in women in educational institutions. The investigation lacks details on the machine learning techniques used to analyze stress in working women. The lack of information about sample size and demography affects generalizability. Understanding accuracy is limited by the lack of information on model variables. The focus on Indian working women in educational institutions may limit the relevance of the findings to other situations.

Gupta, et al. [8] utilized supervised machine learning, specifically support vector machines (SVM), to predict work-life balance for women in the IT industry. Data from 425 participants using a 40-parameter questionnaire was collected for the investigation. They compared ML models, including regression trees and SVM, using metrics like R2, mean absolute error, and mean square error. The optimized SVM model demonstrated the highest accuracy in predicting work-life balance, outperforming a multiple regression model. They suggested SVM as an alternative in organizational behavior studies and provided practical guidelines for ML modeling.

Sindhu, et al. [9] outlined the application of machine learning algorithms - Random Forest, Gradient Boosting, and SVM—for predicting work-life balance, with a specific focus on addressing the needs of individuals, especially women. Its usefulness lies in providing valuable insights and recommendations for effectively managing both professional and personal aspects. Nevertheless, it's important to acknowledge certain limitations, such as the restricted applicability beyond the specified algorithms, potential biases in the dataset, and a predominant focus on women, which could affect the broader relevance to diverse populations.

Shah, et al. [10] focused on analyzing Deep Learning (DL) and Machine Learning (ML) techniques, particularly the implementation of the Artificial Neural Network (ANN) algorithm, which proved promising in understanding work-life balance determinants. The achieved test accuracy of 73.91% indicates how well the ANN model understands the complex dynamics of work-life balance. Although the research offers significant perspectives and possible remedies for tackling issues related to work-life balance, it is imperative to recognize the constraints of machine learning techniques, such as possible biases within the dataset and the incompleteness of capturing complex human experiences. Therefore, for practical application, a balanced interpretation of the results is essential.

Radha, et al. [11] proposed machine learning tools to predict and enhance work-life balance, emphasizing the societal impact of AI. Analyzing data from 12,756 individuals with Random Forest, SVM, and Naive Bayes, it achieved 71.5% predictive accuracy. The authors also highlighted the significance of factors influencing work-life balance, showcasing machine learning's potential to predict and improve this crucial aspect of individuals' quality of life. They gathered data from 12,756 individuals to analyze factors influencing work-life balance. Their research lacks detailed demographic information and comprehensive explanations of factors considered in analyzing work-life balance. Additionally, relying on three classifiers with a 71.5% accuracy raises concerns about limitations in exploring alternative algorithms and the practical applicability of the results in real-world scenarios.

TÜMEN, et al. [12] concentrated on predicting work-life balance based on effort and factors by employing ensemble learning techniques like support vector machines and random forests. They emphasized the effectiveness of machine learning, including ensemble learning, in analyzing and measuring work-life balance, achieving an overall f-score of 86%. However, the authors figure that the xGBoost ensemble learning model performed poorly with the lowest f-score of 69%, suggesting that the choice of the algorithm may impact the accuracy of work-life balance predictions. The study concentrated solely on clustering employees based on attrition levels using effort and work-life balance parameters, omitting consideration of other potential influences on work-life balance. The lack of detailed information on algorithms and parameters raises concerns about result reproducibility and generalizability, and the subpar performance of the xGBoost ensemble learning model, with a low f-score of 69%, suggests the chosen algorithm may not be optimal for predicting work-life balance.

Shankar, et al. [13] addressed the issue of employee attrition in organizations and employed data mining techniques to predict and prevent it. Analyzing current and past employee data, the authors utilize various classification methods, including decision trees, logistic regression, SVM, KNN, random forest, and naive Bayes, to identify common reasons for attrition and make predictions. The emphasis is placed on preventing attrition to mitigate organizational financial losses and reduce human resource costs. The paper lacks a comparison with existing prediction models or studies, hindering the assessment of the proposed methods' effectiveness. Additionally, the limited discussion on specific features, dataset size, and the absence of consideration for external factors impede a comprehensive evaluation and generalizability of the findings on employee attrition. Most previous studies on work-life balance have concentrated on developed Western countries, leaving a gap in understanding in emerging economies like Bangladesh, where supportive structures and policies are limited.

Uddin, et al. [14] addressed the void by examining the impact of the workplace and supervisory support, along with work-life balance policies, on female employees in commercial banks in Bangladesh, revealing their significant influence on achieving better work-life balance. They involve 558 female employees and provide valuable insights for Bangladesh's scholars, professionals, policymakers, and female bankers. They exclusively focus on female employees in commercial banks in Bangladesh raising concerns about the generalizability of its findings to other industries or countries. Additionally, reliance on self-reported data, lack of exploration into specific work-life balance policies, a potentially non-representative sample size, and omitting other influencing factors may limit the study's external validity and comprehensive understanding of work-life balance determinants.

Basak, et al. [15] investigated factors affecting the work-life balance of female university teachers in Bangladesh during the COVID-19 pandemic, based on responses from 210 teachers across public and private universities. Employing various statistical tools, the research highlights the significant impact of the pandemic, on job satisfaction, workplace support, and flexibility, emphasizing the importance of work-life balance policies, day-care centers, and a healthy work environment to enhance productivity and job satisfaction among female university teachers. The study's main focus on female university teachers in Bangladesh raises concerns about generalizability, and the small sample size of 210 participants may impact the reliability and validity of the results. Additionally, reliance on self-reported data, the absence of perspectives from male teachers and other stakeholders, and the omission of potential influencing factors like cultural norms or organizational policies limit the study's comprehensiveness in understanding work-life balance in academia.

Alsaadi, et al. [16] concentrated on employee attrition in human resource analytics, utilizing the employee information value concept to identify critical features such as overtime, total projects, and job level. It compares the performance of classification algorithms, highlighting the decision tree classifier's superior accuracy of 97% in predicting attrition. The study contributes to human resource analytics by providing insights into employee attrition and evaluating the effectiveness of machine learning algorithms for predictive purposes.
The paper lacks comprehensiveness by not considering factors like work-life balance, job satisfaction, or organizational culture in the study of employee attrition. Additionally, the absence of exploration into feature interactions, alternative evaluation metrics, and ethical implications of using machine learning in HR analytics are notable gaps in the paper's approach.

Upon reviewing related papers, it is evident that there is a notable gap in research focusing on working women in academia in Bangladesh. Additionally, no prior studies have investigated predicting work-life balance for women using machine learning in this context. So my main contribution to this research will be:

- Collecting data from various educational institutions in Bangladesh, targeting individuals with different roles at their institutions.
- Examining the factors that contribute to job satisfaction in academia in Bangladesh.
- Predicting if working women can balance home and work using machine learning.

Chapter 3

Research Methodology

The methodology of this research is initiated with the presentation of a dataset description, followed by the execution of data preprocessing steps. Subsequently, various machine learning algorithms and techniques are discussed in detail in this chapter.

3.1 Research Framework

The study's structure and organization, as well as the main elements and procedures involved in carrying out the research, are referred to as the research framework. Starting with the original problem statement and ending with the final results and suggestions, it offers a road map for how the thesis will be conducted. Depending on the research question's nature, its framework's exact components may change.

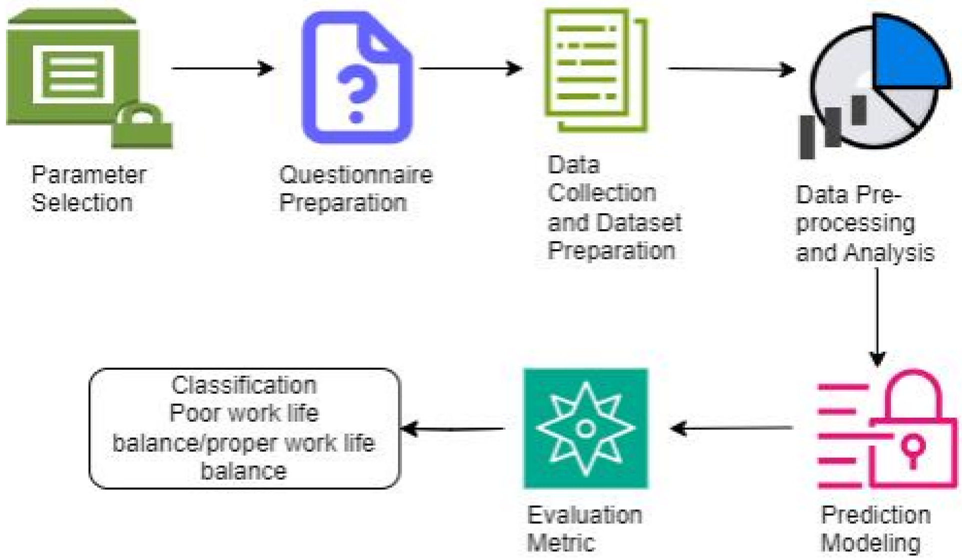


Figure 3.1: Framework for predicting work-life balance

3.1.1 Parameter Selection

The working women, mainly in Bangladesh, make up the study's sample population. The sample frame was composed of women working in academics. A set of characteristics for the research was determined from the literature review and advice from professionals.

Following a thorough literature review to identify the variables influencing working women's ability to maintain a work-life balance, a set of criteria was chosen.

Table 3.1: Parameter of questionnaires

Parameter	**Type**	**Description**
Age	int	Between 20 and 60 years of age.
Marital status	string	Married or unmarried.
Educational Qualification	string	Higher Secondary – Doctoral Degree.
Years of teaching experience in academia	string	It indicates an employee’s skills, skill level, and depth of knowledge in her chosen career.
The designation	string	refers to the lecturer - the professor for the university and assistant teacher, the head teacher for the school, and the college.
Working hours	int	It determines how much time an employee spends engaged in employment-related activities over a predetermined period.
Income	int	It will be different based on the designation.
Management support	string	Mental support from family members.
Health assistance	string	A variety of services and support are offered to employees to help them maintain or improve their health and well-being.
Crèche facility in the Educational institution	string	refers to a daycare or childcare center that provides a safe and nurturing environment for employees' children.
Pregnancy support	string	A variety of services and help that institutions offer to pregnant women to ensure their well-being.

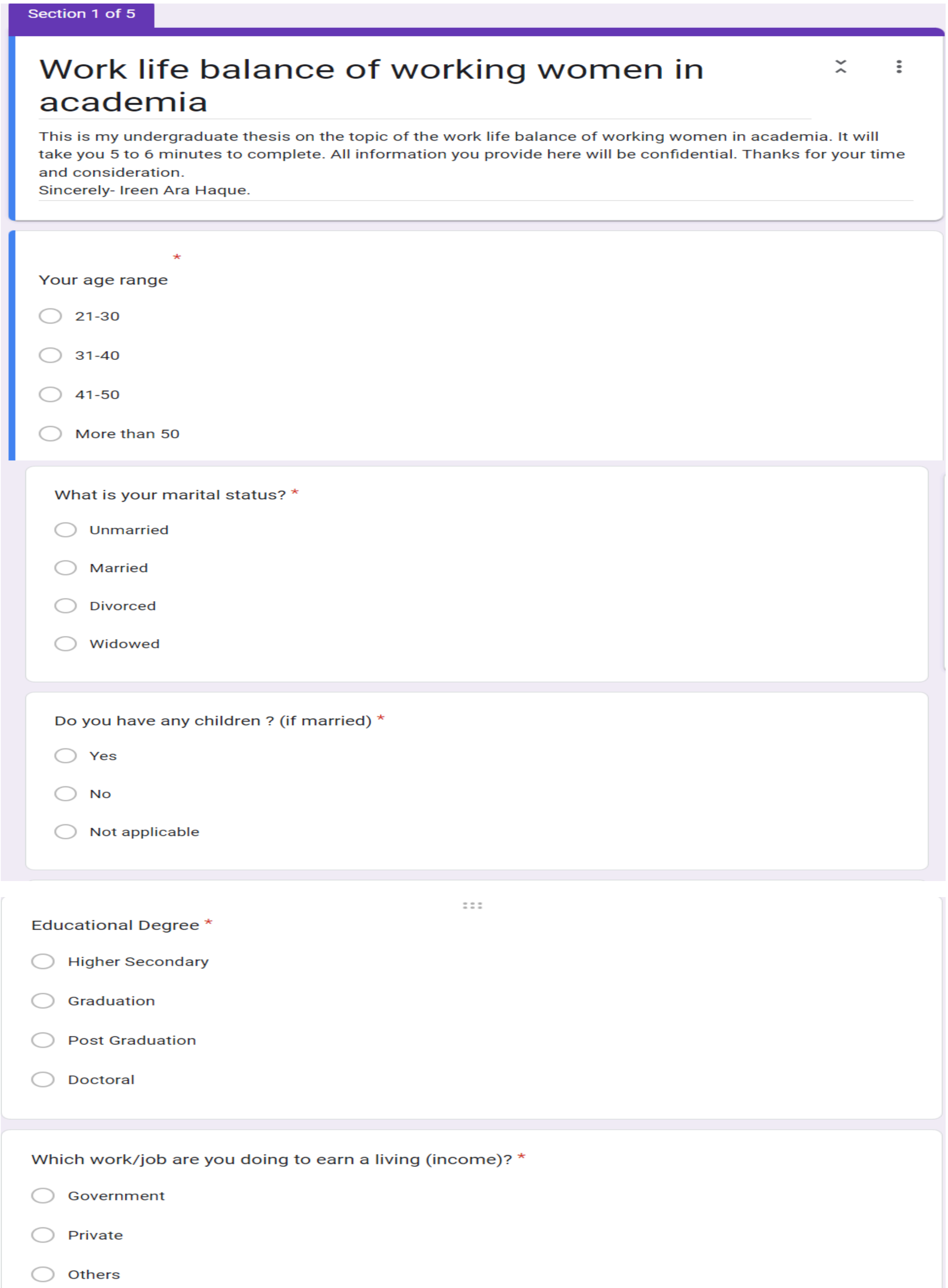
Safety at work	string	Security facilities provided by the educational institution.
Flexibility	string	Employer permits workers to modify work-related factors like hours, location, and responsibilities to suit their needs better.
Motivational support	string	Individuals are given a variety of forms of support, direction, and encouragement by their employers or seniors.
Sharing of household chores	string	refers to the fair division of household chores among family members.
Family expectation	string	Providing economic support to working women in the family.
Family type	string	It can be nuclear or extended family
Number of family members	int	The number of family members will be different for different families.
Sharing of childcare responsibilities	string	Family members significantly assist working women in child-rearing.
Anxiety	string	Whether working women have any anxiety due to work.
Time for self	string	Examining whether working women get any time for themselves.
Time for social activities	string	Determining whether or not working women have any time for social activities.

3.1.2 Questionnaire Preparation

To evaluate the work-life balance of working women, we developed a questionnaire that covered a wide range of topics in both their professional and personal lives. The questionnaires are divided into five sections: demographic data, organizational data, familial data, personal data, and social data. A sample questionnaire that was used as an initial starting point is provided below.

Table 3.2: A Section of questionnaires

Demographic data	**Organizational data**	**Familial data**	**Personal data**	**Social data**
Your age range	How long do you work at your workplace every day?	Family type	About how long do you typically sleep? (hours per night)	Do you find any time for social activities?
What is your marital status?	Are you satisfied with the working hours of your institutions?	How many family members depend on you?	What is your Body Mass Index (BMI) range?	How many times do you donate your time or money to good causes? (per month)
Do you have any children? (if married)	Is the work environment-	How sufficient is your income to cover basic life expenses?	Does your work affect your personal life or relationships with your partner?	Are social prejudices affecting your life (i.e., skin color, marriage, having babies, etc.)?
Educational Degree	Can you complete your task without any delay?	Do family responsibilities add to your workload?	How many hours do you spend every day doing what you are passionate about?	
Which work/job are you doing to earn a living (income)?	Do you have any stress related to higher studies or research?	How much time do you spend on domestic activities? (per day)	How often do you think or worry about work?	



Section 1 of 5

Work life balance of working women in academia

This is my undergraduate thesis on the topic of the work life balance of working women in academia. It will take you 5 to 6 minutes to complete. All information you provide here will be confidential. Thanks for your time and consideration.
Sincerely- Ireen Ara Haque.

Your age range *

- 21-30
- 31-40
- 41-50
- More than 50

What is your marital status? *

- Unmarried
- Married
- Divorced
- Widowed

Do you have any children ? (if married) *

- Yes
- No
- Not applicable

Educational Degree *

- Higher Secondary
- Graduation
- Post Graduation
- Doctoral

Which work/job are you doing to earn a living (income)? *

- Government
- Private
- Others

Figure 3.2: A section of questionnaires

3.1.3 Data Collection and Dataset Preparation

The survey method has been used to collect primary data for the research's objective. The required data is collected from the sample respondents through in-person interviews and a specially designed questionnaire. Based on the study's objectives, a questionnaire was created. The survey includes a few multiple-choice questions, a few binary (YES/NO) inquiries, and statements that make use of the Likert scale method. We created a Work-Life Balance (WLB) dataset using information gathered through primary data collecting.

Timestamp	Your age range	What is your marital status?	Do you have any children ? (if married)	Educational Degree	Which work/job are you doing to earn a living (income)?	Your organization name -
2023/11/09 11:3:	21-30	Married	Yes	Post Graduation	Government	Government' Pimary School
2023/11/10 4:54	21-30	Married	No	Post Graduation	Government	Dhuliapara Govt primary school
2023/11/11 12:2!	21-30	Unmarried	Not applicable	Post Graduation	Government	Primary
2023/11/27 1:08	21-30	Married	Yes	Post Graduation	Private	Daffodil International University
2023/11/27 1:34	21-30	Unmarried	Not applicable	Graduation	Private	Daffodil International University
2023/11/27 4:09	21-30	Married	No	Post Graduation	Private	Daffodil international university
2023/11/27 6:26	21-30	Unmarried	Not applicable	Post Graduation	Private	BGC Trust University, Bangladesh
2023/11/28 1:08	41-50	Married	Yes	Doctoral	Government	Comilla University
2023/11/28 7:06	31-40	Married	Yes	Post Graduation	Private	Jafargonj mir Abdul gafur degree college
2023/11/29 4:49	41-50	Married	Yes	Graduation	Private	Ramkrisnapor degree College
2023/11/30 9:31	41-50	Married	No	Graduation	Government	Noakhali Science and Technology University
2023/11/30 8:29	31-40	Married	Yes	Post Graduation	Government	Bangabandhu Sheikh Mujibur Rahman Science and
2023/12/01 7:36	41-50	Married	Yes	Doctoral	Government	Noakhali Science and Technology University, Bangla
2023/12/01 11:1	31-40	Married	Yes	Post Graduation	Government	Military Institute of Science and Technology
2023/12/01 11:2	31-40	Married	No	Graduation	Government	MIST
2023/12/02 12:0	31-40	Married	No	Graduation	Private	
2023/12/02 12:2	31-40	Unmarried	Not applicable	Graduation	Private	M2SYS Technology
2023/12/02 12:4	31-40	Married	Yes	Post Graduation	Government	Jiban Bima Corporation
2023/12/02 1:04	21-30	Married	No	Doctoral	Private	AIUB
2023/12/02 1:12	21-30	Married	Yes	Doctoral	Private	Military Institute of Science and Technology
2023/12/02 2:13	21-30	Unmarried	Not applicable	Graduation	Private	

Figure 3.3: Sample of dataset

3.1.4 Data Preprocessing

- Reading the CSV file using Dataframe: To read a CSV file using a DataFrame in Python, we used the pandas library.
- Removing Unnecessary columns: At first, we had 48 columns in which some columns were unnecessary. We removed five columns: Timestamp, Unnamed 45, Unnamed 46, Unnamed 47, and Your organization name.

  We used the ‘drop' method to remove these unnecessary columns from the DataFrame in pandas. This gave us 43 columns.

- Removing rows that contain NULL values: To remove rows containing NULL (or NaN) values from the DataFrame, we used the 'dropna( )' method. By removing the null values from the 205 values, we got 162 values.
- Removing unnecessary spaces between text: It helped us to get unique values for each column.
- Converting all the texts to small letters: To convert all text in a specific column of a pandas DataFrame to lowercase, we used the ' str. lower( )' method
- Converting categorical values to numeric values: To convert categorical values to numeric values in a pandas data frame, we used the 'LabelEncoder( )' from Scikit-learn for more flexibility.

3.1.5 Data Analysis

This section represents an overview of data analysis that involves the examination and interpretation of data to uncover meaningful insights, patterns, and trends. It encompasses the application of various statistical and computational techniques to extract valuable information and support decision-making processes.

Sample of Training and Testing data:

S

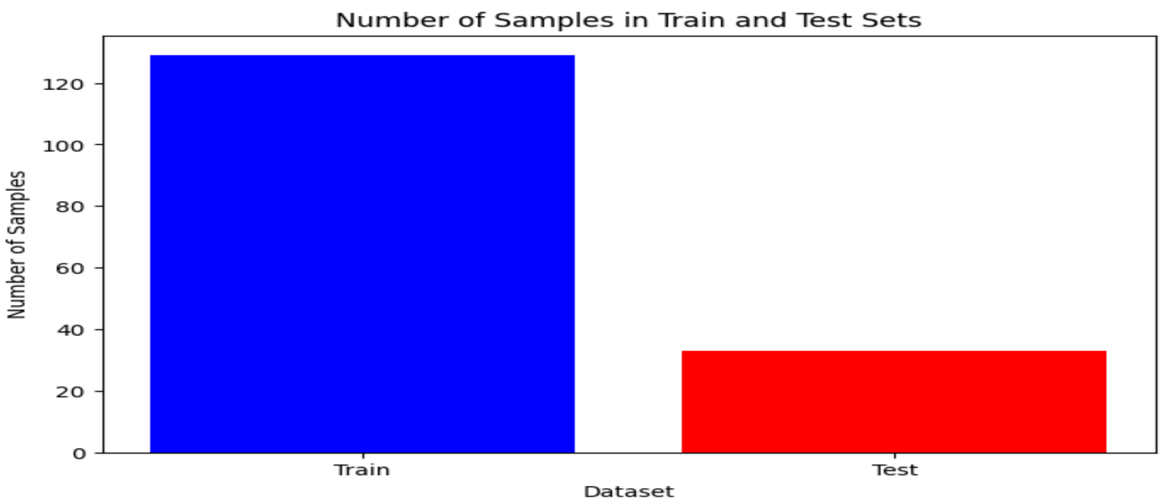


Figure 3.4: Training and Testing Data

The provided figure indicates that in the research conducted, a total of 162 data points were utilized, with 129 data points allocated for training the models and 33 data points reserved for testing the models' performance. This distribution of data ensures that the models are trained on a sufficiently large dataset while still having an independent set of data for evaluating their performance accurately.

Sample of Balance and Imbalance Ratio:

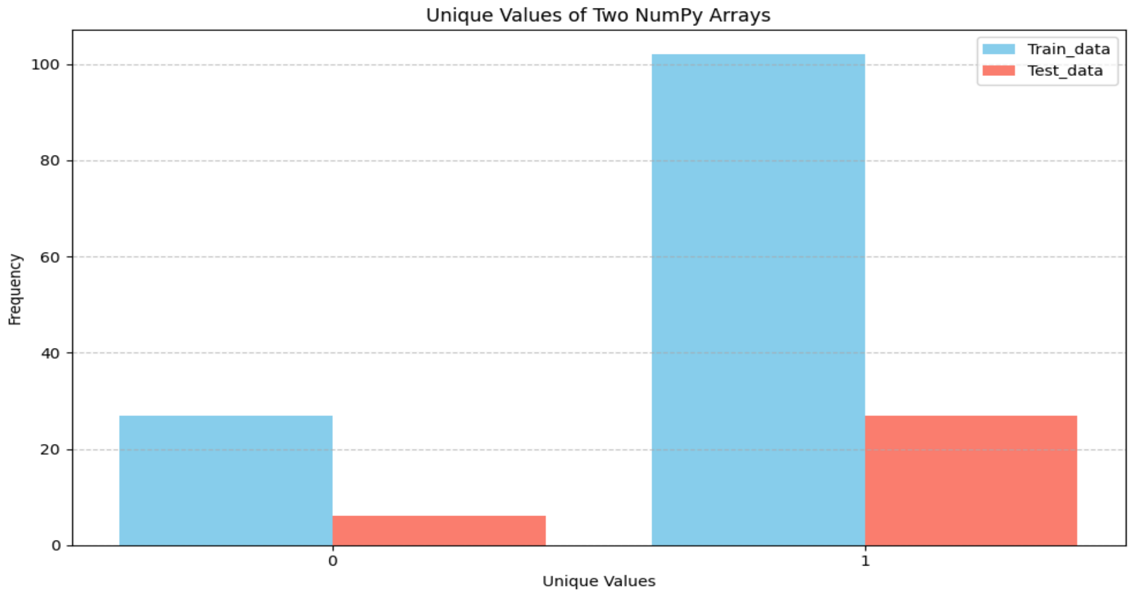


Figure 3.5: Sample of Balance and Imbalance Ratio

The information provided delineates the distribution and balance of data within both the training and testing sets utilized in the research study. In the figure 0 means imbalance and 1 denotes balance. Within the training dataset, a majority of 102 data points are classified as balanced, while a smaller subset of 27 data points are categorized as imbalanced. Similarly, within the testing dataset, the majority comprises 27 balanced data points, whereas 6 data points are classified as imbalanced. This distribution indicates a predominantly balanced representation within both sets, facilitating a comprehensive assessment of the models' performance across diverse scenarios.

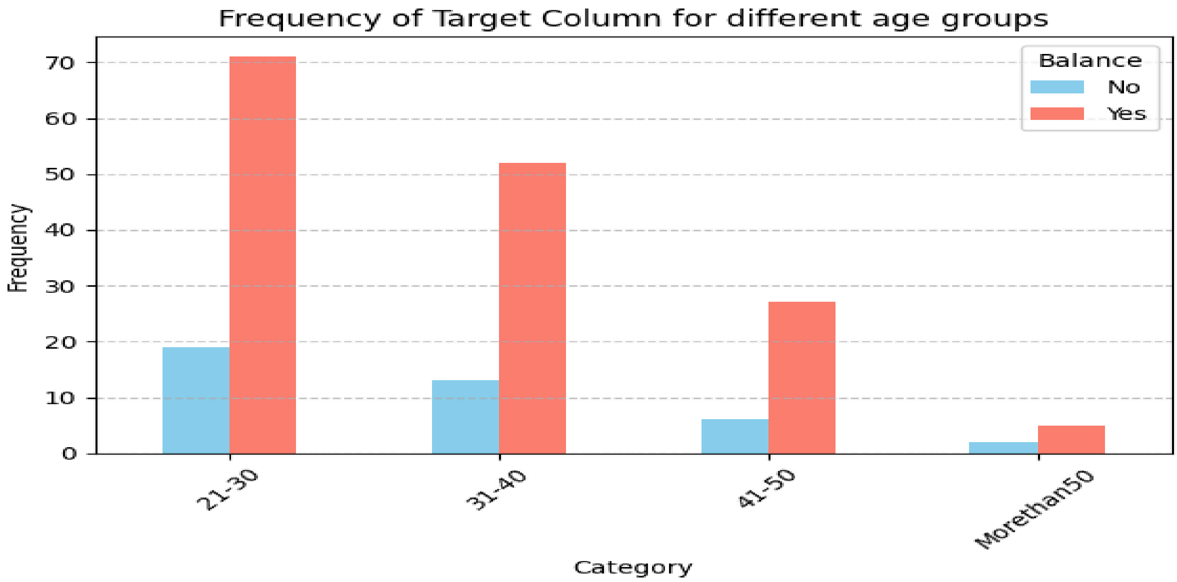


Figure 3.6: Data analysis for different age groups

The graph shows that among employees aged 21 to 30, 72 individuals can maintain a work-life balance, while 18 individuals struggle to do so. Within the age bracket of 31 to 40, 53 employees effectively manage their work-life balance, but 15 individuals can not manage their work-life balance. In the age range of 41 to 50, 27 women employees successfully balance their professional and personal lives, while 8 employees encounter difficulties. For employees older than 50 years, 5 women can maintain a work-life balance, but 2 employees find it challenging to do so.

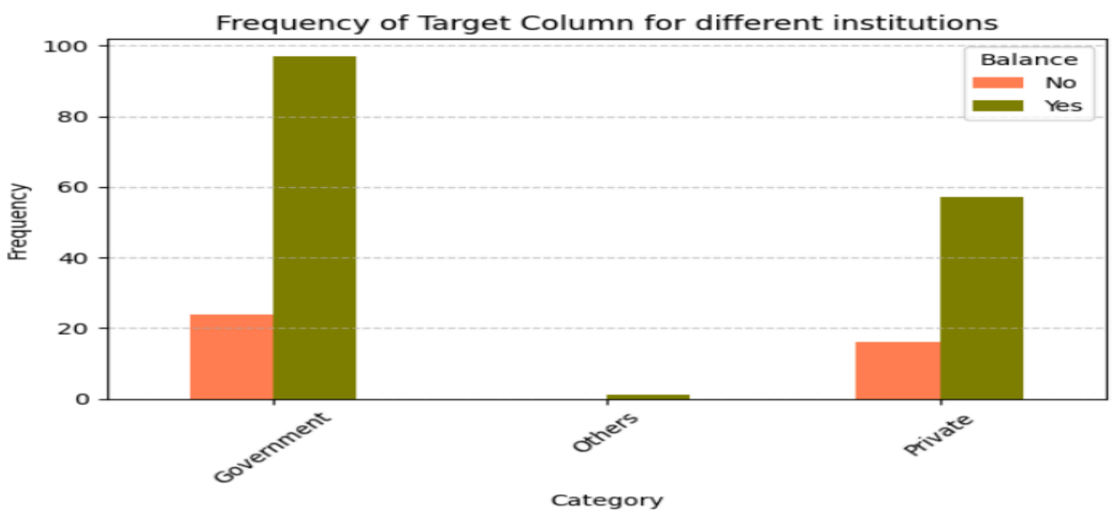


Figure 3.7: Data analysis for different educational institutions

The following figure highlights that in government educational institutions, 90 women employees can maintain a balanced life, while 25 women face challenges in achieving this balance. Within private educational institutions, 58 women employees successfully balance their professional and personal lives, while 15 women struggle to do so. Additionally, in other sectors, 5 women employees demonstrate the ability to balance their work and personal life effectively. Based on the provided information, it can be inferred that women employees in the government sector exhibit a higher capability to balance their work and personal life compared to women in other sectors.

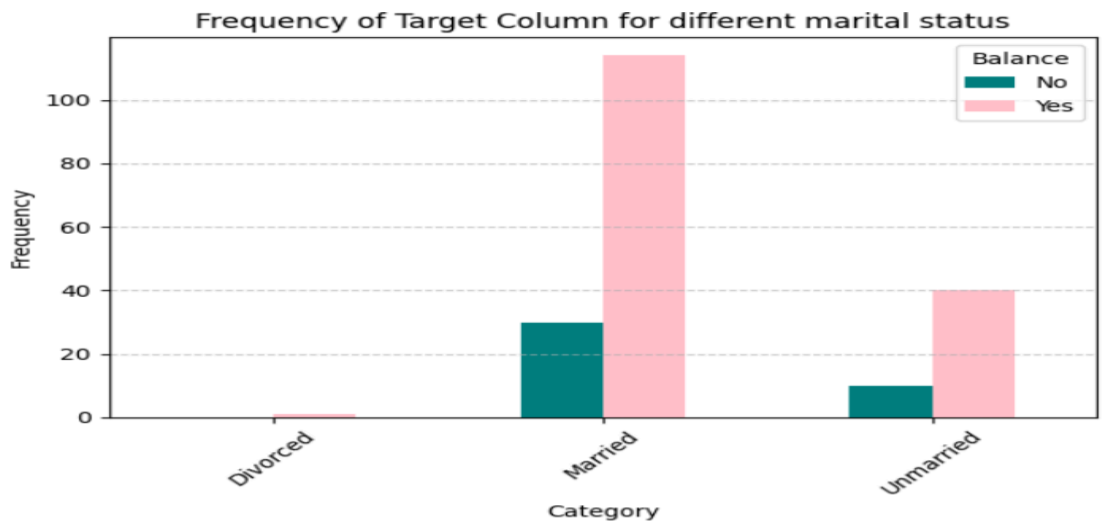


Figure 3.8: Data analysis for different marital statuses

The provided information highlights the work-life balance of married, unmarried, and divorced women employees. Among married women employees, 95 individuals can maintain a balanced life, while 20 individuals struggle to do so. Conversely, among unmarried women employees, 58 individuals effectively balance their professional and personal lives, while 16 individuals find it challenging to do so. Furthermore, the number of divorced women employees is significantly less, and they can maintain a balanced life without facing notable difficulties. This data suggests that married women employees face more challenges in achieving work-life balance compared to unmarried women employees. Additionally, the relatively fewer divorced women employees seem to fare well in managing their work and personal lives effectively.

Feature importance: Feature importance is a crucial aspect in understanding the behavior of machine learning models, as it helps identify which features have the most significant influence on the model's predictions. There are various techniques to assess feature importance, depending on the model used.

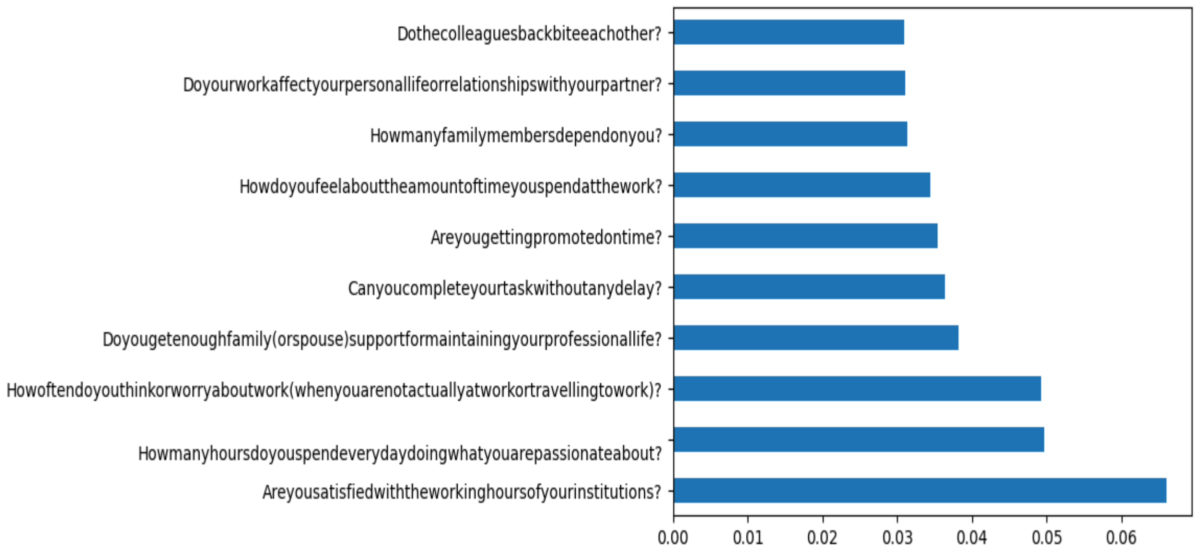


Figure 3.9: Chi-squared test

We use univariate selection using the Chi-squared test, which provides a simple yet effective approach for ranking and selecting features in classification tasks. The Chi-squared test is a statistical test used to determine if there is a significant association between two categorical variables. In the context of feature selection, the figure measures the dependence between each feature and the target variable (or class labels) in the classification problem.

3.1.6 Prediction Modeling

The training dataset was utilized after data analysis to build machine learning models, and the testing dataset was used to assess the learning models built. We suggested utilizing ensemble-based machine learning methods for the project since they are motivated by the ability of ensemble learning methods to generalize across multiple domains (in our case, in Bangladesh), where the distributions are also distinct. For the experiment, we primarily used the Classifiers Logistic Regression, Decision Tree, Naive Bayes, Support Vector Machine Algorithm, Random Forest (RF), and K- Nearest Neighbor(KNN).

3.1.7 Evaluation Metric

In this research, we focused on a classification/prediction problem with two classes (balance and no balance). To measure the prediction performance, we utilized accuracy, precision, recall, and f1-score.

3.2 Classification Approaches

Based on how working women view their work-life balance, we categorized working women into several groups to develop a categorization strategy. Using a dataset of variables and labels, the classification strategy for measuring work-life balance among working women uses machine learning to create a predictive model that can identify whether a woman has a good or poor work-life balance.

3.2.1 Logistic Regression

A statistical analysis technique called logistic regression uses previous observations from a data set to predict a binary outcome, such as yes or no. A logistic regression model uses an analysis of the correlation between one or more pre-existing independent variables to predict a dependent data variable. The technique of logistic regression has grown in significance in the field of machine learning. It enables machine learning algorithms to categorize incoming input based on previous data. The algorithms get more accurate at predicting classes within data sets when new relevant data is added.

3.2.2 Decision Tree

A Decision Tree is a supervised learning method that can be applied to classification and regression issues, however, it is most frequently used to address classification issues. It is a tree-structured classifier, where internal nodes stand in for the dataset's features, branches for the rules of classification, and each leaf node for the result. A decision tree has two nodes: a decision node and a leaf node. Decision nodes are used to make decisions and have numerous branches, whereas Leaf nodes represent the results of those decisions and do not have any additional branches.

3.2.3 Naive Bayes

The Nave-Bayes algorithm is a supervised learning method for classification issues that is based on the Bayes theorem. It is mostly employed in text categorization with a large training set. The Naive Bayes Classifier is one of today's most straightforward and efficient classification algorithms. It aids in developing quick machine-learning models capable of making accurate predictions. It makes predictions based on object probabilities because it is a probabilistic classifier.

$$P(A|B) = \frac{P(B|A)\,P(A)}{P(B)}$$

Where

$P(A|B)$ is Posterior probability: Probability of hypothesis A on the observed event B.

$P(B|A)$ is Likelihood probability: Probability of the evidence given that the probability of a hypothesis is true.

$P(B)$ is Marginal Probability: Probability of Evidence.

$P(A)$ is Prior Probability: Probability of the hypothesis before observing the evidence

3.2.4 Support Vector Machine Algorithm

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms. The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to classify fresh data points in the future quickly. A hyperplane is the name given to this optimal decision boundary. To create the hyperplane, SVM selects the extreme points and vectors. Support vectors, which are used to represent these extreme instances, are what give the Support Vector Machine method its name.

3.2.5 Random Forest

Random Forest is a classifier that uses many decision trees on different subsets of the provided dataset and averages the results to increase the dataset's predicted accuracy. Instead of depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions. Higher accuracy and overfitting are prevented by the larger number of trees in the forest.

3.2.6 K-Nearest Neighbor

One of the simplest machine learning algorithms, based on the supervised learning method, is K- Nearest Neighbor. The K-NN algorithm places the new case in the category that is most similar to the available categories based on the assumption that the new instance and the data are comparable to the examples that are already accessible. The K-NN algorithm is non-parametric, which means it doesn't make any assumptions about the underlying data.

Chapter 4

Result and Discussions

This chapter involves the analysis of experimental results achieved through the utilization of various machine learning algorithms and techniques. The primary objective is to predict, with high accuracy and efficiency, whether the balancing of life by women can be effectively determined. The focus of the chapter is on evaluating the predictive performance of the implemented models in determining work-life balance for women.

4.1 Performance Metrics

We employed many performance metrics to assess the suggested model's effectiveness. The accuracy metric alone may not provide an accurate evaluation of the model's performance because the dataset used in this study was heavily imbalanced. The evaluation metrics employed in this instance are F1 score, accuracy, precision, and recall, which are defined as follows:

- **Accuracy:** Accuracy is a metric that measures how frequently a machine learning model makes correct predictions. The ratio of correctly predicted observations to total observations produces an accuracy value.

$$Accuracy = \frac{Total\ \ Instances}{True\ Positive + True\ Negative}$$

- **Precision:** Precision tells us the fraction of predicted positive results that are true positives. Precision is the number of True Positives divided by the number of True Positives and False Positives.

$$Precision\ = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

- **Recall (Sensitivity or True Positive Rate):** The proportion of actual positive cases that the model accurately predicted is measured by the recall. It evaluates the model's ability to find every relevant case.

$$Recall\ = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

- **F1 - Score:** The harmonic mean of precision and recall is the F1-Score. It is helpful when dealing with imbalanced datasets since it strikes a balance between precision and recall.

$$F1 - Score = \frac{2(Precision*Recall)}{Precision + Recall}$$

Table 4.1: Performance Metrics for Original Dataset

Model Name	Accuracy	Precision	Recall	F1-score
Logistic Regression	81%	86%	93%	89%
Decision Tree	84%	84%	100%	92%
Naive Bayes	**90%**	**90%**	**100%**	**95%**
Support Vector Machine	81%	89%	89%	89%
Random Forest	81%	86%	93%	89%
K- Nearest Neighbor	75%	83%	89%	86%

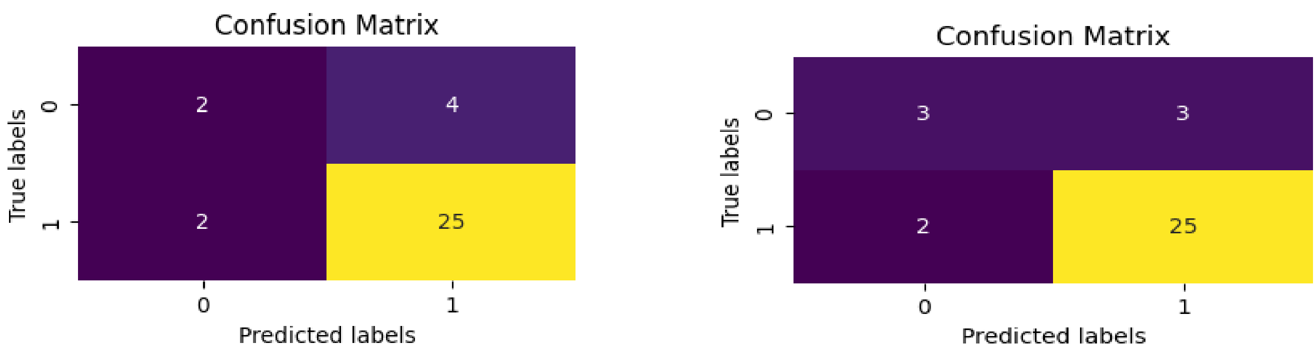
4.2 Confusion Matrix

A Confusion Matrix is a tabular visualization of the ground-truth labels versus model predictions that are often used to describe the performance of a classification model. Each row of the confusion matrix represents the instances in a predicted class and each column describes the instances in an actual class. Confusion Matrix is not precisely a performance metric but sort of a basis on which other metrics evaluate the results. It contains True Positive, True Negative, False Positive, False Negative. A false positive value is lesser which shows fraud not detected cases are low.

Table 4.2: Confusion Matrix

Actual	Predicted	
	Imbalanced	Balanced
Imbalanced	True Negative	False Positive
Balanced	False Negative	True Positive

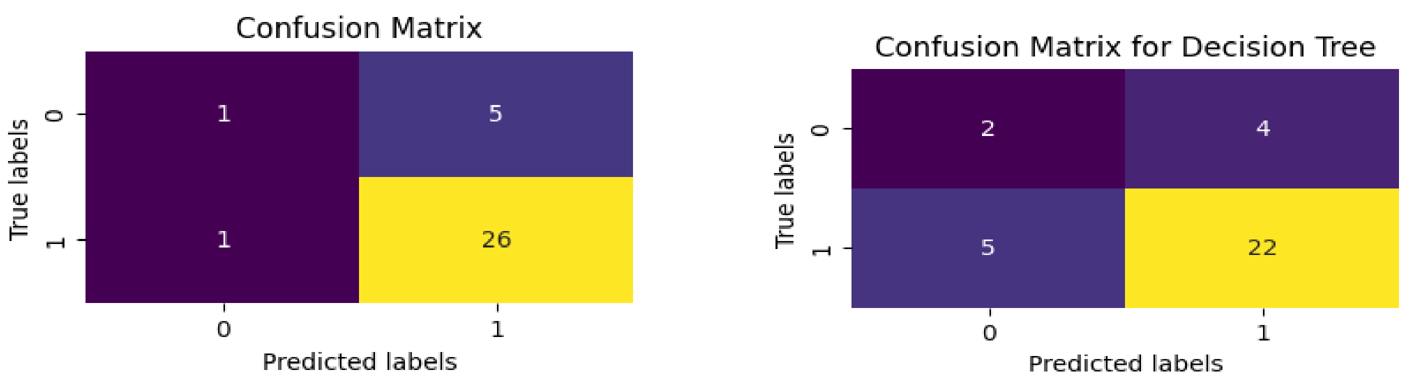
Confusion matrix for Logistic Regression



(a) (b)

Figure 4.1: Confusion matrix of Logistic Regression (a) Original dataset, (b) Oversampling. **Here 0 indicates imbalance and 1 indicates balance**

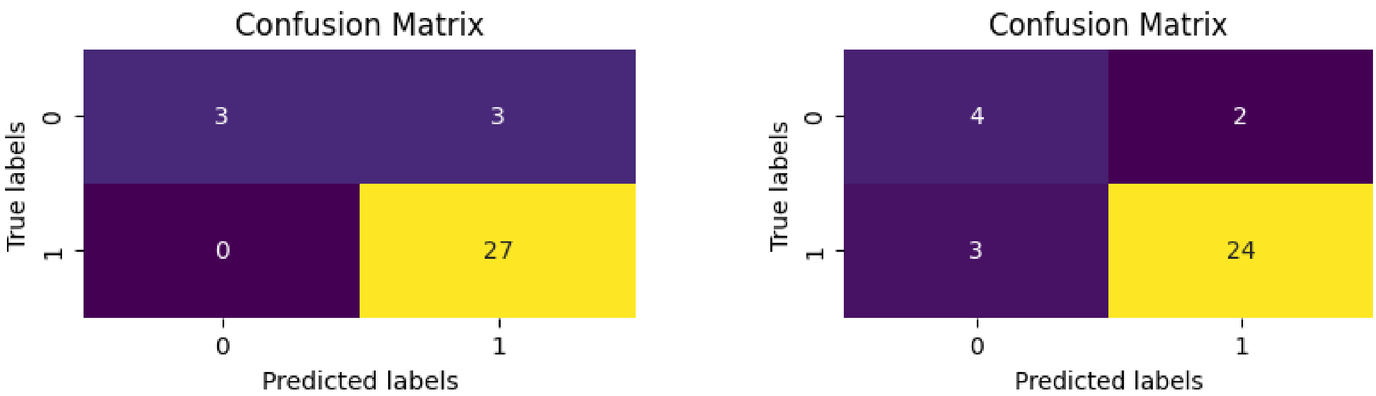
Confusion matrix for Decision Tree



(a) (b)

Figure 4.2: Confusion matrix of Decision Tree (a) Original dataset, (b) Oversampling

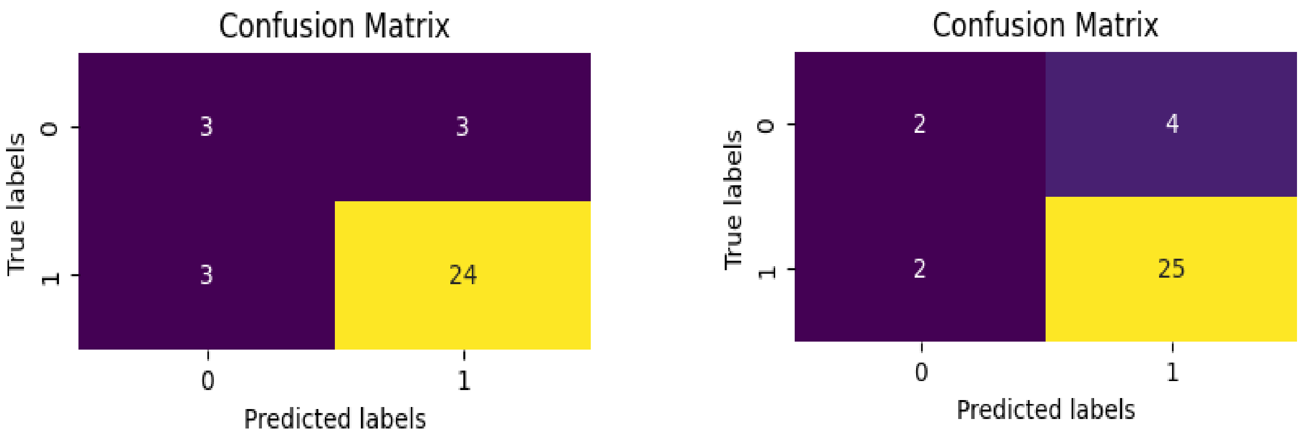
Confusion matrix for Naive Bayes



(a) (b)

Figure 4.3: Confusion matrix of Naive Bayes (a) Original dataset, (b) Oversampling

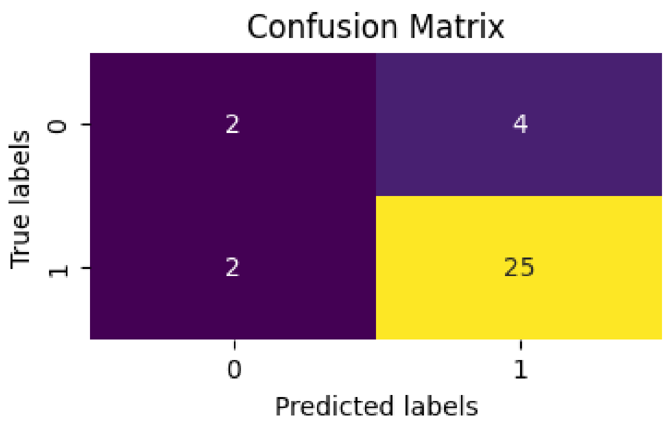
Confusion matrix for Support Vector Machine



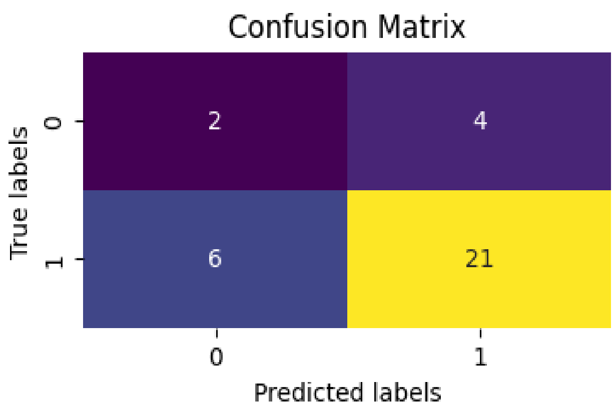
(a) (b)

Figure 4.4: Confusion matrix of Support Vector Machine (a) Original dataset, (b) Over-sampling

Confusion matrix for Random Forest



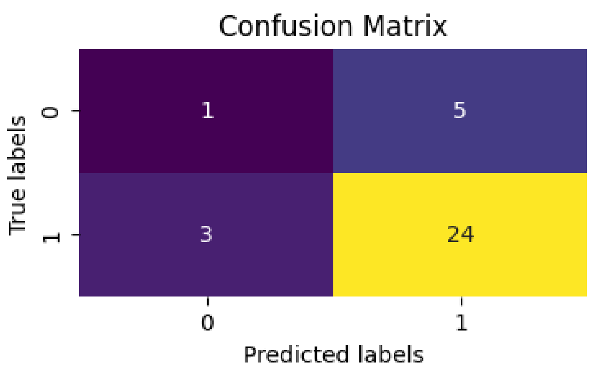
(a)



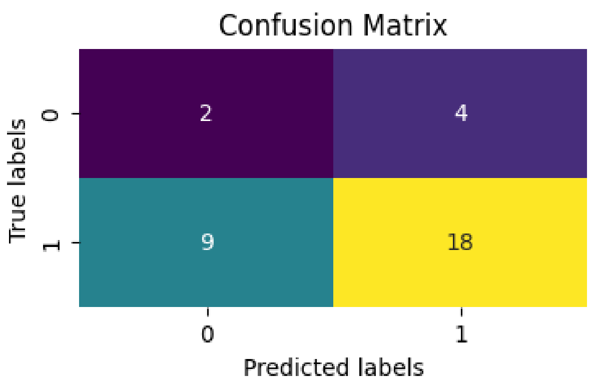
(b)

Figure 4.5: Confusion matrix of Random Forest (a) Original dataset, (b) Oversampling

Confusion matrix for K- Nearest Neighbor



(a)



(b)

Figure 4.6: Confusion matrix of K-nearest Neighbor (a) Original dataset, (b) Oversampling

4.3 Performance (Accuracy, Precision, Recall, F1-score) Evaluation of Classifiers

Accuracy is a metric that measures how frequently a machine learning model makes correct predictions. Precision is the number of True Positives divided by the number of True Positives (TP) and False Positives (FP). Precision is a measure of exactness that determines what percentage of tuples is labeled as positive. Recall is the number of True Positives (TP) divided by the total number of True Positives (TP) and False Negatives (FN). Recall is also known as the True Positive Rate (TPR) or Sensitivity. Recall is a measure of completeness that estimates what percentage of positive tuples are labeled. The F-score, also called the F1-score, is the weighted average of Precision and Recall.F1-score, is the weighted average of Precision and Recall.

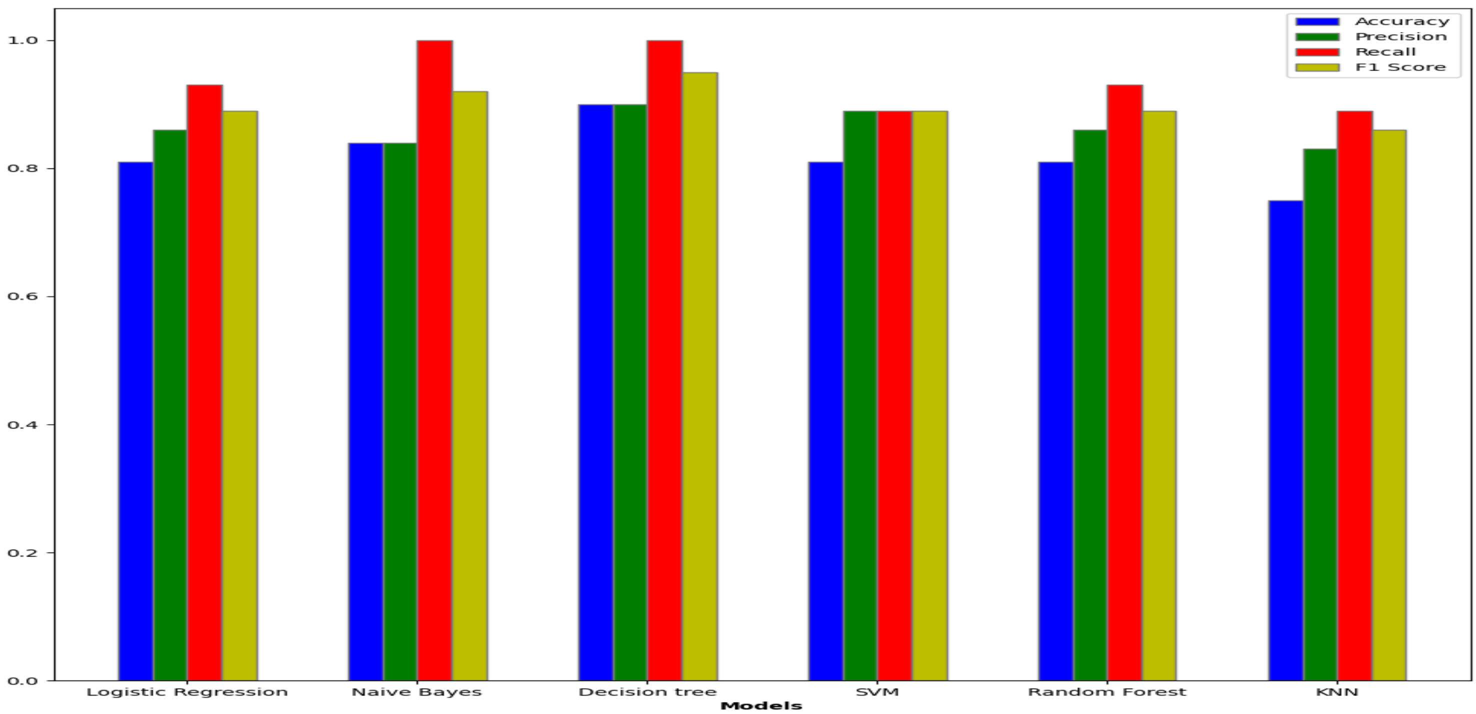


Figure 4.7: Performance (Accuracy, Precision, Recall, F1-score) Evaluation of Classifiers for the Original Dataset

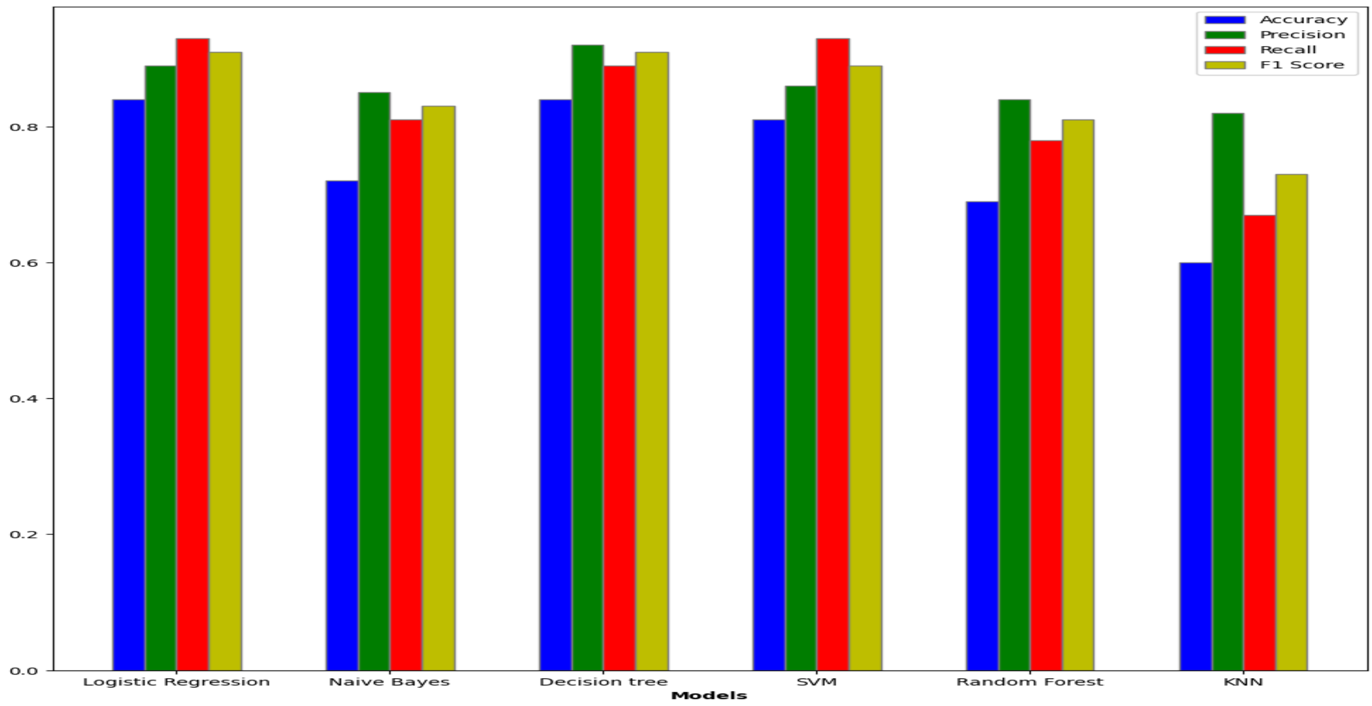


Figure 4.8: Performance (Accuracy, Precision, Recall, F1-score) Evaluation of Classifiers for Oversampling Using SMOTE

Chapter 5

Conclusion and Future Work

In this chapter, we summarize the main findings, discuss potential future directions for further research, and conclude the thesis. We encapsulate the core outcomes and lay the groundwork for potential advancements in the research while bringing the overall thesis to a close. Section 5.1 discusses the summary of the thesis and future works are provided in section 5.2.

5.1 Summary of the Thesis

In this thesis, we investigated the important and complicated topic of work-life balance for working women and we used machine learning approaches to obtain insights and forecast their work-life balance. In this research, we identified the critical elements affecting working women's ability to maintain a work-life balance, including workplace policies, family dynamics, individual preferences, workloads, etc. Using these elements, we created questionnaires, which helped us collect data. After creating questionnaires, we collected data, and the data was analyzed using some models like logistic regression, decision tree, support vector machine, naive Bayes, etc. These models helped us to get accurate results.

In conclusion, this thesis is a crucial step in using machine learning to comprehend and measure the work-life balance of working women. We have applied the six machine learning models stated in the methodology, and the models indicate a high accuracy score for each. Among these models, Naive Bayes provided the highest accuracy. We also used different sampling techniques to get better results. Since our dataset was imbalanced, we couldn't rely solely on accuracy to evaluate classifiers. Therefore, we explored other metrics such as Precision, Recall, F1-score, AUC, etc. We have a better result accuracy of 90%, precision of 100%, and recall 95%, on our original dataset using the Naive Bayes classifier.

5.2 Future Work

There is scope to broaden the dataset, ensuring a comprehensive analysis of work-life balance factors applicable to diverse working women in Bangladesh using machine learning algorithms. Moreover, qualitative insights can enhance the depth of interviews or surveys, capturing subjective experiences. Additionally, there is scope for research in banking, IT, and other industries through collaboration with organizations. Accessing real-time data and integrating policies could provide an integral understanding of the influences on work-life balance.

REFERENCES

[1] Kumari, K.T. and Devi, V.R., 2015. A study on work–life balance of women employees in selected service sectors. *Pacific business review international*, *7*(10), pp.17-23.

[2] Rahaman, A. and Tul-Jannat, K., 2015. Reconciliation of Professional and Personal Life: An Empirical Work Life Balance Study on Context of Women Empowerment. *International Journal of Economics, Commerce and Management*, *3*(9).

[3] Gupta, C., Rao, K.R. and Datta, P., 2022. Support vector machine based prediction of work-life balance among women in information technology organizations. *IEEE Engineering Management Review*, *50*(2), pp.147-155.

[4] Khan, S., Thomas, G., Kunbhar, B.A. and Mohamed, N.H.M., 2022. Impact of Work–Life Balance on Working Women in the Banking Sector. *Administrative Sciences*, *13*(1), p.7

[5] Pawlicka, A., Pawlicki, M., Tomaszewska, R., Choraś, M. and Gerlach, R., 2020. Innovative machine learning approach and evaluation campaign for predicting the subjective feeling of work-life balance among employees. *PloS one*, *15*(5), p.e0232771.

[6] Paigude, S. D., & Shikalgar, S. R. (2022, December). Deep Learning Model for Work-Life Balance Prediction for Working Women in IT Industry. In *Proceedings of the 4th International Conference on Information Management & Machine Intelligence* (pp. 1-8).

[7] Sujithra, M., Velvadivu, P., Rathika, J., Priyadharshini, R. and Preethi, P., 2022, October. 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)* (pp. 1-7). IEEE.

[8] Sindhu, V.H. and Reddy, V.V.R., PREDICT WORK-LIFE BALANCE-BASED EMPLOYEE ATTRITION FOR WOMEN.

[9] Shah, D. and Parekh, M., 2023. Understanding Work-Life Balance: An Analysis of Quiet Quitting and Age Dynamics Using Deep Learning. *International Research Journal of Engineering Technology*, *1*(06), pp.1230-1235.

[10] Radha, K. and Rohith, M., 2021. An Experimental Analysis of Work-Life Balance Among The Employees Using Machine Learning Classifiers. *arXiv preprint arXiv:2105.07837*.

[11] TÜMEN, V. and SUNAR, A.S., 2023. Predicting the Work-Life Balance of Employees Based on the Ensemble Learning Method. *Bitlis Eren Üniversitesi Fen Bilimleri Dergisi*, *12*(2), pp.344-353.

[12] Shankar, R.S., Rajanikanth, J., Sivaramaraju, V.V. and Murthy, K.V.S.S.R., 2018, July. Prediction of employee attrition using datamining. In *2018 ieee international conference on system, computation, automation and networking (icscan)* (pp. 1-8). IEEE.

[13] Uddin, M., Ali, K. and Khan, M.A., 2020. Impact of perceived family support, workplace support, and work-life balance policies on work-life balance among female bankers in Bangladesh. *International Journal of Economics, Management and Accounting*.

[14] Basak, S. and Akter, K., 2022. Work-life balance of female university teachers during COVID-19 pandemic in Bangladesh. *Work*, *14*(4).

[15] Alsaadi, E.M.T.A., Khlebus, S.F. and Alabaichi, A., 2022. Identification of human resource analytics using machine learning algorithms. *Telkomnika (Telecommunication Computing Electronics and Control)*, *20*(5), pp.1004-1015.

[16] Kumari, K.T. and Devi, V.R., 2015. A study on work–life balance of women employees in selected service sectors. *Pacific business review international*, *7*(10), pp.17-23.

.