

**A MACHINE LEARNING APPROACH TO MODEL SUBJECTIVE WELL-
BEING FROM A PSYCHOLOGICAL PERSPECTIVE (PSİKOLOJİ BİLİMİ
YAKLAŞIMIYLA ÖZNEL İYİ HAL DURUMUNUN MAKİNE ÖĞRENMESİYLE
MODELLENMESİ)**

by

Nail ŞENBAŞ, B.S.

Thesis

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE

in

INDUSTRIAL ENGINEERING

in the

GRADUATE SCHOOL OF SCIENCE AND ENGINEERING

of

GALATASARAY UNIVERSITY

Apr 2022

ACKNOWLEDGEMENTS

Foremost, I would like to express my deepest gratitude to my advisor, Assoc. Prof. S. Emre Alptekin for his enduring support through all the stages of writing this thesis. I am grateful for his patience, valuable guidance, and insightful comments. He taught me more than I could ever give him credit for here.

Besides my advisor, I would like to thank Assoc. Prof. Özlem Durmaz İncel who supported this thesis with her academic network and immense knowledge.

I am indebted to my colleague, Aslı Uludoğan, for her tremendous understanding, support, and tolerance over the last few years.

None of this could have happened without my mother and sister, whose love and guidance are with me in whatever I pursue. I will forever be grateful to them.

Most importantly, I wish to thank my loving and supportive wife, Güzin, who provides unending motivation and inspiration for me in the pursuit of this thesis.

Additionally, I do not want to forget to thank my pet Sumo for always being with me throughout long study sessions.

Mar 2022

Nail ŞENBAŞ

TABLE OF CONTENTS

LIST OF SYMBOLS	vi
LIST OF FIGURES	vii
LIST OF TABLES	ix
ABSTRACT	x
ÖZET	xi
1. INTRODUCTION	1
2. BACKGROUND	6
2.1. What Is Well-Being?	6
2.2. The Concepts of Well-Being.....	6
2.3. Subjective Well-Being	7
2.4. The Assessment of Subjective Well-Being in Psychology	8
2.5. The Validity of Assessment Scales for Ground Truth	9
2.6. Machine Learning Methods	10
2.6.1. Machine Learning Description and Techniques	10
2.6.2. Supervised Learning Techniques	11
2.6.3. The Learning Process of Supervised Machine Learning Algorithms	12
3. RELATED WORKS.....	21
3.1. Technologies of Well-Being	21
3.2. Smartphones	22
3.2.1. Active Data	22
3.2.2. Passive Data	23
3.3. Wearable Devices.....	23
3.4. Tangible Devices.....	24
3.5. Internet and Social Media Platforms	24
3.6. Evaluation of Technology in Well-Being	25

3.7. Evaluation of Related Works	27
4. NETHEALTH PROJECT.....	29
4.1. Surveys Data	30
4.2. Smartphone Data	33
4.3. Fitbit Data.....	34
4.4. Existing Research on Nethealth	35
5. SUBJECTIVE WELL-BEING PREDICTION MODEL AND METHODOLOGY	38
5.1. Description of Model	38
5.2. Feature Selection	39
5.3. Smartphone Features	39
5.4. Smartwatch Features	40
5.5. Surveys Features	41
5.5.1. Personality Traits.....	41
5.5.2. Confidence and Control.....	41
5.5.3. Major Event	42
5.6. Data Processing and Principles	43
5.6.1. Input Data Processing	44
5.6.2. Ground Truth Data Processing	45
5.6.2.1. Subjective Happiness Scale	45
5.6.2.2. The Construction of SHS From Surveys from Nethealth	46
5.6.2.3. Generation of Subjective Well-Being Classes	47
5.7. Machine Learning Models	48
5.8. Results	50
5.8.1. Classification Results	50
5.8.2. Regression Results	54
5.8.3. Discussion	55
6. CONCLUSION	57
REFERENCES.....	58
BIOGRAPHICAL SKETCH.....	64

LIST OF SYMBOLS

ANN	: Artificial Neural Network
Adam	: Adaptive Moment Estimation
AUC	: Area Under Curve
BAI	: Beck Anxiety Inventory
BDI	: Beck's Depression Inventory
BFI	: Big Five Personality
BMR	: Basal Metabolic Rate
CES-D	: Centre for Epidemiologic Studies Depression Scale
GSR	: Galvanic Skin Response
HRV	: Heart Rate Variability
KNN	: K-Nearest Neighbors
LDA	: Linear Discriminant Analysis
MAE	: Mean Absolute Error
MET	: Metabolic Equivalent Of Task
MEQ	: Morningness -Eveningness Questionnaire
MSE	: Mean Square Error
PANAS	: Positive and Negative Affect Schedule
PHQ	: Patient Health Questionnaire
PSS	: Perceived Stress Scale
PSQI	: Pittsburgh Sleep Quality Index
PWB	: Psychological Well-Being
QIDS	: Quick Inventory of Depressive Symptomatology
ReLU	: Rectified linear activation function
RMSE	: Root Mean Square Error
RSES	: Rosenberg Self-Esteem Scale
SELSA	: Social and Emotional Loneliness Scale for Adults
SHS	: Subjective Happiness Scale
SPANE	: The Scale of Positive and Negative Experience
SRQ	: Self-Regulation Questionnaire
SRQ-E	: Self-Regulation Questionnaire – Exercise
SRQE_RAI	: Self-Regulation Exercise Relative Autonomy Index
STAI	: The State-Trait Anxiety Inventory
SVM	: Support Vector Machine
SWB	: Subjective Well-Being
SWLS	: The Satisfaction of Life Scale

THE LIST OF FIGURES

Figure 2. 1: Machine learning techniques.....	10
Figure 2. 2: Supervised Learning techniques and examples.....	11
Figure 2. 3: The structure of artificial neural network.....	12
Figure 2. 4: The calculation process of neural network nodes in the forward propagation	13
Figure 2. 5: The learning process of machine learning algorithm by iteration.....	14
Figure 2. 6: The graph of sigmoid function	16
Figure 2. 7: The graph of ReLU function	17
Figure 2. 8: The visual expression of dropout	17
Figure 5. 1: The machine learning model for predicting subjective well-being.....	39
Figure 5. 2: The summary of data process.....	43
Figure 5. 3: The distribution of subjective well-being scores	47
Figure 5. 4: The visual description of the designed neural network model with hyperparameters	51

THE LIST OF TABLES

Table 3. 1: The functionality of popular smartphone apps	22
Table 3. 2: The recent studies on NetHealth project.....	35
Table 5. 1: The Subjective Happiness Scale Questionnaire	45
Table 5. 2: The equivalent questions used for Subjective Happiness Scale	46
Table 5. 3: Equally interval SWB Model with 5-fold cross validation	52
Table 5. 4: Equally interval SWB Model with 10-fold cross validation	52
Table 5. 5: Slightly Balanced SWB Model with 5-folds cross validation	53
Table 5. 6: Slightly Balanced SWB Model with 10-folds cross validation	53
Table 5. 7: The Summary of Machine Learning Performances.....	54
Table 5. 8: The Results of Regression Models	55

ABSTRACT

Recent advances in pervasive computing enable the collection of personal health-related data using diverse sensors in the everyday-life environment. However, human behavior modeling and analysis, particularly the quantification of subjective well-being, is still challenging, as there are variations in its definition and measurement. The psychology literature defines different perspectives on subjective well-being, such as hedonic and eudaimonic. In this work, we propose a model for predicting an individual's subjective well-being from the psychological perspective using her/his daily activities collected via smart wristbands, social relationships monitored through smartphones, and personality traits data from surveys. The model is applied to the NetHealth study, a heterogeneous data set of 577 student participants from the University of Notre Dame. We developed a multi-class classifier based on commonly accepted machine learning algorithms. The results enable us to predict an individual's well-being with almost 80% accuracy. We show the feasibility of a pervasive application as a personalized well-being assistant.

ÖZET

Günlük hayatta yaygın olarak kullanılan teknoloji ve cihazların gelişmesi, birçok sensör yardımıyla kişilerin sağlıklarıyla ilgili verilerin toplanmasını sağlamaktadır. Bununla birlikte, insan davranışı modellemesi ve analizi, özellikle de öznel iyi oluşun nicelleştirilmesi, tanımı ve ölçümünde farklılıklar olduğu için zorluğunu korumaktadır. Psikoloji literatürü, öznel iyi oluş için hedonik ve eudaimonik gibi farklı bakış açıları altında tanımlamalar içerir. Bu tez çalışmasında, akıllı bileklikler aracılığıyla toplanan günlük aktiviteler, akıllı telefonlar aracılığıyla izlenen sosyal ilişkiler ve anketlerden elde edilen kişilik özellikleri verileri kullanılarak bireyin öznel iyi oluşunu psikoloji perspektifinden tahmin etmeye yönelik bir model yapılmıştır. Model, Notre Dame Üniversitesi'nden 577 öğrencinin katılımıyla oluşan heterojen bir veri seti olan NetHealth çalışmasına uygulandı. Yaygın olarak kabul gören makine öğrenimi algoritmaları kullanılarak bir sınıflandırma tahmin modeli geliştirilmiştir. Sonuçlar, bir bireyin öznel iyilik halini yaklaşık %80 doğrulukla tahmin etmemizi sağlamaktadır. Günlük hayatın her anında kullandığımız teknolojilerle kişiselleştirilmiş bir iyi hal asistanı uygulamasının bir anlamda fizibilitesini göstermektedir.

1. INTRODUCTION

Well-being is a longstanding term that has been discussed for many centuries, from ancient times to the present time. Happiness, life satisfaction, mental health, and other terms have been used to describe it through time (Diener, Oishi, & Tay, 2018; Luhmann & Intelisano, 2018). Because it has existed for a long time, but it is also a living term whose definition has progressed, transformed many times, and diversified into various concepts due to the complexity of well-being (Luhmann & Intelisano, 2018). The extended definitions may appear more confusing and complex. On the contrary, the diversity of concepts for well-being creates new visions and approaches that help researchers to investigate thoroughly. Subjective well-being (SWB) has an exponential growth in the amount of research among the concepts of well-being. The main reason for interest may be that subjective well-being is one of the best representative forms of well-being in a legitimate way due to its subjectivity and principles (Diener, Lucas, & Oishi, 2018a).

The nature of well-being also brings its complexity to the assessment process. The diverse concepts of well-being give opportunities to measure the well-being of individuals. With the advancement of well-being research in psychology, validated questionnaires have made the assessment of subjective well-being more applicable, as some examples show (Diener, 1994; Diener et al., 2009, 2010; Luhmann & Intelisano, 2018; Pavot, 2018):

- The Scale of Positive and Negative Experience (SPANE)
- Positive and Negative Affect Schedule (PANAS)
- Psychological Well-Being (PWB)
- The Satisfaction of Life Scale (SWLS)
- Subjective Happiness Scale (SHS)

However, technological developments create new opportunities for data collection in an expanded and continuous way. The most useful source of data collection is smartphones because almost the whole population has interactions with them on a regular basis in their daily lives. Smartphones' physical sensors detect movement, and the global positioning system (GPS) records location data. The smartphone continuously records social interactions like SMS, calls, and social apps like Instagram and Facebook. The patterns and frequency of smartphone application usage also provide individual data all the time. The new trending device, the smartwatch is also getting involved rapidly in our lives day by day. The physical sensors in smartwatches are able to measure heart rate, skin response, temperature and additionally detect physical activities such as running, walking, sleeping, etc. to measure the physical activity level, calories burned, the number of steps, the quality of sleep and can be extended for further measurements (Woodward et al., 2020). In this manner, continuous data collection from individuals may support to assess well-being and, furthermore, provide crucial and detailed features to monitor subjective well-being on a broader scale than during the limited questionnaire time.

On the other hand, machine learning algorithms have improved capabilities to model new prediction systems in many aspects due to the growth in data collection. The level of development in machine learning technology is beyond the objective industrial sectors and is available to model subjective, complex human-based phenomena in psychology (Glenn & Monteith, 2014; S. Liu et al., 2019; Woodward et al., 2020). Existing research provides machine learning approaches for stress, depression, and anxiety detection in medical psychological problems and emotion, mood detection etc. for non-medical features (Glenn & Monteith, 2014). The detection models utilize mainly the daily data sources from smartphones, wearable devices, and social media platforms, or alternatively tangible devices (Woodward et al., 2020).

This study proposes machine learning approaches for predicting individuals' subjective well-being from a psychological perspective. The training process of machine learning algorithms is based on the rich dataset of the NetHealth project. The input of the prediction model can be categorized into three different sources from the NetHealth dataset. In the first category, social interaction data is continuously gathered from

smartphones. In the second category, activity and physical data are also continuously taken from the smartwatch at the same time. In the third category, psychological surveys are used to provide a basis for the characteristic features of each individual, and the psychological questionnaires set the ground truth for subjective well-being measurement.

Existing research mostly aims at the detection of poor mental well-being, which often leads to physiological problems like depression and anxiety (Glenn & Monteith, 2014; S. Liu et al., 2019; Woodward et al., 2020). Within the scope of SWB, two-component metrics are conceived: positive effects include happiness and enjoyment, while negative effects include stress, depression, and sadness (Jebb, Morrison, Tay, & Diener, 2020). Thus, not having depression or anxiety does not mean having an unnecessarily high SWB. The Study of Ed Diener (Jebb, Morrison, Tay, & Diener, 2020) discusses the findings from globally representative cross-sections of 166 nations with more than 1.7 million individuals. It concludes that subjective well-being, which is defined from 20 to 80 years of age, has a change variation pattern with only minimal differences in negative affect across the life span.

In contrast, decreases in positive affect were more significant (Kushlev, Drummond, & Diener, 2020). The positive effects such as happiness and enjoyment would be considered more determinant factors than negative effects such as depression and stress regarding SWB. In addition, the analyses have stated that SWB is linked to increased survival in both healthy and physically ill samples (excluding mental illness such as depression) and short-term immune functioning (Chida & Steptoe, 2008). One meta-analysis of well-being studies (Howell, Kern, & Lyubomirsky, 2007) concluded that the effect of well-being is proven by empirical metrics in both short- and long-term health across. More especially, these researchers examine the effects of well-being on ill-being (e.g., mental illness) and healthy participants separately and prove that well-being proceeds positively on health regardless of ill-being's negative effects (Alvarez-Lozano et al., 2014).

In this study, we reviewed cutting-edge pervasive well-being technologies and evaluated the tools and techniques in terms of effectiveness, capabilities, benefits, and drawbacks. We explore the various concepts of well-being, the advancement of subjective well-being

in psychology literature, as well as the measurement of subjective well-being through the validation of subjective well-being scales. Based on existing research, we determine the forms of daily pervasive data that may serve to assess subjective well-being. Distinctively, a wealthy, multi-dimensional, and long-term dataset of NetHealth is analyzed and processed for the machine learning models. The main contribution of this research is that we present a machine learning model for predicting a personally tailored subjective well-being from a psychological standpoint using data from her/his daily activities collected via smart wristbands, social relationships monitored by smartphones, and survey data on personality traits.

In order to create an applicable and meaningful base for both engineering and psychology literature, this proposed system has significant design principles and advantages, which are as follows:

- The ground truth of the model is based on the validated subjective well-being questionnaire. The machine learning model is trained and optimized to predict the level of subjective well-being, which is determined by the real individual cases.
- In the design of the machine learning model, the feature selection process is completely designed with respect to the existing psychology literature. Each feature of the machine learning model is chosen from the large dataset only if it has a relationship with subjective well-being and is demonstrated by statements from various psychological studies. To appeal to the psychological perspective, not only the model's output but also its input or explanatory data are chosen.
- The data requirements for the model are designed to be practical and user-friendly. The machine learning model works with the daily individual data which is collected passively by smartphone and smartwatch, except only a few psychological surveys are applied at the beginning of the model. Hence, the prediction model is available for use at any time and updated automatically by the continuously passive data.
- The machine learning model makes predictions by using the personality surveys and private daily data of individuals. This way of working enables the model to be personalized. Additionally, it is promising for further personalization.

The rest of the thesis is organized as follows. In Chapter 2, the well-being concepts are presented to create a clear view. The subjective well-being concept is defined in detail by psychological terminology. The modern validated questionnaires are presented and examined to determine the substitutional options. Then the commonly used classification machine learning methods, which are applied in the methodology section, are explained in the design structures. In Chapter 3, the state-of-art technologies in well-being and the related psychological phenomena are presented. The existing related studies are examined, and the performance and capabilities of pervasive machine learning methods are evaluated by comparing them. In Chapter 4, the rich NetHealth data set is examined in respect of its features, the data gathering principles, and the number of participants. The studies that have used the NetHealth dataset are evaluated. In Chapter 5, the subjective well-being model is presented in detail. The data preprocessing, the feature selection, and the input-output data process based on the psychology research are presented. Then the applied machine learning methods' performances are shown, compared in the results. In Chapter 6, the performance of machine learning models is summarized, the findings are discussed, and the thesis is concluded.

2. BACKGROUND

2.1. What Is Well-Being?

Well-being is a fact that concerns every person in society; it is desired to be improved and pursued at a reasonable level in every part of the life span. That is because people have higher expectations for a longer lifetime and a higher quality of life along with it. As a result, interest in measuring well-being has exploded in recent years, guided by demand for societal transformation with a contemporary vision for health that prioritizes well-being at the center (VanderWeele et al., 2020).

In ancient history, well-being or "happiness" was a longstanding term in psychology, and its complexity creates many definitions. In the scope of the science of happiness and well-being, Luhmann's research recounts the paradigm shifts, provides a review of past and current well-being theories, longitudinal empirical studies, and essential classifications for theoretical approaches (Luhmann & Intelisano, 2018). The review article by Ed Diener brings the modern key concepts of well-being with definitions together, such as subjective, psychological, hedonic, emotional, eudaimonic well-being, quality of life, etc. (Diener, Lucas, & Oishi, 2018b).

2.2. The Concepts of Well-Being

Three main approaches to well-being are primarily known in the psychology literature. The first one is the evaluative approach, which defines well-being as the satisfaction status of a person's life by self-evaluation and determination (VanderWeele et al., 2020). The second one is the eudaimonic perspective, which focuses on self-realization and fulfilling a sense of purpose (VanderWeele et al., 2020). The third approach, defined as the hedonic perspective that determines well-being in respect of the degree of happiness

or the frequency of experiences of pleasure and pain (VanderWeele et al., 2020). Ed Diener reported that “Subjective well-being reflects an overall evaluation of the quality of a person's life from their perspective” (Diener, Lucas, et al., 2018b). Based on the publication of *Well-Being: The Foundation of Hedonic Psychology* (“Well-Being: The Foundations of Hedonic Psychology,” 1999), subjective well-being (SWB) has been associated with the hedonistic approach to well-being.

2.3. Subjective Well-Being

Among the concepts, subjective well-being shows considerable growth in the field based on the exponential increase of articles and writings. Subjective well-being (SWB) refers to the extent to which a person believes or feels that their life is going well and reflects an overall evaluation of the quality of a person's life from their perspective (Diener, Lucas, et al., 2018b). Starting from this point of view, every individual reacts to life circumstances in respect of their evaluation, particularly depending on personality, social values, and even physical and mental health conditions. Thus, the subjectivity of this concept is more realistic compared to objective approaches and non-personalized concepts. Despite this acknowledgment that SWB is not synonymous with well-being, SWB can be considered as one of the best conventional concepts of well-being and, furthermore, may be used as an easily measurable fact for overall well-being.

Subjective well-being (SWB) is a highly active area of research. There has been a rapid increase in the number of articles and books with over 150,000 publications in the last decade (Diener, Lucas, et al., 2018b). As this review clarifies, the research on SWB remains of interest in both academic and applied sciences. Nevertheless, there are still a large number of questions in need of research (Diener, Lucas, et al., 2018b). This focus is on mental illnesses such as depression and anxiety that people are trying to avoid. Still, it would not necessarily mean being mentally excellent or happy. Subjective well-being is a broad and general concept that includes not only the absence of mental illness but also a higher level of life satisfaction and a better evaluation of an individual's own life (Wood & Joseph, 2010). The advancement of research on subjective well-being guides scientists to understand the key components and the concepts of well-being. In light of

these advancements, more knowledge about well-being allows applied studies to design many interfaces and models to improve the quality of people in society (Diener & Ryan, 2009).

2.4. The Assessment of Subjective Well-Being in Psychology

The advancement of SWB has reformed from the various definition concepts to mainly validated measurement approaches over 10 years (Diener, Lucas, et al., 2018b; Pavot, 2018). In Ed Diener's (Diener, Lucas, et al., 2018b) research, it is demonstrated that new self-report measures give high consistency and test reliability, mainly over short time intervals. Additionally, Luhmann's study (Luhmann & Intelisano, 2018) reviews empirical evidence from long-term studies that SWB concepts are not a stable fact after different circumstances and the high level of effectiveness of SWB interventions.

In Diener's and Ryan's research (Diener & Ryan, 2009), one of the theories for SWB is concluded as well-being may be measured by summing up the positive and negative affects in respect of a person's evaluation. In this view, positive or "happy" events will increase an individual's well-being. In contrast, the negative or "sad" events will decrease the well-being (Diener, 1994). Consequently, the term "happiness" is not only an inevitable factor but also a significant indicator of measurement.

Based on the thoughtful engagement of the UK's Measuring National Well-Being Programme, the subjective happiness scale (SHS) which consists of four particular questions, is used for ongoing well-being surveys, and successful results are obtained for mental well-being in large-scale populations. (VanderWeele et al., 2020). In our study, similar questions would be used to assess the individuals' subjective well-being, and the survey results would be labeled with the continuous passive data from hybrid interfaces. It is considered a hedonistic well-being measurement. In Luhmann's research, it is suggested that most psychological theories relate to aspects of happiness and well-being that can be counted as components of subjective well-being (SWB) (Luhmann & Intelisano, 2018). In Ed Diener's research, happiness is considered as one of the significant feelings for SWB measurement (Diener, Lucas, et al., 2018b). In addition, he

has built the main basis of SWB on the answer to the question of whether happiness can be measured since he considers SWB as an empirical science. (Larsen & Eid, 2008). Thus, the subjective happiness scale (SHS) is used for SWB measurement in Diener's research for new-wellbeing measurement. Along the same line, Lyubomirsky and Lepper used the same four questions (SHS) to be able to measure SWB precisely as we did in our research (Diener et al., 2010; Lyubomirsky & Lepper, 1999).

2.5. The Validity of Assessment Scales for Ground Truth

The short scales like SHS with 4-items or the scales of the UK's Measuring National Well-Being Programme would be helpful and applicable for larger populations. The practical advantage is reported in the research (Lyubomirsky & Lepper, 1999). The SHS is simple to use, and its brief form provides accurate assessment with low cost or respondent effort. However, only if the validity of scales exists, it is reasonable to apply the practical and short scales instead. The validity of the happiness scales as a measure of well-being and global mental health has been examined, and the convergence with other happiness instruments allow us to discriminate the different levels of happiness and prove that the happiness scale results have a correlation with commonly used mental health assessments (Fordyce, 1988). This research showed that happiness measures are acceptable for well-being. However, meaningful coordination with psychology is essential for the validity of any measurement. The research of (Howell, Rodzon, Kurai, & Sanchez, 2010) reports that "The results, taken as a whole, support the reliability, validity, and generalizability of the Internet format of the Satisfaction with Life Scale (SWLS), Positive and Negative Affect Schedule (PANAS), and Subjective Happiness Scale (SHS)". Another research is focused on the validity of SHS reports as the Subjective Happiness Scale shows excellent psychometric properties (Lyubomirsky & Lepper, 1999). The SHS has a high correlation with other happiness and well-being measurements, and the construct is theoretically and empirically supported. (Lyubomirsky & Lepper, 1999). As a result, the SHS scale has validity for measuring SWB, and it is used as the primary guidelines to generate an SWB output in this study.

2.6. Machine Learning Methods

2.6.1. Machine Learning Description and Techniques

Machine learning is a form of data analytics science that gives computers the ability to learn. The distinctive technique of machine learning from other algorithms is that it uses computational methods to learn without a specifically determined model. Machine learning algorithms have common structures that can improve their performance directly from the data. Hence, they can be applied to many problems without explicitly programming them. Machine learning algorithms have two fundamentally different types of techniques: unsupervised and supervised learning. The primary distinction between these techniques is the input data; whether it has a labeled response is unknown.

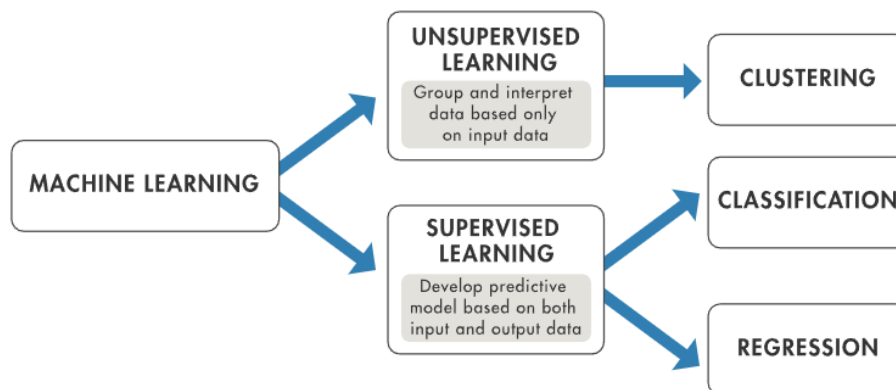


Figure 2. 1 Machine learning techniques¹

Unsupervised Learning: This machine learning technique uses the input data without labeled responses. It is trained to explore any hidden or intrinsic patterns in the data. The most common utilization of this technique is clustering, which is the grouping of data based on the hidden patterns.

¹ <https://ch.mathworks.com/discovery/machine-learning.html#how-it-works>

Supervised Learning: This technique builds a prediction model that has a mapping from input data to the response variable. It trains the weights in the model to generate reasonable response predictions for the new input data. Supervised learning has two different categories: classification and regression.

2.6.2. Supervised Learning Techniques

Classification model: It is a supervised learning technique to classify the examples in the data. The input data is labelled with the response variable in discrete numbers that define different classes or categories regardless of the numerical value. Mostly, the classification models calculate the probability of belonging to each class based on the input and improve their performance based on the number of examples available.

Regression model: It is another supervised learning technique if the data has a continuous response variable regarding its numerical value. The model trains its weights to fit the data with the best hyper-plane and gives predictions with the least amount of error as possible.

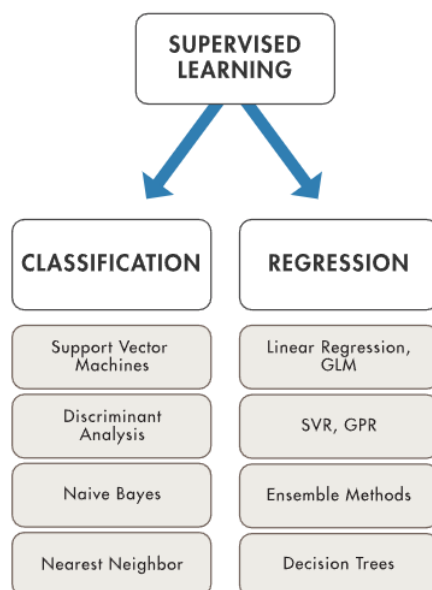


Figure 2. 2. Supervised Learning techniques and examples²

² <https://ch.mathworks.com/discovery/machine-learning.html#how-it-works>

Artificial Neural Network: It is a subset of machine learning algorithms. A neural network has a structure of consecutive layers that consist of interconnected nodes. Each node has an associated weight and activation function. The neural network updates the weights in the nodes to improve the accuracy of the model based on the input data with labeled responses.

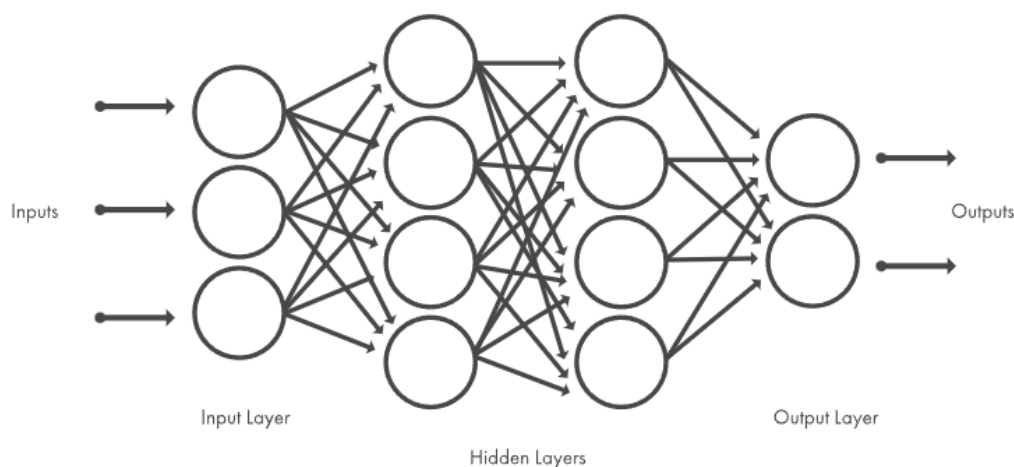


Figure 2. 3. The structure of artificial neural network³

2.6.3. The Learning Process of Supervised Machine Learning Algorithms

After data processing and feature selection, many widely used machine learning algorithms for classification are applied to the processed data to build prediction models. The fundamental working principle is mostly common in supervised machine learning models including artificial neural networks. As explained in the following, there are simply four steps for the learning process of algorithms:

The first step is the forward propagation, where the features of input data are used to start the series of consecutive calculations with randomized weights by various activation

³ <https://ch.mathworks.com/discovery/neural-network.html>

functions in respect of the structure of the machine learning algorithm. The results of the calculation series are the outputs in the output layer and they end the forward propagation process.

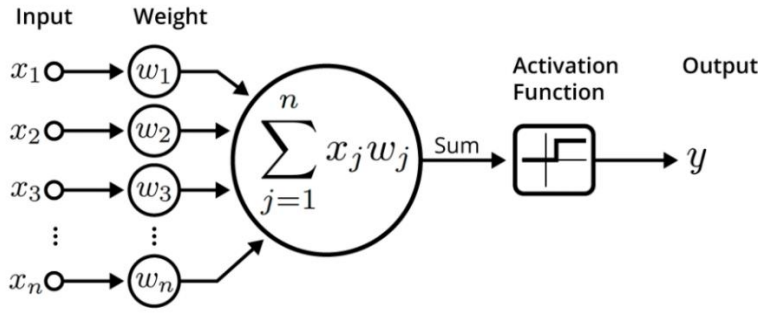


Figure 2. 4. The calculation process of neural network nodes in the forward propagation

$$\text{Activation output of node} = f\left(\sum_{j=1}^n x_j w_j + b_j\right) \quad (2.1)$$

The second step is the evaluation of the predictions. Each prediction which is calculated with the input (x) of labelled data (x,y) is compared to the ground truth of (y). The wrong predictions are considered errors of the machine learning algorithms, and the average error is measured by the defined multi-label classification loss function in the algorithm.

$$J = -\frac{1}{N} \sum_{n=1}^N \sum_{i=1}^k [T_{ni} \log(y_{ni}) + (1 - T_{ni}) \log(1 - y_{ni})] \quad (2.2)$$

(The multi-label classification cost function where N and k are the number of observations and classes, respectively)

The third step can be named as backpropagation through time. The backpropagation process works the gradient of the multi-label classification loss function with respect to each weight. The share of the error (misclassification) for each weight in the algorithm is measured by chain rule in the reverse order of forward propagation calculations.

$$\text{Gradient for weights} = \frac{\partial J}{\partial w_n} \quad (2.3)$$

The final step is the learning part of the model. The weights in the algorithm are updated by the shared error, which is measured in the third step. This learning process is repeated many times to minimize the loss function of the model. The iterative process ends when the cost function converges in a local minimum. Thus, the machine learning model is trained with the examples of data and would perform better in prediction.

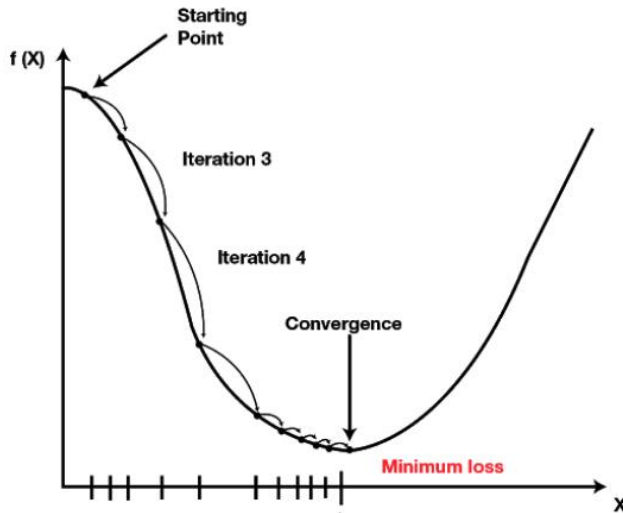


Figure 2. 5. The learning process of machine learning algorithm by iteration

The optimization algorithms are used to minimize the loss function. Two of optimization algorithm is explained in the following:

Gradient Descent Optimizer: It is a commonly used optimization algorithm for the learning process of machine learning models. It is an iterative process wherein in each iteration, the weights in the model are updated by the multiplication of the learning rate () and the first order derivative of the cost function in respect of each weight.

$$\text{Update } w_n = w_n - \alpha \cdot \frac{\partial J}{\partial w_n} \quad (2.4)$$

Adaptive Moment Estimation (Adam) Optimizer: It is the extended form of gradient descent optimizer in an efficient way. In this optimization process, the algorithm computes the adaptive learning rates for weights in respect of the first- and second-degree gradients. By the change in individual learning rates, the algorithm can minimize the cost function rapidly and avoid being stuck in a local minimum which may not be best optimized point.

The terms used in machine learning models are explained as follows:

- **Sigmoid activation function:** A non-linear function with a ‘S’ curve that evolved from logistic function. The range of sigmoid function is between 0 and 1. In the neural network models, the sigmoid function is used as non- linear activation function in the node which gives outputs from the weighted sum of inputs.

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (2.5)$$

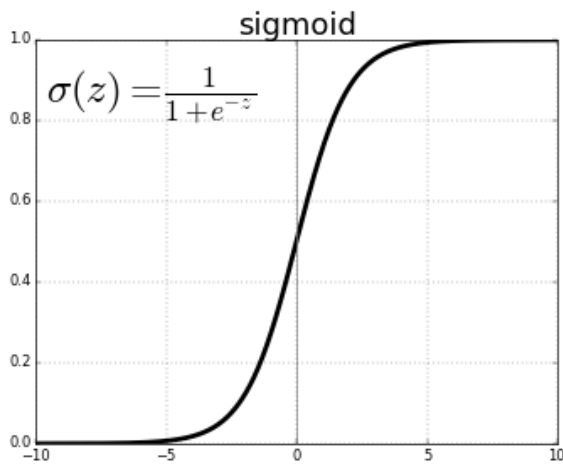


Figure 2. 6. The graph of sigmoid function

- Rectified linear activation function (ReLU): A piecewise function that outputs the directly proportional values when the input is positive, otherwise outputs zero. It is used mainly in the hidden layers of neural network models. Because it helps models to overcome the vanishing gradient problem, makes optimization of models faster and better.

$$R(z) = \max(0, z) \quad (2.6)$$

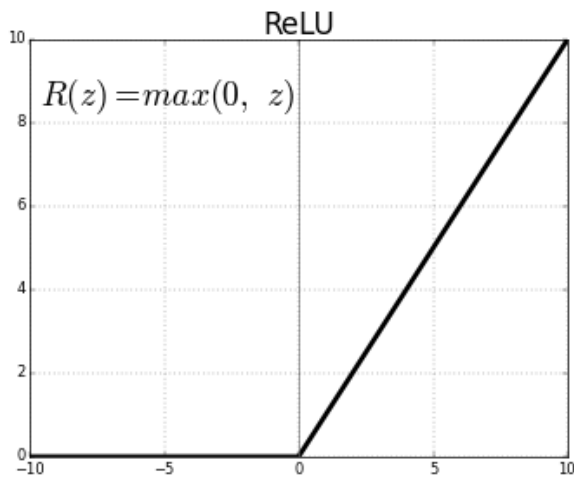


Figure 2. 7. The graph of ReLU function

- Dropout: It is a computationally cheap and effective way of regularization. In the training process, some of the layer outputs are ignored in the iterations randomly with specified probability.

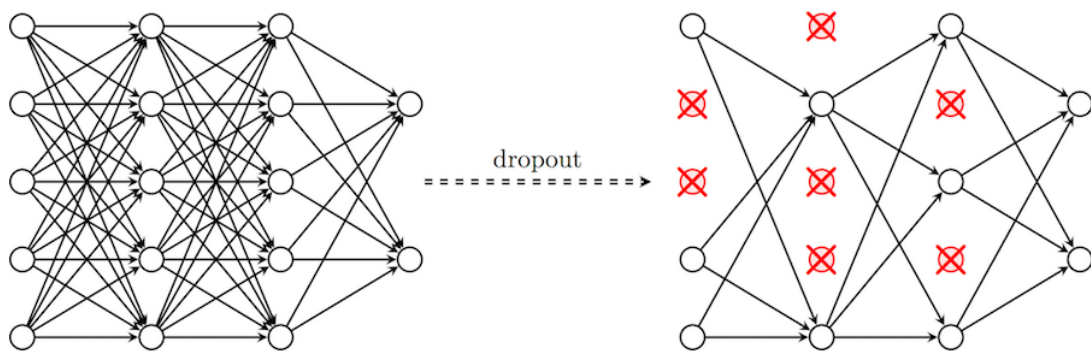


Figure 2. 8. The visual expression of dropout

- Root Mean Square Error (RMSE): It is a widely used measure as a single number for evaluating the model's performances, especially supervised machine learning applications. Numerically, RMSE is the standard deviation of the residuals and

measures the error of a model in prediction. So, it indicates how concentrated the data is around the prediction line of model.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2.7)$$

- Mean Absolute Error (MAE): It is another commonly used error parametric for evaluating the regression models. MAE is the average the magnitude of the prediction errors which calculates the metric linearly.

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2.8)$$

- Mean Square Error (MSE): It is the error parametric for regression models and calculated as the mean of the squared differences between predictions and ground truth values. MSE is the transformed form of RMSE which is the square root of MSE.

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2.9)$$

Among the machine learning algorithms, Ensemble Boosted Trees, Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA) are selected to present in respect of their better performance. Additionally, Artificial Neural Network model is designed for classification. The hyperparameter optimization is done to have better performance and evaluated in Section 5.

The descriptions of selected classifiers are as follows:

- **Ensemble Boosted Trees:** The model is a supervised learning algorithm that constructs a sequence of decision trees for classification or regression. The sequence of trees are the forms of training sets with different weights. Instead of weighted classification errors in the bagging tree, the algorithm uses pseudo-loss as a measure of classification accuracy for the similar idea of increasing the weight of misclassified examples. By this step, the algorithm focuses on the training examples, which are more difficult to predict the correct class. The final model is obtained by minimizing the specific loss function with a sequential of trees which have class prediction.
- **Support Vector Machine:** SVM is a supervised machine learning algorithm that uses labelled data for classification or regression. The classification of SVM sort the data into two categories by linear, quadratic, or other non-linear functions. SVM optimizes the hyperplanes to separate the groups by the gaps as much as possible in respect of the pattern of the training set.
- **K-Nearest Neighbors:** KNN is one of the fundamental and basic machine learning algorithms for both supervised and unsupervised learning methods. The model calculates the similarity between the examples in respect of classes. Then, in the prediction process, the model calculates the similarity of new data to each class and predicts the class that has the most similar examples.
- **Linear Discriminant Analysis:** LDA is a supervised classification method that reduces the dimension of data by forming the features as a linear combination.

The model uses reduced dimensionality to maximize the distance between classes, minimize the variation of each class, and generate a prediction model of two or more classes in respect of the training set.



3. RELATED WORKS

The advancement of well-being technologies is summarized in this section of the research, which intends to explore the use of innovative interfaces to analyze existing and emerging technologies as well as how they are used to measure and improve well-being. It continues with a discussion of the performances, limitations, and achievements that current technological developments provide for well-being.

3.1. Technologies of Well-Being

In terms of assessing well-being, traditional methods work in the way of self-reporting, people recall their emotions and stresses in daily life which are tracked in order to assist identify stressful factors (S. Liu et al., 2018, 2019) and use questionnaires such as the Positive and Negative Affect Schedule (PANAS) (Faust, Jiménez, et al., 2018), Quick Inventory of Depressive Symptomatology (QIDS) (Hone, Jarden, Schofield, & Duncan, 2014) and the Patient Health Questionnaire (PHQ-9) (Gray-Little, Williams, & Hancock, 1997), Satisfaction with Life Scale (SWLS) (Howell et al., 2010). There are validated ways to assess mental well-being in psychology and applied by clinical visits (Woodward et al., 2020). However, the traditional way of assessing well-being through questionnaires is only available to apply in clinic appointments. Due to the infrequent evaluation of mental health, doctors have difficulties detecting the symptoms of diseases in a limited amount of time. Therefore, clinicians are challenged to decipher the possible manifestation of symptoms and disease trajectory (Woodward et al., 2020). Technology advancements allow for the collection of a multitude of data that may be used to closely monitor individuals' mental health and suggest technologies to increase their mental well-being. This would be helpful since it would increase access to mental health resources (Kushlev et al., 2020). In this section, the state-of-the-art modern technological advancements for well-being are categorized, discussed, and evaluated.

3.2. Smartphones

From data source varieties, the smartphone has a wide range of positions that are classified as “active” and “passive” data regarding users’ interactions. Active data is obtained from self-reports, app records, and changes in mood feedback from the users. On the other hand, passive data is any available data gathered from the device's records and sensors.

3.2.1. Active Data

The interaction with smartphones happens by the mHealth apps. Many apps have been developed to modernize and advance existing practices of recording mental well-being (Woodward et al., 2020). The top popular mental health and well-being apps (Woodward et al., 2020), with a summary of their functions and capabilities explained in the following.

Table 3. 1. The functionality of popular smartphone apps

Apps List	Functionality
Well mind	Designed by National Health Services, monitor feelings and give advices for well-being
Dayio	Reasonable way to motinor moods by using self-reports and able to examine the factors of variation in moods
MoodPath	Records feelings, moods and provides mental health assesments after two weeks. The long-term usage may serve rich data to apply statistical analysis
Calm	Guide users to have better relaxation, calm and decrease the stress in a various applicable advices
Headspace	The major source is meditation to deal with anxiety and depression. Provide user-friendly meditation techniques to relax, descriptive statistics to track moods easily

Chatbot programs are designed to make individuals feel as if they were talking with a human being (A. & John, 2015), so the requirement to self-report is provided on a

continual basis. Additionally, chatbots are able to mimic social human reactions through the conservation of advanced natural language algorithms. Thus, they can continuously gather emotional and personal information about the mental state of patients (Lee, Oh, & Choi, 2017). Wysa (Inkster, Sarda, & Subramanian, 2018) and Woebot are the existing examples of chatbots that participants evaluate as helpful and encouraging, resulting in higher well-being conditions (Fitzpatrick, Darcy, & Vierhile, 2017).

3.2.2. Passive Data

The smartphone usage data is automatically taken from the device, so it does not require feedback or active recording. Therefore, it has a great variety of raw data types and gives opportunities for different detection purposes. Examples of the detection of mental well-being include the following: Smartphone passive data is used for monitoring of emotion and stress by measuring speech analysis (Glenn & Monteith, 2014), mood state recognition (depression) by activity and mobility (Grünerbl et al., 2014), the mood in bipolar disorder by usage patterns and app selection (Alvarez-Lozano et al., 2014); mood and well-being by frequency and duration of calls and SMS, email, web browsing, location (Likamwa, Liu, Lane, & Zhong, 2013), stress by sleep patterns and social contexts (Mohr, Zhang, & Schueller, 2017).

3.3. Wearable Devices

Due to their long-term usability and user-friendly interface, smartwatches and wrist-worn sensors are widely accepted in social life. They generally have non-invasive physiological sensors like GPS, accelerometer, heart rate sensor, and additional sensors in particular types. One of the main measures of stress is Galvanic Skin Response (GSR) which directly correlates with the sympathetic nervous system (Gravenhorst, Muaremi, Tröster, Arnrich, & Grünerbl, 2013; Setz et al., 2010; Sharma & Gedeon, 2012). The other main measurement is the Heart Rate Variability (HRV) which uses the changes in heartbeats to detect stress level (Acharya, Joseph, Kannathal, Lim, & Suri, 2007). By way of an accelerometer sensor, the intensity of physical activity is used to detect depression, which is assessed with the Patient Health Questionnaire-9 (PHQ-9) (Vallance et al., 2011).

3.4. Tangible Devices

Tangible devices are the interfaces that get digital information when users interact physically. They provide the new technological ways to traditional assessments, which give availability for individuals to track their well-being simultaneously (Woodward et al., 2020). For example, Emoball (Bravo, Hervás, & Villarreal, 2015) is a tangible device that enables individuals to squeeze the ball when they have feelings, and the tangible device (ball) records their mood through the physical sensors. Mood sprite (Barker & van der Linden, 2016) is an example of a tangible device designed for patients to deal with anxiety and stress and provide another way to feel relaxed. For stress detection, another example of a tangible device is Grasp (Guribye & Gjørseter, 2018). Users could squeeze whenever they felt stressed, and the tangible device would also measure the pressure applied.

3.5. Internet and Social Media Platforms

Recent studies have demonstrated increasing evidence that natural language processing can assess well-being through users' internet and social media posts. In this group, the keywords and phrases in Twitter posts of users are processed for depression detection (de Choudhury, Gamon, Counts, & Horvitz, 2013; Glenn & Monteith, 2014; Nadeem, 2016; Wongkoblap, Vadillo, & Curcin, 2017). Facebook posts (Glenn & Monteith, 2014; Grünerbl et al., 2014; Wongkoblap et al., 2017), for subjecting well-being (Bollen, Gonçalves, Ruan, & Mao, 2011; Hao, Li, Gao, Li, & Zhu, 2014; P. Liu, Tov, Kosinski, Stillwell, & Qiu, 2015). Another popular alternative is using photos and videos from social media platforms to infer mental well-being (Woodward et al., 2020). The shared photos on Instagram may serve well-being machine learning models to distinguish facial expressions relating to mental problems such as anxiety, depression, and more serious disorders like bipolar (Manikonda & Choudhury, 2017).

3.6. Evaluation of Technology in Well-Being

The massive use of well-being apps shows their exponential growth in usage. The reason for rising interest is their constant accessibility in individuals' lives at any time. On the other hand, tangible interfaces allow individuals to react physically to different stress levels and also record their emotional variations (Woodward et al., 2020). However, one of the main problems with well-being apps is that they have not been tested in scientifically rigorous research studies. There is no scientific evidence supporting its use, despite its having millions of users (Woodward et al., 2020). Their functionality is more focused on the practical approach of helping users improve their well-being when necessary rather than assessing well-being.

On the other hand, smartphones can passively collect a vast amount of data, such as location, motion, and phone usage. It provides continuous and longitudinal data in a timeline, which is extremely valuable and vital for machine learning algorithms to train. When machine learning models are trained by the smartphone's active and passive data, they may gain the ability to predict emotions with 70% accuracy (Woodward et al., 2020). In other important research (Bogomolov, Lepri, Ferron, Pianesi, & Pentland, 2014) they proposed StressSense, a machine learning model to predict the stress level by using human voice in social conversations. Their model has an accuracy of 81% for indoor and 76% for outdoor environments. Another research states that the model detects daily stress levels with a 73% accuracy using hybrid data model, such as big five personality, social interactions like the number of calls and messages, additionally the weather condition (Bogomolov et al., 2014).

The small wearable devices use physical sensors to measure GSR or HRV continuously. This ability could be evolved to measure anxiety problems and the status of well-being with high accuracy (Woodward et al., 2020). For example, the wearable device records at 5-minute intervals and the predictions for the stress level result in an accuracy of 97.4%. It means small wearable devices are reasonable interfaces for detecting mental states (Woodward et al., 2020). However, physiological signals are not directly linked to well-being or mental health. They are considered as significant features to detect perceived

stress levels, so the additional sensors can contribute the accuracy for assessing well-being (Woodward et al., 2020).

The research of (Matthews & Doherty, 2011) and (Niemantsverdriet & Versteeg, 2016) has stated that individuals prefer physical (tangible) devices because they think the physical contact allows them to express their emotional experiences in a better way. In contrast, the tangible devices lack the capability to provide continuous and rich data sets, which is a major requirement for machine learning models. The studies on tangible devices are mostly on a small scale, so they remain primarily unproven and the results are lacking in statistical analysis. In this manner, digital interfaces are available to provide the necessary requirements of data for machine learning models and have many advantages over tangible devices.

Among the social media platforms, Twitter has also been used for natural language processing. The depression detection model can achieve 81% accuracy by using a dataset of 2.5 million tweets in a one-year timeline (Woodward et al., 2020). Similarly, one of the social media platform “Facebook” posts used similar machine learning models. The prediction for depression can achieve 69% accuracy in respect of medical reports from over 650 patients (Woodward et al., 2020).

The hybrid models have multi-source data from smartphones and wearable devices on the same longitudinal timeline. These types of models have approximately 75% accuracy for stress detection by gathering continuous daily activity data, phone usage traits, and the results show hybrid models have more efficiency and higher accuracy than other single source models (Sano & Picard, 2013). Similarly, a wrist sensor along with a mobile app and self-reported PHQ-8 and PHQ-4 depression scores were used. The more physical sensors, the more accurate depression detection in respect of the study results (Kroenke, Spitzer, Williams, & Löwe, 2009).

3.7. Evaluation of Related Works

In addition to traditional assessments such as questionnaires for well-being, technological advancement gives new opportunities to assess and improve well-being. Technologies in well-being are utilized by smartphones, wearable devices, tangible devices, and social media platforms on the Internet. The examples are classified as medical and non-medical projects in tables regarding used technology, measured items, and aim (Glenn & Monteith, 2014). Non-medical studies are focused on detecting stress, measuring mood, emotional states to improve well-being (Glenn & Monteith, 2014) .

However, they commonly lack well-being measurement or validation. Worst of all, they do not have any scientific examination and evaluation by comprehensive research (Glenn & Monteith, 2014; Woodward et al., 2020). Medical studies make use of mental well-being to detect mainly depression and anxiety. They have prediction models with validated questionnaires for measurements such as the Patient Health Questionnaire (PHQ-4, PHQ-8), Quick Inventory of Depressive Symptomatology (QIDS) (S. Liu et al., 2019). Most of the existing studies on mental well-being have several limitations (S. Liu et al., 2019).

Firstly, most are conducted on a limited number of individuals with limited time. A limited number of individuals results in the lack of statistical power and personalization of studies. The limited short-term period cannot meet the requirement for data for machine learning approaches. While machine learning improves itself for well-being estimation, it essentially requires a sizeable, labeled dataset to initially train the models, which can be challenging to obtain (Woodward et al., 2020).

Secondly, focusing on a single type of data results in lower prediction accuracy for predicting well-being concepts due to complexity (Glenn & Monteith, 2014; Woodward et al., 2020). When pervasive continuous data is labelled with self-reports in the longitudinal timeline, the machine learning algorithms would be trained more well and make predictions more accurately but in this multi type data approach is depends on the high feedback ratios of the users' self-reports (Woodward et al., 2020). In order to go

beyond the limitations of existing studies, The NetHealth Project data from the University of Notre Dame is used as the source of a machine learning model to predict subjective well-being.



4. NETHEALTH PROJECT

The NetHealth Project is a multi-year longitudinal study of social networks and health and conducted by the center for Network Science and Data at the University of Notre Dame. The study collected smartphone data, wearable sensor (Fitbit) data, and regularly various surveys data (self-reposts) from approximately 700 undergraduate student participants from 2015 to 2019 (S. Liu et al., 2019). NetHealth participants were polled on their whole of education programs, activity level, mental well-being, sleep habits, as well as personal and family demographics (Faust, Hachen, Lizardo, & Chawla, 2018). Social network data was collected from a smartphone app that records students' gather from calls and SMS (Glenn & Monteith, 2014). Fitbit data would be summarized as physical activities (step counts, physical activity levels) and sleep habits (bedtime and sleep duration). The researchers apply periodic surveys that gather individuals' demographics, education, mental health, sleep quality and personality traits (S. Liu et al., 2018).

- **Communication Data:** Event (time-stamped) records of all communication events logged on a study participant's smartphone. This is the information used to construct social networks (arc and edge lists) of participant's communicative interactions with other people.
- **Basic Survey Data:** 8 waves of data collected through surveys completed online by participants every 3-4 months.
- **Fitbit Data (Daily Activity and Daily Sleep Data):** Daily data on a variety of measures computed from Fitbit sensors such as number of steps, active minutes, average daily heart rate, minutes in various heart rate zones, beginning and ending times of sleep periods, sleep duration, and sleep interruptions.
- **Calendar:** The data collection lasted 194 weeks. We provide a weekly calendar for researchers detailing when classes began, break weeks, holidays, etc.

4.1. Surveys Data

The applied surveys in the NetHealth project are listed as follows:

- Big Five Personality
- Self-Esteem (Rosenberg Scale)
- Trust
- Self-Regulation Questionnaire (SRQ)
- Self-Regulation Questionnaire – Exercise (SRQ-E)
- Social and Emotional Loneliness Scale for Adults (SELSA)
- Self-Efficacy
- Self-Regulation
- Sedentary Time
- The State-Trait Anxiety Inventory (STAI)
- Centre for Epidemiologic Studies Depression Scale (CES-D)
- Beck Anxiety Inventory (BAI)
- Pittsburgh Sleep Quality Index (PSQI)
- Morningness -Eveningness Questionnaire (MEQ)
- Beck's Depression Inventory (BDI)
- The Perceived Stress Scale (PSS)

Big Five Personality is a self-report survey with 44-questions to investigate the personality of individuals in quantifiable psychological traits. The “Big Five” is referred to the five broad personality traits which are extraversion, agreeableness, openness, conscientiousness and neuroticism. The explanations of these traits as follows.

- Openness is related to the curiosity for the new ideas, experience or change. People who have a high level of openness tend to be creative, imaginative, open to trying new things. The crucial aspect is that they enjoy the struggle of new experiences, challenges and satisfied with the steps for the new visions and

progress. In contrast, the people who have a low level of openness dislike changes, resist new ideas and experiences, prefers the ordinary ways of their own.

- Conscientiousness refers to being organized, having controlled and goal-oriented behaviors. High level of conscientiousness represents the personality which tend to plan ahead in detail in order to finish important tasks in scheduled time. Additionally, this trait also refers awareness of status ideally, having an organized mindset which is even prepared to contingency. Low level of conscientiousness is failing to complete necessary tasks, not taking care of things and much more letting the life flow.
- Extraversion is described as sociability, talkativeness and expressiveness of emotional feelings. People who have the high level of extraversion feel better and energetic in social events and meetings. They are easy-going to make new friends, connections, like making conversations and enjoy with their wide social network. People who have the low level of extraversion prefers solitude, enjoys spending time by themselves. They have difficulty in starting conversations and making new social network. Furthermore, attending to social events a lot may feel them exhausted and they require a lonely period to recharge.
- Agreeableness refers to the trust, empathy, altruism and affectionateness. People who are high in agreeableness more cooperative, emphatic, helpful and enjoys contributing to the happiness of other people. In contrast, people who are low in agreeableness don't care about the feelings of others, focused only themselves, tend to be more competitive instead of cooperative and may manipulate people toward their purposes.
- Neuroticism is related to the emotional instability, anxious and sadness. The people who have the high level of neuroticism feel stressful, experiences shift in mood rapidly and anxiety. They have difficulty to relax after emotional events. However, the low level of neuroticism means stable emotional state, relax and resilience to deal with stress.

The NetHealth study applies Big Five Personality surveys to the participants at the beginning of the first wave, updates the measurement of traits not only for the new participants but also the existing participants in the study. The Big Five Personality survey

has 44 questions with options (1-Strongly Disagree, 2-Disagree, 3-Neither Agree nor Disagree, 4- Agree, 5-Strongly Agree). Then the study calculates the personality trait measurements as openness, conscientiousness, extraversion, agreeableness, neuroticism in the range of [1,5] for every participant and records in waves of the basic survey data. Self- Esteem is determined by The Rosenberg Self-Esteem Scale which is well-accepted and widely used survey. This scale has a purpose to investigate the participants' the general feelings about themselves in 10 item questions with options (1-Strongly Disagree, 2-Disagree, 3-Neither Agree nor Disagree, 4- Agree, 5-Strongly Agree). The self-esteem score is quantified in the range of [0,50] by this scale and recorded in the 7 out of 8 waves.

Self-Regulation (Exercise) is known as SRQ-E which has 16-item questions with 7-degree options between “1-Not at all true, 4- Somewhat true, 7-Very true”. This scale investigates each participants' reason the why they would exercise regularly among a variety of reasons. The results of scale are recorded outputs as External Self-Regulation (SRQE-Ext), Introjective Self-Regulation (SRQE—Introj), Identified Self-Regulation (SRQE-Ident), Intrinsic Self-Regulation (SRQE-Intrin) those have range [1-7]. Additionally, Self-Regulation Exercise Relative Autonomy Index (SRQE_RAI) is also measured with the same scale with a range [(-12) – (+12)].

Self-Regulation (SRQ) is the capability of people of developing and progressing through a planned behavior consistently to reach their purposes. This measurement shows the degree of self-discipline that have 12-item questions with options (1-Strongly Disagree, 2-Disagree, 3-Neither Agree nor Disagree, 4- Agree, 5-Strongly Agree) and in the range of [1-5].

Major Event is a single question as “Have you experienced any major life changes in the past six months?”. The major event is classified as “1” when the major life change happened in last six months, otherwise “0” for non-existence of major event. The major life changes include the events which affects the participants dramatically as follows.

- new relationship
- breakup

- divorce
- marriage
- end-friendship
- family issues
- someone sick or dead
- pet-dead
- property loss
- severe illness

4.2. Smartphone Data

In the NetHealth study, the smartphone data is derived from the records of calls and different types of text messages which are gathered from Android and iPhones. In the 4-year (8 waves) period, the obtained huge communication data has approximately 60 million records of communication event as the logs of voice calls and text messages records. The collection of data works both for Android and iPhone. The logs of communication events are retrieved through a smartphone application that installed on the participants' smartphones. The Android phones sent the logs to the NetHealth servers every day. iPhones have backup system that applied periodically by the participants computers and the computer applications processed the backup data and send the communication event data to the NetHealth servers. However, the problem of the call data records from Android phones, there is no call records of Android users. The similar problem comes up for the WhatsApp text messages. Fortunately, over the 95% of the records obtained from iPhone users and similar ratio occurs for iPhone users among all participants. Additionally, the number of SMS events are about 53 million and it consists of overwhelmingly iMessage. In the USA, iMessage is an equivalent sort of WhatsApp text message those both works by the internet connection. We can conclude that iPhone users in the study prefers iMessage for texting among the text message types based on the logs of text message events. The communication data labelled each participant for every event in detail with the date and time, the day of the week, outgoing or incoming, the type of event like call, SMS, WhatsApp and also the duration of calls or the length of the SMS.

4.3. Fitbit Data

In the NetHealth study, FitBit is a wearable device like smartwatch and the participants are asked to wear them all the time. FitBit device has two physical sensors to compute many measures of activity. One of the sensors is a 3-axis accelerometer and the other one is the heart rate sensor. FitBit is able to compute the number of steps and climbed floors daily by the accelerometer sensor data. Furthermore, it is able to classify the level of physical activity for every non-sleeping minute of the day in four states: sedentary, lightly active, fairly active and very active. This FitBit data of 4-year long necessarily contains a huge amount of data in a various type for each participant and it is classified into two groups as daily active data and daily sleep data. The FitBit data stores the daily total minutes of each level of physical activity zones for every participant. From the heart rate sensor, the FitBit records the heart rate minute by minute and aggregates the participants' daily mean heart rate and its standard deviation. Additionally, it calculates the participants' daily minutes of burned calories in four categories (low range, fat burn, cardio, peak) based on the metabolic equivalent of task (MET) and basal metabolic rate (BMR). The crucial factor of using daily activity data is the compliance score which indicates the percentage of each day that a participant was wearing a Fitbit in minute basis. The Fitbit data are more reliable and efficient by recording this compliance factor for every day. Fitbit also uses its sensors to determine whether a participant is sleeping or not. They can have some sleep periods more than one and it contains every sleep period with details such as date, time to go to bed, duration of sleep period, sleep efficiency etc.

NetHealth study data from the University of Notre Dame have been demanded for recent research which applies machine learning approaches (Hone et al., 2014). The NetHealth data represents many participants which is much more than other existing studies. The NetHealth project includes a much longer timeline than the data of the existing studies so it may be considered more reliable (S. Liu et al., 2019). The most of studies have long time periods of measurement as once in a year or more. It would be a reasonable period for life-satisfaction or similar concepts which these studies have commonly focused on. SWB was measured annually in most studies, which is contrary to the principles of subjective well-being concept. The measurement of SWB is based on the detection of

variation of SWB in short terms. So, the SWB studies would more efficient prediction models with higher accuracy when the data gathering and applied surveys are more frequent. However, the longitudinal studies with short measurement periods to detect on changes in well-being are very rare because of the high level of requirements (Luhmann & Intelisano, 2018).

4.4. Existing Research on Nethealth

The multi-dimensionality and reliability of NetHealth data is the main reason for its selection as a source in various studies. With the FitBit data of the NetHealth study, the variation of activity levels only on the weekends may correlate with the traits. It implicitly proposes that the real time activity data of individuals may contribute for the measurement of well-being (Faust, Hachen, et al., 2018). One pervasive study in Notre Dame University (S. Liu et al., 2020) examines dynamic social network data and exploits the relationship between social interactions and mental health to predict individuals' likelihood of being depressed or anxious from. Other studies on NetHealth dataset are shown in the following:

Table 3. 2. The recent studies on NetHealth project

Author	Year	Study	Findings
Faust, L., D. Hachen, O. Lizardo, and N. V. Chawla	2018	Quantifying Subjective Well-Being Using Trends in Weekend Activity	Measuring subjective well-being by the linear relationship between activity level and big five personality
Faust, Louis, Priscilla Jiménez, et al.	2018	Long-Term Compliance Habits: What Early Data Tells Us	Identifying long-term compliance habits and the feasibility of detection

Faust, Louis, Priscilla Jiménez-Pazmino, et al.	2019	What 30 Days Tells Us About 3 Years: Identifying Early Signs of User Abandonment and Non-Adherence	Identify the stability of adherence behaviors, feasibility of their early detection
Faust, Louis, Rachael Purta, David Hachen et al.	2017	Exploring Compliance: Observations from a Large-Scale Fitbit Study	Machine learning approach to discover of the influence compliance for studies
Faust, Louis, Cheng Wang, David Hachen, Omar Lizardo, and Nitesh V. Chawla.	2019	Physical Activity Trend eXtraction: A Framework for Extracting Moderate-Vigorous Physical Activity Trends from Wearable Fitness Tracker Data	Developing a framework of activity trends from smartwatch data
Fridmanski, E., Wood, M.L., Lizardo, O. and Hachen, D.	2020	Clustering in a newly forming social network by subjective perceptions of loneliness	Determining the clusters in networks based on subjective perceptions of loneliness
Hossain, A., and C. Poellabauer	2016	Challenges in Building Continuous Smartphone Sensing Applications	Exploring the variety of technical and non-technical challenges and solutions
Liu, Shikang, David Hachen, Omar Lizardo et al.	2018	Network Analysis of the NetHealth Data: Exploring Co-Evolution of Individuals' Social Network Positions and Physical Activities	Finding the links between individuals' social network, health-related behaviors, and big five personality

Shikang Liu, David Hachen, Omar Lizardo et al.	2020	The power of dynamic social networks to predict individuals' mental health	Prediction model depression and anxiety by using participants' social network
Wang, Cheng, Omar Lizardo, and David S. Hachen	2021	Using Fitbit Data to Examine Factors That Affect Daily Activity Levels of College Students	Modulating the dynamics of physical activity levels by social and behavioral factors

5. SUBJECTIVE WELL-BEING PREDICTION MODEL AND METHODOLOGY

5.1. Description of Model

Individuals' daily continuous data from smart wristbands, social interactions tracked by cellphones, and personality traits data from surveys are used in this research to develop a machine learning approach to predict individualized subjective well-being. In regard of personalization, the model utilizes validated psychological surveys to gather information on physical characteristics, personality traits (Big Five Inventory), confidence, and control. Daily continuous data such as physical activity, sleep duration, and heart rate patterns from wearables, as well as social interactions from smartphones, are used to support the machine learning model. Finally, the machine learning model is trained with subjective well-being assessments derived from self-report questionnaires completed on a regular basis over period that synchronized with the pervasive collected data. The model is visualized by the Figure. The following section explains why features were chosen for the machine learning model based on psychological studies, and consequently how the data was processed on the NetHealth dataset.

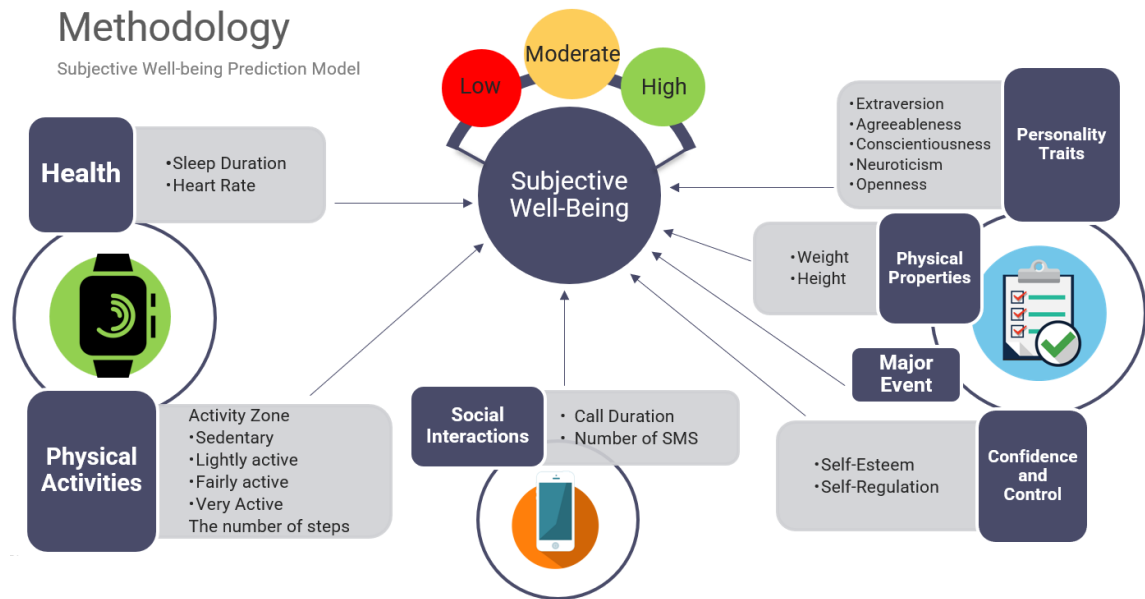


Figure 5. 1. The machine learning model for predicting subjective well-being

5.2. Feature Selection

There are wide-range data available in the NetHealth study dataset. Therefore, the elimination of features is done among the different categories in NetHealth data regarding potential indicators of SWB. This selection process has some principles which serve the aim of this study based on the psychological research.

5.3. Smartphone Features

Social Interactions

The second principle is that the prepared data used in the model include the daily data of an individual. The continuous daily data as much as possible would result in an advantage for the model's performance. However, the more complex daily information would harm both the practicality and simplicity of the model, as mentioned in the purpose of this research. In the NetHealth study, many detailed categories of data are gathered from

smartphones and FitBit devices daily, hourly, and even secondly. There should be a balance between these two opposite points. Another issue is the privacy of personal data. In NetHealth, all individuals' data is tagged as "egoid," so they are anonymized in respect of individuals. However, if the categories of data used in the model are detailed in personality, it would make individuals avoid and be a restriction for the usability of the model. For these reasons, the duration of calls and the number of SMSes (i-message, MMS, and WhatsApp) are taken from the smartphone data in communication. These two types of data would give a social interaction of individuals as input for the prediction model.

5.4. Smartwatch Features

Physical Activities

Physical activity and sleep duration are explicitly chosen to investigate the effects on mental well-being in the (Gothe et al., 2020) research. Sleep duration, heart rate, the activity time across four different groups (sedentary, lightly active, fairly active, very active) are used in the model from the FitBit device data. Participants with higher step counts on weekend activities were happier and resulted in having higher levels of SWB (Faust, Hachen, et al., 2018). Given the correlations among these factors, we propose weekend activity be considered as a measure of informing one's subjective well-being (Faust, Hachen, et al., 2018). The number of steps can be used as an indicator of SWB. However, the effect of physical activities may change among individuals in respect of mental or physical conditions. Intervention effects of exercise self-regulation on physical exercise are evaluated, and Exercise Self-Regulation Questionnaire (SRQ-E) is used for the evaluation (Fleig, Lippke, Pomp, & Schwarzer, 2011). When the questions of the SRQ-E survey are examined, it gives the purpose of individuals for physical activities, the attitudes towards them mentally, and the reaction for physical activities in SWB perspective. The weight and height of individuals are also added to the model to support the physical activity category of data from a physical condition perspective. In this manner, physical activity data would be a more meaningful and more effective predictor for every individual by SRQ-E information.

5.5. Surveys Features

5.5.1. Personality Traits

The first principle is that the most known features about well-being in psychology literature are taken from the data for the ground truth of SWB. Personality characteristics can be significant predictors of SWB. The biggest connections with life happiness were found to be between neuroticism and extraversion. The relationships between agreeableness and conscientiousness and particular well-being components were modest, although they were smaller than those between neuroticism and extraversion (Diener, Lucas, et al., 2018b). Extroversion and neuroticism are the most consistently and strongly related to well-being (Diener, Lucas, et al., 2018b). Extraversion has been shown to predict positive effects while negative effects are strongly indicated by neuroticism (Diener & Ryan, 2009). A widely accepted way to measure the personality traits of any individual is known Big Five Inventory (BFI). The BFI used in research is a wide used validated scale used to measure individual's five personality traits as extraversion, agreeableness, openness, conscientiousness, and neuroticism. Individuals are asked to assess their level of agreement or disagreement with each statement and similar conclusions are shown in the (Howell et al., 2010) research that examined the robust findings that the BFI personality traits predict SWLS scores. BFI is also commonly used as an explanatory variable with other SWB scales. Positive affect and negative affect have been consistently correlated with each of the BFI personality factors; however, extraversion have the highest degree of impact on determining of positive affect, and neuroticism has the similar role for negative affect. The SWLS can be tested for validity when administered as existed survey versus as an Internet survey (Howell et al., 2010).

5.5.2. Confidence and Control

In [8], the authors made a review for the evidence that other traits such as self-esteem and optimism also are moderately associated with SWB judgments. It is reported that self-esteem is a necessary emotional state with personality traits that affect SWB in the HUPPERT & SO and Deiner's models (Gray-Little et al., 1997; Hone et al., 2014). In the

NetHealth study, self-esteem is measured by the Rosenberg Self-Esteem Scale (RSES), which is defined in (Gray-Little et al., 1997). Furthermore, the self-esteem measured by RSES is considered one of the predictors for well-being and reported that low self-esteem was the best predictor of increases in sadness (Ciarrochi, Heaven, & Davies, 2007).

5.5.3. Major Event

All the features examined above have reasonable effects on the SWB, and they are used as inputs in the model to predict SWB. The possible threats for the model have become more visible in the grand scheme of things, especially from the life events. Firstly, the classic adaptation is described in the SWB perspective in (Luhmann & Intelisano, 2018) research as "Together, these studies indicate that SWB is relatively stable but changeable. For most people, their SWB fluctuates around a stable baseline that is partly heritable, which is consistent with classical theories of adaptation". Individuals are bound by the "hedonic treadmill," that explains the process through which recent changes in living circumstances momentarily raise an individual's subjective standard before adapting to the current changes (Diener & Ryan, 2009). Studies on loss, marriage, and divorce show that more sophisticated theories of adaptability are viable. According to bereavement research, the loss of a spouse produces a significant drop in life satisfaction for family members, which can take up to seven years to recover from (Diener & Ryan, 2009). Studies on marriage and divorce stated that people often recover from adverse life events and experience well-being after extremely positive events (Diener & Ryan, 2009).

Research on SWB produce strong evidence that the individuals who had experienced major changes in their lives, have significantly different level of SWB in average (Luhmann & Intelisano, 2018). For example, individuals who had experienced a divorcement period are tend to have lower level of SWB compared to married people do (Diener & Ryan, 2009). Thus, life circumstance and experienced events also have significant effects on SWB in circumstances of stable personality and the level of SWB over time (Diener, Lucas, et al., 2018b).

In the light of the literature, it would be reasonable to conclude that the significant events in life have a dominant effect on SWB compared to the changes in the features selected earlier in the study. If hedonic adaptation is inevitable, even significant life events should not have any lasting impact on SWB (Luhmann & Intelisano, 2018). At this point, some research stated that individuals' well-being can be changed after experiencing positive life events like marriage and recover from negative events such as the death of a spouse (Diener & Ryan, 2009). It shows that significant events must be considered the temporary but dominant effect on the SWB. The solution for crucial "major event" threat to the model comes with a question in the surveys of the NetHealth study: "Have you experienced any major life changes in the past six months (e.g., loss of a loved one, a break-up, divorce, marriage)?" Hence, a significant life change in any individual in the recent past would protect the prediction model's performance from dominant effects on SWB and make it more reliable.

5.6. Data Processing and Principles

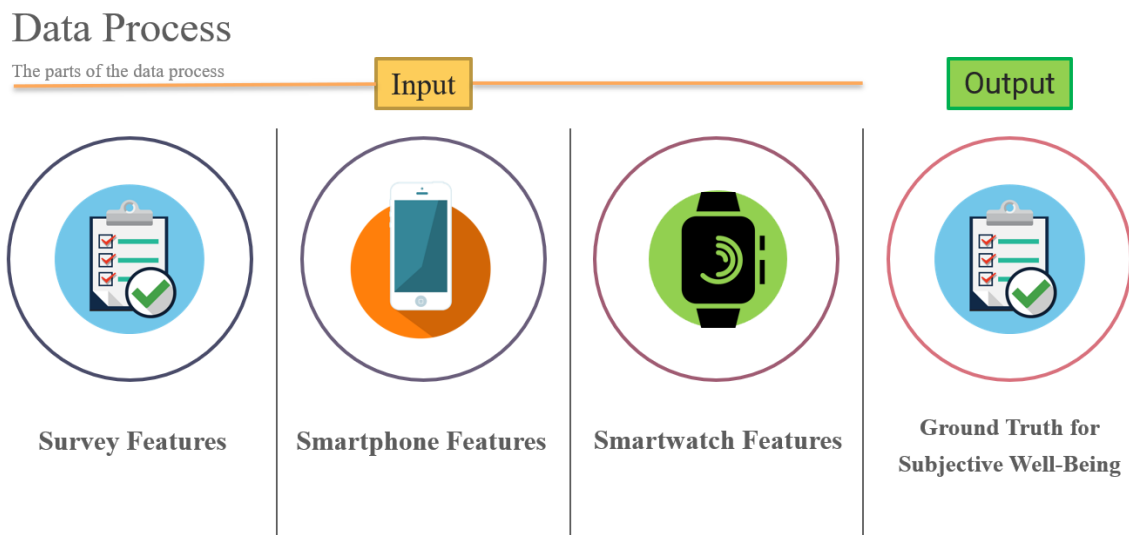


Figure 5. 2.The summary of data process

5.6.1. Input Data Processing

The NetHealth data has to be processed data cleaning to prepare data for the prediction model. The BFI elements, Self Esteem and SRQ-E values in the NetHealth study are measured by their valid appropriate scales mentioned in the section before. Weight and height are eliminated from outliers that are too far from being possible for a human being. Finally, the major event is added to the model as given in the dataset as binary.

BFI scale evaluation has stability in a short time as considered in (Specht, Egloff, & Schmukle, 2011) and (Deci & Ryan, 2008). This approach is applied in (S. Liu et al., 2018): NetHealth project applied personality surveys at every wave, so we consider the personality characteristics which were determined by big five personality surveys are stable in short time scales. However, BFI properties of individuals are updated by the new results of BFI from the next surveys in the waves. So, in theory, some participants' personalities might change over the three years. (S. Liu et al., 2018). Thus, the BFI blanks in the seasons are completed with the last season before if the same individual has BFI measurement by using stability.

The examples are prepared for the machine learning models as supervised learning methods require (paired). Thus, every season in NetHealth gives paired data for the model. The duration of calls, the number of SMSs, sedentary time, lightly active time, fairly active time, very active time, sleep duration, heart rate, the number of steps is taken as a daily average between seasons. The standard deviations of these features are also added to give the ability to model to distinguish the different distributions that would result in other effects on SWB. Every individual has different survey dates in seasons, so the average and standard deviations of features are calculated and paired with every participant's survey dates for data consistency. For the accuracy of averages and standard deviation, the examples are chosen if they have %80 or higher comply recent that is the ratio of using Fitbit device in one day.

5.6.2. Ground Truth Data Processing

The validity of subjective happiness scale (SHS) is shown that happiness is an essential measurement for SWB in the background section. SHS is used as guidance for the SWB measurement.

5.6.2.1. Subjective Happiness Scale

Table 5. 1. The Subjective Happiness Scale Questionnaire

For each of the following statements and/or questions, please circle the point on the scale that you feel is most appropriate in describing you.									
1. In general, I consider myself:									
not a very happy person	1	2	3	4	5	6	7	a very happy person	
2. Compared with most of my peers, I consider myself:									
less happy	1	2	3	4	5	6	7	more happy	
3. Some people are generally very happy. They enjoy life regardless of what is going on, getting the most out of everything. To what extent does this characterization describe you?									
not at all	1	2	3	4	5	6	7	a great deal	
4. Some people are generally not very happy. Although they are not depressed, they never seem as happy as they might be. To what extent does this characterization describe you?									
not at all	1	2	3	4	5	6	7	a great deal	

To score the scale, reverse code the 4th item (i.e., turn a 7 into a 1, a 6 into a 2, a 5 into a 3, a 3 into a 5, a 2 into a 6, and a 1 into a 7), and compute the mean of the 4 items (Lyubomirsky & Lepper, 1999).

5.6.2.2. The Construction of SHS From Surveys from Nethealth

In order to use SHS, the following answers from the NetHealth surveys are evaluated in equivalent forms of questions in SHS:

Table 5. 2. The equivalent questions used for Subjective Happiness Scale

The right of the statement to indicate how one generally feels					
1. I am happy					
almost never	1	2	3	4	almost always
2. I wish I could be as happy as others seem to be					
almost never	1	2	3	4	almost always
3. Taken all together, how would you say things are these days -- would you say that you are thrilled, pretty happy or not too happy?					
almost never	1	2	3	4	almost always

A new approach to assessing hedonic well-being taps emotional states without the overlay of evaluation. New approaches significantly reduce this problem by having individuals report relatively brief and recent periods (Steptoe, Deaton, & Stone, 2014). To follow this advancement, the answers for the question "Taken all together, how would you say things are these days -- would you say that you are thrilled, pretty happy or not too happy?" is also prepared for the model. This addition makes the SWB measurement neither general feeling nor momentary evaluation. Thus, it would be a sound output for the model since SWB is considered recently time as defined.

5.6.2.3. Generation of Subjective Well-Being Classes

The distribution of SWB scores in respect of SHS answers of participants are as follows.

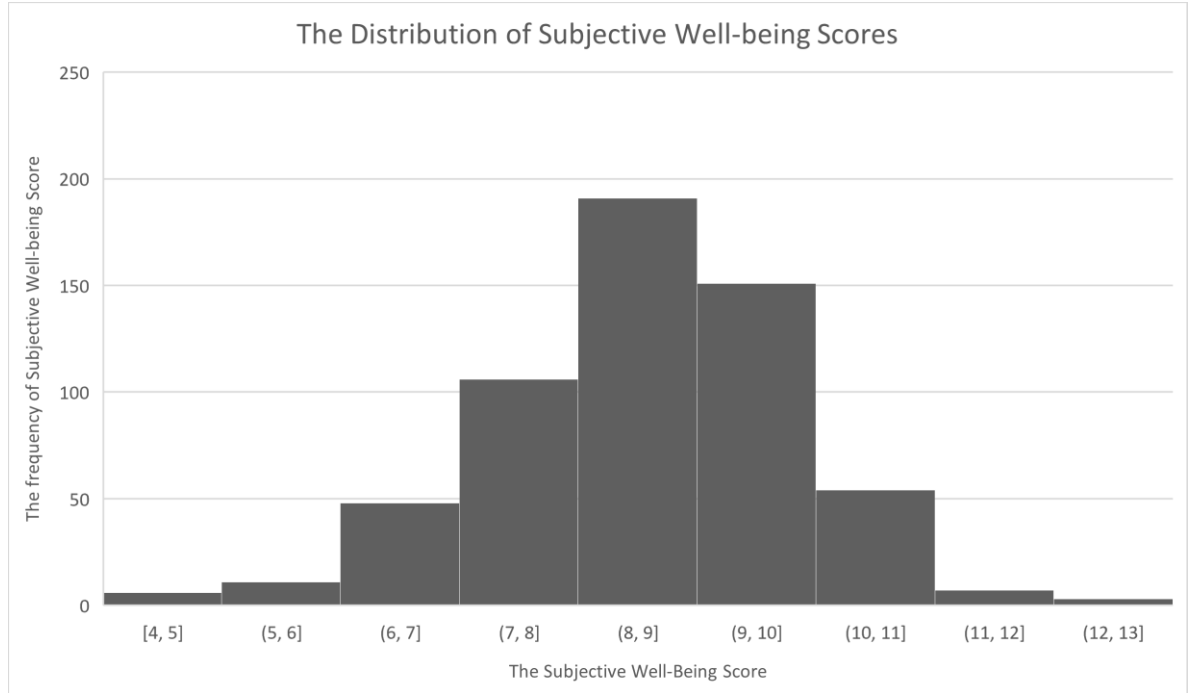


Figure 5. 3. The distribution of subjective well-being scores

The distribution of subjective well-being scores is approximately symmetric and normal. It means extreme SWB scores are rare and the most of the SWB scores occur in the center of the distribution. However, the imbalanced distribution of every possible SWB score complicates the generation of SWB classes. To avoid misevaluating the machine learning models and deal with possible imbalance problem and, the generation of SWB classes is split into two categories by determining different thresholds for each level of SWB. First category is equal interval generation which has thresholds between levels in order to create approximately equal ranges. This method of grouping is a standard way of grading for the commonly used various scales. Second category is the balanced generation which creates slightly balanced distribution of SWB scores in respect of the number of observations in each level of SWB. By this grouping method, there are more balanced ground truth for SWB levels. Thus, the machine learning models would have more effective training process and ability to distinguish between the levels of SWB.

Equal Interval Generation of SWB Levels

The SHS scores of individuals are divided into 3 level of SWB groups as low, moderate, and high level by approximately equal intervals. The classification thresholds are as follows.

SHS score ≤ 7 as the low level of SWB (Observation: 65)

$7 < \text{SHS score} \leq 10$ as the moderate level of SWB (Observation: 448)

$10 < \text{SHS score} \leq 15$ as the high level of SWB (Observation: 64)

Slightly Balanced Generation of SWB Levels

SHS score ≤ 8 as the low level of SWB (Observation: 171)

$8 < \text{SHS score} \leq 10$ as the moderate level of SWB (Observation: 342)

$10 < \text{SHS score} \leq 15$ as the high level of SWB (Observation: 64)

5.7. Machine Learning Models

The many of common-utilized machine learning algorithms are applied for classification in this study. After the data processing in the Section 5.2 and Section 5.3, the input data and SWB scores of individuals as output data is paired in the end of process. The sources of input data are mostly available and gather huge amount of data but the part of data with low compliance rate of wearing FitBit or meaningless calls and messages (zero duration or length) etc. are eliminated in this study. The subjective well-being scores of participants are prepared from the surveys which are applied in every 3-4 months (every wave). Only if both input data and SWB scores of individuals are available at the same date periods, they are paired and generate the examples of trainable data for supervised machine learning models. After all data process in the Section 5.2 and Section 5.3, the available processed data has 577 examples which is relatively low in amount compared to the commonly used datasets for machine learning models. Hence, the training process would be deteriorated, and machine learning algorithms would perform worse in prediction. To overcome the low number of data examples, the many widely used of

machine learning classifiers are applied in this study by using Scikit-Learn, Keras libraries in Python programming language and MATLAB. The applied machine learning classifiers are as follows:

- Trees
 - Fine
 - Medium
 - Coarse
- Linear Discriminant
- Naive Bayes
 - Gaussian
 - Kernel
- Support Vector Machine (SVM)
 - Linear Kernel
 - Quadratic
 - Cubic
 - Fine Gaussian
 - Medium Gaussian
 - Coarse Gaussian
- K-Nearest Neighbors (KNN)
 - Fine
 - Medium
 - Coarse
 - Cosine
 - Cubic
- Ensemble
 - Boosted Trees
 - Bagged Trees
 - Subspace Discriminant
- Neural Network

Based on the high performance in Section 5 (Results), we selected the four of the machine learning classifiers such that Boosted Tress, Quadratic Support Vector Machine, Cubic K-Nearest Neighbors and Linear Discriminant Analysis.

There are different techniques of validation in literature. The commonly used one of them is separating the data into training, validation or test subsets. By the rule of thumb, in the normal size dataset is separated as 80% training subset and 20% validation and in the large dataset the ratios shift to 95% training subset and %5 validation. However, our dataset is too small to separate into two subsets according to rule of thumb as mentioned. Additionally, the one or two of the SWB levels have relatively a smaller number of examples. The validation set which comes from separation may not include the examples of each SWB levels. Thus, the validation of the machine learning performances would not be reliable. The other common validation way is k-fold cross-validation. In this method, data is split into k folds. One of these folds is for testing and the rest (k-1) folds are for training process. The machine learning model is trained with k-1 training fold and the performance of model is measured by the one testing fold. This performance measurement is repeated k times at each time by selecting different test fold. Then the validation result is calculated as taking average of all performance measurements. Consequently, the validation of model does not depend on only one test subset and the performance is measured in respect of every fold of dataset. Hence, when dataset is small, the validation of model would be more reliable and reasonable.

5.8. Results

5.8.1. Classification Results

The neural network for classification is designed to predict the SWB classes of individuals. The structure of neural network consists of 7 full-connected layers which has one input layer, one output layer and 5 hidden layers. Input layer is created as 33 dimensional features set those processed in the Section 5.3.1. 5 hidden layers are organized with different number nodes as 16, 32, 64, 64 and 16 respectively. The activation function of hidden layers is rectified linear activation function (ReLU). In the

output layer surely have 3 nodes which use sigmoid activation function to calculate the probability of belonging to each level SWB. The neural network is trained by Adam optimization algorithm with 0.00001 of learning rate in respect of the classification loss function as categorical cross entropy. Particularly, the accuracy metric is the categorical accuracy which is commonly used in classification neural network models. Due to the small amount of the examples, the dataset is split into training and test set as the 80% and 20% respectively. To overcome the overfitting problem in neural network model, dropout function is applied in some of hidden layers with the ratios of 0.1 or 0.2. In the training process, the training set accuracy and validation set accuracy is balanced with dropout to avoid overfitting or bias in the neural network model. Consequently, the optimized neural network model achieves the accuracy of 80% in the training set and 75% validation set.

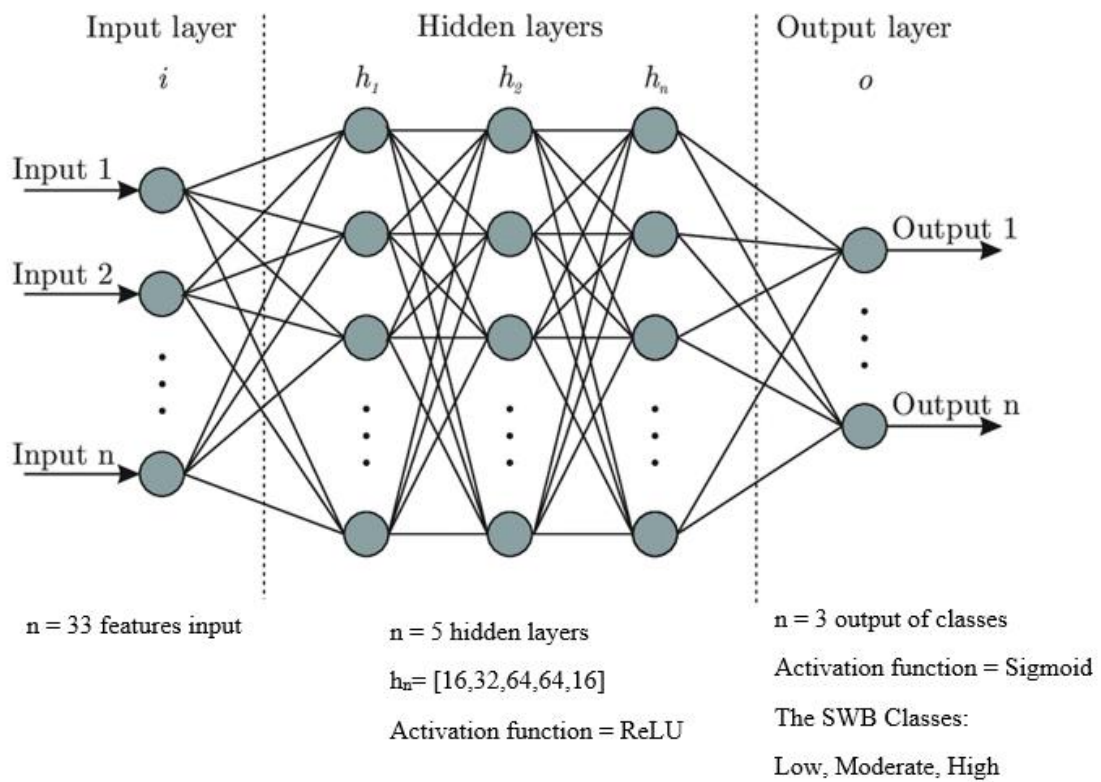


Figure 5. 4. The visual description of the designed neural network model with hyperparameters

The machine learning methods which are listed in Section 5.4 are applied in processed data and the classification accuracy is measured by both the ground truths as equal interval classes of SWB levels and slightly balanced classes of SWB levels. In order to solve imbalance and the small dataset problem for performance measurement, the accuracy of models is measured by the 5-fold and 10-fold cross validations. The four best results of machine learning methods are selected to compare their performances are explained as follow:

Table 5. 3. Equally interval SWB Model with 5-fold cross validation

Machine Learning Model	Accuracy (%)	Area Under Curve		
		Low	Moderate	High
Ensemble Boosted Tree	75.9%	0.51	0.48	0.51
Quadratic SVM	75.6%	0.65	0.52	0.60
Cubic KNN	77.5%	0.55	0.50	0.49
Linear Discriminant Analysis	76.4%	0.58	0.51	0.47

Table 5. 4. Equally interval SWB Model with 10-fold cross validation

Machine Learning Model	Accuracy (%)	Area Under Curve		
		Low	Moderate	High
Ensemble Boosted Tree	74.7%	0.56	0.48	0.46
Quadratic SVM	72.8%	0.65	0.54	0.56
Cubic KNN	67.2%	0.59	0.50	0.47

Linear Discriminant Analysis	71.1%	0.60	0.52	0.46
------------------------------	-------	------	------	------

Table 5. 5. Slightly Balanced SWB Model with 5-folds cross validation

Machine Learning Model	Accuracy (%)	Area Under Curve		
		Low	Moderate	High
Ensemble Boosted Tree	56.7%	0.64	0.58	0.45
Quadratic SVM	58.2%	0.67	0.60	0.58
Cubic KNN	59.1%	0.63	0.58	0.47
Linear Discriminant Analysis	58.2%	0.68	0.62	0.52

Table 5. 6. Slightly Balanced SWB Model with 10-folds cross validation

Machine Learning Model	Accuracy (%)	Area Under Curve		
		Low	Moderate	High
Ensemble Boosted Tree	58.8%	0.63	0.60	0.52
Quadratic SVM	58.1%	0.66	0.60	0.58
Cubic KNN	56.3%	0.59	0.55	0.48
Linear Discriminant Analysis	58.1%	0.67	0.60	0.50

The machine learning models which are applied in equally interval SWB classes achieve 77.5% with Cubic KNN method and the other methods ensemble boosted trees, quadratic support vector machine and linear discriminant analysis have similar accuracy results to Cubic KNN in the 5-fold cross validation. In the 10-fold cross validations, ensemble boosted tree method has the best accuracy 74.7% and other methods are close classification accuracies except Cubic KNN method. The Cubic KNN loses its stability due to the decrease in classification accuracy in the 10-fold cross validation compared to the 5-fold cross validation. The final classification results are shown in the following

Table 5. 7. The Summary of Machine Learning Performances

Machine Learning Model	Accuracy (%)
Cubic KNN	77.5%
Linear Discriminant Analysis	76.4%
Ensemble Boosted Tree	75.9%
Quadratic SVM	75.6%
Artificial Neural Network	75.0%

5.8.2. Regression Results

The ground truth classes for machine learning methods are constructed by thresholds in Section 5.3.2.3. The thresholds are determined with the intention of equal ranged classes and balanced number of examples. The separation of SWB scores into three classes as low, moderate and high subjective well-being classes may be appropriate for interpreting the scores in a practical approach. However, the classification of SWB scores may not

serve the machine learning training process. Thus, the regression models are applied to the processed input data labelled with original SWB scores of individuals as shown in Section 5.3.2.3.

Table 5. 8. The Results of Regression Models

Regression Model	RMSE	MSE	MAE	R-Squared	Corr. Coef.
Linear Regression	1.27	1.60	0.96	0.07	0.27
Linear SVM Regression	1.25	1.56	0.95	0.09	0.30

The distribution of SWB scores is not in the uniform form. In spite of non-uniform distribution of scores, the positive R-Squared values indicates that the regression models have better performances compared to the constant prediction for every example in the dataset. The Linear Support Vector Machine for Regression model outperforms the Linear Regression model by the RMSE of 1.25, MSE of 1.56 and correlation coefficient of 0.30 as shown in the table.

5.8.3. Discussion

As linear SVM regression is the best model in regression, the RMSE can be considered a small error 1.25 compared to the range of SWB as 9 which derived from the observations are distributed in [4-13]. However, RMSE may be found small due to the imbalanced data scores and most of the predictions are concentrated on the SWB score 9. Additionally, the r-squared value of model is 7% is very low that indicates the weak relationship between the variation of input data and the SWB scores.

The machine learning methods which are trained with slightly balanced SWB classes decrease in accuracy due to the shift in the distribution of level classes. The accuracy results are so similar between 5-fold and 10-fold cross validations except the cubic KNN.

The same instability case of cubic KNN occurs in the slightly balanced SWB classes. The classification methods perform worse in accuracy. On the other hand, the area under curve (AUC) values of each of SWB classes are mostly higher in the slightly balanced SWB compared to the equal interval SWB. The higher performance in AUC values indicates the ability of machine learning methods of distinguishing between classes. The more examples on the classes which have less examples and the more balanced distribution denotes that the machine learning methods are able to be trained with the higher degree of separability.

The regression model losses its ability to predictions as the distance of SWB score increases from the score 9. Compared to the regression results, classification results have stronger proofs with increasing AUC values those indicates the ability of separation for classes. The result states that the machine learning methods have high potential for better prediction performance in the case of more available examples in the training set.

Limitations

The primary reason of limited performance would be the insufficient number of examples. The dataset is rich in many aspects, but it needs shorter term for subjective well-being measurements to create more labelled data which is required for supervised machine learning models. Other limitation due to the lack of enough examples is personalization of prediction model. After data process, there is no opportunity to design a machine learning model for prediction on each participant's well-being in a sequential way. In this reason, sequential machine learning models cannot be used in methodology. NetHealth study is applied on university students as participants. Therefore, the sample represents a population which consists of a specific interval of age. Theoretically, this property of sample should be considered in terms of generalization of model. However, NetHealth is a rare dataset which have a broad range of variability in participants, surveys, data sources, high number of participants as well as the long range of time. The number of examples is not obstacle to use machine learning models for accurate predictions on individuals' well-being from a psychology perspective.

6. CONCLUSION

Subjective well-being has a sphere of impact in all aspects of an individual's life. In the light of current studies, loss in subjective well-being across the lifespan mainly arise from the positive effects such as happiness and enjoyment, as opposed to negative effects such as depression, stress and sadness. It indicates that even whether you have a mental illness (depression, anxiety, etc.) or a fatal physical illness, improving your subjective well-being is essential for health. Even though quantifying subjective well-being is challenging due to the complexities of its definition and assessment, the advancement of pervasive technology has resulted in the availability of prediction models of validated subjective well-being measurements. Our study applies commonly used machine learning models for classification and regression and provides a comparison of performances. In the end, we offer a machine learning model can predict an individual's well-being from a psychology perspective almost with an 80% accuracy by using data from her/his daily activities collected via smart wristbands, social relationships monitored by smartphones, and personality traits data from surveys.

For future research, the study can be extended with primarily the larger datasets which includes more frequent subjective well-being scores to improve the performances of models. The more examples of labelled data would also give opportunity to apply the complex machine learning models those explores the data patterns in timeline. The proposed model can be trained with the different validated subjective well-being scales as the ground truth for response variable. Practically, the machine learning model is adequate to develop as a personalized well-being assistant by providing smartphone and wearable device data. In conclusion, our research provides a machine learning model regarding psychological perspective and shows the feasibility of a pervasive application as a personalized well-being assistant.

REFERENCES

- A., S., & John, D. (2015). Survey on Chatbot Design Techniques in Speech Conversation Systems. *International Journal of Advanced Computer Science and Applications*, 6. <https://doi.org/10.14569/IJACSA.2015.060712>
- Acharya, U. R., Joseph, P., Kannathal, N., Lim, C., & Suri, J. (2007). Heart rate variability: A review. *Medical & Biological Engineering & Computing*, 44, 1031–1051. <https://doi.org/10.1007/s11517-006-0119-0>
- Alvarez-Lozano, J., Osmani, V., Mayora, O., Frost, M., Bardram, J., Faurholt-Jepsen, M., & Kessing, L. V. (2014). Tell Me Your Apps and I Will Tell You Your Mood: Correlation of Apps Usage with Bipolar Disorder State. *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2674396.2674408>
- Barker, M., & van der Linden, J. (2016). *A sprite in the dark: supporting conventional mental healthcare practices with a tangible device*.
- Bogomolov, A., Lepri, B., Ferron, M., Pianesi, F., & Pentland, A. (2014). Daily stress recognition from mobile phone data, weather conditions and individual traits. *MM 2014 - Proceedings of the 2014 ACM Conference on Multimedia*, 477–486. Association for Computing Machinery. <https://doi.org/10.1145/2647868.2654933>
- Bollen, J., Gonçalves, B., Ruan, G., & Mao, H. (2011). Happiness Is Assortative In Online Social Networks. *Artificial Life*, 17, 237–251. https://doi.org/10.1162/artl_a_00034
- Bravo, J., Hervás, R., & Villarreal, V. (2015). Ambient intelligence for health first international conference, AmiHEALTH 2015 Puerto Varas, Chile, December 1-4, 2015 proceedings. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 9456.
- Chida, Y., & Steptoe, A. (2008). Positive Psychological Well-Being and Mortality: A Quantitative Review of Prospective Observational Studies. *Psychosomatic Medicine*, 70, 741–756. <https://doi.org/10.1097/PSY.0b013e31818105ba>

- Ciarrochi, J., Heaven, P. C. L., & Davies, F. (2007). The impact of hope, self-esteem, and attributional style on adolescents' school grades and emotional well-being: A longitudinal study. *Journal of Research in Personality*, 41(6), 1161–1178. <https://doi.org/10.1016/j.jrp.2007.02.001>
- de Choudhury, M., Gamon, M., Counts, S., & Horvitz, E. (2013). *Predicting Depression via Social Media*. Retrieved from www.aaai.org
- Deci, E. L., & Ryan, R. M. (2008). Hedonia, eudaimonia, and well-being: an introduction. *Journal of Happiness Studies*, 9(1), 1–11. <https://doi.org/10.1007/s10902-006-9018-1>
- Diener, E. (1994). Assessing subjective well-being: Progress and opportunities. *Social Indicators Research*, 31(2), 103–157. <https://doi.org/10.1007/BF01207052>
- Diener, E., Lucas, R. E., & Oishi, S. (2018a). Advances and open questions in the science of subjective well-being. *Collabra: Psychology*, Vol. 4. University of California Press. <https://doi.org/10.1525/collabra.115>
- Diener, E., Lucas, R. E., & Oishi, S. (2018b). Advances and open questions in the science of subjective well-being. *Collabra: Psychology*, Vol. 4. University of California Press. <https://doi.org/10.1525/collabra.115>
- Diener, E., Oishi, S., & Tay, L. (2018). Advances in subjective well-being research. *Nature Human Behaviour*, 2(4), 253–260. <https://doi.org/10.1038/s41562-018-0307-6>
- Diener, E., & Ryan, K. (2009). *Subjective well-being: a general overview*.
- Diener, E., Wirtz, D., Biswas-Diener, R., Tov, W., Kim-Prieto, C., Choi, D., & Oishi, S. (2009). *New Measures of Well-Being*. https://doi.org/10.1007/978-90-481-2354-4_12
- Diener, E., Wirtz, D., Tov, W., Kim-Prieto, C., Choi, D. won, Oishi, S., & Biswas-Diener, R. (2010). New well-being measures: Short scales to assess flourishing and positive and negative feelings. *Social Indicators Research*, 97(2), 143–156. <https://doi.org/10.1007/s11205-009-9493-y>
- Faust, L., Hachen, D., Lizardo, O., & Chawla, N. v. (2018). Quantifying subjective well-being using trends in weekend activity. *Proceedings - 2018 IEEE International Conference on Healthcare Informatics, ICHI 2018*, 123–129. Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ICHI.2018.00021>
- Faust, L., Jiménez, P., Hachen, D., Lizardo, O., Striegel, A., & Chawla, N. v. (2018). *Long-term Compliance Habits: What Early Data Tells Us*. Retrieved from <http://arxiv.org/abs/1804.04256>
- Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a

- Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial. *JMIR Mental Health*, 4, e19. <https://doi.org/10.2196/mental.7785>
- Fleig, L., Lippke, S., Pomp, S., & Schwarzer, R. (2011). Intervention effects of exercise self-regulation on physical exercise and eating fruits and vegetables: A longitudinal study in orthopedic and cardiac rehabilitation. *Preventive Medicine*, 53, 182–187. <https://doi.org/10.1016/j.ypmed.2011.06.019>
- Fordyce, M. W. (1988). A review of research on the happiness measures: A sixty second index of happiness and mental health. *Social Indicators Research*, 20(4), 355–381. <https://doi.org/10.1007/BF00302333>
- Glenn, T., & Monteith, S. (2014, December 1). New Measures of Mental State and Behavior Based on Data Collected From Sensors, Smartphones, and the Internet. *Current Psychiatry Reports*, Vol. 16. Current Medicine Group LLC 1. <https://doi.org/10.1007/s11920-014-0523-3>
- Gothe, N. P., Ehlers, D. K., Salerno, E. A., Fanning, J., Kramer, A. F., & McAuley, E. (2020). Physical Activity, Sleep and Quality of Life in Older Adults: Influence of Physical, Mental and Social Well-being. *Behavioral Sleep Medicine*, 18(6), 797–808. <https://doi.org/10.1080/15402002.2019.1690493>
- Gravenhorst, F., Muaremi, A., Tröster, G., Arnrich, B., & Grünerbl, A. (2013). Towards a Mobile Galvanic Skin Response Measurement System for Mentally Disordered Patients. In *Proceedings of the 8th International Conference on Body Area Networks, BodyNets 2013*. <https://doi.org/10.4108/icst.bodynets.2013.253684>
- Gray-Little, B., Williams, V. S. L., & Hancock, T. D. (1997). An Item Response Theory Analysis of the Rosenberg Self-Esteem Scale. *Personality and Social Psychology Bulletin*, 23(5), 443–451. <https://doi.org/10.1177/0146167297235001>
- Grünerbl, A., Osmani, V., Bahle, G., Carrasco, J., Oehler, S., Mayora, O., ... Lukowicz, P. (2014). Using Smart Phone Mobility Traces for the Diagnosis of Depressive and Manic Episodes in Bipolar Patients. In *ACM International Conference Proceeding Series*. <https://doi.org/10.1145/2582051.2582089>
- Guribye, F., & Gjøster, T. (2018). Tangible Interaction in the Dentist Office. *Proceedings of the Twelfth International Conference on Tangible, Embedded, and Embodied Interaction*, 123–130. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/3173225.3173287>
- Hao, B., Li, L., Gao, R., Li, A., & Zhu, T. (2014). Sensing Subjective Well-Being from Social Media. *ArXiv*, *abs/1403.3807*.
- Hone, L. C., Jarden, A., Schofield, G. M., & Duncan, S. (2014). Measuring flourishing: The impact of operational definitions on the prevalence of high levels of wellbeing. *International Journal of Wellbeing*, 4(1), 62–90. <https://doi.org/10.5502/ijw.v4i1.4>
- Howell, R. T., Kern, M. L., & Lyubomirsky, S. (2007). Health benefits: Meta-analytically determining the impact of well-being on objective health outcomes.

- Health Psychology Review*, 1(1), 83–136.
<https://doi.org/10.1080/17437190701492486>
- Howell, R. T., Rodzon, K. S., Kurai, M., & Sanchez, A. H. (2010). A validation of well-being and happiness surveys for administration via the internet. *Behavior Research Methods*, 42(3), 775–784. <https://doi.org/10.3758/BRM.42.3.775>
- Inkster, B., Sarda, S., & Subramanian, V. (2018). An Empathy-Driven, Conversational Artificial Intelligence Agent (Wysa) for Digital Mental Well-Being: Real-World Data Evaluation Mixed-Methods Study. *JMIR Mhealth Uhealth*, 6(11), e12106. <https://doi.org/10.2196/12106>
- Jebb, A. T., Morrison, M., Tay, L., & Diener, E. (2020). Subjective Well-Being Around the World: Trends and Predictors Across the Life Span. *Psychological Science*, 31(3), 293–305. <https://doi.org/10.1177/0956797619898826>
- Kroenke, K., Spitzer, R., Williams, J., & Löwe, B. (2009). An Ultra-Brief Screening Scale for Anxiety and Depression: The PHQ–4. *Psychosomatics*, 50, 613–621. <https://doi.org/10.1176/appi.psy.50.6.613>
- Kushlev, K., Drummond, D. M., & Diener, E. (2020). Subjective Well-Being and Health Behaviors in 2.5 Million Americans. *Applied Psychology: Health and Well-Being*, 12(1), 166–187. <https://doi.org/10.1111/aphw.12178>
- Larsen, R. J., & Eid, M. (2008). *Ed Diener and the Science of Subjective Well-Being*.
- Lee, D., Oh, K.-J., & Choi, H.-J. (2017). The chatbot feels you - a counseling service using emotional response generation. *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, 437–440.
- Likamwa, R., Liu, Y., Lane, N., & Zhong, L. (2013). MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In *MobiSys 2013 - Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services*. <https://doi.org/10.1145/2462456.2464449>
- Liu, P., Tov, W., Kosinski, M., Stillwell, D., & Qiu, L. (2015). Do Facebook Status Updates Reflect Subjective Well-Being? *Cyberpsychology, Behavior and Social Networking*, 18, 373–379. <https://doi.org/10.1089/cyber.2015.0022>
- Liu, S., Hachen, D., Lizardo, O., Poellabauer, C., Striegel, A., & Milenković, T. (2018). Network analysis of the NetHealth data: exploring co-evolution of individuals' social network positions and physical activities. *Applied Network Science*, 3(1). <https://doi.org/10.1007/s41109-018-0103-2>
- Liu, S., Hachen, D., Lizardo, O., Poellabauer, C., Striegel, A., & Milenković, T. (2020). The power of dynamic social networks to predict individuals' mental health. *Pacific Symposium on Biocomputing*. *Pacific Symposium on Biocomputing*, 25, 635–646.

- Liu, S., Vahedian, F., Hachen, D., Lizardo, O., Poellabauer, C., Striegel, A., & Milenkovic, T. (2019). *Heterogeneous network approach to predict individuals' mental health*. Retrieved from <http://arxiv.org/abs/1906.04346>
- Luhmann, M., & Intelisano, S. (2018). *Hedonic Adaptation and the Set Point for Subjective Well-Being*.
- Lyubomirsky, S., & Lepper, H. S. (1999). *A MEASURE OF SUBJECTIVE HAPPINESS: PRELIMINARY RELIABILITY AND CONSTRUCT VALIDATION*.
- Manikonda, L., & Choudhury, M. (2017). *Modeling and Understanding Visual Attributes of Mental Health Disclosures in Social Media*. <https://doi.org/10.1145/3025453.3025932>
- Matthews, M., & Doherty, G. (2011). In the Mood: Engaging Teenagers in Psychotherapy Using Mobile Phones. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2947–2956. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/1978942.1979379>
- Mohr, D. C., Zhang, M., & Schueller, S. M. (2017). Personal Sensing: Understanding Mental Health Using Ubiquitous Sensors and Machine Learning. *Annu. Rev. Clin. Psychol*, 13, 23–47. <https://doi.org/10.1146/annurev-clinpsy-032816>
- Nadeem, M. (2016). *Identifying Depression on Twitter*.
- Niemantsverdriet, K., & Versteeg, M. (2016). Interactive Jewellery as Memory Cue: Designing a Sound Locket for Individual Reminiscence. *Proceedings of the TEI '16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction*, 532–538. New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2839462.2856524>
- Pavot, W. (2018). *The Cornerstone of Research on Subjective Well-Being: Valid Assessment Methodology*.
- Sano, A., & Picard, R. W. (2013). Stress Recognition Using Wearable Sensors and Mobile Phones. *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*, 671–676.
- Setz, C., Arnrich, B., Schumm, J., Marca, R. la, Tröster, G., & Ehlert, U. (2010). Discriminating Stress From Cognitive Load Using a Wearable EDA Device. *IEEE Transactions on Information Technology in Biomedicine*, 14(2), 410–417. <https://doi.org/10.1109/TITB.2009.2036164>
- Sharma, N., & Gedeon, T. (2012). Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, 108(3), 1287–1301. <https://doi.org/10.1016/j.cmpb.2012.07.003>
- Specht, J., Egloff, B., & Schmukle, S. C. (2011). *Stability and Change of Personality Across the Life Course: The Impact of Age and Major Life Events on Mean-Level*

- and Rank-Order Stability of the Big Five 377 SOEPpapers on Multidisciplinary Panel Data Research*. Retrieved from www.diw.de
- Step toe, A., Deaton, A., & Stone, A. (2014). Subjective Wellbeing, Health, and Ageing. *Lancet*, 385. [https://doi.org/10.1016/S0140-6736\(13\)61489-0](https://doi.org/10.1016/S0140-6736(13)61489-0)
- Vallance, J., Winkler, E., Gardiner, P., Healy, G., Lynch, B., & Owen, N. (2011). Associations of objectively-assessed physical activity and sedentary time with depression: NHANES (2005-2006). *Preventive Medicine*, 53, 284–288. <https://doi.org/10.1016/j.ypmed.2011.07.013>
- VanderWeele, T. J., Trudel-Fitzgerald, C., Allin, P., Farrelly, C., Fletcher, G., Frederick, D. E., ... Kubzansky, L. D. (2020, April 1). Current recommendations on the selection of measures for well-being. *Preventive Medicine*, Vol. 133. Academic Press Inc. <https://doi.org/10.1016/j.ypmed.2020.106004>
- Well-being: The foundations of hedonic psychology. (1999). In D. Kahneman, E. Diener, & N. Schwarz (Eds.), *Well-being: The foundations of hedonic psychology*. New York, NY, US: Russell Sage Foundation.
- Wongkoblap, A., Vadillo, M. A., & Curcin, V. (2017). Researching Mental Health Disorders in the Era of Social Media: Systematic Review. *J Med Internet Res*, 19(6), e228. <https://doi.org/10.2196/jmir.7215>
- Wood, A. M., & Joseph, S. (2010). The absence of positive psychological (eudemonic) well-being as a risk factor for depression: A ten year cohort study. *Journal of Affective Disorders*, 122(3), 213–217. <https://doi.org/10.1016/j.jad.2009.06.032>
- Woodward, K., Kanjo, E., Brown, D., McGinnity, T. M., Inkster, B., MacIntyre, D., & Tsanas, T. (2020). Beyond Mobile Apps: a Survey of Technologies for Mental Well-being. *IEEE Transactions on Affective Computing*. <https://doi.org/10.1109/TAFFC.2020.3015018>

BIOGRAPHICAL SKETCH

Nail Şenbaş graduated from Bornova Anatolian High School, he began studying Civil Engineering at Bogaziçi University. He received B.S. degree in 2015. He has been studying M.Sc. at Galatasaray University Industrial Engineering Department since 2017 while working as lecturer in private sector. He has been working as Teaching Assistant in Contemporary Technology University.