PREDICTING WORK-LIFE BALANCE: A MACHINE LEARNING APPROACH WITH LIFESTYLE AND BEHAVIOR-RELATED VARIABLES

KIMBERLY ZIMMERMAN

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

STUDENT NUMBER

2053785

COMMITTEE

Prof.dr.ir. Pieter Spronck
Dr. Giacomo Spigler

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

January 15th, 2024

WORD COUNT

7519

ACKNOWLEDGMENTS

I would like to express my heartfelt gratitude to my supervisor, Pieter Spronck, for his tremendous guidance during the duration of this project. In addition, my greatest gratitude goes to my family, whose steadfast support was invaluable in navigating this journey.

PREDICTING WORK-LIFE BALANCE: A MACHINE LEARNING APPROACH WITH LIFESTYLE AND BEHAVIOR-RELATED VARIABLES

KIMBERLY ZIMMERMAN

Abstract

In today's workforce, creating a harmonious Work-Life Balance (WLB) is critical for both individual well-being and corporate effectiveness. This work investigates WLB prediction using advanced machine-learning techniques and a dataset enriched with lifestyle and behavioral characteristics. By using Support Vector Regression, Multiple Linear Regression, and Multiple Layer Perceptron this study dives into modeling and prediction of WLB using Lifestyle and behavioral features. The dataset used included 15,972 survey responses from the Authentic-Happiness.com global work-life survey, which assesses how people shape their lifestyles, habits, and behaviors across dimensions such as Healthy Body, Healthy Mind, Expertise, Connection, and Meaning to maximize overall life satisfaction. The dataset underwent several pre-processing steps, with a focus on data transformation to address the dataset's uneven distribution of several attributes. The Quantile Transformation method was used to adress skewness of several attributes. The models were initially trained on the complete dataset, followed by a reduction based on permutation scores that indicated predominant features affecting the models predictive performance. The results demonstrated that various factors had a substantial impact on the predictive performance of the models. Notably, the models depended on BMI Range, Sufficient Income, and Donation as primary determinants. Furthermore, the findings showed that Support Vector Regression (SVR) and Multiple Layer Perceptron (MLP) models beat Multiple Linear Regression when it comes to predicting Work-Life Balance (WLB) utilizing lifestyle and behavioral attributes.

1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

The dataset utilized in this thesis is publicly available on Kaggle and can be accessed through the linked repository: https://www.kaggle.com/datasets/ydalat/lifestyle-and-wellbeing-data/data. The acquired data contains no observable or logical link that may allow real individuals to be recognized. This study required no data collection, and the data utilized remains the property of the original owner both during and after the completion of this thesis. The author acknowledges having no legal claim to this data. All figures presented in this thesis are self-made by the author. The software tools employed in this thesis are detailed in section Section 4.9.

2 INTRODUCTION

In response to evolving demographics, increased work hours, and dynamic work environments, scholars and practitioners delve into understanding the intricate relationship between work and life (Helmle et al., 2014). Increased demands for multitasking and proficiency in the workplace now put a strain on workers, causing a disconnect between work and personal life (Poulose & Sudarsan, 2017). Work-life balance (WLB) has grown in importance as a means of overcoming personal obstacles and adjusting to rapid industry changes (Shah & Parekh, 2023). Achieving a clear work-life balance (WLB) is difficult due to the concept's ambiguity (Paigude & Shikalgar, 2022). There's no universally accepted definition or assessment method for work-life balance (WLB) making it difficult to evaluate and comprehend (Paigude & Shikalgar, 2022).

To define WLB, this study uses the definition proposed by Haar et al. (2014). WLB is defined as an individual's assessment of how well their many life roles are balanced. Accomplishing WLB is a result of a two-dimensional interaction between organizational and individual factors such as lifestyle and behavior. Lifestyle decisions, habits, and behavior all play an important influence in defining an individual's capacity to balance work and personal life (Shah & Parekh, 2023). As a result, it is essential to examine lifestyle and behaviors as relevant variables in predicting WLB (Shah & Parekh, 2023). This study explores how lifestyle and behavioral factors contribute to predicting and enhancing WLB using various machine-learning approaches.

2.1 *Scientific and Societal Relevance*

Establishing a harmonious work-life balance is important given its direct impact on employee health, job satisfaction, and overall productivity (Bhadana et al., 2022). Imbalances can lead to job dissatisfaction and reduced efficiency, affecting employee well-being (Shah & Parekh, 2023). Existing studies in the field have been focused on identifying the factors that contribute to WLB in organizational settings. In contrast, this study anticipates and investigates the impact of lifestyle-related variables on understanding and predicting WLB. By moving away from these traditional organizational-centric analyses, new insights can be gathered on the possible impact of individual lifestyle choices and behaviors on achieving a harmonious WLB. Using machine learning to predict and enhance WLB can alleviate the major concerns brought on by a lack of WLB, improving employee well-being, job satisfaction, and overall quality of life (Pawlicka et al., 2020). Furthermore, it can help shape policies and improve workplace dynamics, resulting in a more balanced and happy work environment for employees (Munyeka & Maharaj, 2022).Individuals can also benefit from tailored recommendations to improve their daily routines and make well-informed decisions to improve their overall quality of life.

2.2 *Research Aim*

In order to predict someone's subjective perception of work-life balance (WLB) based on lifestyle and behavioral elements, this study starts by defining the key variables that affect WLB. These parameters are used as features to predict a person's subjective WLB. The following research question will be addressed:

> *How accurately can machine learning models predict an individual's subjective feeling of Work-Life Balance (WLB) using lifestyle and behavior-related variables?*

The following sub-questions were developed to support the Main Research Question:

RQ1 *Which lifestyle and behavior-related variables significantly impact predicting an individual's subjective feeling of Work-Life Balance (WLB)?*

This sub-question focuses on identifying the predominant lifestyle and behavioral attributes that play a significant role in determining an individual's WLB.

RQ2 *How does the predictive accuracy of Support Vector Regression (SVR), Multiple Linear Regression (MLR), and Multilayer Perceptron (MLP)* *differ when utilizing lifestyle and behavior-related variables to predict an individual's subjective feelings of Work-Life Balance (WLB)?*

This sub-question focuses on the comparative performance of different machine learning models when using lifestyle and behavior-related variables for WLB prediction. Understanding how these models perform with lifestyle and behavior-related variables will help in selecting the most suitable approach for accurate WLB prediction. Three different machine-learning models will be compared.

The research will involve multiple steps, beginning with a comprehensive analysis of different models utilizing the entire dataset. Following that, hyperparameter tuning and cross-validation will be used to optimize the model's performance. Following that, the dataset will be refined by identifying the most prominent features in response to RQ1. This comparative analysis seeks to identify the best-performing model.

The upcoming section examines related literature on the topic, providing a complete overview of existing studies and theories relevant to predicting Work-Life Balance (WLB).

3 RELATED WORK

This section of the paper is organized as follows: it begins with a clear definition of Work-Life Balance (WLB) (3.1), then moves on to the traditional WLB research methods and the identified features impacting WLB (3.2), and the section finishes with the Machine Learning (ML) methods for WLB prediction and their performance, including the identified predominant factors in WLB prediction (3.3).

3.1 *Work-Life Balance*

There is no universally accepted definition of WLB. Early researchers defined WLB as the amount of satisfaction people have when they can perform in both their personal and professional lives with little to no role conflict (Clark, 2000). Grzywacz and Carlson (2007) defined WLB as the joint attainment of role-related expectations in the work and family domains by an individual and their role-related partners. A more modern definition is given by Devadoss and Minnie (2013), who define WLB as the degree to which individuals have control over the timing, location, and manner in which they engage in work activities. It indicates people's capacity to manage and adjust their work schedules and responsibilities so that they align with their personal lives and preferences. In this study, we work with a more conceptual definition of WLB, given by Haar et al. (2014), who define WLB as an individual's assessment of how well their many life roles are balanced. This definition is utilized in this study because it recognizes the subjective nature of WLB and emphasizes the individualized perspective that molds one's impression of balance in the context of their specific life circumstances. Balance, particularly in the context of WLB, embodies a subjective sense of harmony between the domain of life and work (Nilawati et al., 2019). To have a positive experience in all areas of life, effective allocation of resources such as energy, time and dedication are essential (Nilawati et al., 2019). Individual behavior and actions balance aspects of one's personal and organizational life (Poulose & Sudarsan, 2017).

Maintaining WLB is important for working professionals since it has been linked to a variety of advantages including enhanced mental and physical health, lower stress levels, higher job satisfaction, increased productivity, and overall well-being (Bhadana et al., 2022). In contrast, a lack of WLB can lead to burnout, decreased job satisfaction, lower employee productivity, strained personal relationships, and detrimental effects on physical and mental health (Shah & Parekh, 2023).

3.2 *Traditional Methods*

Different researchers have measured the concept of WLB in various ways. Early studies based their measurements on time, stress, and behavior-based disputes in the work and family domains (Poulose & Sudarsan, 2017). However, as the field expanded, following studies began to focus on additional characteristics of WLB, such as work overload, shift work, reduced work support, role conflicts, and role ambiguity when studying WLB (Poulose & Sudarsan, 2017).

Statistical methods are among the most widely utilized methodologies and serve a crucial role in determining the WLB (Paigude & Shikalgar, 2022). Various statistical techniques have been used to study WLB and identify predominant factors influencing WLB. For instance, researchers Munyeka and Maharaj (2022), used several statistical methods to find the most important issues affecting the WLB of female employees in a South African telecommunications company. Their study utilized a self-designed questionnaire and conducted Cronbach's alpha analysis and factor analysis to identify the most influencing challenges affecting WLB. Gender stereotypes, flexibility, and time management, tasks involving dependent adults or children, work/home conflicts, a sense of accomplishment, and skill acquisition were among the 6 most important challenges female IT professionals face in the telecommunications industry.

Poulose and Sudarsan (2017) also utilized statistical methods to investigate the influence of work-related factors such as work overload and work support on the WLB of employees and its significant impact on work satisfaction in the healthcare sector. To identify influential factors and relationships affecting WLB, they used statistical approaches such as principal component analysis, factor analysis, correlation analysis, and mediation analysis. The results demonstrated that work satisfaction had a negative relationship with work overload and a positive relationship with work support. Furthermore, different aspects of WLB were found as mediators in the relationship between work support and job satisfaction.

3.3 *Machine Learning Methods*

Although machine learning has been advancing for several years, it has only recently been used for organizational behavior research and social science applications (Gupta et al., 2022). This section of the literature delves into several machine-learning approaches used in the context of WLB. Recent studies that have made significant contributions to the field of organizational behavior are highlighted in this section.

The limited use of Machine Learning (ML) in organizational behavior research and social science applications, notably in the context of WLB studies, provided the context for Gupta et al. (2022) research. Researchers Gupta et al. (2022) combined statistical and ML techniques to predict workers' subjective feelings of WLB. They used factor analysis and multiple regression analysis to find the most predominant factors affecting the WLB of working women in IT in India. Specifically, the principal component analysis (PCA) was used for dimensionality reduction. They then used the multiple regression analysis to find the most significant factors affecting WLB. Support Vector Machine (SVM) was utilized to compare ML to linear approaches. Support Vector Machine (SVM) outperformed linear techniques, with self-care time, sociability, daily work hours, family support, workplace flexibility, gender equality, and workplace health support all having a substantial impact on WLB.

Paigude and Shikalgar (2022) contributed significantly to organizational behavioral research. While using the same dataset as researchers Gupta et al., they used different deep learning models to predict WLB, specifically Long Short-Term Memory (LSTM) and Multilayer Perceptron (MLP). These models were developed, and their performance was compared against one another to see which model produced better results. They also carried out factor analysis and multiple regression analysis to find the most predominant factors affecting WLB. The same predominant factors as in the study of Gupta et al. were discovered. Four of these components are related to the corporate structure. The average number of hours worked per day is one of the most influential criteria to examine in relation to WLB in the organizational context. Followed by workplace flexibility, gender equality, and support for health-related infrastructure in the workplace. The other components are related to the personal and family aspects. Several variants of Multilayer Perceptons (MLP) and Long Short-Term Memory (LSTM) were modeled and their performance was evaluated using the Root Mean Squared error. The results demonstrated that both the LSTM and MLP models accurately predicted WLB, with the LSTM outperforming the MLP and linear models. Both deep learning models performed better than the linear and nonlinear models used in the research by Gupta et al.

Researchers Shah and Parekh (2023)also used deep learning techniques to predict WLB. They focused on understanding the determinants of work-life balance with a specific focus on the concept of quiet quitting and they also aimed to establish whether different attributes pose varying degrees of importance in establishing work-life balance across different age groups. Their study compromised 15,977 survey responses encompassing 25 attributes related to work and personal life.

Their research identified several factors contributing to the concept of quiet quitting and ultimately the WLB of individuals. Factors such as inadequate sleep, high daily stress levels, and lack of recognition were the most determinant factors. Their study employed an Artificial Neural Network (ANN) which was able to deliver high accuracy with a test accuracy of 73.91%. Emphasizing how well the ANN model is able to capture the dynamics of WLB which are complicated.

Researchers Shah and Parekh (2023) assessed the effect of Work from home on the WLB of IT employees using regression analysis. Easiness to manage current workload, satisfaction with job performance, compensation for work from home, and peaceful working environment were used as predictors to predict workers WLB. Their study reached an R^2 of 0.459.

Work-Life Balance (WLB) research has traditionally relied on Questionnaires, a popular and versatile instrument, supplemented by various statistical methodologies, to discover the multiple factors impacting employees' WLB. Support Vector Regression (SVR), Multiple Linear Regression (MLR), Artificial Neural Network (ANN), and Multilayer Perceptron (MLP) have recently been introduced for predicting WLB with high accuracy. While these methods have primarily been used in organizational contexts, this study takes a different approach, focusing on individual elements such as lifestyle, habits, and behavior, all of which play an important role in determining one's ability to balance work and personal life. Support Vector Regression (SVR), Multiple Linear Regression (MLR), and Multilayer

Perceptron (MLP) will be used for predicting WLB in this study as these methods have delivered accurate results for predicting WLB.

4 METHOD

Chapter 4 describes the methodology used to address the research questions. Starting with a description of the dataset used for this study. Followed by the key findings of an exploratory analysis of the dataset and the several preprocessing steps performed on the dataset to make it ready for further analysis. This section also gives an overview of the different methods used to predict WLB.

4.1 *Dataset Description*

For this project, the Lifestyle and Wellbeing dataset is used. The dataset is publicly available on Kaggle (Dalat, 2021). This dataset consists of 15972 survey responses from the Authentic-Happiness.com global work-life survey, making it globally accessible. The survey evaluates how well individuals shape their lifestyle, habits, and behaviors to maximize their overall life satisfaction along 5 dimensions, namely Healthy body, Healthy mind, Expertise, Connection, and Meaning.

The dataset is downloaded in CSV format and each row of the dataset represents a participant. The survey was conducted from July 2015 until February 2020. This dataset contains 24 features, which are listed in Table 1. Table 1 also gives the data type of the different features within the dataset. The dataset has ordinal and categorical variables such as AGE, GENDER, 'BMI RANGE' and SUFFICIENT INCOME.

Table 1: Dataset Description

Feature	Data Type	Range
Timestamp	datetime	July 2015 up to February 2020
FRUITS_VEGGIES	int64	Between 0 and 5
DAILY_STRESS	object	Between 0 and 10
PLACES_VISITED	int64	Between 0 and 10
CORE_CIRCLE	int64	Between 0 and 10
SUPPORTING_OTHERS	int64	Between 0 and 10
SOCIAL_NETWORK	int64	Between 0 and 10
ACHIEVEMENT	int64	Between 0 and 10
DONATION	int64	Between 0 and 5
BMI_RANGE	int64	1: below 25, 2: above 25
TODO_COMPLETED	int64	Between 0 and 10
FLOW	int64	Between 0 and 10
DAILY_STEPS	int64	Between 0 and 10
LIVE_VISION	int64	Between 0 and 10
SLEEP_HOURS	int64	Between 0 and 10
LOST_VACATION	int64	Between 0 and 10
DAILY_SHOUTING	int64	Between 0 and 10
SUFFICIENT_INCOME	int64	1: insufficient, 2: sufficient
PERSONAL_AWARDS	int64	Between 0 and 10
TIME_FOR_PASSION	int64	Between 0 and 10
WEEKLY_MEDITATION	int64	Between 0 and 10
AGE	object	Less than 20, 21 to 35, 36 to 50, 51 or more
GENDER	object	male, female
WORK_LIFE_BALANCE_SCORE	float64	Sum of different categories

4.2 *Exploratory Data Analysis*

To gain a better understanding of the different features within the dataset and to identify potential transformations needed several visualizations were conducted. Histograms, crosstabs, violin plots, and correlation matrices were drafted to inspect the categorical features, to better understand the distribution of the different features, and to understand the correlation of different features within the dataset with the target feature. Some key results are given in this section and further analysis is given in the appendix section A (page 36).

The target feature within this study is WLB, which is represented by the feature WORK LIFE BALANCE SCORE, this feature is the sum of the five different dimensions measured within the survey. According to the survey's designers, a poor score is less than 550, a favorable score is greater than 680, and an exceptional score is greater than 700.

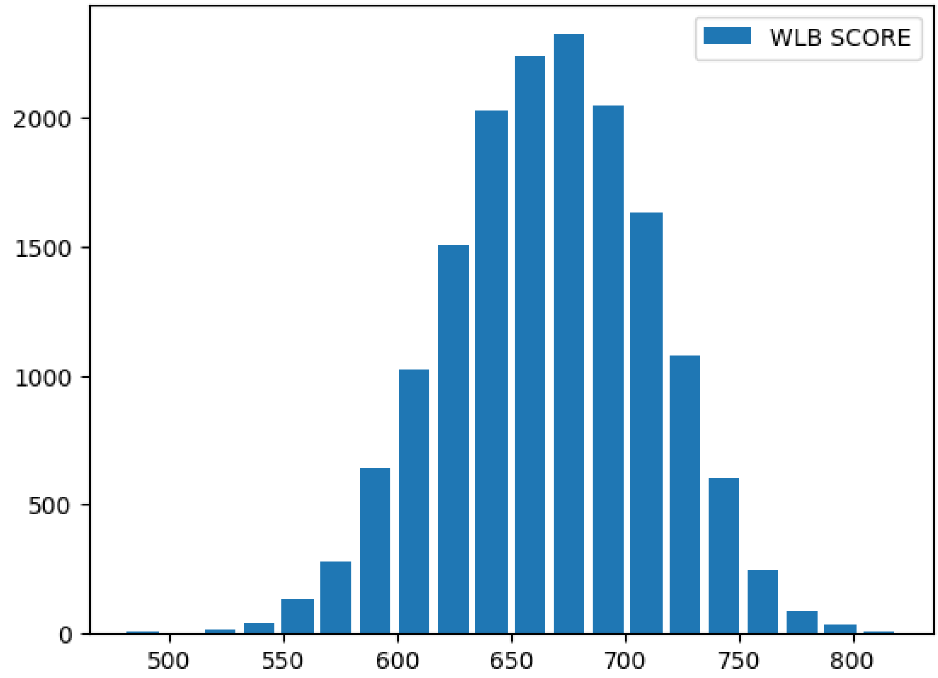

Figure 1: Work-life Balance Distribution.

Figure 1 gives the distribution of the WLB feature. The histogram visualization highlights that a significant portion of the WLB scores lies between 635 and 700, respondents within this range have a moderate to high level of WLB. The WLB features follows a quite normal distribution, with scores that are relatively even and balanced. However, the distribution does slightly lean to the left, indicating that more respondents have scores on the lower end of the scale.

GENDER	**AGE**				**TOTAL**
	Less than 20	**21 to 36**	**36 to 50**	**51 or more**	
FEMALE	6.67%	22.17%	18.54%	14.34%	61.72%
MALE	4.72%	16.07%	10.61%	6.88%	38.28%
TOTAL	11.39%	38.24%	29.14%	21.22%	100%

Table 2: Crosstab Age and Gender

A gender imbalance was detected in the dataset during exploratory data analysis. The dataset includes 61.72% female respondents and 38.28% male respondents. Furthermore, the crosstab in Table 2 shows that most survey participants are working adults, primarily between the ages of 21 and 36, and between the ages of 36 and 50.

To further inspect these categorical features in relation to the target feature, a violin plot was created. The violin plot is given in Figure 2.

Males and females have very similar WLB distributions in the age group "less than 20," with comparable median values and interquartile ranges. The age group 51 or more also shares similar median WLB values for both females and males. However, for the other age groups, the median values are different, with females having higher median WLB scores than males. The largest age group “21 to 35”, has a left-tailed distribution for both genders indicating that that are some quite lower values that are pulling the distribution to the left. Another observation is that the distribution of Work-Life Balance scores is consistent across age groups, with the exception of the 36 to 50 age group, which shows different patterns.

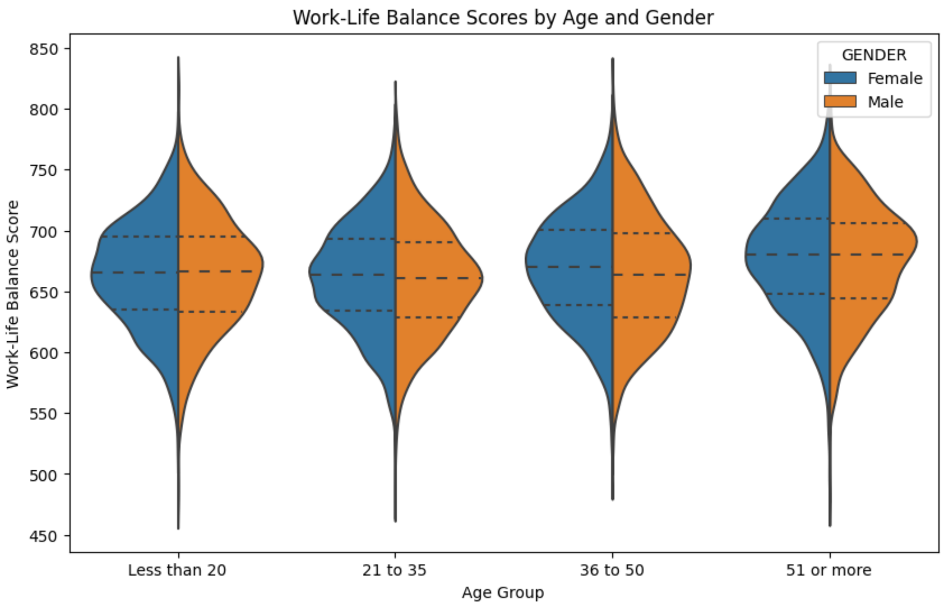

Figure 2: Work-life Balance score by Age and Gender.

The distribution of the other features was also inspected, there were no negative values. This is because the survey contains counts and ratings therefore the values are either zero or positive. However, the distribution of several features was skewed. Achievement, Live vision, Lost vacation, Flow, and Time for passion were highly skewed. A more in-depth analysis of these features is given in Appendix A (page 36). The transformation of these features is given in Section 4.3.3.

Correlations were also drawn to assess whether the different lifestyle and behavioral attributes have a correlation with the target feature. The correlation matrix is given in Figure 3.

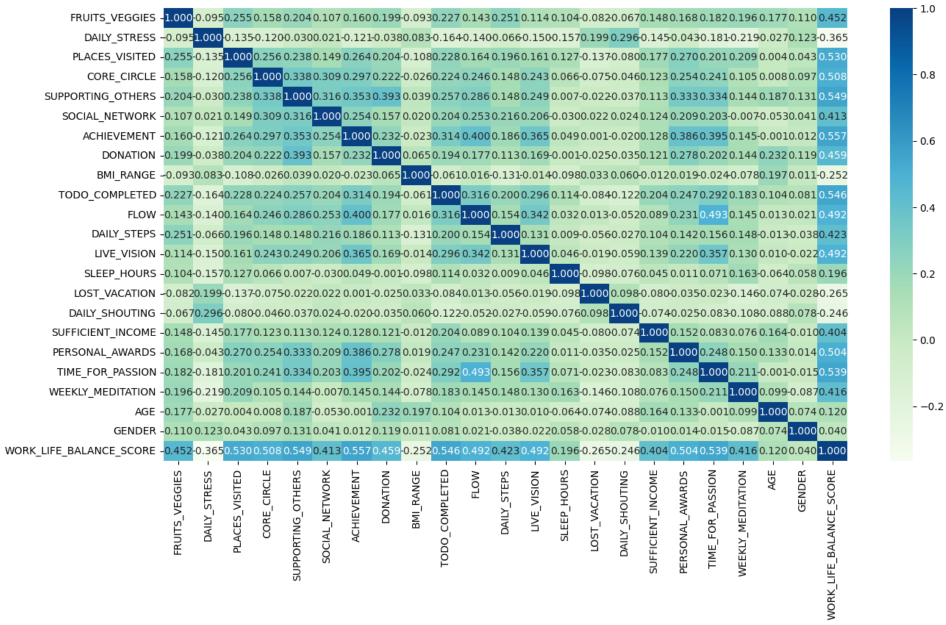

Figure 3: Correlation Matrix.

The correlation matrix revealed that several features have a positive moderate correlation with the target feature. The feature achievement had the highest correlation with the work-life balance feature among the different features. However, Gender and Age had a very low correlation with the target feature. The lack of a significant correlation of Gender and Age with Work-life balance indicates that changes in Age or Gender do not appear to influence variations in work-life balance.

4.3 *Dataset Preprocessing & Cleaning*

4.3.1 *Categorical Features Transformation*

During the preprocessing phase, categorical variables were encoded and mapped. The categorical variable gender was converted into a binary form using one-hot encoding, with "Female" represented as 1 and "male" as 0. The age feature was also encoded, but to ensure the order within this categorical variable, a mapping was used. The age groups were encoded as 0 for "less than 20", 1 for "21 to 35", 2 for "36 to 50", and 3 for "51 or more".

4.3.2 *Handling Missing Data & Feature Removal*

The dataset was inspected for missing values, only 1 row of the Daily stress feature had a missing value. This was addressed by imputing the missing value with the median value of the feature. This method was chosen because there was only 1 row that was missing and because the median value is representative of the typical daily stress level in the dataset. The timestamp feature was also dropped from the dataset because it will not be used within the analysis.

4.3.3 *Data transformation*

The dataset consists of survey responses that contain counts and ratings therefore there were no negative values. However, the distribution of several features was quite skewed with some features having a positive skewness and some a negative skewness. To reduce the skewness of these features 3 different skewness methods were tested. Log transformation, Quantile transformation, and Yeo-Johnson transformation were applied, the results of these transformations are given in Table 3.

Table 3: Skewness Transformations

Features	**Pre-transformation**	**Log transformation**	**Quantile Transformation**	**Yeo-Johnson Transformation**
Achievement	0.63	-0.52	0.004	-0.06
Live vision	0.74	-0.23	-0.04	-0.06
Lost vacation	0.92	0.43	0.27	0.25
Flow	0.87	-0.34	-0.02	-0.03
Time for passion	0.87	-0.21	-0.05	-0.03
Sleep hours	-0.36	-0.95	-0.01	0.05
To do completed	-0.36	-1.36	0.02	-0.22

The Quantile and Yeo-Johnson transformations were the best-performing methods, these methods reduced the skewness effectively of these features bringing some of the scores closer to zero. The Quantile transformation gave the best results therefore this method was applied to these features to reduce the skewness.

4.4 *Algorithms*

4.4.1 *Support Vector Regression*

Support vector regression (SVR) is a popular choice for both linear and nonlinear regression prediction and curve fitting (Parbat & Chakraborty, 2020). Support Vector Regression extends binary classification from Support Vector Machines to perform regression estimations (Panagopoulos et al., 2019). SVR aims to find a hyperplane that best represents the relationship between the input features and the output variable while minimizing the prediction error (Agustina et al., 2018).

The equation of the hyperplane in SVR is given by:

$$f(x) = \langle w, x \rangle + b$$

The SVR model has important parameters to consider such as the choice of kernel function, the regularization parameter C which balances the trade-off between fitting the data and keeping the weights small (Agustina et al., 2018). Gamma and epsilon are other important aspects of SVR that can impact the models performance (Bagheripour et al., 2015). Table 4 gives the list of hyperparameters tested for the SVR model using a grid-search.

Hyperparameter	**Values**
KERNEL	'linear', 'poly', 'rbf', 'sigmoid'
C	0.1, 1, 10, 100
EPSILON	0.001, 0.01, 1
GAMMA	0.001, 0.01, 1

Table 4: SVR Hyperparameter Values

4.4.2 *Multiple Linear Regression*

Multiple Linear Regression (MLR) models the relationship between a dependent variable y and multiple independent variables $x_1, x_2, \ldots, x_p$ using a linear equation (Trunfio et al., 2022). MLR extends the simple linear regression model, which involves only one single exploratory variable (Trunfio et al., 2022). It assumes a linear combination of the input features with coefficients $b_0, b_1, \ldots, b_p$ and an error term e (Brown, 2009). The MLR model was used to predict the value of the dependent variable (WLB) based on several independent variables (lifestyle and behavioral variables). The equation of the MLR model is given by:

$$y = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p + e$$

4.4.3 *Multilayer Perceptron (Neural Network)*

Multilayer Perceptron (MLP) often called "back-propagation" network is a type of artificial neural network with multiple layers, consisting of an input layer, one or more hidden layers, and an output layer (Elansari et al., 2023). Neurons in each layer are connected to neurons in the subsequent layer with associated weights, neurons in the same layer are not connected (Ramchoun et al., 2016). The choice of layers numbers and neurons in each layers and connections is known as the architecture problem and the main goal is to optimize it for a suitable network with sufficient parameters and good generalizations (Elansari et al., 2023). For the MLP model a small architecture consisting of an input layer, a hidden layer with 64 neurons and output layer. A simple model architecture was chosen considering the datasets small feature set and to balance the models complexity given a limited set of features. Several hyperparameters were tested for this architecture, these parameters are given in Table 5.

Table 5: MLP Hyperparameter Values

Hyperparameter	**Values**
Batch Size	32, 64, 128, 256
Learning Rates	0.0001, 0.0015, 0.001, 0.002
Activation Function	'relu', 'elu', 'selu', 'tanh', 'sigmoid', 'linear'
Dropout Rate	0.0, 0.1, 0.2, 0.3
Optimizers	Adam, Nadam

4.5 *Feature Importance*

To be able to identify predominant features affecting the models predictive performance, permutation importance was performed. The sklearn function *permutation_importance* was used to calculate the permutation importance scores for the different models. Permutation scores were generated after training and fitting the models. Permutation importance scores assess how a model's performance changes when the values of a certain feature are randomly shuffled. By analyzing the impact of each feature's variation on the overall predictive capabilities, this technique aids in identifying features that significantly contribute to the model's accuracy. Given the variety of models used for predicting WLB in this study, it is critical to use a method that is compatible with multiple models, allowing for adaptation and interpretability for a meaningful comparison of detected aspects. In this study we use Mean Absolute error (MAE) as a measure to quantify these variations.

4.6 *Splitting & Scaling Dataset*

The dataset was splitted into training, validation and test set. The training dataset was used to train the model, the validation dataset was used for hyperparameter tuning and the test set was used to evaluate the models performance after all hyperparamters were defined. The training, validation and test dataset were also standardized using the he sklearn function *StandardScaler*. Standardization converts the features to have a mean of zero and a standard deviation of one, reducing the impact of different scales and facilitating machine learning model convergence and performance.

4.7 *Evaluation Metrics*

MAE is used as evaluation metric to evaluate the performance of the different models used in this study. MAE is used because it is easy to interpret because it represent the average size of errors between the expected and the true response values. This makes it easy to understand the magnitude of errors for each model.

For the SVR model, MAE scores are generated for each fold, and the final test model is also evaluated using the MAE metric. The MLR model performance is also evaluated using the MAE metric. For the MLP model MAE scores are also generated for each fold when hyperparameter tuning. Mean Squared Error (MSE) is used as the loss function during training for the MLP model. MSE calculates the average squared difference between the anticipated and true response values and is an important factor in optimizing the model's performance. MAE is used to evaluate the training, validation, and test sets.

Different cross-validation folds were employed in the MLP and SVR models. Following extensive hyperparameter testing, it was determined that a 3-fold cross-validation technique was best suited to the MLP model. This choice was taken to establish a balance between computing requirements, time constraints, and the necessity for efficiency optimization based on the hyperparameters studied.

4.8 *Workflow*

Figure 4 depicts the workflow for this study. The procedure is divided into several steps, beginning with an exploratory data analysis to get insights into the dataset's properties. Several pre-processing steps were then undertaken to prepare the data for modeling. The exploratory data analysis revealed skewed features, forcing a focus on data transformation through testing with various methodologies. The dataset was then partitioned and the features were scaled. Following that, the models were trained and optimized. The models were evaluated, and then permutation scores were generated. Finally, the models' performance was compared using the dominating attributes identified by permutation scores.

4.9 *Softwares*

The dataset preprocessing, cleaning, and modeling stages were carried out using Python 3.10.12 in Google Colab. The following libraries were used:

- Numpy (Harris et al., 2020)
- Pandas (pandas development team, 2023)
- Matplotlib (Hunter, 2007)
- Seaborn (Waskom, 2021)
- Scikit-learn (Sklearn) (Pedregosa et al., 2011)
- Keras with Tensorflow (Chollet et al., 2018)

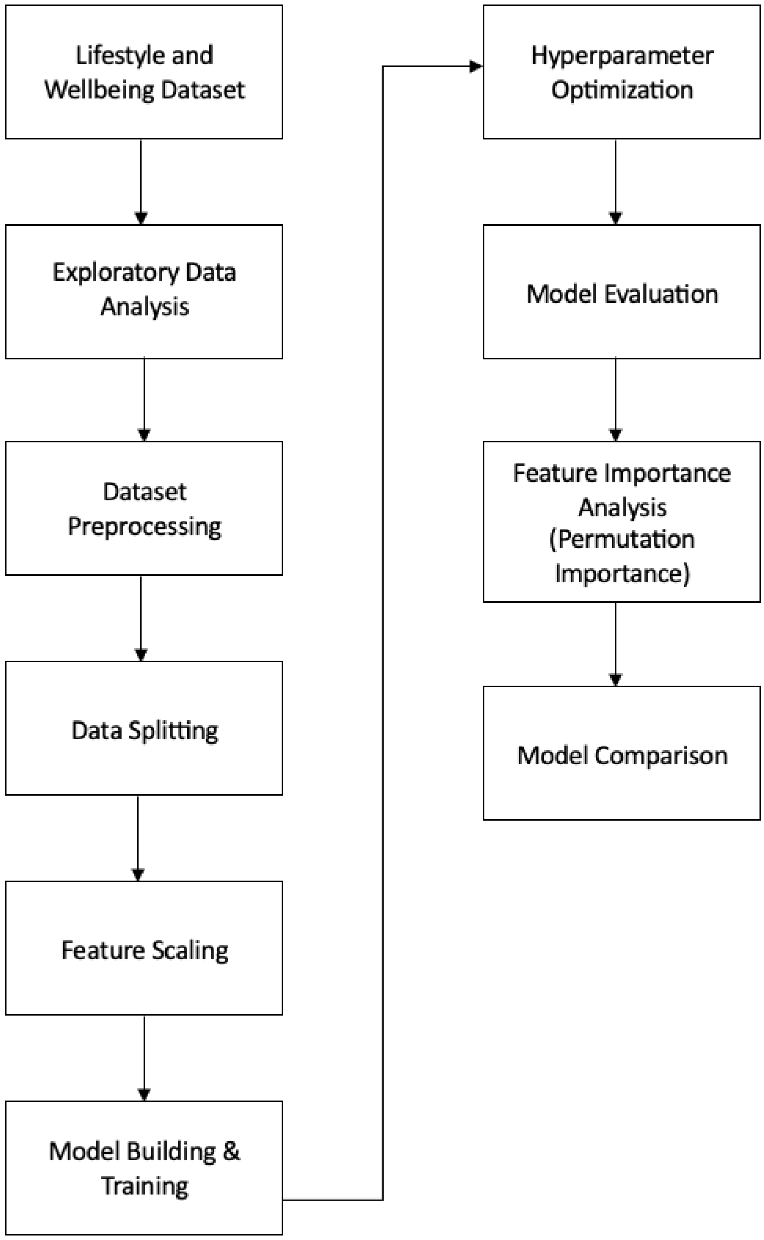

Figure 4: Methodology Workflow

5 RESULTS

The main goal of this thesis is to find the predominant lifestyle and behavioral factors affecting one's subjective feeling of WLB. Three different models were trained initially using the entire feature set, and subsequently the dataset was reduced by incorporating the predominant features for model comparison.

5.1 *Hyperparameter tuning and Cross-Validation Results*

5.1.1 *SVR Hyperparameter tuning and Cross- validation results (With entire dataset)*

A grid search was performed to find the best set of hyperparameters for the SVR model. The combination of hyperparameters that gives the lowest MAE on a 5-fold cross-validation is given in Table 6. The combination of hyperparameters reached an MAE value of 0.981.

Table 6: SVR Best Hyperparameter Values

Hyperparameter	**Values**
KERNEL	rbf
C	100
EPSILON	0.1
GAMMA	0.01

5.1.2 *MLR Hyperparameter tuning and Cross- validation results (With entire dataset)*

For the MLR model used in this thesis, no hyperparameters were tuned.

5.1.3 *MLP Hyperparameter tuning and Cross- validation results (With entire dataset)*

For the MLP model a 3 fold cross validation was used, the parameters that yield the best results are given in Table 7. The model worked with Nadam as optimizer and Sigmoid as activation function for the hidden layer.

Table 7: MLP Best Hyperparameter Values

Hyperparameter	Values
Batch Size	32
Learning Rates	0.0001
Activation Function	'sigmoid'
Dropout Rate	0.0
Optimizers	Nadam

5.1.4 *SVR Hyperparameter tuning and Cross- validation results (With Predominant Features)*

The best performing combination of hyperparameters achieved an MAE score of 10.485, on a 5-fold cross-validation. The best combination of hyperparameters is given in Table 8. The model worked with an rbf kernel.

Table 8: SVR Best Hyperparameter Values

Hyperparameter	Values
KERNEL	rbf
C	100
EPSILON	0.001
GAMMA	0.01

5.1.5 *MLP Hyperparameter tuning and Cross- validation results (With Predominant Features)*

A 3-fold cross-validation was used to train the MLP model. The parameters that yieled the best results on this reduced dataset are given in Table 9. The model worked with Nadam as optimizer and elu as activation function for the hidden layer.

Table 9: MLP Best Hyperparameter Values

Hyperparameter	Values
Batch Size	32
Learning Rates	0.0015
Activation Function	'elu'
Dropout Rate	0.0
Optimizers	Nadam

5.2 *Model Performance on Test Data*

This section displays the performance of fine-tuned models on test data utilizing the entire feature set. All models with the best hyperparameters on the training data were retained after the optimal sets of hyperparameters for each model were identified using cross-validation. The results of the different models are given in Table 10. Out of the 3 different models, the Multiple Layer Perceptron gave the best results with an MAE score of 0.5969. The Support Vector Regressor also performed well on the test dataset with a MAE score of 0.9063. The Multiple Linear Regression (MLR) model, on the other hand, had comparatively higher errors. The MLR model had a prediction error of roughly 2.8790 units from the true value on average.

Table 10: Model Performance

Models	**Mean Absolute Error**
Support Vector Regressor	0.9063
Multiple Linear Regression	2.8790
Multiple Layer Perceptron	0.5969

Figure 5 and 6 gives a visualization of the predicted versus the actual values of the best-performing models utilizing the entire feature set.

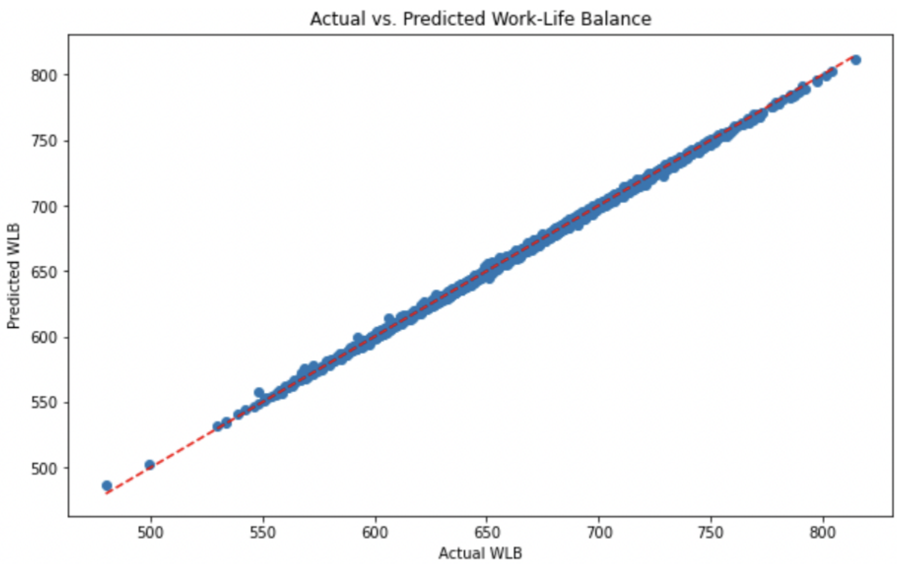

Figure 5: Actual vs. Predicted values SVR.

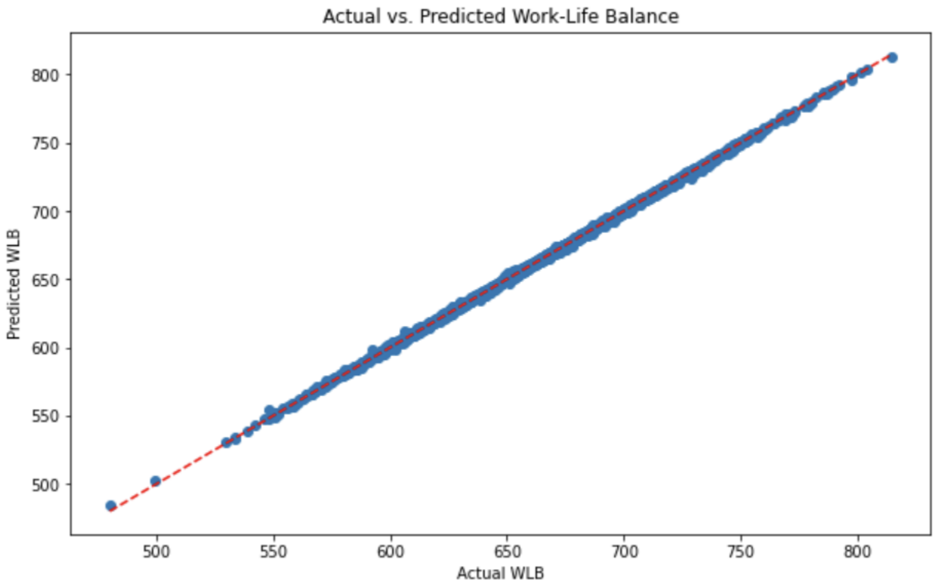

Figure 6: Actual vs. Predicted values MLP.

5.3 *Model Learning*

This section delves into the training and learning of the different models. The best-performing model is the MLP model, which trains and generalizes very well. The smooth decline in the MAE value for both the training and validation set is given in Figure 7.

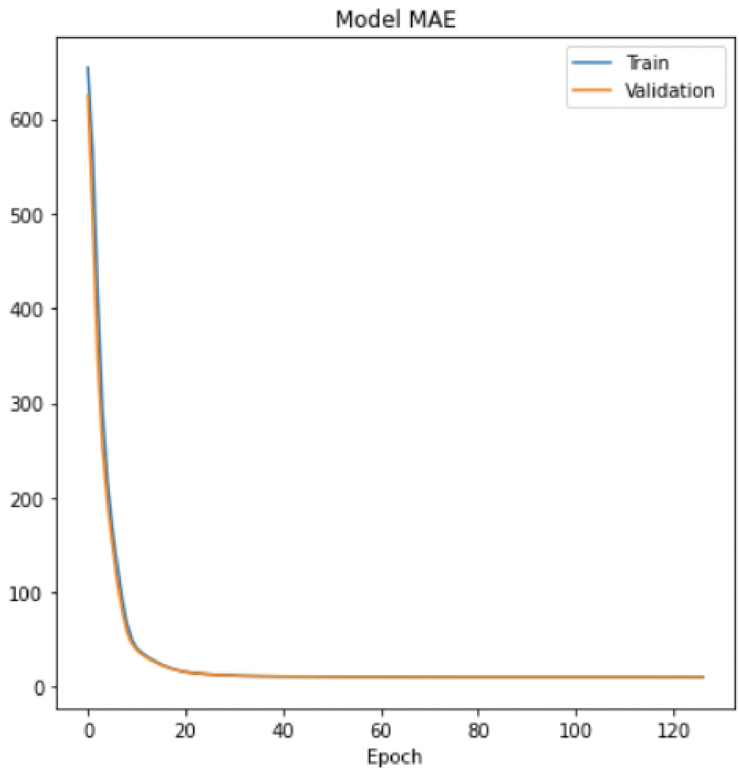

Figure 7: MLP Final Training MAE

The SVR learning curve is given in Figure 8. From the visualization it can be seen that there is a decline in the MAE value for both the training and validation set. However, the model performs better on the training set than on the validation set. The less gradual drop in MAE for the validation set shows that the model may fail to generalize effectively, particularly with smaller training samples, indicating a potential overfitting problem.

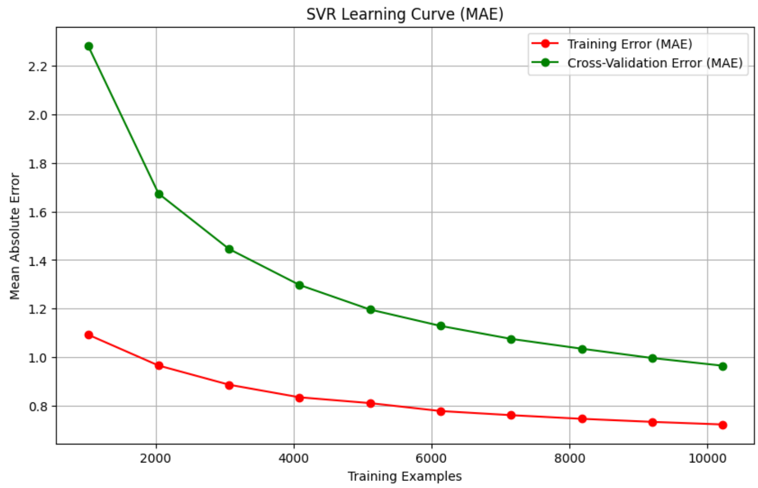

Figure 8: SVR Learning curve

A descriptive table of residuals was generated to inspect the performance of the MLR model during testing. An inspection of the residuals in Table 11, illustrates areas where the model has more difficulties. Looking at the overall distribution of residuals it is visible that the model does have a slight tendency to underestimate. The spread of residuals is relatively moderate, with a few instances of larger errors, both positive and negative.

Table 11: DESCRIPTIVE TABLE OF RESIDUALS MLR Model

	Actual	Predicted	Residual	Residual%
COUNT	3195	3195	3195	3195
MEAN	667.25	667.30	-0.04	0.4
STD	44.58	44.44	3.65	0.33
MIN	480	487.28	-9.10	0.000001
25%	636.50	636.67	-2.69	0.17
50%	667.50	668.12	-0.33	0.37
75%	698.50	699.15	2.18	0.60
MAX	814.50	802.11	19.06	2.64

5.4 *Feature Importance*

One of the primary objectives of this thesis is to identify the most predominant features affecting WLB predictions. Permutation feature importance analysis was performed for the best performing models, SVR and MLP, in order to determine the most influential features in predicting Work-Life Balance (WLB).

The permutation scores for the SVR mmodel are shown in Figure 9. For the SVR model, BMI RANGE stands out as the predominant feature. Changes in BMI RANGE significantly affect the models performance. SUFFICIENT INCOME, DONATION, SUPPORTING OTHERS and PLACES VISITED were the also in the top five most predominant features affecting the models predictive performance. AGE and GENDER again have lower importance values indicating a smaller impact on the model's MAE.

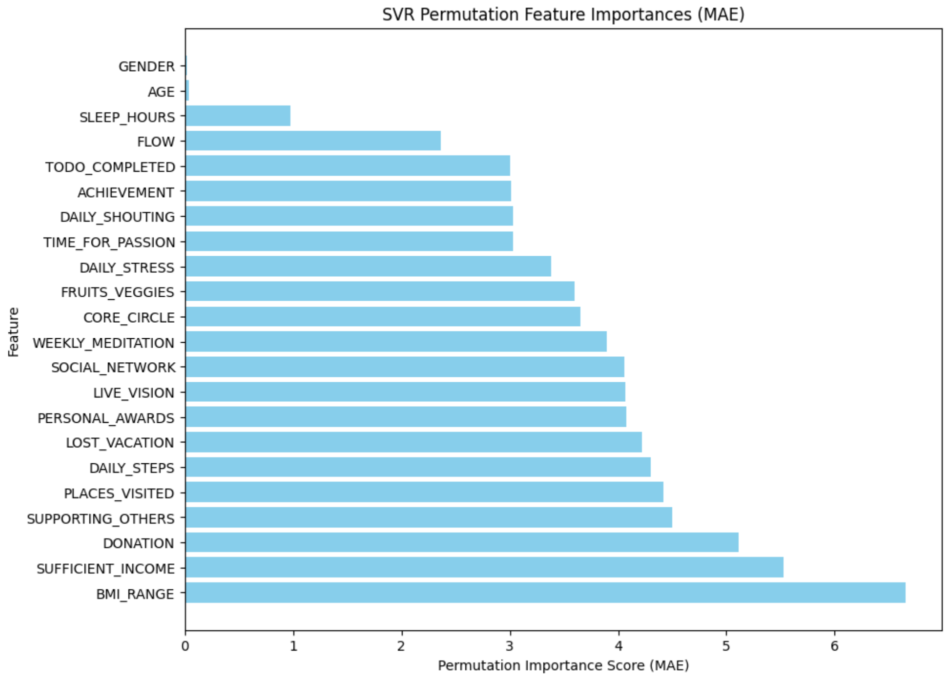

Figure 9: SVR Permutation feature Importance scores

For the MLP model BMI RANGE, SUFFICIENT INCOME, DONATION, LOST VACATION and PLACES VISITED were the most predominant features, a shuffle in these features significantly affect the model's performance. For the MLP model, AGE and GENDER have the least importance, similar to the SVR model. The results are given in Figure 10.

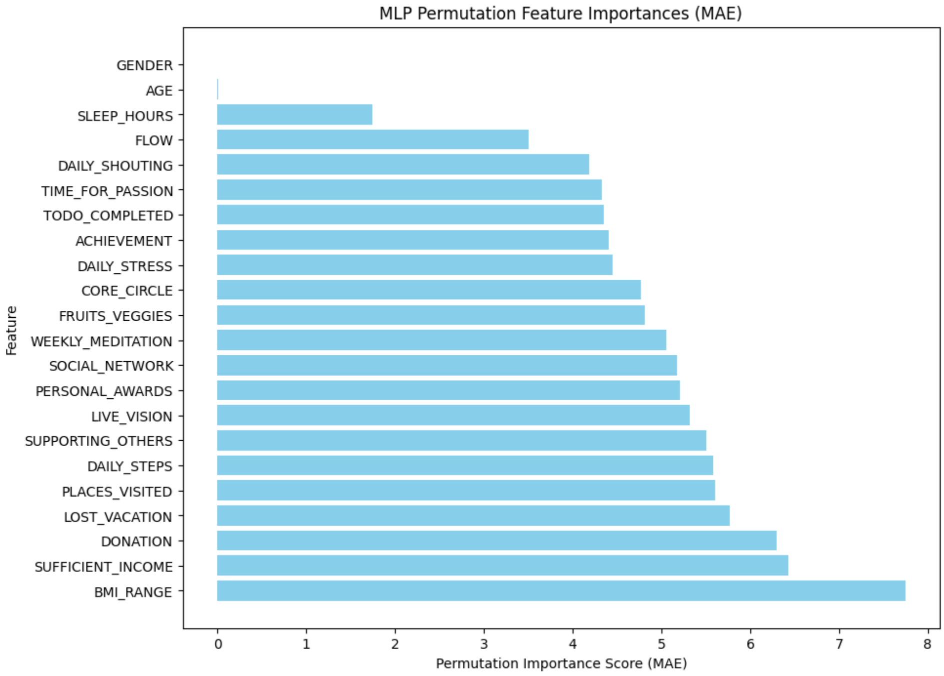

Figure 10: MLP Permutation feature Importance scores

When comparing the ranges of permutation scores it can be seen that the predictors in the MLP model have a larger effect than those in the SVR model. BMI RANGE, SUFFICIENT INCOME and DONATION consistently are the most influential features across all the 3 models. The models shared a set of 13 common features, all of which were important for the models considering the permutation scores. The common set of features were BMI RANGE, SUFFICIENT INCOME, DONATION, LOST VACATION, SUPPORTING OTHERS, DAILY STEPS, PLACES VISITED, LIVE VISION, SOCIAL NETWORK, PERSONAL AWARDS, WEEKLY MEDITATION, CORE CIRCLE AND FRUITS VEGGIES.

5.5 *Model Performance Reduced feature set*

The 13 common features were used to train the 2 best performing models, MLP and SVR. The results are given in Table 12.

Table 12: Model Comparison

	Mean Absolute Error (Original)	Mean Absolute Error (13 features)	Difference
SVR	0.9063	10.5057	9.5994
MLP	0.5969	10.6056	10.0087

The SVR model outperformed the MLP model when using the reduced feature set, achieving an MAE score of 10.5057 versus 10.6056. However, both models did have similar performance when using the reduced feature set. Figure 11 and 12 gives a visualization of the predicted versus the actual values of the SVR and MLP model with the reduced feature set.

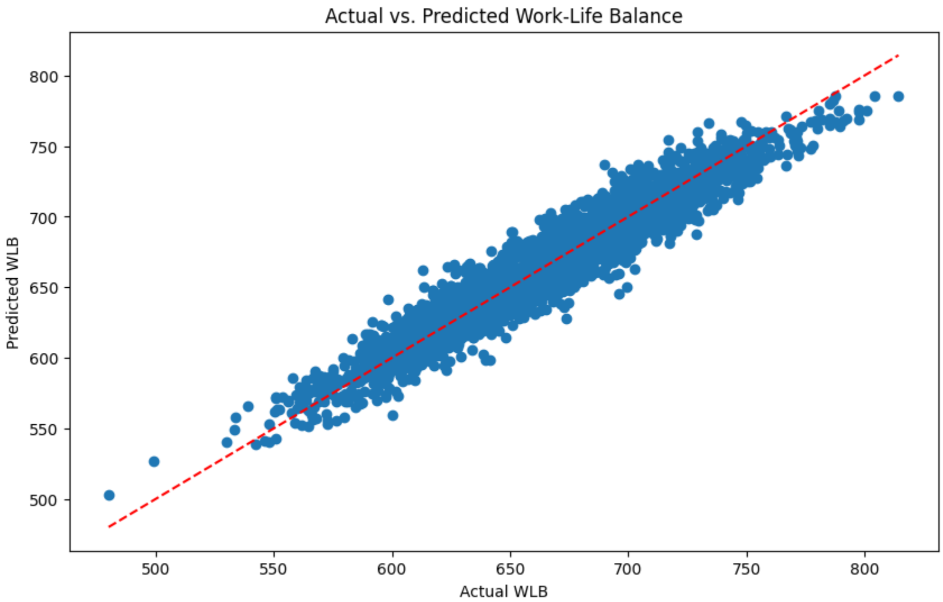

Figure 11: Actual vs. Predicted values SVR Reduced feature set

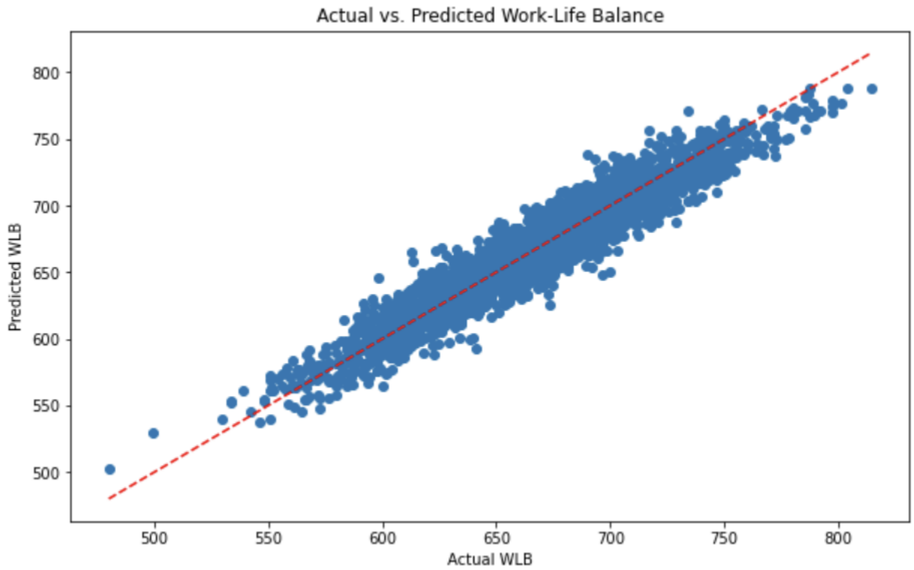

Figure 12: Actual vs. Predicted values MLP Reduced feature set

When comparing the results of the reduced feature set to the original dataset, MAE scores increased when working with the reduced feature set. When using the whole dataset, the models demonstrated better prediction of WLB scores.

6 DISCUSSION

The primary goal of this study is to measure how accurately machine learning models can predict an individual's subjective feeling of WLB using lifestyle and behavior data. The study findings are provided in Chapter 5, and Chapter 6 will go deeper into the interpretation of these findings, explaining potential consequences and implications.

6.1 *Results Discussion*

When observing the performance of the models using the entire dataset, it becomes clear that the MLP model has the best performance in terms of MAE values. The MLP model outperforms other models with an MAE value of 0.5969. However, both MLP and SVR do perform well with relatively low MAE values. Noteworthy, is that the Models do depend on the whole dataset to make accurate WLB predictions when using the lifestyle and behavioral data. There is a large difference in prediction power when the entire feature set is used against when only the predominant features are used. When going from the entire feature set to the reduced set, the MAE value for the SVR Model increases significantly from 0.9063 to 10.5057. Similarly, when using the smaller feature set, the MLP model's MAE increases from 0.5969 to 10.6056.

The most important features influencing an individual's subjective WLB were identified using permutation feature importance scores. The findings revealed numerous features that have a significant impact on the WLB variable. Across all models, the most important characteristics were BMI RANGE, SUFFICIENT INCOME, and DONATION. Features such as SUPPORTING OTHERS, PLACES VISITED AND LOST VACATION were also important contributors for the other models. A notable observation is the models' shared reliance on a single set of top 13 features, which exert the most influence. However, the permutation scores differed amongst models, indicating that each model places a varied emphasis on significant contributors in shaping WLB results. Out of the SVR and MLP models the permutation scores for the MLP model were significantly higher, MLP models are more complex than linear models such as SVR, allowing them to capture complex patterns, non-linear dependencies, and complex interactions in data. The MLP model's high permutation scores demonstrate its ability to exploit a broader collection of features and uncover subtle correlations, especially in cases where linear models may struggle to detect non-linear patterns. Recognizing these subtle differences in feature relevance not only improves our academic understanding, but also provides practical insights for designing interventions to improve Work-Life Balance.

BMI RANGE is the top predictor of WLB across the different models based on the permutation scores. The negative association between BMI RANGE and WLB as given in the correlation matrix in Figure 3 implies that when an individual's BMI score increases their WLB score tends to decrease. In other words, people with greater BMIs are more likely to have lower levels of perceived WLB. This negative correlation suggests that those with higher BMI ranges may perceive difficulties in achieving a satisfactory perception of WLB. One possible explanation for the connection is that people with higher BMIs are more vulnerable to numerous health problems, affecting both their physical well-being and overall life satisfaction. Individuals with higher BMIs may have difficulty balancing many life roles which hinders them in reaching a balance between the domains of life and work. However, it is also important to consider that an unbalanced work-life scenario may contribute to health issues, such as a higher BMI. This implies that addressing work-related problems becomes critical for those with higher BMI ranges in order to enhance their perceived WLB.

The second most important feature is SUFFICIENT INCOME. SUFFICIENT INCOME has a positive correlation with the WLB feature. Indicating if one's perception of their income is satisfactory their WLB will likely increase. This specific feature assesses one's gratitude for their income level. By investigating whether one's income is sufficient to cover basic life expenses, such as the cost of housing, food, education, and health care. A potential explanation for Sufficient Income being a top predictor for WLB is its association with financial security. Having sufficient income not only enables individuals to meet their basic needs but also facilitates the management of job responsibilities and personal interests. This financial stability encourages a more balanced and harmonious WLB.

DONATION is also in the top 3 predictors for WLB. This feature has a positive correlation with the target feature. This suggests that the more time individuals dedicate to giving back, the higher the likelihood that their WLB score will increase. The main focus of this feature is to assess the individuals contribution to good causes, including time and money. One possible explanation for DONATION being one of the top predictors of WLB is that actively participating in charitable activities, whether through time or monetary donations, promotes a sense of fulfillment and purpose, which can have a good impact on an individuals overall perception of WLB.

The features AGE and GENDER did not influence the different models performance when considering the imputations scores. This implies that individual habits and behaviors have a more significant impact on perceived WLB than demographic attributes. Possible reasons for the limited impact of AGE and GENDER could be that individual choices, lifestyle preferences, and behavioral patterns collectively overshadow the influence of age and gender in determining people's perceptions of WLB. The findings highlight the importance of focusing on behaviors and habits when understanding and predicting subjective WLB, as opposed to demographic factors.

6.2 *Comparison with the literature*

The studies presented in the literature concerning predicting WLB, have been focused on studying WLB in the organizational context using corporate expects. Lifestyle and behavioral factors have not been used to study WLB. This study bridges this gap in the literature by studying different lifestyle and behavioral factors to understand how on can improve their WLB by shaping their lifestyle and behavioral actions to increase their WLB. The outstanding performance of the MLP model when using the entire feature set demonstrated in this study confirms the success of deep learning techniques in forecasting WLB, as established in recent years. The literature also highlighted that Support Vector models outperform MLR models and this is also sustained in this study considering that the SVR model outperforms the MLR model when using the entire feature set to predict WLB.

6.3 *Limitations and Future Research*

The dataset used in this study has limitations, particularly due to the survey's voluntary nature. Voluntary involvement may add a selection bias by recruiting individuals with a specific interest in WLB subjects. This survey is being performed on an internet platform that is accessible to people all around the world, regardless of their geographical location. Moreover, the observed gender imbalance within the dataset, with a higher number of female participants relative to males, may be influenced by factors such as the voluntary nature of the survey and the diverse accessibility of the online platform.

The study found that numerous lifestyle and behavioral factors have a considerable impact on WLB and play an important role in determining an individual's subjective perception of WLB. Furthermore, when these features were used, the models displayed great predictive performance. Recognizing the survey's global accessibility, a future research suggestion is to confine the focus to a single target group or industry. Researchers can find industry-specific patterns, difficulties, and effective interventions by going deeper into a specific context, giving tailored insights for enhancing work-life balance within various professional domains.

Despite its limitations, the study provides useful insights into the interaction of lifestyle, behavioral characteristics, and subjective perceptions of WLB. Recognizing these factors openly adds dimension to the continuing conversation about work-life balance and lays the path for future research.

6.4 *Relevance*

As previously mentioned, studies have focused mainly on studying organizational factors to determinant individuals WLB, disregarding the impact of individual aspects that might affect an individuals perception of WLB. This study highlights that numerous lifestyle and behavioral features have an important affect on an individuals WLB and therefor should be considered when predicting WLB.

The relevance of this study extends beyond the academic world to practical applications in a variety of settings. Organizations might use the findings of this study to fine-tune tactics targeted at improving employees' work-life balance. Recognizing the importance of individual lifestyle and behavioral characteristics enables employers and policymakers to conduct customized interventions that promote a more supportive and balanced work environment.

Furthermore, the study is relevant to those attempting to maximize their own WLB. Individuals who are aware of the impact of personal behaviors can make informed decisions to improve overall well-being and develop a better Work-Life Balance (WLB).

7 CONCLUSION

This section addresses the research questions of this study. The main focus of the study was to identify predominant lifestyle and behavioral features affecting an individuals subjective level of WLB. This study also compared the performance of different Machine Learning models when predicting WLB using lifestyle and behavioral features.

7.1 *Research Question 1-Which lifestyle and behavior-related variables significantly impact predicting an individual's subjective feeling of Work-Life Balance (WLB)?*

This research question focused on identifying the predominant features affecting an individuals WLB. The models permutation importance scores revealed that several features affect the different models performance. The top 3 predominant features across the three models are BMI RANGE, SUFFICIENT INCOME AND DONATION. For the SVR model SUPPORTING OTHERS and PLACES VISITED were also in the top five predictors influencing the SVR model's effectiveness when predicting WLB. LOST VACATION and PLACES VISITED were also top features with great permutation importance scores for the MLP model. Overall BMI RANGE, SUFFICIENT INCOME, DONATION, LOST VACATION, SUPPORTING OTHERS, DAILY STEPS, PLACES VISITED, LIVE VISION, SOCIAL NETWORK, PERSONAL AWARDS, WEEKLY MEDITATION, CORE CIRCLE AND FRUITS VEGGIES are the main features influencing an individuals subjective level of WLB.

7.2 *Research Question 2-How does the predictive accuracy of Support Vector Regression (SVR), Multiple Linear Regression (MLR), and Multilayer Perceptron (MLP) differ when utilizing lifestyle and behavior-related variables to predict an individual's subjective feelings of Work-Life Balance (WLB)?*

This sub-question focuses on the comparative performance of the different models when utilizing lifestyle and behavioral attributes as features.

The MLP model outperforms both the SVR and MLR models when using the entire feature set to predict WLB, with an MAE score of 0.5969. Notably, the SVR model also performs well, with an MAE score of 0.9063. When the feature set is reduced to the 13 primary features that both models share, the SVR model beats the MLP model, achieving a lower MAE score of 10.5057 compared to the MLP model's MAE score of 10.6056.

Reducing the feature set reduces the predictive accuracy of the best-performing models, emphasizing the need of using the entire feature set collectively for accurate WLB predictions with minimal errors.

7.3 *Main Research Question-How accurately can machine learning models predict an individual's subjective feeling of Work-Life Balance (WLB) using lifestyle and behavior-related variables?*

The findings highlight that both MLP and SVR models perform very well when predicting an individuals subjective level of WLB when all lifestyle and behavioral factors are used as features. These models demonstrated a great performance with low error margin when utilizing these features as predictors. The results also emphasize the need to include all lifestyle and behavioral factors as features for good predictive performance for both MLP and SVR models. This highlights the importance of a broad feature set in improving models' ability to accurately predict an individual's subjective level of WLB when various lifestyle and behavioral factors are used as features.

The final conclusion of this study is that usage of lifestyle and behavioral features as predictor have led to good predictive performance of an individuals subjective feeling of WLB.

REFERENCES

Agustina, S. D., Bella, C., Ramadhan, M. A., et al. (2018). Support vector regression algorithm modeling to predict the availability of foodstuff in indonesia to face the demographic bonus. *Journal of Physics: Conference Series*, *1028*(1), 012240.

Bagheripour, P., Gholami, A., Asoodeh, M., & Vaezzadeh-Asadi, M. (2015). Support vector regression based determination of shear wave velocity. *Journal of Petroleum Science and Engineering*, *125*, 95–99.

Bhadana, J., Saxena, N., & Bhatia, A. (2022). Uttar pradesh academics' occupational stress, organisational work environment and work-life balance: A quantitative study. *SA Journal of Human Resource Management*, *20*, 1639.

Brown, S. H. (2009). Multiple linear regression analysis: A matrix approach with matlab. *Alabama Journal of Mathematics*, *34*, 1–3.

Chollet, F., et al. (2018). Keras: The python deep learning library. *Astrophysics source code library*, ascl–1806.

Clark, S. C. (2000). Work/family border theory: A new theory of work/family balance. *Human relations*, *53*(6), 747–770.

Dalat, Y. (2021). Lifestyle and wellbeing data. https://www.kaggle.com/datasets/ydalat/lifestyle-and-wellbeing-data

Devadoss, A. V., & Minnie, J. B. (2013). A study of personality influence in building work life balance using fuzzy relation mapping (frm). *Indo-Bhutan International Conference on Gross National Happiness*, *2*, 211–216.

Elansari, T., Ouanan, M., & Bourray, H. (2023). Modeling of multilayer perceptron neural network hyperparameter optimization and training.

Grzywacz, J. G., & Carlson, D. S. (2007). Conceptualizing work—family balance: Implications for practice and research. *Advances in developing human resources*, *9*(4), 455–471.

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

Haar, J. M., Russo, M., Suñe, A., & Ollier-Malaterre, A. (2014). Outcomes of work–life balance on job satisfaction, life satisfaction and mental health: A study across seven cultures. *Journal of vocational behavior*, *85*(3), 361–373.

Harris, C. R., Millman, K. J., Van Der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., et al. (2020). Array programming with numpy. *Nature*, *585*(7825), 357–362.

Helmle, J. R., Botero, I. C., & Seibold, D. R. (2014). Factors that influence perceptions of work-life balance in owners of copreneurial firms. *Journal of Family Business Management*, *4*(2), 110–132.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in science & engineering*, *9*(03), 90–95.

Munyeka, W., & Maharaj, A. (2022). A quantitative study on salient work-life balance challenge (s) influencing female information and communication technology professionals in a south african telecommunications organisation. *SA Journal of Human Resource Management*, *20*, 1874.

Nilawati, F., Umar, N., Kusdi, R., & Zainul, A. (2019). The influence of work life balance and organizational pride on job satisfaction and its impact on organizational citizenship behavior in five and four star hotels employee. *Russian Journal of Agricultural and Socio-Economic Sciences*, *91*(7), 191–196.

Paigude, S. D., & Shikalgar, S. R. (2022). Deep learning model for work-life balance prediction for working women in it industry. *Proceedings of the 4th International Conference on Information Management & Machine Intelligence*, 1–8.

Panagopoulos, O. P., Xanthopoulos, P., Razzaghi, T., & Şeref, O. (2019). Relaxed support vector regression. *Annals of Operations Research*, *276*, 191–210.

pandas development team, T. (2023, December). *Pandas-dev/pandas: Pandas* (Version v2.2.0rc0). Zenodo. https://doi.org/10.5281/zenodo.10426137

Parbat, D., & Chakraborty, M. (2020). A python based support vector regression model for prediction of covid19 cases in india. *Chaos, Solitons & Fractals*, *138*, 109942.

Pawlicka, A., Pawlicki, M., Tomaszewska, R., Choraś, M., & 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), e0232771.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, *12*, 2825–2830.

Poulose, S., & Sudarsan, N. (2017). Assessing the influence of work-life balance dimensions among nurses in the healthcare sector. *Journal of Management Development*, *36*(3), 427–437.

Ramchoun, H., Ghanou, Y., Ettaouil, M., & Janati Idrissi, M. A. (2016). Multilayer perceptron: Architecture optimization and training.

Shah, D., & Parekh, M. (2023). Understanding work-life balance: An analysis of quiet quitting and age dynamics using deep learning.

Trunfio, T. A., Scala, A., Giglio, C., Rossi, G., Borrelli, A., Romano, M., & Improta, G. (2022). Multiple regression model to analyze the total los for patients undergoing laparoscopic appendectomy. *BMC Medical Informatics and Decision Making*, *22*(1), 1–8.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

APPENDIX A

Several lifestyle and behavioral attributes were visualized, and some of them were quite skewed. The Achievement feature illustrated in Figure 13, was highly skewed to the right with peaks around achievement scores of 2 and 3. This indicates that most respondents recorded that over the 12 past months, they reached 2 to 3 achievements that they were proud of.

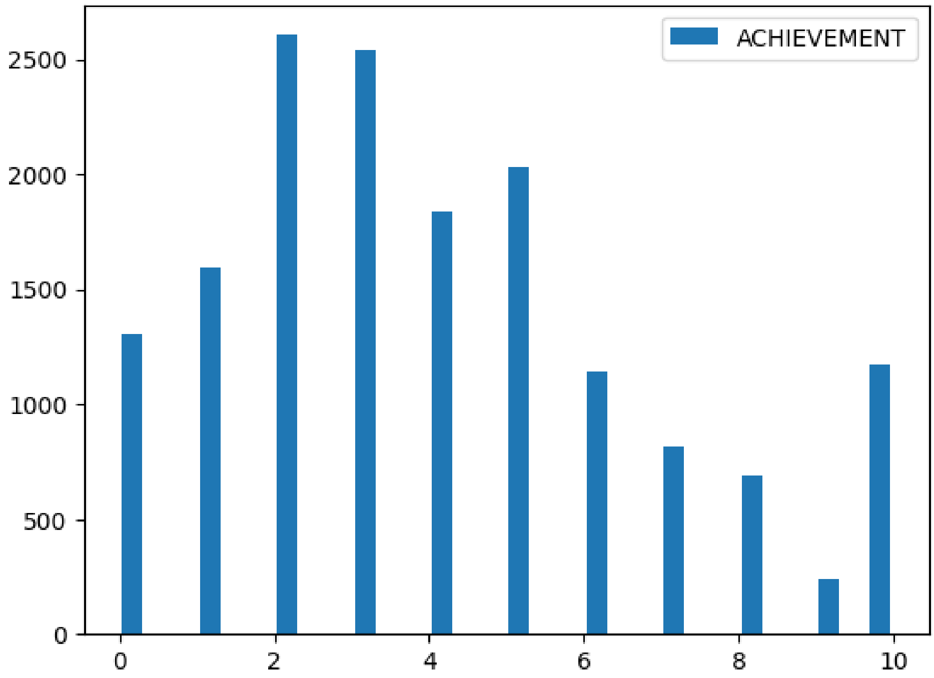

Figure 13: Achievement Distribution

The Flow feature visualized in Figure 14, was also highly skewed to the right with some high peaks at 1 to 2 flow ratings. This indicates that many of the respondents have a low flow level, many respondents experience flow 1 up to 2 hours a day, and some of the respondents also feel it 5 hours a day. Flow refers to being able to be fully immersed in performing an activity, you become energetic, fully focused, and have extreme enjoyment.

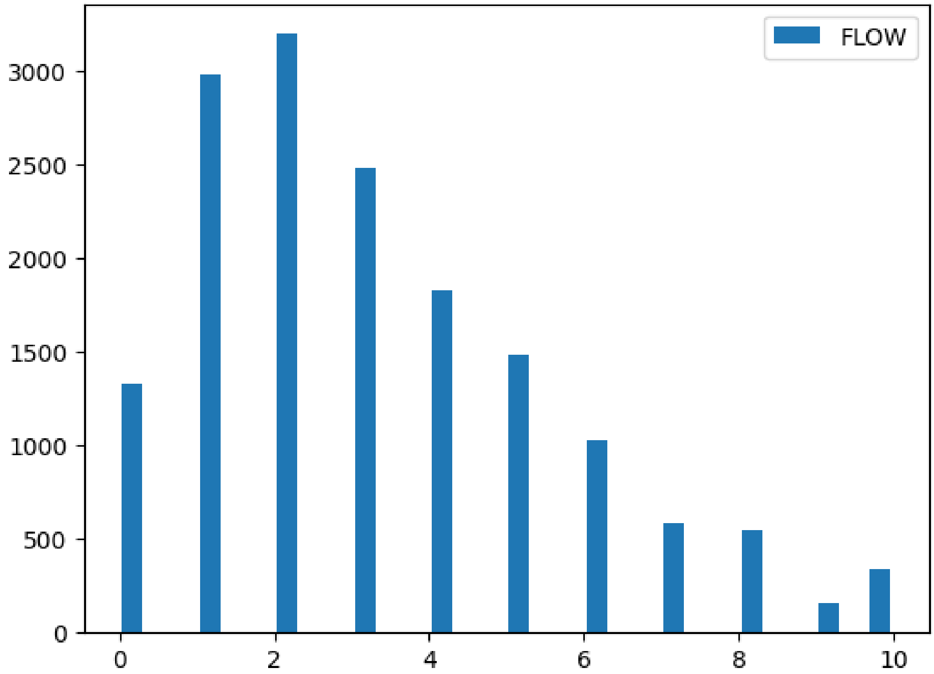

Figure 14: Flow Distribution

The Live vision feature was also skewed to the right with its highest peak at a live vision rating of 5. This feature distribution is given in Figure 15. The Live vision feature assesses whether respondents have a clear live vision planned out for them, the results indicated that most respondents have a clear plan for the next 5 years. Some respondents are focused on immediate goals, with plans focused on 1 year ahead, while others do not have any plans at all. Some respondents also have long-term plans that extend up to 10 years into the future.

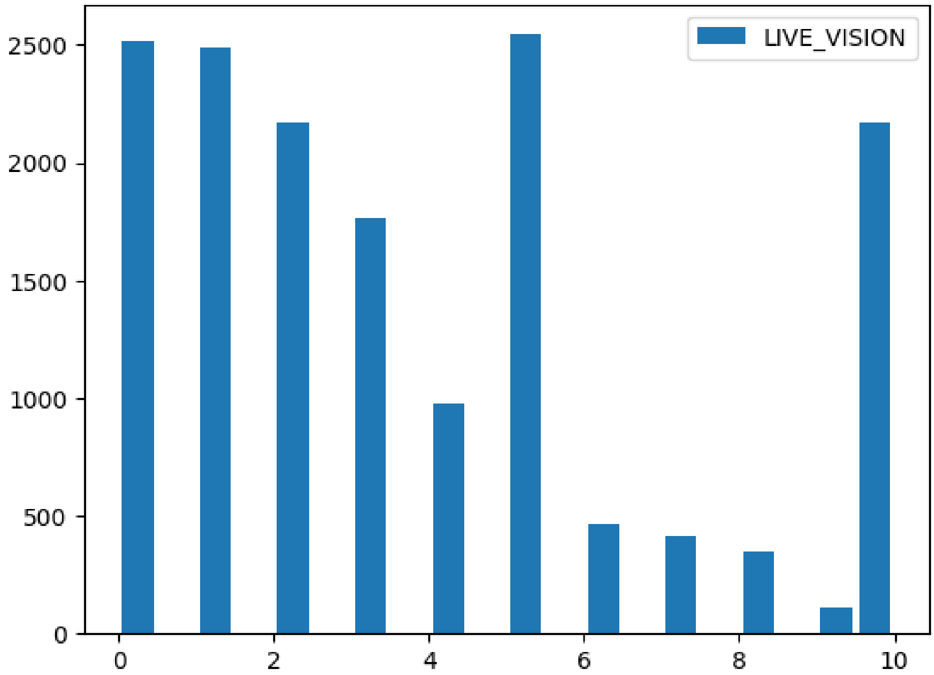

Figure 15: Live Vision Distribution

The feature Time for passion also has a right-skewed distribution with peaks at a score of 1 and 2 hours a day. The distribution is given in Figure 16. Most respondents have about 1 to 2 hours a day to spend on things they love doing for themselves, there are also respondents that have more time to do things for themselves given that there are large values that pull the distribution to the right. The lost vacation feature illustrated in Figure 17, has its highest peak at 0 days, meaning that most respondents use up all their vacation days, there are also respondents who do give up days but most of them make sure they take them all up.

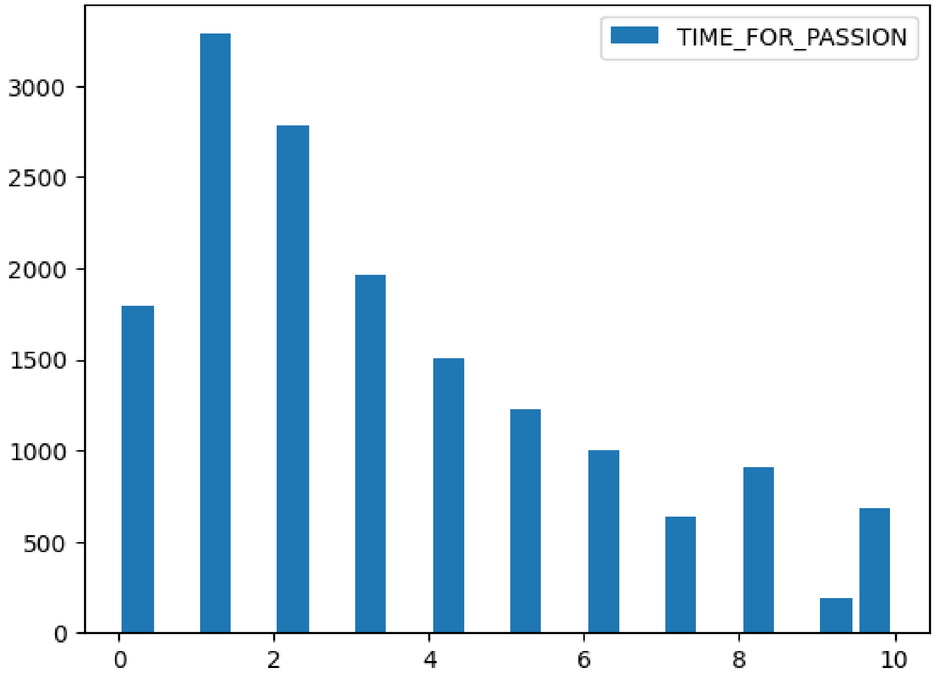

Figure 16: Time for passion Distribution

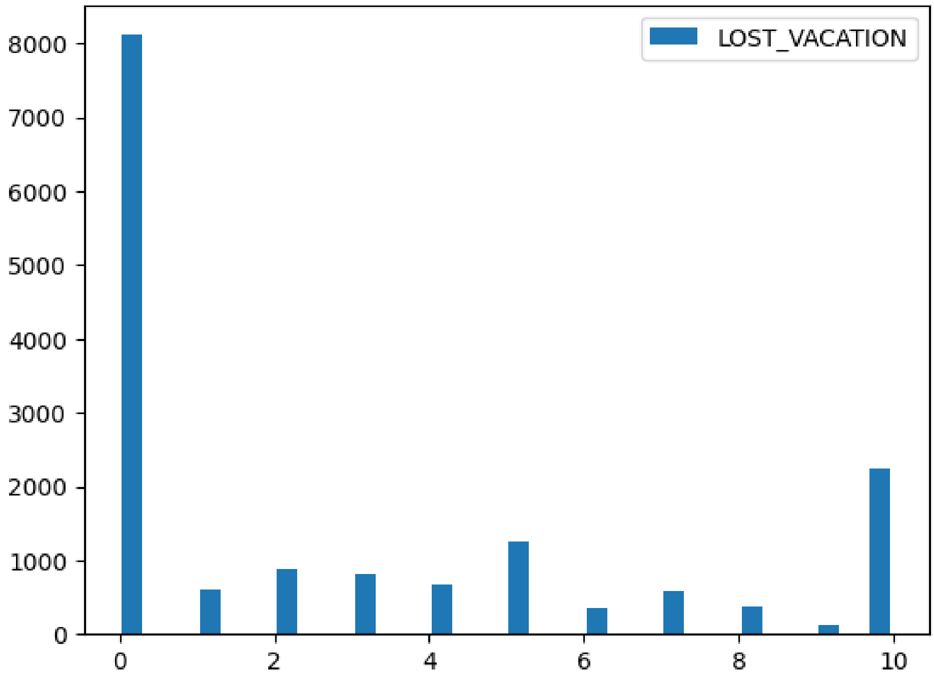

Figure 17: Lost vacation Distribution

The sleep hours and to-do completed features were skewed to the left with a longer tail to the left, their distribution is given in Figures 18 and 19. The sleep hours feature has a peek at 7 hours of sleep, indicating that most respondents sleep on average 7 hours a day during a typical workday and weekend. The feature to do completed has peaks within the distribution at a score of 7 and 8, indicating that the participants are able to complete their weekly goals, work, and personal-related tasks quite well.

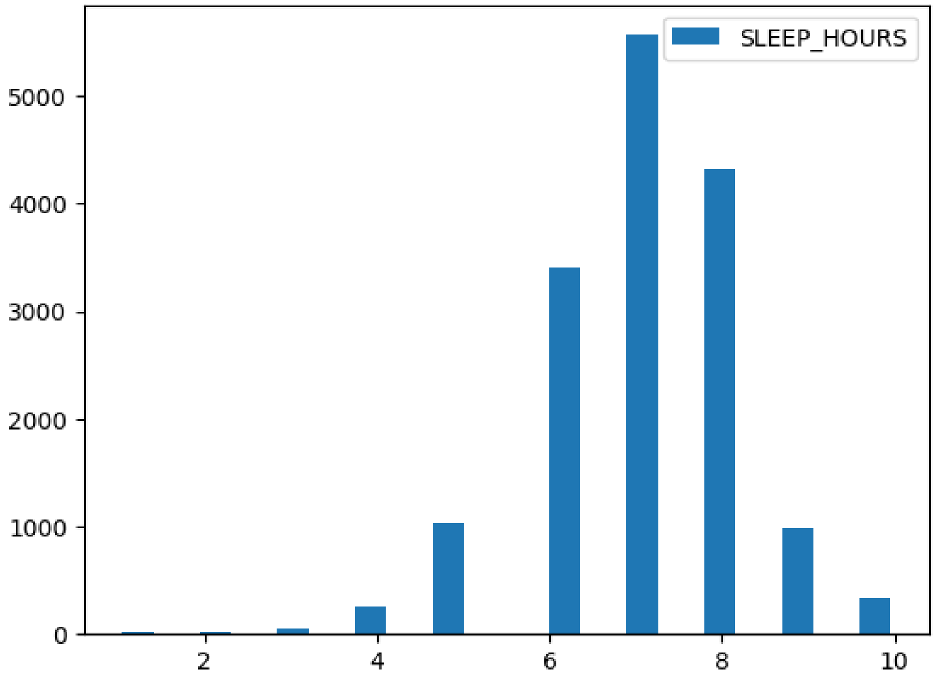

Figure 18: Sleeping hours Distribution

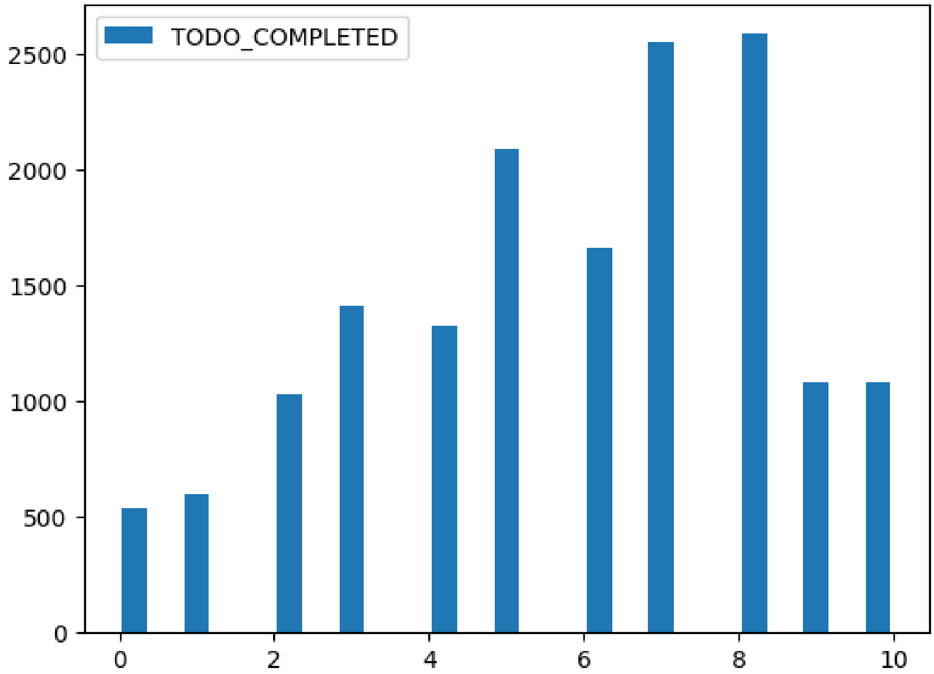

Figure 19: To-do Completed Distribution