

**BC3409: AI in Accounting & Finance Report**

**Prepared for: Dr. Teoh Teik Toe**

**Seminar Class 3, Team 1**

|  |  |
| --- | --- |
| **Chan Shi Hui Amanda** | **U1910759K** |
| **Harvinder Singh Chahal** | **U1910668J** |
| **Lee Wei Jie James** | **U1810181K** |
| **Tan Jing Yu, Denis** | **U1810101D** |
| **Valerie Chang Yingwen** | **U1810071D** |

**TABLE OF CONTENTS**

[**EXECUTIVE SUMMARY 1**](#_Toc68809219)

[**1 INTRODUCTION 1**](#_Toc68809220)

[**1.1 Business Problem Statement 1**](#_Toc68809221)

[**1.2 Significance of the Business Problem 2**](#_Toc68809222)

[**1.3 Expected Business Outcome 3**](#_Toc68809223)

[**2 ANALYTICS APPROACH 3**](#_Toc68809224)

[**3 DATA PREPARATION AND UNDERSTANDING 3**](#_Toc68809225)

[**3.1 Data Set Description 3**](#_Toc68809226)

[**3.2 Exploratory Data Analysis 4**](#_Toc68809227)

[**3.3 Data Preprocessing 6**](#_Toc68809228)

[**3.3.1 Data Cleaning 6**](#_Toc68809229)

[**3.3.1.1 Removal of Missing Values 6**](#_Toc68809230)

[**3.3.1.2 Removal of Duplicates 6**](#_Toc68809231)

[**3.3.1.3 Data Type Conversion 7**](#_Toc68809232)

[**3.3.1.4 Removal of Rows with Negative FWBscore 7**](#_Toc68809233)

[**3.3.2 Discretization of the Target Variable 7**](#_Toc68809234)

[**3.3.3 Feature Selection 8**](#_Toc68809235)

[**3.3.3.1 Removal of Unique ID Column 8**](#_Toc68809236)

[**3.3.3.2 Removal of Columns Used to Adjust for Demographic and Poverty Differences 8**](#_Toc68809237)

[**3.3.3.3 Removal of Direct Survey Questions for FWBscore 9**](#_Toc68809238)

[**3.3.3.4 Removal of Secondary Survey Score Columns 9**](#_Toc68809239)

[**3.3.3.5 Removal of Survey Items with Incomplete Base 9**](#_Toc68809240)

[**3.3.3.6 Removal of Survey Items with Sensitive, Vague, or Region Specific Questions 9**](#_Toc68809241)

[**3.3.3.7 Removal of Survey Items with Specific Financial Knowledge Questions 10**](#_Toc68809242)

[**3.3.4 Data Transformation 10**](#_Toc68809243)

[**3.3.5 Data Partitioning 11**](#_Toc68809244)

[**4 MODEL DEVELOPMENT AND EVALUATION 12**](#_Toc68809245)

[**4.1 Model Development 12**](#_Toc68809246)

[**4.1.1 Choice of Machine Learning Algorithms 12**](#_Toc68809247)

[**4.1.2 Hyperparameter Tuning 12**](#_Toc68809248)

[**4.2 Model Evaluation 13**](#_Toc68809249)

[**4.2.1 Confusion Matrix Explanation and its Tradeoffs 13**](#_Toc68809250)

[**4.2.2 Choice of Evaluation Metrics 14**](#_Toc68809251)

[**4.2.3 Baseline F1-Score 15**](#_Toc68809252)

[**4.2.4 Comparison of Evaluation Metrics 16**](#_Toc68809253)

[**5 PROPOSED BUSINESS RECOMMENDATIONS 16**](#_Toc68809254)

[**5.1 Recommendation 1: Application of the Proof of Concept to Actual Dataset and Identify Target Group to Assist 16**](#_Toc68809255)

[**5.1.1 Policy Level Implementation 16**](#_Toc68809256)

[**5.1.2 Ministry and Individual Level Implementation 17**](#_Toc68809257)

[**5.1.3 Subgroup Level Implementation 17**](#_Toc68809258)

[**5.2 Recommendation 2: Design of Graphical User Interface 18**](#_Toc68809259)

[**5.3 Recommendation 3: Feedback Collection for Accuracy and Longevity 18**](#_Toc68809260)

[**6 LIMITATIONS 18**](#_Toc68809261)

[**6.1 Limitations of Data 18**](#_Toc68809262)

[**6.2 Limitations of Models 19**](#_Toc68809263)

[**6.2.1 Tradeoff Between Interpretability and Predictability 19**](#_Toc68809264)

[**6.2.2 Difficulty in Predicting Social and Dynamic Systems 19**](#_Toc68809265)

[**7 FUTURE WORK 20**](#_Toc68809266)

[**7.1 Experiment with Reduced and Random Feature Selection 20**](#_Toc68809267)

[**7.2 Augmentation of Target Variable 20**](#_Toc68809268)

[**7.3 Shift in Focus to Other Aspects of Happiness Index 20**](#_Toc68809269)

[**8 CONCLUSION 20**](#_Toc68809270)

[**REFERENCES 21**](#_Toc68809271)

[**APPENDICES 22**](#_Toc68809272)

[**Appendix A: Details of the CFPB study and the FWBscore 22**](#_Toc68809273)

[**Appendix B: Questions of the Financial Well-being Scale 23**](#_Toc68809274)

[**Appendix C: Hyperparameters Tuned for all the 7 Algorithms 24**](#_Toc68809275)

[**Appendix D: Formulas of the 4 Key Evaluation Metrics 25**](#_Toc68809276)

[**Appendix E: Illustration of Graphic User Interface 26**](#_Toc68809277)

# **EXECUTIVE SUMMARY**

This report aims to identify solutions and business recommendations to help the Ministry of Social and Family Development (MSF) better identify people with poor financial well-being which tends to lead to unhappiness.

Singapore is reputed for its high Gross Domestic Product (GDP) per capita, life expectancy, and integrity. Yet, it has a relatively low happiness level at 31st ranking in the world. To improve happiness in Singapore, our team has chosen to target an aspect that the Singapore Government can likely influence and intervene in. The aspect is financial well-being in which freedom to make life choices is a significant factor.

We adopted a data-driven approach to analyze the personal data that the government collects to aggregate a more holistic judgment of one’s financial well-being through the usage of machine learning and artificial intelligence techniques. Data preprocessing steps such as data cleaning, discretization of the target variable (“At risk” and “Not at risk”), and feature selection were performed followed by data transformation before feeding into the machine learning models. Following that, we developed 7 machine learning models for binary classification: logistic regression, decision tree classifier, random forest classifier, XGBoost classifier, support vector classifier, Bernoulli naïve Bayes classifier, and neural network. A 70:30 train-test split was used to partition the data set for evaluation purposes. Hyperparameter tuning was performed on the training data set to extract the optimal performance out of the models.

The 7 machine learning models were evaluated using various evaluation metrics such as accuracy, precision, recall, and F1-Score. F1-Score was chosen as the final criteria for model selection to account for imbalanced data and the cost of misclassification. All the models managed to beat the baseline F1-Score which was calculated to be at 0.3333. Support vector classifier emerged as the winner providing the highest F1-Score at 0.7194. As such, the support vector classifier was chosen to be recommended to MSF.

A few business recommendations were proposed to utilize the machine learning model. Firstly, ideas were developed to facilitate the application of the proof of concept to the actual dataset and identify the target group to assist. Secondly, to simplify the workflow for MSF’s transition into interacting with our machine learning model, a graphical user interface was proposed. They can directly export the collected data into the system and the candidates who belong to the “At risk” group along with the relevant information will be generated automatically. Thirdly, a longitudinal study can be conducted to evaluate the efficacy of the intervention by MSF.

The limitations of the data and models were discussed. Limitations of models include the tradeoff between interpretability and predictability and the difficulty in predicting social and dynamic systems. Suggested future work includes survey questions reduction using a randomized approach to reduce survey fatigue, change of target variable to happiness or health scores, and shift of focus from freedom to make life’s choices to social support and generosity, and the other factors which performed poorly when used to determine the Happiness Index in Singapore.

# **1 INTRODUCTION**

## **1.1 Business Problem Statement**

Singapore is internationally recognized for its booming economy. We have amongst the world’s highest GDP per capita, healthy life expectancy, and integrity in government and businesses. Furthermore, according to the Programme for International Assessment (PISA), we also deliver one of the world’s best education systems. Yet, Singapore is only ranked 31st in terms of happiness according to the World Happiness Report (WHR) 2020. The top spots were secured by the Nordic countries, with Finland taking 1st place.



Figure 1 Maslow's hierarchy of needs

In Singapore today, the government has extended various sorts of financial aid to help people suffering from poverty meet their daily needs for food, shelter, and education. These factors form the basis of happiness according to Maslow’s hierarchy of needs as shown in Figure 1 ([Dr. Saul McLeod](https://www.simplypsychology.org/saul-mcleod.html), 2009). Despite these government interventions, it seems that general happiness has not increased.

Out of the six factors that contributed to the Happiness Index score, Singapore ranked 1st and 2nd in GDP per capita, healthy life expectancy, and Perception of integrity in society. However, we are only ranked 14th in freedom to make life choices which definitely can be improved (Figure 2).

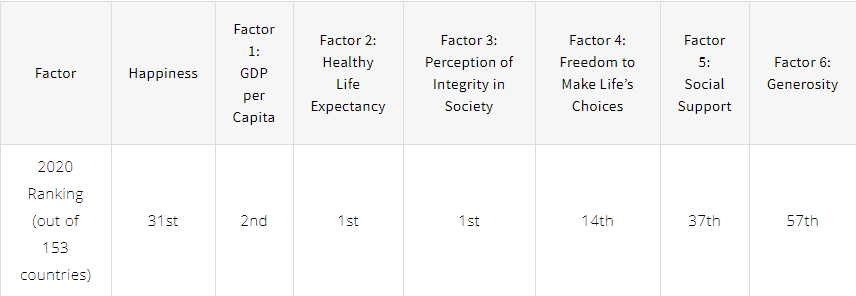


Figure 2 Singapore’s Happiness Index score (Simon Leow, 2020)

The freedom to make life’s choices can be linked to a concept known as financial well-being. According to the Consumer Financial Protection Bureau (CFPB), financial well-being is defined as a state of being wherein a person can fully meet their financial obligations, can feel secure in their financial future, and can make choices that allow them to enjoy life (and thus be happy).

A recent study by the University of Purdue (Andrew T. Jebb, Louis Tay, Ed Diener, Shigehiro Oishi, 2018) also supports the fact that happiness is impacted by financial well-being, beyond just income level and basic day-to-day living expenses. By improving a person’s financial well-being, they will be able to have greater freedom to make life choices, and hence their happiness.

Therefore, it becomes critical that we can identify people who do not have this financial freedom and are hence unhappy, yet remain undetected by current measures for financial aid from the Singapore Government.

As analytics consultants to the Ministry of Social and Family Development (MSF)/Community Care Endowment Fund (ComCare), our team aims to study the Financial Well-Being Survey dataset to identify the people who are struggling to obtain financial well-being. This allows us to assist policymakers in easier identification of people who are unhappy and highlight potential reasons behind their struggles.

## **1.2 Significance of the Business Problem**

Overall happiness is a subjective measure that is affected by a myriad of factors, thus it would be unrealistic to tackle the whole nationwide population of unhappiness. This is why we have decided to use financial well-being as a proxy for happiness as this would mean that our target group of people’s level of happiness can be partially fulfilled by financial aid.

Studies have shown that poor financial well-being can adversely impact one’s physical, mental, and social health which can lead to other negative impacts such as poorer job performance, lower productivity, and absenteeism, all of which could lead to unhappiness (Human Resources Director). The OCBC Financial Wellness Index (OCBC, 2020), which is a measurement of Singapore’s financial wellness, has dropped from 63 in 2019 to 61 in 2020, especially due to the COVID-19 pandemic. With the current pandemic, more people are likely to have poor financial well-being which may result in decreased happiness levels.

Given that Singapore’s main resource is its population, the government aims to bridge the income gap to ensure that all Singaporeans are given equal opportunities. The improvement of financial well-being will not only improve these people’s happiness but also their productivity and health.

Therefore, identifying these people with poor financial well-being will allow the government to more efficiently allocate resources such as money, time and effort, towards those who require financial aid the most to achieve greater happiness. Thus, we need a systematic approach to identify these people through machine learning and artificial intelligence techniques.

## **1.3 Expected Business Outcome**

Through the machine learning models developed by our team, we hope to improve the financial well-being and happiness of individuals in Singapore. This can be done by highlighting individuals that have low financial well-being so that more attention can be given to them, by better adjusting policies and budgeting to make them happier. Additionally, by identifying the underlying causes of their poor financial well-being, we can help to alleviate their pain points and concerns by recommending them for specialized programs to help them get back on their feet.

# **2 ANALYTICS APPROACH**

Today, government agencies have access to a great deal of personal data (both financial and non-financial data) of their citizens, such as phone and traffic records, health records (including genetic records), spending habits, water and electricity consumption, and even online browsing activities.

In Singapore, citizens receive financial aid from various Ministries (MSF, MOM, MOE) according to their per capita income as well as housing type. Yet, we know that financial well-being is unable to be captured fully through basic financial measures such as income or housing. This grey area requires a more sophisticated solution through a data-driven approach.

Using machine learning and artificial intelligence techniques, we aim to leverage data analytics to sieve through the wide array of personal data that the government collects to aggregate a better and more holistic judgment of a person’s financial well-being and hence extend financial help to those struggling to obtain financial well-being and happiness.

# **3 DATA PREPARATION AND UNDERSTANDING**

## **3.1 Data Set Description**

To ensure that our model design would be able to cater to the types of citizen profile data that would be used in real-life implementation, the dataset selected for this project was one that the team deemed to have sufficient nexus with the actual data that is likely to be available. The main criteria for a suitable dataset were that it needed to have a qualifying variable to determine financial well-being, as well as a variety of other variables commonly found in government databases.

The dataset we have selected is from a survey conducted by the Consumer Financial Protection Bureau (CFPB) in 2015. The dataset consists of 217 columns and 6395 rows. The full list of variables is provided in Appendix A. The column “FWBscore” is a discrete variable that indicates a person’s financial well-being on a scale of 0-100.

“Financial Well-Being Survey Data” from Kaggle:

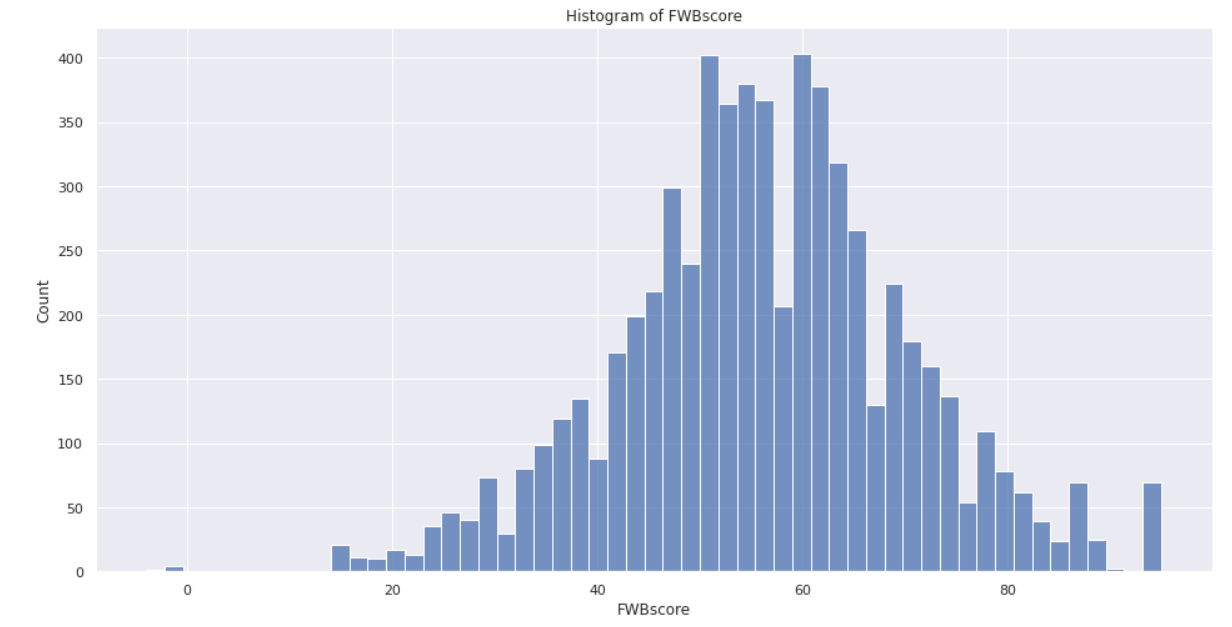
<https://www.kaggle.com/anthonyku1031/nfwbs-puf-2016-data>

The “FWBscore” scale was determined via an extensive research study involving cognitive interviews, factor testing, and psychometric testing to accurately measure a person’s financial well-being according to four key elements: (i) control over day-to-day finances, (ii) capacity to absorb financial shock, (iii) financial freedom to make choices to enjoy life, and (iv) being on track to meet financial goals. The result was a 10 question scale-based scoring system that accorded each person in the dataset a holistic and comprehensive indicator of their financial well-being. Further elaboration on the survey and methods can be found in Appendix B.

In addition to asking the 10 questions directly related to computing the FWB Score, indirect questions such as financial knowledge, education level, income and employment, family history, financial habits, demographic information were also asked in the survey.

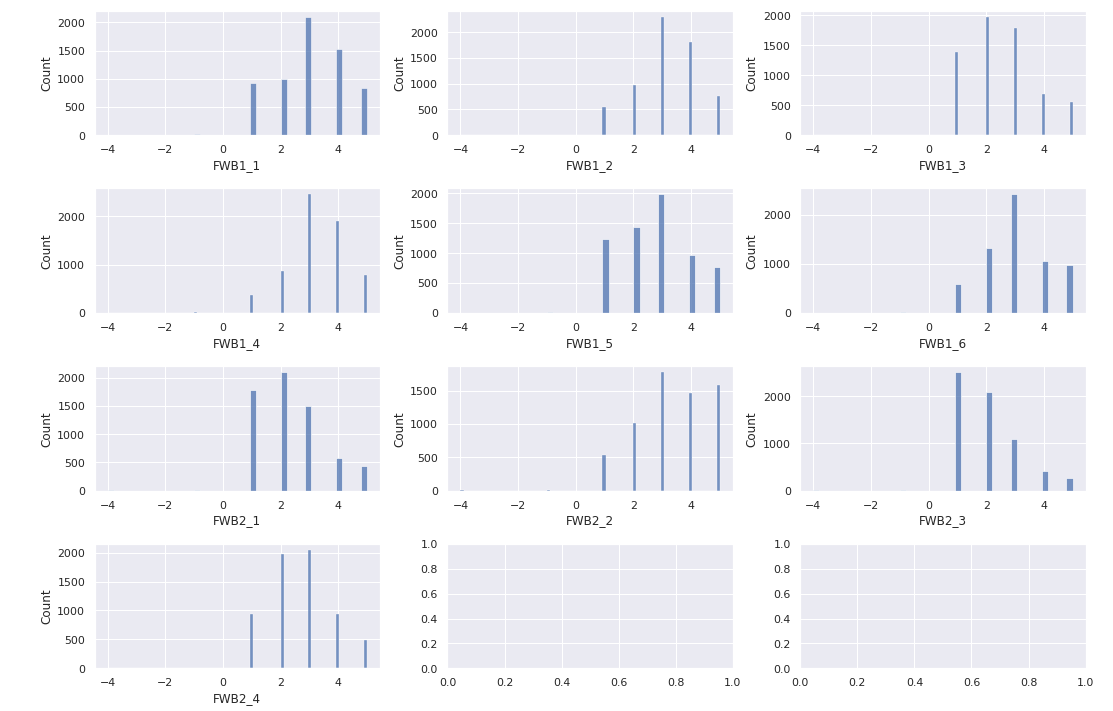
This dataset provides us with a multitude of personal data available to a government agency as well as a highly accurate proxy for actual financial well-being “FWBscore”. This would allow our team to develop predictive models and test the accuracy of utilizing personal data to predict financial well-being.

## **3.2 Exploratory Data Analysis**

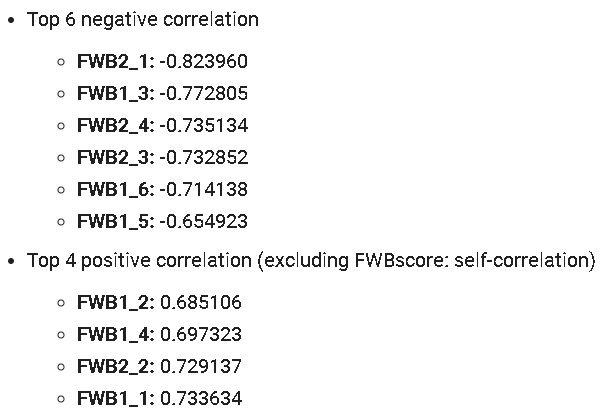


The original distribution of FWBscore is roughly normally distributed with values ranging from 14-95, mean of 56, and standard deviation of 14. This is to be expected given the extensive research that has gone into formulating the FWBscore scale as well as the relatively large sample size that has been adjusted for minority group representation. Scores related to social science applications are usually normally distributed as well.

Since the FWBscore is on a scale of 0-100, the negative values are not valid scores. They are arbitrarily encoded to represent certain statuses. Specifically, “-4” indicates that the respondent’s answer was not written to the database and “-1” indicates that the respondent refused to have the score calculated. Also, when transforming the target variable into a binary variable, later on, we may have to rebalance the dataset such that the predictive power of the model is not hampered.



The 10 features (survey questions) used to compute the target variable FWBscore were explored as well using histograms. FWB1\_4 and FWB2\_2 are concentrated towards the last 3 responses (somewhat/very well/completely). FWB1\_3, FWB2\_1, FWB2\_3 and FWB2\_4 are concentrated towards the first 3 responses (never/rarely/sometimes). Other features are roughly normally distributed with not much skew. The overall sentiment is that the respondents are quite optimistic about their finances in terms of security and freedom of choice for the present and future. However, the FWBscore's normal distribution would slightly disagree with the respondents' optimism.



All the feature correlations to FWBscore were calculated and sorted. As seen from above, the top 6 negative correlation and top 4 positive correlation are all the direct questions asked in the Financial Well-Being Survey to compute FWBscore.

## **3.3 Data Preprocessing**

### **3.3.1 Data Cleaning**

To ensure that the data set is fit for real-life decision-making purposes for financial intervention, the data itself needs to be of high quality. Without quality data, sub-optimal machine learning models might be trained and provide poor or biased predictions. Data cleaning helps to achieve this objective. This is an iterative process and generally has no end to it. We can only do it on a best efforts basis.

The original dataset we are using is already quite comprehensive and has already been cleaned by the CFPB once. Hence, our data cleaning process will mainly consist of the following steps: removal of missing values, removal of duplicates, data type conversion, and removal of rows with negative FWBscore. In summary, we removed 5 rows and 66 columns to produce a final dataset with 6389 rows and 151 columns.

#### 3.3.1.1 Removal of Missing Values

Most machine learning models are not able to internally handle missing values. A few exceptions are tree-based models such as decision trees and random forest. Therefore, the removal or imputation of missing values serves as an important data preprocessing step before feeding the data to the machine learning models. Fortunately, our data set was specifically curated for a financial well-being research program spearheaded by the Consumer Financial Protection Bureau (CFPB). Therefore, the collection of survey data was well executed and contained no missing data. No action was required on our part to remove any missing values.

#### 3.3.1.2 Removal of Duplicates

Duplicates usually refer to rows that have completely identical data. This requires that one row has the same values for every column in the exact similar order of another row. Duplicated rows might be due to erroneous repeated entries or they might be real data that happen to have identical values.

In the first case, these repeated data should be removed as they can lead to misleading results during model evaluation. They exist due to nonrandom sampling and the machine learning model will overfit to this set of repeated data. Moreover, if one of the duplicated rows is in the training set and the other row(s) are in the testing set, the model should theoretically predict these duplicated data perfectly and result in an overly optimistic test performance score.

In the second case, duplicated data that are organic can happen sometimes. This is possible in our data set which consists of survey items. However, given that there are over a hundred different questions, it should be unlikely that any 2 rows are the same.

Our approach involves checking for the column ‘PUF\_ID’ which refers to the public use file ID. This ID variable is an integer uniquely identifying each survey respondent and should not contain any duplicates. We did not find any duplicated ‘PUF\_ID’ in our data set. We accept that the second case of duplicated rows could happen for our data set and we do not attempt to remove rows where 2 respondents answered all survey questions similarly.

#### 3.3.1.3 Data Type Conversion

Data types of columns need to be converted to the right format to be appropriately and efficiently used by most machine learning models. Using the wrong data type can lead to biased interpretations by the models. The majority of data can be classified into 5 main types: numerical data, categorical data, time-series data, text data, and image data.

For our dataset, all survey items related columns will be converted to categorical data types. The unique primary key “PUF\_ID” will remain as an integer data type. The discrete score columns “FWBscore”, “FSscore”, “LMscore” will also remain as integer data types. The float score columns “KHscore”, “finalwt” will remain as float data types.

The rationale for converting survey items related columns to categorical data type are as follows. Even though the survey items are mainly composed of n-point Likert scale data (e.g. 1, 2, 3, 4, 5), they also contain an arbitrary value “-1” which indicates refusal by the respondent to answer the survey item. A way to handle this would be to remove all the “-1” values and treat the remaining data as either interval or ordinal. However, we decided to respect the respondents’ choice in refusing to answer any specific survey question. The fact that there was such an option provided by the CFPB team further reinforces its validity and ability to provide valuable information. Removal of these values might introduce bias into our models. Therefore, the data cannot be converted to interval or ordinal format as they are not strictly ranked anymore with the inclusion of the “-1” value. Our final decision would be to convert these survey items-related columns into nominal data.

#### 3.3.1.4 Removal of Rows with Negative FWBscore

As shown in the exploratory data analysis, the FWBscore is roughly normally distributed and can be represented using integer data type. More specifically, the score is only valid from the 0-100 range. Therefore, we have decided to remove 2 specific values “-4” and “-1” from the column. “-4” indicates that the response was not written to the database. “-1” indicates that the respondent refused to have his or her financial well-being scale score computed. These 2 negative values are arbitrarily assigned to represent these 2 categories of respondents’ results and hold no meaningful purpose. It makes little sense to assign these negative values to respondents as “0” is the lower bound of the FWBscore. After the removal of these negative values, only positive values remain in the FWBscore column.

### **3.3.2 Discretization of the Target Variable**

Rather than assign every respondent a Financial Well-Being score, we are more interested in identifying those with a significantly lower score that requires our help. Hence, we decided to alter the target variable (FWBscore) of our dataset from a discrete integer score of 1-100 to a binary variable which indicates those who are at financial risk and those who are not.

In Singapore, 10-14% of the population struggles with severe financial issues. Using the bottom 14% poverty line as a benchmark, we add a buffer of 6% to make up the bottom 20th percentile of population data as our “At risk” group. The remaining population will naturally belong to the “Not at risk” group requiring minimal financial intervention.

### **3.3.3 Feature Selection**

The data set that we are using consists of a rather high number of columns relative to rows. Keeping all these columns might result in the curse of dimensionality where the machine learning models are not designed to handle such high dimensionality. In addition to weaker performance, the models will also take a longer time to train depending on the algorithms used. Running time that increases exponentially with the number of columns fed to the model will cause a problem for us. Therefore, feature selection serves to remove redundant or non-informative columns from the data set.

#### 3.3.3.1 Removal of Unique ID Column

The “PUF\_ID” column is a unique primary key and provides no value for prediction purposes. No trend can be picked up by the machine learning models as every single instance has different values. There will be little to no correlation between this column and the target column. It does not help the model to differentiate between “At risk” or “Not at risk” groups.

#### 3.3.3.2 Removal of Columns Used to Adjust for Demographic and Poverty Differences

Two columns “sample” and “fpl” have been removed.

The “Sample” column contains 3 possible response values: 1, 2, 3. These correspond to the subpopulation that the respondent belongs to. The value “1” stands for the general population. The value “2” stands for the oversampled population who are 62 years old and above. The value “3” stands for the oversampled population who are underrepresented with regard to race/ethnicity and poverty.

The “fpl” column refers to the federal poverty level or poverty line. It is a measure of income issued every year by the Department of Health and Human Services (HHS). The column contains 3 possible response values: 1, 2, 3. The values “1” , “2”, “3” stand for <100% FPL, 100%-199% FPL, and 200%+ FPL respectively.

It was shown that while perceptions of financial well-being by working age and older adults were fairly consistent, minor systematic differences could still exist. Therefore, the scale development process to come up with the financial well-being score took these differences into account. The FWBscore of respondents who belong to underrepresented groups would be adjusted to arrive at a score that could be compared with all the other respondents. No separate interpretation of the score is required from the end-user.

#### 3.3.3.3 Removal of Direct Survey Questions for FWBscore

The CFPB Financial Well-Being Scale contains 10 questions which were selected through a state-of-the-art process involving a series of cognitive interviews, factor analysis, and three rounds of psychometric testing. Although these questions are rigorously selected, respondents might not always be willing to answer them as they are very specific and sensitive. A better way would be to use indirect questions to determine the respondent's FWB Score. Therefore, we will remove the features corresponding to the 10 questions used to compute the FWB Score.

#### 3.3.3.4 Removal of Secondary Survey Score Columns

There are 4 secondary survey score columns in this data set: “FSscore”, “LMscore”, “KHscore”, “finalwt”. These scores are derived from the related set of survey items. For example, FSscore is derived from the columns prefixed with “FS” which are the questions asked to compute the score itself. As a result, the score would have a very high correlation with the questions. Keeping just the columns representing the questions would be more appropriate as they are reliable predictors of the final score computed.

Moreover, excluding these secondary survey scores as predictor variables would be realistic when the machine learning model is deployed in real life. To calculate these scores, the full set of questions related to the specific score have to be asked. This would limit the efficacy of the feature selection process as we would be forced to retain a lot of variables that might not be that useful to predict our target of interest, which is the FWBscore. Therefore, we would only retain FWBscore as the target variable while all the other scores would not be used as the predictor variables.

#### 3.3.3.5 Removal of Survey Items with Incomplete Base

Survey items with incomplete bases are removed. The word “Base” here refers to who was asked for each survey item. For example, the column “VALUERANGES” corresponds to the following question: “If you were to sell your home today, what do you think it would be worth?”. The base for this question consists of those who answered “1” to another question: “Which one of the following best describes your housing situation”. The value “1” indicates that these people own their homes instead of renting or do not currently own any housing at all. Therefore, we only retain features where Base = "All" so the questions apply to any respondent and not only a subset of those who answered relevant preceding questions.

#### 3.3.3.6 Removal of Survey Items with Sensitive, Vague, or Region Specific Questions

Survey items with sensitive, vague, or region-specific questions are removed.

Sensitive questions tend to be avoided or answered less truthfully. For example, the column “HOUSERANGES” corresponds to the question: “About how much do you pay for your home each month?”. This can be rather sensitive as it requires providing a range of values. Moreover, this number can also vary widely depending on which country the respondent is from. Given that we intend to use these data to formulate a model to be used in Singapore, this question is not apt. The other columns relate to questions such as the amount of savings, age, generation, highest education received, gender, household income, and marital status. Our team is focused on minimizing the usage of such personal or sensitive questions. Instead, we prefer questions that are more subjective and encourage the respondents to answer the questions more truthfully.

Vague questions are broad, ill-defined and generally fail to focus on a specific topic. For our data set, the column “CONNECT” requires the respondents to rate themselves on a scale of 0-100 regarding how they perceive their psychological connectedness. We deem this survey item to be vague and deviate from the other survey items which are generally focused on asking a specific question instead of a score directly. Moreover, the score has 101 levels for the related column and will result in high dimensionality problems when performing feature encoding.

Region-specific questions are survey items that can only be accurately answered by those who live in a specific region. This data set focuses on the financial well-being of those residing in the United States. Therefore, some survey items can only be answered by those who live there. For example, the “MILITARY” column corresponds to the statement: “Current/former member of US Armed Forces or spouse/dependent of service member”. Since we intend to extend this survey to Singaporeans, this column will not be relevant to them. Other columns relate to survey items requiring information on race/ethnicity, metropolitan statistical area (MSA) status, census region, and census division. Our team removed all these columns to make the survey more location agnostic.

#### 3.3.3.7 Removal of Survey Items with Specific Financial Knowledge Questions

There are a total of 12 specific financial knowledge questions in this data set and another 12 related columns documenting whether the respondent answered the questions correctly. The questions were presented as multiple-choice questions. The topics were related to compound interest, inflation, return on savings, stocks, mutual funds, and a variety of other topics. Our team prefers a more subjective information-seeking survey rather than an objective one. Those financial knowledge questions are testing for specific factual knowledge which does not align with our survey goals. Therefore, we removed these columns.

### **3.3.4 Data Transformation**

Machine learning models implemented using Scikit-Learn and Keras require all the X features and y labels to be numeric. This means that categorical columns need to be cast into some form of numbers before the machine learning models can be trained. Given that we have established the fact that the X features are considered nominal data instead of interval or ordinal, a few nominal encoding techniques are viable.

A popular technique used is one-hot encoding. It is suitable for nominal data where no relationship exists between different levels. For every unique label in each column, one-hot encoding creates a new binary column for that label. Using the column “SWB\_1” as an example, it has 9 unique labels: -4, -1, 1, 2, 3, 4, 5, 6, 7. Therefore, 9 new columns (dummy variables) will be created to represent these unique labels. If a row contains “7” as a value for “SWB\_1”, the new dummy variable representing “7” will be filled with the value 1 while all the other dummy variables will be filled with the value 0.

Dummy encoding is similar to one-hot encoding but it aims to avoid the issue of dummy variable trap. It is a scenario where 1 independent variable can be predicted with the help of other independent variables. This leads to multicollinearity where there is a dependency between the independent variables. Using the same column “SWB\_1” as an example, one-hot encoding will result in 1 dummy variable with the value 1 and the others 0 for every row. This means that all the dummy variables are perfectly correlated with each other. As long as we know 1 dummy variable is turned on (value = 1), the other related dummy variables for the column are assumed to be turned off (value = 0). Multicollinearity is a problem for machine learning models like linear regression and logistic regression where highly correlated variables result in unreliable and unstable estimates of regression coefficients. Dummy encoding drops one of the dummy variables to avoid the dummy variable trap. The dropped column implicitly acts as the reference group and the coefficients of the remaining dummy variables are interpreted with respect to this reference group.

Binary encoding initially converts the categorical feature into a numeric data type. The numbers are then converted into binary numbers. Each binary value (0/1) in the binary number forms a new binary column. The row with the longest binary number dictates how many new binary columns are needed to encode the feature. This ensures that all the bits of that row can be represented.

The three nominal encoding techniques described above are considered by our team and we decide to move ahead with binary encoding. Firstly, one-hot encoding is unsuitable for logistic regression due to the dummy variable trap which is one of the machine learning models that we intend to use for our classification task. Secondly, one-hot encoding will result in a very sparse matrix as our survey item columns are of high cardinality. The sparse matrix contains a lot of zero values and wastes considerable space. Thirdly, both one-hot encoding and dummy encoding create many binary columns. This can result in high dimensional data where the number of features can exceed the number of observations and lead to the curse of dimensionality. Binary encoding serves as a middle ground where it is more memory efficient than the other encoding schemes.

For the label (“FWBscore”), a simple label encoding scheme was used. The “At risk” group was represented with the value “1” while the “Not at risk” group was represented with the value “0”. No assumption of order is made with regard to the numbers.

### **3.3.5 Data Partitioning**

Data partitioning is required for a train-test split evaluation of our machine learning models. This helps to prevent the problem of overfitting. If we train and evaluate our models on the same data, the models would have seen all the data and achieve overly optimistic performance scores. A separate set of data is required to independently evaluate our models on unseen data for a robust estimate of their classification accuracies. The procedure requires splitting our data set into 2 subsets: training dataset and testing dataset. For the split ratio, we use the common 70:30 train-test split. This means that 70% of the data will be used for training the machine learning models while 30% of the data will be used to evaluate the performance of the models.

# **4 MODEL DEVELOPMENT AND EVALUATION**

## **4.1 Model Development**

### **4.1.1 Choice of Machine Learning Algorithms**

The choice of machine learning algorithms to use depends on the business problem and analytics approach. Given that there is a need to predict which candidates belong to the “At risk” or “Not at risk” group and the labels already exist through the CFPB survey, a supervised learning approach would be appropriate. Furthermore, the prediction problem belongs to that of a classification task since we are predicting categories instead of continuous values. Therefore, the machine learning algorithms chosen should be able to support classification tasks.

There are countless machine learning algorithms developed over the years and all of them have their advantages and disadvantages. Generally, there is a tradeoff between interpretability and predictability. Highly interpretable models tend to be linear with well-defined relationships and easy to compute. Models with high predictive power tend to have non-linear relationships and high computation time. Algorithms can also be tree-based, kernel-based, or distance-based. Ensemble methods such as bagging and boosting can be applied to a base model for improved performance as well, usually reducing bias or variance.

To balance between exploring sufficient machine learning algorithms and avoiding excessive computation time, a total of 7 different algorithms are chosen. They are logistic regression, decision tree classifier, random forest classifier, XGBoost classifier, support vector classifier, Bernoulli naive Bayes classifier, and neural network.

### **4.1.2 Hyperparameter Tuning**

Hyperparameters have to be pre-defined before training any machine learning model as they cannot be learnt during training time. [Hyperparameter tuning](https://towardsdatascience.com/hyperparameter-tuning-c5619e7e6624) seeks to determine the optimal combination of hyperparameters that allows the machine learning model to maximize its performance. The process of choosing the optimal combination of hyperparameters is not a straightforward task. A balance needs to be struck between searching for a wide range of hyperparameters and minimizing search time. There are multiple ways to tune hyperparameters. The two most popular ways are grid search and randomized search.

Grid search creates a grid of possible values for the hyperparameters. Each search iteration attempts a different combination of hyperparameters in a specific order. It fits the model on all the possible combinations of hyperparameters as defined by the user. If there is a huge combination of hyperparameters, computation time will be very long.

R[andom search](https://www.jmlr.org/papers/volume13/bergstra12a/bergstra12a.pdf) also creates a grid of possible values for the hyperparameters. However, each search iteration tries a random combination of hyperparameters from the grid in no specific order. It fits the model for a number of combinations as decided by the user. If 100 iterations are chosen, the searching process is terminated after 100 times even though the entire grid has not been searched yet. The obvious advantage of random search is the faster computation time. A disadvantage is a possibility that the optimal set of hyperparameters might not be discovered.

A mix of both methods will be utilized for the 7 machine learning algorithms. Grid search will be used for algorithms with only a few hyperparameters such as logistic regression, support vector classifier, and Bernoulli naive Bayes classifier. Random search will be used for the other algorithms with more hyperparameters such as decision tree classifier, random forest classifier, XGBoost classifier, and neural network.

For both grid search and random search, the number of cross-validation trials for each selected combination of hyperparameters will be set at 3. This is to maximize the data that is used for evaluating each combination of hyperparameters and minimize the variance. For the random search, the number of iterations will be set at 100. Refer to Appendix C for the hyperparameters tuned for all 7 algorithms.

## **4.2 Model Evaluation**

### **4.2.1 Confusion Matrix Explanation and its Tradeoffs**

|  |  |  |  |
| --- | --- | --- | --- |
| **True Label** | **Not at risk**  **(0)** | True Negatives  (TN) | False Positives  (FP) |
| **At risk**  **(1)** | False Negatives  (FN) | True Positives  (TP) |
|  | | **Not at risk**  **(0)** | **At risk**  **(1)** |
| **Predicted Label** | |

A confusion matrix can be thought of as a contingency table for evaluating the classification performance of a machine learning model. Essentially, it is a N x N table or matrix where N is the number of classes we are trying to predict. For our data set, there are 2 classes: “At risk” and “Not at risk” encoded by 1 and 0 respectively. Therefore, the value of N is 2 and the table will have 2 rows and 2 columns with a total of 4 values.

For this data set, the true label refers to the correct answer of “FWBscore” in the test data set. It is the ground truth used for evaluating the machine learning models. The predicted label is the prediction output by the machine learning models. The prediction will be either 0 or 1 since it is a binary classification problem. A comparison will be made between the true label and predicted label for each row in the test data set. Various calculations can be made with the comparisons. Specific to the confusion matrix, 4 metrics can be easily calculated. They are true positives, true negatives, false positives, and false negatives.

True positives are predicted values (“At risk”) correctly predicted as actual positives (“At risk”). True negatives are predicted values (“Not at risk”) correctly predicted as actual negatives (“Not at risk”). False positives are predicted values (“At risk”) incorrectly predicted as actual positives (“At risk”). False negatives are predicted values (“Not at risk”) incorrectly predicted as actual negatives (“Not at risk”).

True positives and true negatives are correctly classified predictions. If both true positives and true negatives are both 1 on a normalized scale, the machine learning model is perfect and made no prediction errors at all. For this project, it means all the “At risk” candidates are successfully detected and further financial intervention can be planned for them. There are also no “Not at risk” candidates misclassified as “At risk”. This ensures that all the resources provided by MSF are not wasted on candidates that do not require help. False positives and false negatives are incorrectly classified predictions. For this project, false positives and false negatives are both important and have high costs associated with them. As mentioned earlier, false positives generally take up resources that could be provided to those “At risk” candidates who need help. False negatives mean that the workers at MSF would fail to detect the “At risk” candidates and early financial intervention cannot be provided to them. Therefore, a balanced approach is required to reduce both false positives and false negatives to an acceptable level.

### **4.2.2 Choice of Evaluation Metrics**

The choice of evaluation metrics is an important decision to make when evaluating the trained machine learning models. Our team will consider 4 key evaluation metrics commonly used by practitioners. They are accuracy, precision, recall, and F1-Score.

Accuracy is probably the most common evaluation metric used as it is simple to calculate and interpret. It is essentially the proportion of predictions the machine learning model predicted correctly. It is bounded within a range of 0 to 1 with 0 being the worst score and 1 being the best score. Generally, the accuracy score is used when the true positives and true negatives are more important and the class distribution is balanced.

Precision is the proportion of true positives out of those total predicted positives. It is a suitable evaluation metric when the costs of false positives are high.

Recall is the proportion of true positives out of those total actual positives. It is a suitable evaluation metric when the costs of false negatives are high.

F1-Score is the harmonic mean of precision and recall. It is a suitable evaluation metric when the costs of false positives and false negatives are both high. For imbalanced data sets, F1-Score is also better than the accuracy score. It ignores true negatives in its calculation. A data set with a high proportion of one class tends to have an inflated accuracy score as the machine learning model can simply predict the majority class all the time.

Given that the data set is imbalanced with an 80% negative class and both false positives and false negatives are relatively important, our team will use F1-Score as the evaluation metric for choosing the final machine learning model. However, all 4 evaluation metrics will be calculated to examine the tradeoffs between different machine learning models. Refer to Appendix D for the formula of these 4 key evaluation metrics.

### **4.2.3 Baseline F1-Score**

A baseline F1-Score acts as a lower bound to evaluate the performances of the machine learning models. The baseline model will be defined as the dummy predictor where it will predict the candidate to be “At risk” or 1 all the time. A trained model is usually expected to perform better than the baseline model. Otherwise, there is no reason to use the model.

Given that the baseline model predicts 1 all the time, the precision will be the marginal probability P(y = 1) which equates to 0.2. This is because 20% of the labels belong to the positive class and they will be classified correctly. The other 80% of the labels belong to the negative class and they will naturally be classified wrongly. Specifically, the true positives will be 0.2 and the false positives will be 0.8. Plugging these values into the precision formula will result in the value 0.2.

On the other hand, the recall will be equal to 1. Since the baseline model predicts 1 all the time, it can find out all the “At risk” candidates. It will never predict 0 or “Not at risk”. Therefore, false negatives will not exist and equate to 0 all the time. Specifically, the true positives will be 0.2 and the false negatives will be 0. Plugging these values into the recall formula will result in the value 1.

With the precision and recall values for the baseline model, we can calculate the baseline F1-Score. The calculation is shown below.

|  |
| --- |
| Baseline F1-score = 2 \* (precision \* recall) / (precision + recall)  Since precision = 0.2 and recall = 1,  Baseline F1-score = 2 \* (0.2 \* 1) / (0.2 + 1)  Baseline F1-score = 0.4 / 1.2 = **0.3333** |

Therefore, the benchmark F1-Score is 0.3333 and the model is required to beat this score.

### **4.2.4 Comparison of Evaluation Metrics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Accuracy** | **Precision** | **Recall** | **F1-Score** |
| **Logistic Regression** | 0.8821 | 0.7477 | 0.6279 | 0.6826 |
| **Decision Tree Classifier** | 0.8086 | 0.5158 | 0.8424 | 0.6398 |
| **Random Forest Classifier** | 0.8472 | 0.5816 | 0.8656 | 0.6957 |
| **XGBoost Classifier** | 0.8811 | 0.6916 | 0.7416 | 0.7157 |
| **Support Vector Classifier** | 0.8649 | 0.6194 | 0.8579 | 0.7194 |
| **Bernoulli Naive Bayes Classifier** | 0.8279 | 0.5494 | 0.8191 | 0.6577 |
| **Neural Network** | 0.8816 | 0.7778 | 0.5788 | 0.6637 |

The evaluation metrics table above details the accuracy, precision, recall, and F1-Score for all 7 trained machine learning models. All the scores were evaluated based on the test data set. Logistics regression has the highest accuracy at 0.8821. The neural network has the highest precision at 0.7778. Random forest classifier has the highest recall at 0.8656. The support vector classifier has the highest F1-Score at 0.7194.

As observed, all the machine learning models managed to beat the baseline F1-Score of 0.3333. Given that we will be using the F1-Score as the evaluation metric to discriminate between the best machine learning models, the final model chosen by our team will be the support vector classifier.

# **5 PROPOSED BUSINESS RECOMMENDATIONS**

## **5.1 Recommendation 1: Application of the Proof of Concept to Actual Dataset and Identify Target Group to Assist**

In line with our objectives outlined in Section 1, we believe that our solutions can help policymakers identify the target group of people that are “At risk” of low financial well-being leading to unhappiness. Input variables can be changed to better suit the local context and include any available information that the SingPass database may already contain: MediSave coverage, ethnocultural details, to even BTO status. Once these “At risk” people are identified, the government can offer financial aid tailored to their needs.

### **5.1.1 Policy Level Implementation**

Our findings can impact policy decisions on many fronts and across different ministries. Our team believes that the greatest benefit can be brought about through budget reallocation in a strategic manner towards this target group. We want to be able to improve their financial well-being and happiness without compromising the welfare of others who still require finances for basic needs. This will translate to re-budgeting on areas outside of ComCare.

One of the current measures the Singapore Government has introduced is the SG bonus. This occurs when the Government has a budget surplus for the fiscal year and decides to issue a one-time bonus to all Singaporean adults above 21. The last incident of this was in 2018, in which all Singaporean adults received $100-300 with no income cap.

However, not everyone may elicit the same response to this measure, especially for those who are better off in terms of financial well-being which may not have much impact on them. Someone who makes $100,000 a year may not appreciate that extra $100 as much as someone who makes $20,000 a year.

As such, the Singapore Government could optimize the effectiveness of the SG bonus by channeling the funds towards the target “At Risk” group whose happiness can be impacted by more financial measures. This will allow them to give out these bonuses to those who require it more rather than giving it out to everyone when some people do not even require it.

### **5.1.2 Ministry and Individual Level Implementation**

After identifying the “At risk” individuals, the Ministry of Social and Family Development can assign dedicated financial planners to help these people plan their finances properly. While bonuses payout and fine-tuned policies may be able to help the “At risk” individuals in the short term, proper financial planning is required so that these people will be able to eventually achieve the financial freedom required to make their own life choices. Having a group of financial planners specialized in helping these “At risk” individuals will better aid them in the long term as the financial planners are well-versed in the various support schemes that the “At risk” individuals can apply for to achieve financial freedom. These volunteers can be sought from various financial institutions as a form of pro-bono work.

### **5.1.3 Subgroup Level Implementation**

People who are identified with a low FWB score are bound to show similar traits. Some of these traits include poor financial knowledge, pessimism, inability to pay utility bills, and a poor outlook on life. As such, focused solutions need to be tailored to bridge the gap between this “At risk” group and the general population. This will raise their FWB score, suggesting greater freedom to make life choices and hence increase the happiness index of Singapore.

In the case of individuals with poor financial knowledge, a recommendation would be to provide classes to educate them on financial terms, investment strategies, saving habits, and various bank schemes. This relevant knowledge would allow them to effectively manage their finances and directly improve their FWB score.

As both pessimism and a poor outlook on life are aspects of mental health, the government should schedule appointments with psychologists. This provides an avenue to talk about their problems and develop a better mindset, indirectly improving their FWB score.

Lastly, individuals who are unable to pay their utility bills can be offered alternative options such as alternative billing methods, taking a loan, or extending the deadline for payment, which may potentially improve their ability to manage their finances and hence their FWB score.

## **5.2 Recommendation 2: Design of Graphical User Interface**

Although the solution is functional on its own, we recognize that the users of this solution (officers of MSF) may have little to no experience with machine learning modeling or programming. Hence, we can both improve user experience as well as enhance efficiency by automating the modeling process and implementing it via a graphical user interface (illustrated in Appendix E).

The initial population database will be implemented with our aid. However, we have designed a graphical user interface to allow the officers to key in new entries, i.e. new converted citizens, into the system. The system will go through the same data preprocessing and machine learning modeling process on the entire new database as mentioned above and identify individuals that belong to the “At risk” group again. The system will then generate a list of individuals and their respective particulars into a document for the officers.

Officers can also go back in the system to update the status of financial aid for the “At risk” group, and track the progress and effectiveness of our intervention, much like a doctor’s patient log. This allows for the database to remain relevant even many years later.

## **5.3 Recommendation 3: Feedback Collection for Accuracy and Longevity**

Before extending help to the identified target persons, policymakers can engage them in a general survey to collect data on their general happiness level. After each year of financial aid, evaluation of the usefulness of the solution can be performed by getting these target individuals to gauge their happiness. This will be similar to a longitudinal study where we observe the same individual over different periods to assess the efficacy of our targeted financial intervention.

# **6 LIMITATIONS**

While we are confident in our models’ abilities to predict one’s financial well-being, there are some limitations.

## **6.1 Limitations of Data**

Firstly, the collected data reflects the mindset of US citizens. Translating the results to reflect the situation of Singaporeans may not be accurate as there could be a vast difference between how the two groups interpret the various questions due to the contrasting culture and background.

Secondly, the utilization of the Likert scale amplifies this ambiguity and leaves plenty of room for broad interpretation. For example, despite being in the same financial situation, what one person may consider as a luxury good may not be the same for the other person, causing these individuals to score different ratings on the scale.

Thirdly, respondents may have different interpretations of the definition of financial well-being. Hence, this may affect the “correctness” of their scoring scale. Furthermore, since this is a self-administered survey, there may be some limitations on the credibility of the scores that they provide.

As such, the above factors may affect the accuracy of our models due to the limitations of data in terms of various interpretations and the level of truthfulness provided by respondents.

## **6.2 Limitations of Models**

### **6.2.1 Tradeoff Between Interpretability and Predictability**

The optimal course of action would be to predict which individuals are financially risky and identify the underlying reasons for each individual. When these cases are highlighted to the people working at MSF, they will understand better how to assist these individuals through proper financial planning and lifestyle changes.

However, as with every machine learning model, there is a trade-off between interpretability and predictability. In our case, we place a greater emphasis on models that can accurately predict whether an individual is at financial risk. This means we are not able to identify the specific cause(s) of financial risk for each case as highly accurate models tend to be black box in nature with low interpretability.

We have chosen this route as we believe predicting correctly is more crucial. These individuals may be in dire circumstances and desperately require aid. Identifying the cause(s) is secondary and can be left to be determined by MSF who possess the domain expertise to ensure that these individuals are well taken care of. Overall, our team has adopted a human-in-the-loop approach where the relevant personnel has to interact with the machine learning models for the optimal action to take.

### **6.2.2 Difficulty in Predicting Social and Dynamic Systems**

Predicting human behavior or scores related to human characteristics such as FWBscore is an inherently difficult task. Individuals tend to exhibit random or unpredictable behavior as compared to physical systems such as machinery and equipment. When a social system we hope to predict constantly changes, the validity of prediction remains a question mark. A more challenging aspect of this project is that the policymakers from MSF will interact with the machine learning predictions and act on the respondents accordingly. We are dealing with a dynamic system where the respondents’ reactions to finding out that they belong to the “At risk” group will probably change their behavior. As a result, this might partially invalidate the machine learning model’s future prediction. Unlike weather forecasting where the prediction won’t affect the weather itself, the prediction of dynamic systems such as an individual’ FWBscore will create a feedback loop due to the respondent’s reaction. Public policymakers at MSF will have to grapple with such issues when deploying the machine learning model.

# **7 FUTURE WORK**

## **7.1 Experiment with Reduced and Random Feature Selection**

Despite performing the feature selection phase, there is still a high number of features relative to the number of samples. This can result in the curse of dimensionality and overfitting issues. To resolve these issues, a subset of the features (e.g. 10%) can be randomly selected to train the model. This serves to reduce the computation time during the training process.

An experiment can also be designed to test whether any specific features (survey questions) are important in predicting the FWBscore class. The experiment will be run multiple times where each trial uses a different combination of features. If the evaluation metric does not fluctuate excessively, it might indicate the possibility that the survey questions asked need not be extremely specific as long as sufficient questions are asked. Using fewer questions also helps to combat survey fatigue when the respondents become bored, tired, or uninterested in the survey and begin to perform at a substandard level.

## **7.2 Augmentation of Target Variable**

Currently, the four elements of financial well-being are significantly related to monetary capabilities such as the ability to purchase goods and pay down debts. While important, these represent an incomplete or misleading picture regarding the role that money plays in our lives. Money is treated as the main factor in determining financial well-being based on the four elements. A change can be made where money plays more of a supporting role rather than the leading role to explore the other factors that impact happiness. The machine learning problem can be augmented where instead of predicting the FWBscore, other scores such as happiness or health can be predicted. The hypothesis can be tested regarding the utilization of money as a moderating factor, not the main factor, for the happiness or health of a person. Additional data which are less monetary-focused can be collected such as life experience-based questions. This can shed light on whether an individual’s joy and contentment come from leading a fulfilling life with experiences that may or may not be attained with money.

## **7.3 Shift in Focus to Other Aspects of Happiness Index**

Finally, we can shift our focus from freedom to make life’s choices to social support and generosity, and the other factors which performed poorly when used to determine the Happiness Index in Singapore.

# **8 CONCLUSION**

In conclusion, the machine learning model facilitates the data-driven approach to identify people in Singapore who might experience higher levels of unhappiness due to poor financial well-being. We accomplish this by first obtaining a complete and cleaned dataset through various data preprocessing steps. The processed and transformed data was fed into the machine learning models. The support vector classifier performed the best based on F1-Score and was recommended as the model of choice for the Ministry of Social and Family Development (MSF).

Through the data-driven approach, our solution overcomes human biases and provides accurate and timely information for better decision-making. In the short and medium term, the machine learning model would help MSF to better identify citizens who have poorer financial well-being which leads to unhappiness, and provide them aid. Fine-tuning of policies serves as a long-term measure to improve the happiness index of Singapore.

# **REFERENCES**

Andrew T. Jebb, Louis Tay, Ed Diener, Shigehiro Oishi. (2018). LETTERShttps://doi.org/10.1038/s41562-017-0277-0© 2018 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.1Department of Psychological Sciences, Purdue University, West Lafayette, IN, USA. 2Department of Psychology, University of Vi. Retrieved April 5, 2021, from https://www.nature.com/articles/s41562-017-0277-0.epdf

Dr. Saul McLeod. (2009, December 29). Maslow's Hierarchy of Needs. Retrieved April 5, 2021, from https://www.simplypsychology.org/maslow.html

Human Resources Director. (n.d.). Why is financial wellness so important? Retrieved April 5, 2021, from https://www.hcamag.com/asia/specialisation/financial-wellness/why-is-financial-wellness-so-important/178498

OCBC. (2020). The OCBC Financial Wellness Index 2020 is 61, down 2 points from last year. Retrieved April 5, 2021, from https://www.ocbc.com/simplyspoton/financial-wellness-index.html

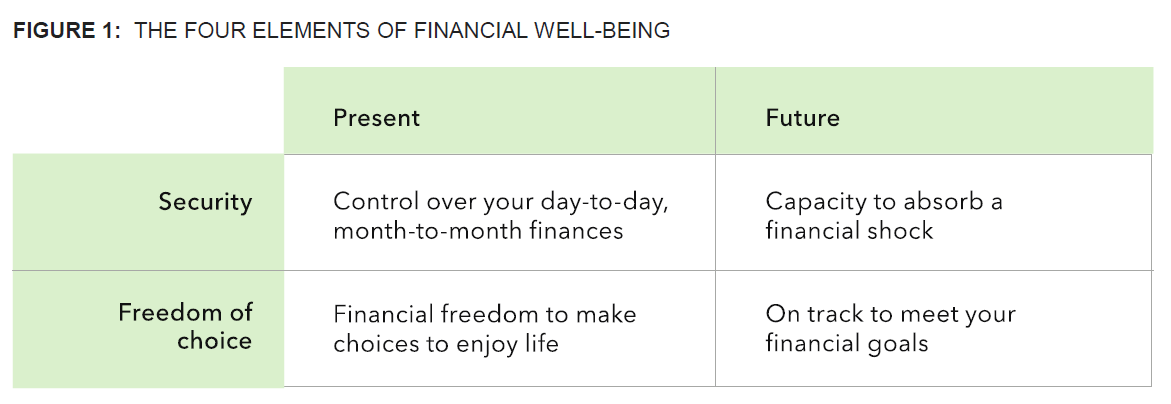
Simon Leow. (2020, July 8). Why Is Singapore Not Happier? Retrieved April 5, 2021, from https://happinessinitiative.sg/why-is-singapore-not-happier/

# 

# **APPENDICES**

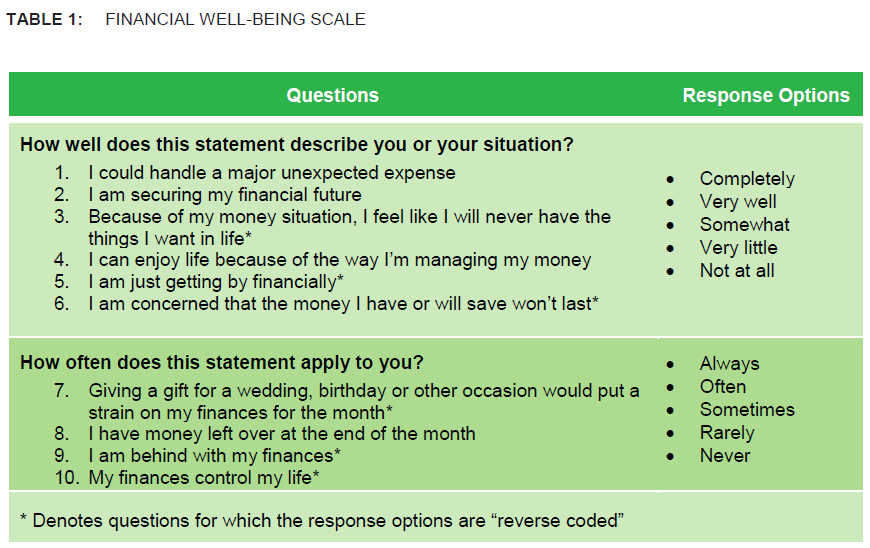
## **Appendix A: Details of the CFPB study and the FWBscore**

In 2015, the CFPB led a rigorous research effort to develop a consumer-driven definition of financial well-being. They tested a set of questions – a “scale” – to measure financial well-being. The scale is designed to allow practitioners and researchers to accurately and consistently quantify, and therefore observe, something that is not directly observable – the extent to which one’s financial situation and financial capability that they have developed provide them with security and freedom of choice.



The development of the scale involves asking the respondents questions that are directly related to their financial well-being and utilizes an item theory response (IRT) scale to provide a financial well-being score (FWB Score). The FWB Score is a standardized number between 0 and 100 that corresponds to the respondent’s estimated level of financial well-being. Knowledge of this score can provide an estimation of whether an individual is susceptible to potential financial risk or is already facing one. It should be noted that the FWB Score is arbitrary and has little meaning when interpreted on its own. The score will have more context when compared longitudinally for the same respondent in determining the change in perception of financial well-being. The majority of the score will lie in the middle similar to a normal distribution. Extreme scores close to the lower or upper limit will be quite rare.

## **Appendix B: Questions of the Financial Well-being Scale**

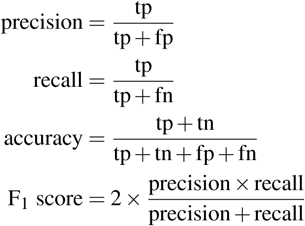


## **Appendix C: Hyperparameters Tuned for all the 7 Algorithms**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Logistic Regression** | **Decision Tree Classifier** | **Random Forest Classifier** | **XGBoost Classifier** | **Support Vector Classifier** | **Bernoulli Naive Bayes Classifier** | **Neural Network** |
| C | criterion | n\_estimators | max\_depth | C | alpha | epochs |
| class\_weight | max\_depth | criterion | learning\_rate | kernel |  | dense\_nparams |
|  | min\_samples\_split | max\_depth | n\_estimators | gamma |  | init |
|  | min\_samples\_leaf | min\_samples\_split | booster | class\_weight |  | batch\_size |
|  | max\_features | min\_samples\_leaf | gamma |  |  | dropout |
|  | min\_impurity\_decrease | max\_features | min\_child\_weight |  |  |  |
|  | class\_weight | min\_impurity\_decrease | subsample |  |  |  |
|  |  | class\_weight | colsample\_bytree |  |  |  |
|  |  |  | colsample\_bylevel |  |  |  |
|  |  |  | colsample\_bynode |  |  |  |
|  |  |  | reg\_alpha |  |  |  |
|  |  |  | reg\_lambda |  |  |  |
|  |  |  | scale\_pos\_weight |  |  |  |

## 

## **Appendix D: Formulas of the 4 Key Evaluation Metrics**



## **Appendix E: Illustration of Graphic User Interface**

