University of Essex
Department of Mathematical Sciences

MA981: DISSERTATION

Migraine Attack Prediction

Khwanchanok Chumkhun

Supervisor: Dr Felipe Maldonado

December 9, 2022
Colchester

Contents

List of Figures

List of Tables

Abstract

Migraine is a headache disorder that affects multiple aspects of a patient's life, including employment, education, interpersonal relationships, and mental health. There are methods for preventing and reducing migraine symptoms, especially in chronic migraines, despite the lack of knowledge about the exact causes of migraine attacks. It is crucial to determine the migraine prognosis in advance to minimize all migraine-related impacts. By implementing Random Forest on the Smartphone Headache Diary (SHD) dataset, the study shows that migraine can be predicted in advance before the attack with an accuracy of 95.13%, 95.56% of precision, 94.66% of recall, and 95.11% of F1-score. An AUC score of 0.9845 indicates that the model distinguishes between migraine attacks and non-migraine attacks. Among migraine triggers, stress is the most influential, whereas smoking is the least influential. Further studies can be conducted to improve the accuracy of the prognosis and to be able to predict migraine attacks earlier based on the results of this study.

CHAPTER 1

Introduction

Originally, this project was a collaboration with Happyr Health, an organization that supports teens with migraines by matching them with behavioural exercises. This project aims to study potential correlations between migraine attacks and their triggers, such as low mood, stress level, heart rate, and so on. But because the data for this work cannot be provided by the company, this project was cancelled. However, an alternative solution was provided, so this dissertation is able to continue. The published dataset called the Smartphone Headache Diary (SHD) that aims at the same goal as this project is analyzed in this paper. Since the data is in Korean and disorganised, translation, which is one of the challenging parts of this work, and data manipulation are mandatory before starting any investigation. So this study can be conducted using this SHD dataset as an alternative solution.

According to the World Health Organization, migraine is a primary headache disorder that develops when a deep-brain mechanism is triggered, inflammatory substances are released near nerves and blood vessels inside the head, causing pain [18]. Migraines are characterized by moderate to severe headaches, usually on one side of the head, and are mostly accompanied by nausea and sensitivity to light and sound [14]. Attacks usually last from a few hours to two or three days and occur anywhere between once a year and once a week. Attacks usually last less time and show more abdominal symptoms in children. The most common age range for migraine sufferers is between 35 and 45 years old. Due to hormonal influences, women typically experience a 2:1 increase in migraine risk compared to men [18]. Migraine differs from common headaches and tension headaches in that, despite the fact that both cause pain, migraines are neurological conditions that are real medical conditions, much like asthma, diabetes, or epilepsy [1].

An international classification system has defined the most common types of migraine as migraine without aura and migraine with aura. A migraine without an aura is the most common type of migraine. There is no warning sign that a migraine attack is about to begin for this migraine type. In about one out of every three migraine sufferers, there is an aura associated with the migraine. Most frequently, the warning indicator is a visual symptom, such as blind spots or seeing flashing lights [27].

Migraine is more than just a headache, it has an impact on many facets of a person's life, including employment, education, interpersonal relationships, and mental health. According to research conducted by The Migraine Trust, 71% of migraineurs feel that their mental health has been significantly impacted, and 60% of migraineurs think that their relationship with their spouse or partner has been significantly affected. One in ten people will suffer from a migraine, which affects between 12% and 15% of their lives [1]. A total of 43 million days from work and school are predicted to be lost annually by people in the UK due to migraines [26]. It was reported in a paper published in 2006 titled "The Economic Cost of Brain Disorders in Europe" that migraines cause the European economy to lose €27 billion annually due to lost productivity and working days [1]. It is still unclear why migraines occur, but it is believed that they result from abnormal brain activity, which alters the chemical balance of the brain, the function of blood vessels, and nerve signalling. Although the exact cause of these changes in brain activity is unknown, it is possible that certain genes predispose to migraines [5].

To reduce economic losses and migraine patient suffering, it is essential to examine the cause and method of migraine cure. In spite of the fact that there is no known cure for migraines yet, there are a variety of ways to prevent them and reduce their symptoms, especially for chronic migraines. So the migraine prognosis is essential to controlling all the impacts of migraine. There haven't been many studies looking at migraine prediction using machine learning models, even though the correlation between migraine attacks and the triggers, including hormonal factors, emotional factors, physical aspects, dietary elements, environmental factors, and medical characteristics, has been statistically analyzed.

This study aims to comprehend the significance of how much each trigger affects migraine attacks and to develop a machine learning model for migraine sufferers to predict the acute attack so they can prepare for preventive treatment.

This paper consists of 5 chapters. First is this Introduction chapter, where the motivation and background of this study are introduced. Next, multiple papers related to this study are presented in the Literature Review chapter. In chapter 3, all methodologies implemented in this study are explained. The results obtained from chapter 3 are visualized and discussed in the Results and Discussion chapter. As a final step, the Conclusion chapter summarizes all findings and prospective future work. In this paper, all experiments are conducted using Python.

CHAPTER 2

Literature review

Migraine is a chronic illness with a poorly understood aetiology. Although there isn't currently a known treatment for migraines, there are a number of precautions and prophylactic management options for chronic forms that can help relieve symptoms. Understanding the triggers of the migraine and taking preventive measures is currently the most effective way to deal with a migraine [5]. Numerous research studies have been conducted on multiple datasets in order to understand the implicated migraine triggers.

Patrícia Timy Fukui1 et al. (2008) [10] examined a dataset representing 200 consecutive migraine patients, 162 (81%) female patients with an average age of 37.0 years old and 38 (18%) male patients with an average age of 40.7 years old. Based on group of triggers analysis, there were no significant differences between females and males regarding the dietary factor (84.5%), which was the most common migraine factor group, and included fasting (63. 5%), the most commonly reported trigger. With a presence of 75.5%, sleep was the second most important group. Lack of sleep was the most reported from this sleep group with 61.5%. 68.5% of respondents cited the environment, while 65% listed stress as a trigger. And there were the third and fourth groups of factors mentioned in the study of Patrícia Timy Fukui1 et al [10].

In 2018, the study of Siirtola, P. et al. [24], utilized data from wearable sensors to predict migraine attacks early. A wrist-worn Empatica E4 sensor was used to collect data on acceleration, galvanic skin response, blood volume pulse, heart rate and heart rate variability, temperature, and sleep time from seven study subjects. From these measurements, only sleep time data were analyzed in this study. An analysis of two different classifiers was conducted; linear discriminant analysis (LDA) and quadratic discriminant analysis (QDA). Based on personal model results, it can be concluded that QDA classifiers deliver better recognition accuracy than LDA (with an average accuracy of 84.1% vs. 70.2%). As a consequence, this study shows that migraine attacks can be predicted in a quadratic manner rather than a linear manner [24].

Some research was focused on specific triggers, for instance, the 2008 review study of Alessandro Panconesi [19], focused on how influential alcohol is as a migraine trigger factor. In both the general population and migraine clinics, alcohol is a significant migraine trigger, according to a review summarising retrospective studies in various countries. 10% of migraine patients reported that alcohol triggers their migraines frequently, while almost a third of migraine patients reported that alcohol triggers their migraines occasionally. Alcohol, especially wine, is a significant trigger of migraines in various countries, with a range of 6.1% to 1.4%. There may be a connection between this and the average level of alcohol consumption in the country. On the other hand, as a result of sophisticated statistical analysis, a prospective study examining a range of migraine-related factors concluded that nutrition (including alcoholic beverages) plays a limited role in migraine precipitation, in contrast to a previous report by the same study group. Additionally, there were notable trends showing a decrease in the frequency of migraines and other types of headaches as alcohol consumption increased. There is no conclusive evidence to support a consistent connection between alcohol consumption and migraine or tension headaches, according to a number of studies. The author suggested that the insufficiency of association may be explained by a change in habits following the experience of these factors provoking attacks; that is, headache sufferers may tend to avoid alcohol in order to prevent aggravating their headaches. The intake of alcohol can, however, precipitate migraine attacks and tension-type headache attacks in some individuals [19].

Several studies were conducted based on the typical type of migraine, such as the research of AW Hauge et al. in 2010 [13]. According to the International Classification of Headache Disorders, Second Edition, 629 Danish patients who had been diagnosed with migraine with aura (MA), were included in this study. A questionnaire with space for free text and a list of 16 trigger factors considered to be relevant migraine triggers (including stress, intense emotional influences, menstruation or a break from the pill, sleep, alcohol consumption, etc.) was sent to the participants. Patients were differentiated based on whether their attacks had an aura or not. 347 of the 522 patients who returned the questionnaire had recent MA attacks. Compared to men, women reported more triggers. Patients who experienced both MA and MO attacks listed more MO-specific trigger factors than MA-specific ones. In total, 80% of participants who were experiencing attacks (278/347) stated that at least one trigger provoked their migraines, and 67% (187/278) of them said they understood at least one trigger frequently or always caused a MA attack. Migraine without aura (MO) attacks occurred concurrently in 41% (113/278) of cases. The most common trigger factors included stress (during stress), bright light, intense emotional influences, stress (following stress), and too much or too little sleep. Among all studied triggers, stress has been identified as the most frequent trigger for MA, MO, and migraines in general. The study made a distinction between migraine triggers during stressful situations (which affected 59% of patients) and triggers following stressful situations (which affected 70% of patients). Other triggers have been declared in more inconsistent ways, aside from this one factor's consistency. Sunlight and other powerful lights were the second most frequently reported trigger factors in this study. Though frequently cited as a major trigger, only 36% of the women in this study mentioned experiencing migraine attacks triggered by hormonal changes. In comparison to other studies that only looked at migraine patients regardless of attack types, the data from this study also revealed a significantly lower frequency of weather changes as a trigger factor for MA attacks [13].

The dataset utilized in this paper was gathered in order to analyze triggers of headaches, including migraine, for the research of Jeong-Wook Park et al. The data were collected using a Smartphone Headache Diary application (SHD) and the data were published along with the research in 2016. According to a study by Jeong-Wook Park et al. [20] in 2016, the most frequent causes of headaches for patients on those days were stress, exhaustion, lack of sleep, hormonal changes, and changes in the weather. The likelihood of a headache factor was 57.7%, 55.1%, 48.5%, and 46.5% for stress, sleep deprivation, fatigue, and other triggers, respectively. In comparison to non-migraine headaches, travelling (odd ratios [OR]: 6.4), hormonal changes (OR: 3.5), noise (OR: 2.8), alcohol (OR: 2.5), overeating (OR: 2.4), and stress (OR: 1.8) were all significantly associated with migraines. The preventive medication, migraines were more frequently brought on by hormonal changes or noise. Without taking preventive medication, migraines were more common in headaches triggered by stress, overeating, drinking, and travelling, but this was not noticeable when taking preventive medication. Despite the fact that several statistical analyses were conducted in this paper, the study was not focused only on migraine, and the model to predict migraine was not introduced [20]. So, in addition to the research of Jeong-Wook Park et al., this paper will concentrate on analyzing triggers for migraine attacks instead of general headaches, and also the machine learning model to predict migraine attacks early will be presented.

There are papers studying this published dataset. Rebecca S. Zhu and Yan Luo [31] applied three machine learning models, inclusive of the Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) models, which achieved considerable results with 99.13%, 96.73%, and 99.13% accuracy, respectively. However, by comparing the paper of Rebecca S. Zhu and Yan Luo with the original research of Jeong-Wook Park et al., the misuse of the dataset is recognized. The symptom that the data owner utilized to clarify whether the headache was a migraine or not was used as one of the triggers in the paper by Rebecca S. Zhu and Yan Luo. And this can unconditionally lead to an incorrect result, as can be clearly explained by the very high accuracy of the model, which is almost 100% [31].

Another paper by Rebecca Zhu [32], but this time working with Rucha Dave, worked on the same dataset. The logistic regression model was implemented for migraine prediction using the number of trigger occurrences as an input, which gained an accuracy of 95%. Another model that was created in this research was a random forest, which was utilized to understand the relationship between migraine attacks in accordance with the helping factors. And the best accuracy produced by the random forest model was 89%. But repeatedly, the data utilization of Rebecca Zhu and Rucha Dave was different from one of the data owners. The non-migraine headache cases were considered migraine cases in their study, and again, this can cause the inaccuracy of the model [32].

CHAPTER 3

Methodology

The purpose of this chapter is to describe all the methodologies that were incorporated into the study. This section begins with an explanation of data manipulations (dataset description and data cleaning). Next, an exploratory data analysis (EDA) is conducted to understand and visualize the descriptive statistics of the data. As a final section, Model and feature importance, the algorithms required to build a migraine prediction model are discussed, including balancing the dataset, machine learning models (Random Forest), and evaluation methods for migraine prediction models.

3.1 Dataset manipulation

The data used in this study was gathered from a mobile application called the Smartphone Headache Diary (SHD), where 62 participants from a neurology outpatient clinic at a university hospital kept a three-month log of their daily activities in relation to headache trigger factors and characteristics [21]. Participant recruitment took place between September 2014 and January 2015 for participants who met the inclusion criteria, which are[21]:

1. between the ages of 19 and 55, with or without auras migraines,
2. a monthly average of 2 to 14 headache days,
3. consistent headache symptoms for at least a year before the survey, and
4. owning a mobile device with a personal platform capable of running the SHD.

A dataset excluded the records revealed secondary causes of headaches and participants that were unable to complete questionnaires for at least 50% of the assessment period. A total of 113 patients were recruited from two centres at the beginning of the study. 30 patients withdrew from the study before its end. Of the remaining 83 patients, only 62 participants maintained a diary for at least half of the research period[21]. Two datasets are analyzed in this study; patient information and migraine diary.

A patient information dataset contains the personal information of 62 migraine patients, including their registration numbers, genders, ages, and BMIs. All 62 participants recorded their gender and age, but 10 of their BMIs were not logged. Before further analysis, all missing values were replaced by the mean calculated for each gender group.

The migraine diary dataset contains 4,579 records of headache diaries. In order to accurately distinguish migraine cases from normal headache cases, several survey questions were asked, such as sensitivity to sound, odour, or light, nausea, vomiting, and headaches on only one side of the head. Alongside a column indicating whether the patients experienced the migraine attack or not, this dataset also includes one column to clarify the duration of the headache, another column to point out if the patients had any signs before the migraine attack; one column to show the number of triggers they had got before getting a migraine, and plausible migraine triggers, inclusive of 'Stress', 'ExcessiveSleep', 'SleepDeprivation', 'Exercise', 'NotExercise', 'Fatigue', 'Menstruation', 'Ovulation', 'EmotionalChanges', 'WeatherChanges', 'Overillumination', 'Noise', 'InadequateLighting', 'Odors', 'Alcohol', 'Fasting', 'Overeating', 'Caffeine', 'Smoking', 'CheeseChocolateConsumption', 'Traveling', 'NoPreventiveDrug' and 'Others'. The initial datasets were in Korean, and before proceeding with any data manipulation processes, one of the challenging parts of this study was to translate them into English. In addition, 3,481 entries were missed from the migraine attack column. 3,480 missing records refer to the days that participants had no headache, and only one record implies that the patient had had a headache but not a migraine. Accordingly, all of these 3,481 missing values are reinstated as non-migraine. In addition, 1 record was missing all trigger columns. However, since the number of triggers is shown as 0, all missing triggers are modified to be 'No trigger' for all columns. The 2 datasets are combined before further research is conducted, whereupon an exploratory data analysis process is utilised to understand the true significance of the data.

3.2 Exploratory data analysis

Starting with univariate descriptive statistics, an approach to summarize the main characteristics of only each variable independently. First, the data on patient information is analyzed; 51 of the 62 participants, or 82.26%, are female. On the other hand, only 11 patients, or 17.74%, are male. As demonstrated in the boxplot below, among participants who meet the first inclusion criteria, the minimum age is 20 and the maximum is 56 years old. The average age of patients is 37.71 ± 8.64 years old. The average BMI is 22.41, with a minimum and maximum BMI of 16.4 and 32.5, respectively. There are 3 outliers shown in the BMI boxplot, but in this case, they do not imply any incorrect data input, so these 3 outliers are kept for the next process. These boxplots are utilsed to visualize the statistical information, so the insight meaning of the data can be explored. To understand detail (exact statistical values explained before), the describe function in Python is executed in addition to these boxplots, as described before, and is included in Table A.1 in the appendix. The histogram is used to check the distribution of each feature, which can also be seen in Figure A.1 in the appendix.

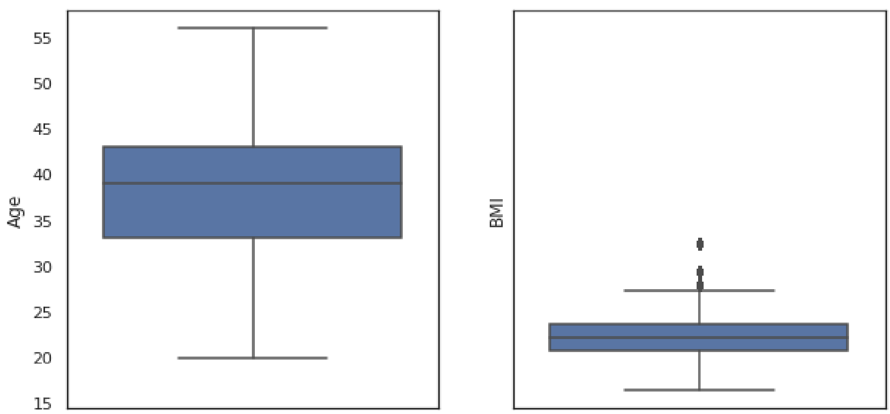

Figure 3.1: Univariate descriptive statistics for Age and BMI

Thereafter, the migraine diary dataset is studied. Figure 3.2 shows the extreme imbalance in the dataset. Migraine attacks are the minority class in this dataset. In a total of 4,579 diary entries, there are only 336 records, or 7.34%, of migraine attacks, while non-migraine cases account for 92.66% of all data. Among 336 migraine records, 82.33%, or 280 entries, reveal at least one trigger (which is for example, stress, excessive sleep, menstruation or smoking) before the migraine attack, whereas there are only 56 cases of migraine attacks without triggers. This suggests that triggers might be able to be used to explain migraine attacks. So we will investigate more using bi-variate analysis to see the relationship between each input feature and migraine attacks.

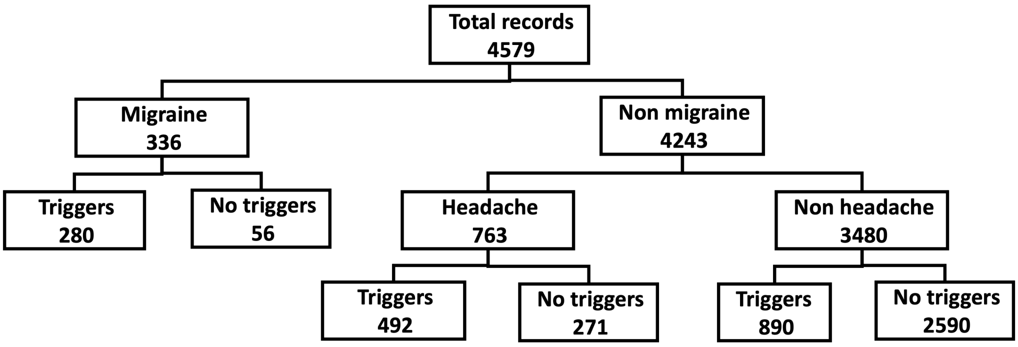

Figure 3.2: Migraine Diary Dataset

After combining the patient data dataset and the migraine diary dataset, the next analysis is carried out. To understand whether gender is significantly associated with migraine attacks or not, the stack graph is used to describe the proportion of each gender regarding migraine attack cases. As we can see from Figure 3.3, 8.1% of all females are afflicted by migraine, while only 3.9% of all males are affected. This can illustrate that gender is related to migraine attacks, and females share more potential to have migraine than males. In addition to gender, other features are also studied, and the results are attached in the appendix (Figure A.2).

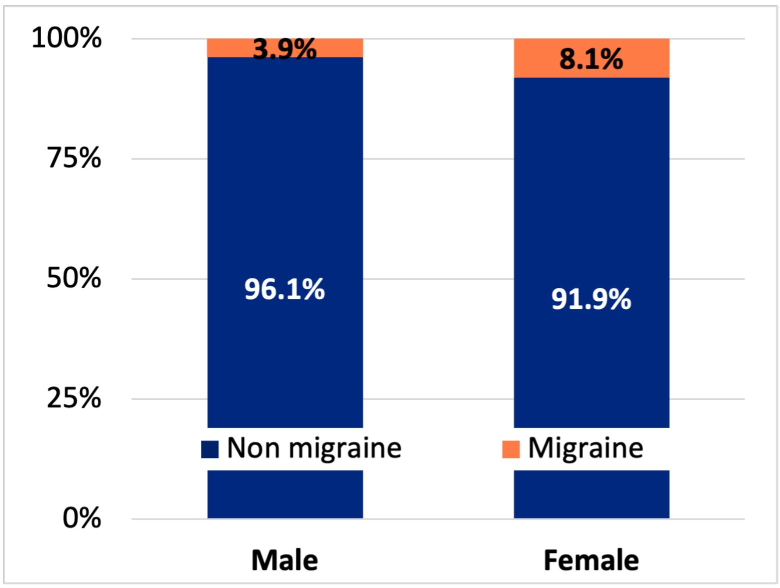

Figure 3.3: Proportion of migraine attack by Gender

Next, after filtering only migraine cases, the proportion of each trigger is visualised in Figure 3.4 . 'PreventiveDrug', 'Stress' and 'SleepDeprivation' revealed the most in migraine attack cases, with ratios of 51.79%, 36.01%, and 24.40%, respectively. On the other hand, 'Excercise', 'IndequateLighting' and 'Smoking' show the smallest proportions, which are 1.49%, 1.19%, and 0.60%, respectively. While their percentages are most prevalent among migraine cases, it is yet unable to be concluded that "PreventiveDrug", "Stress", and "Sleep Deprivation" are the most significant triggers of migraine attacks; further investigation is required to determine this.

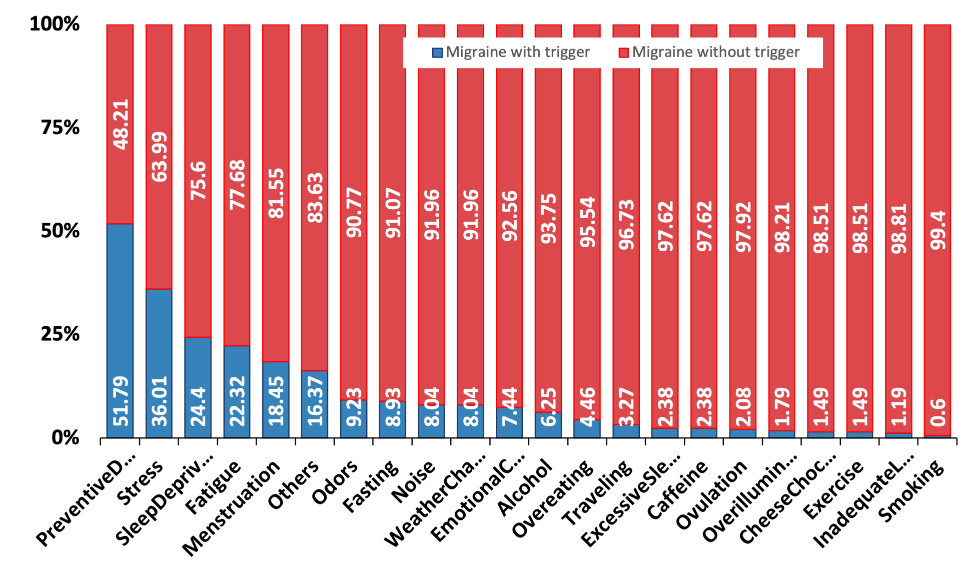

Figure 3.4: Proportion of each trigger for migraine cases

Statistical analysis of categorical data can be performed using Pearson's chi-square test. A significant difference between the experimental and expected data can be determined by this tool [9]. The chi-square test can be classified into two types: the chi-square goodness of fit test and the chi-square test of independence. While the chi-square test of independence is used to determine whether two categorical variables are related to one another, the chi-square goodness of fit test can be used to determine whether or not the frequency distribution of a categorical variable deviates from expectations [28]. In order to determine the association between migraine triggers and migraine attacks, the chi-square test of independence is utilized in this study. The contingency table was applied to calculate the Chi-square and P-value. The P-value will be utilised to understand the correlation between the label (migraine attack) and all binary categorical features.

As shown in Table 3.1, the Chi-square of 203.35.18 and P-value of $3.89e^{-46}$ clearly explain that stress is the trigger that correlates to migraine attacks the most. Others are the second corresponding trigger to migraine attacks, with a P-value of $5.23e^{-42}$. The third trigger that influences migraine attacks is 'Menstruation' with the P-value of $2.93e^{-38}$. With their P-value greater than 0.05, 'Caffeine', 'CheeseChocolateConsumption', 'Overillumination', 'Exercise', 'Ovulation', 'ExcessiveSleep,' and 'Smoking' do not indicate the statistical significance and demonstrate strong support for the null hypothesis that there is no relation between those triggers and migraine attacks. When building the model, this information can be used to decide which features to include or leave out. This Table 3.1 only contains the top 5 features with the highest P-value and features that provide more than 0.05 of a P-value, the full version of this table is also available at Table A.2 in the appendix.

Table 3.1: Migraine vs trigger association table

Trigger	Migraine		Non migraine		Chi square	DOF	P_value
	W/T	W/O	W/T	W/O			
Stress	121	215	417	3826	203.35	1	$3.89e^{-46}$
Others	55	281	99	4144	184.43	1	$5.23e^{-42}$
Menstruation	62	274	139	4104	167.27	1	$2.93e^{-38}$
SleepDeprivation	82	254	318	3925	109.56	1	$1.22e^{-25}$
Odors	31	305	54	4189	103.78	1	$2.26e^{-24}$
Caffeine	8	328	199	4044	3.33	1	0.06803583
CheeseChocolateCons	5	331	24	4219	2.87	1	0.09015349
Overillumination	6	330	32	4211	2.87	1	0.09026721
Exercise	5	331	124	4119	1.85	1	0.17435612
Ovulation	7	329	53	4190	1.09	1	0.29591663
ExcessiveSleep	8	328	64	4179	1.02	1	0.31256749
Smoking	2	334	44	4199	0.25	1	0.61883958

In addition to the association check using the P-value from the Chi-square, the correlation matrix, which in the background uses the Pearson correlation coefficient, the most popular method to determine a linear correlation [29], is also studied. As expected, Figure 3.5, the partial confusion matrix (Figure A.3, the full version of this matrix is appended in the appendix) illustrates that there is no significant linear correlation between each feature and the label (migraine attacks). However, what we found here is that the correlation matrix revealed that 'Stress' is the most highly correlated feature of migraine attacks, as well as using

Chi-square. Same as for other triggers, 'Other', 'Menstruation', and 'SleepDepriation' are the second, third, and fourth correlated features to the label, for both approaches. And the order of the remaining features for both statistical methods (Chi-square and correlation matrix) is similar. This can ensure that these 2 approaches give the same result to clarify which feature relates to the label the most.

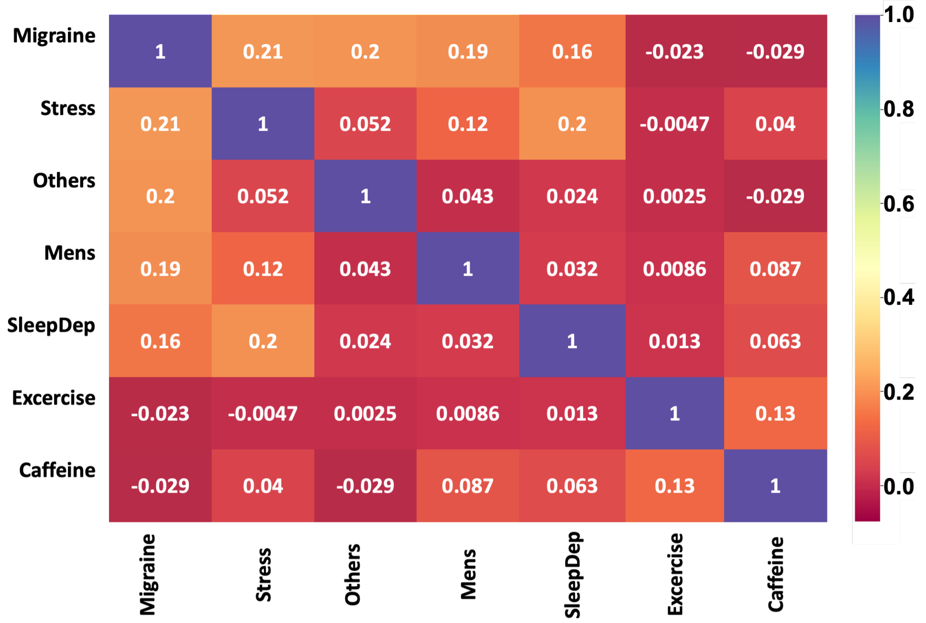

Figure 3.5: Correlation matrix

Other than all the EDA conducted above, this study also combines triggers in the same group into one column, for instance, combining 'SleepDepriation' with 'ExcessiveSleep' to be only 'Sleep', and then proceeding with the EDA again to confirm whether there is a significant change or improvement. Moreover, as a result of the combination of triggers, the dataset becomes less complex and the questionnaire for the survey is reduced, which may reduce the cost of setting up the survey. A benchmarking of the model performance and cost is essential for the practical application of the model.

Table 3.2 reveals that the individual factors of 'Stress' and Others (no combining) are still the greatest triggers of migraine attacks. A combination of 'Menstruation' and 'Ovulation' keeps 'Hormonal_changes' in the same position (the third), but the P-value of this set ($5.15e^{-35}$) is lower than that of 'Menstruation' ($2.93e^{-38}$) individually. Similarly, excessive sleep and sleep deprivation both correlate less with migraine than only 'SleepDeprivation'. A small increase in the P-value of 'Odors', this group's highest score, is seen for 'Environment' which includes 'Odors', 'InadequateLighting', 'Overillumination', and 'Noise'. Taking this result into consideration, we still cannot conclude that merging triggers is a good idea or not. This will be verified by using these combined features to build the model and check how it performs. The correlation matrix (see Figure A.4 in the appendix) is also applied to these combined triggers, similar to the analysis of the individual triggers. In this instance, the result is similar to that of individual triggers, there is no significant linear relationship, but the correlation levels rank similarly to the Chi-square test.

Table 3.2: Migraine vs combined triggers association table

Trigger	**Migraine**		**Non migraine**		**Chi square**	**DOF**	**P_value**
	W/T	**W/O**	**W/T**	**W/O**			
Stress	121	215	417	3826	203.35	1	$3.89e^{-46}$
Others	55	281	99	4144	184.43	1	$5.23e^{-42}$
Hormonal_changes	69	267	185	4058	152.41	1	$5.15e^{-35}$
Sleep	90	246	382	3861	104.58	1	$1.51e^{-24}$
Environment	55	281	166	4077	102.48	1	$4.36e^{-24}$
EmotionalChanges	25	311	62	4181	56.55	1	$5.47e^{-14}$
Physical_Fatigue	78	258	493	3750	37.3	1	$1.02e^{-09}$
Dietary	68	268	412	3831	35.66	1	$2.35e^{-09}$
Traveling	11	325	21	4222	30.76	1	$2.92e^{-08}$
PreventiveDrug	174	162	1679	2564	18.78	1	$1.47e^{-05}$
WeatherChanges	27	309	209	4034	5.54	1	0.01858291
Smoking	2	334	44	4199	0.25	1	0.61883958

3.3 Model and Feature importance

In this study, one of the most powerful machine learning models, Random Forest, is implemented in order to predict migraine attacks. But since there is an extreme imbalance on the utilised dataset, before applying the model, data balancing process is required. Then the model implementation is conducted, and after that, several evaluation methods are applied in order to understand the performance of each model. Finally, the feature importance approaches are executed to confirm how much each feature contributes to the model result. In this section, methodologies used in this study are explained.

3.3.1 SMOTE for Imbalanced Data Handling

Regarding the analysis in the EDA section, a significant imbalance among the classes was notified (Migraine: 7.34%, Non-Migraine: 92.66%). In order to develop a classification model, it is essential to have a balanced data set. It is possible for the classifier to be biased in favour of the majority class if the imbalanced data is fed into the model, simply because there is not enough information to learn about the minority class [25]. Oversampling and undersampling are the two main techniques to overcome this data imbalance problem. Since the size of the dataset examined in this paper is not large and the migraine class, which is the main focus of this study, belongs to the minority class, oversampling called "Synthetic Minority Over-sampling Technique" (SMOTE) was applied in order to increase the migraine class and make the dataset balanced before implementing the model. The main concept of SMOTE is to introduce synthetic examples rather than simply replicating instances of the minority class. By interpolating between various minority class instances that are located within a specified neighbourhood, new data is produced [8].

3.3.2 Random Forest Model

According to the EDA section, the dataset used in this study is not particularly large. So, a machine learning model is selected over a deep learning model since the data might not be sufficient to train the model properly and it would cause overfitting and instability of the result. A supervised machine learning model called "Random Forest" was utilised in this study, in order to predict the possibility of migraine attacks. Random forest is a powerful supervised machine learning model for both classification and regression problems [12]. It is an ensemble method that consists of decision trees trained on different samples through the bagging method (short for bootstrap aggregating), in which one learning algorithm is used for all predictors. They are trained using a random sample of the training set, and replacement is used during sampling (pasting is the practice of sampling without replacement) [30]. The decision trees' habit of overfitting their training data can be corrected by the random forest concept, which improves generalizability and robustness and produces better predictive performance than any one of the individual learning algorithms could. When the random forest is applied to classification tasks, the class with the majority of votes is selected as the output. As for the regression tasks, the result or output is the mean or average prediction of the individual trees [23].

In order to understand the random forest algorithm, it is necessary first to comprehend the decision tree concept since it is a fundamental component of the random forest algorithm [30]. In both classification and regression problems, the decision tree involves a split of the input dataset with regard to the target variable's purity, one dataset is divided into two purer datasets. The split criterion is predetermined in order to formulate optimization problems, and it is recursively segmented until it meets the predetermined stopping condition at the last node of trees, which are commonly called leaf nodes or leaves [30].

To train a decision tree in Python, Scikit-Learn uses the Classification And Regression Tree (CART) algorithm. To begin with, the algorithm separates the training set into two subsets based on a single feature k and a threshold tk. The pair (k, tk) that produces the purest subsets (weighted by their size) is selected [12]. Recursively, once the training set has been split into two parts, it splits the subsets and sub-subsets using the same logic. Whenever it reaches the recursive stopping criteria (defined by max_depth, min_samples_split, min_samples_leaf, min_weight_fraction_leaf, and max_leaf_nodes), it stops recursing. Purity can be conceptualized as a group's degree of homogeneity. However, homogeneity can mean various things depending on the mathematical foundation the decision tree is built upon. Gini Index and Information Entropy are the 2 most popular backbones for impurity measuring in decision tree decisions [12].

The Gini Index is a metric used to determine how frequently a randomly selected element would be misclassified. That means it is preferable to have an attribute with a lower Gini index [11]. When a node is pure, or when every element belongs to a single distinct class, the Gini Index has a minimum value of 0, so this node won't be divided again. The best split is therefore determined by the features with the lowest Gini Index. Furthermore, it is maximized when the probabilities for the two classes are equal [2].

The Gini impurity formula [12]:

$$G_i = 1 - \sum_{k=1}^{n} P_{i,k}^{\ 2} \tag{3.3.1}$$

Where $(P_{i,k})$ is the ratio of class k instances among the training instances in the ith node [12].

Entropy is a gauge of information that shows how disordered the features are in relation to the target. Similar to the Gini Index, the feature with the lowest entropy represents the ideal split, when the probability of the two classes is equal, it reaches its maximum value, and when a node is pure, the entropy is at its lowest point, which is zero [2].

The Entropy formula [12]:

$$H_i = \sum_{k=1}^{n} P_{i,k}\, log(P_{i,k}) \tag{3.3.2}$$

Where $(P_{i,k}) \neq 0$ [12]

The use of Gini impurity or entropy, in most cases, does not significantly alter the outcome: they produce trees that are similar. Since computing Gini impurity is a little quicker, it is selected to be a default value in Scikit-Learn. When they diverge, however, Gini impurity tends to isolate the most prevalent class in its own branch of the tree, whereas entropy tends to produce slightly more balanced trees [12].

In Scikit-Learn, not only are there recursive stopping criteria and impurity measurement methods, as mentioned in the previous section, but there are also other hyperparameters that can be adjusted to improve model accuracy and also to eliminate overfitting [15], including:

1. n_estimators: defines the number of decision trees that will be created.
2. max_depth: indicates how far apart our trees can split.
3. min_samples_split: specifies the minimal number of samples that must be in each node after it has split from its parent node.
4. max_samples: specifies the maximum number of samples from the parent dataset that will be taken into account when bootstrapping the decision trees.
5. max_features: determines the maximum number of features to consider during bootstrapping.
6. bootstrap: is a boolean variable (True or False) used to disable or enable the bootstrap.

3.3.3 Feature Importance

Starting with a built-in feature importance called Gini importance, also known as mean decrease impurity, is calculated from the Random Forest structure. Essentially, each feature is measured for its impact on decreasing impurities (the feature with the biggest impact is selected) internal node), and the importance of each feature is calculated by the average of all trees in the forest. Relative values should be considered when using this method. One of the biggest advantages of this method is its speed of computation since all the values are calculated at the time of Random Forest training. There are some disadvantages to the method, such as its tendency to emphasize numerical (continuous) features and categorical features with high cardinality (select them as important). Additionally, when selecting correlated features, it may ignore the importance of the second feature which may lead to incorrect conclusions [22].

By computing the contributions of each feature to the prediction of an instance x, SHAP (SHapley Additive exPlanations) aims to explain the prediction of an instance x. Using coalitional game theory, SHAP computes Shapley values. Players in a coalition are the feature values of a data instance. We can calculate Shapley values by dividing the "payout" (= the prediction) among the features. In addition to being an individual feature value, a player can also be a group of feature values, such as a value in a tabular graph [17].
SHAP specifies the explanation as [17]:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \tag{3.3.3}$$

where g is the explanation model, $z' \in \{0,1\}^M$ is the coalition vector, M is the maximum coalition size and $\phi_j \in \mathbb{R}$ is the feature attribution for a feature j.

3.4 Evaluation method

Evaluation techniques are used after the models have been trained in order to examine their performance. In the normal distribution (balanced) of classification data, accuracy is normally used to calculate classification performance, which is calculated by dividing the number of correct predictions by the number of predictions. However, imbalanced classification problems cannot be measured by accuracy as a performance metric. Due to the class imbalance, even unskilled models can achieve 90% or 99% accuracy when the examples of the majority class outweigh those of the minority class, which typically, the majority class is referred to as a negative outcome, and the minority class is referred to as a positive outcome. In an imbalanced classification problem, instead of focusing on how much accuracy the model can predict, the target is more on minimizing the incorrect predictions [4]. By examining the confusion matrix, which also reveals which classes are correctly and incorrectly predicted as well as the kinds of errors that are being made, it is possible to better understand how well a predictive model performs [4].

Accuracy equation [6]:

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{3.4.1}$$

Recall, precision, and the F1-score can be calculated using information from the confusion matrix, and they can demonstrate the proportion of incorrect (false) predictions.

A recall statistic, also known as a sensitivity statistic, is the percentage of correct positive prediction within a total number of positive examples. Alternatively, it is the number of true positives divided by the number of true positives plus false negatives [4].

Recall equation [12]:

$$Recall = \frac{(TP)}{(TP + FN)} \tag{3.4.2}$$

While the precision describes the proportion of correct positive prediction by the model and total cases of positive prediction. It is determined by dividing the number of positive examples that were correctly predicted by the total number of examples that were predicted positively [4].

Precision equation [12]:

$$Precision = \frac{TP}{(TP + FP)} \tag{3.4.3}$$

The precision, based on their definitions, is appropriate when the focus is on minimizing false positives. The recall, meanwhile, is appropriate when the focus is on minimizing false negatives [4].

To summarize the performance of a model, the F1 score (also known as the F-Measure) is widely applied. With the F1 score, precision and recall can be combined into a single measurement. Precision and recall are not sufficient on their own to explain everything. There is a possibility of excellent precision with a poor recall score, or the opposite, terrible precision with an excellent recall. It is possible to express both concerns with one score by using the F1 score [4].

F1-score equation [12]:

$$F1 - score = 2\frac{precision * recall}{precision + recall} = \frac{2TP}{2TP + FP + FN} \tag{3.4.4}$$

where:

- TP = True Positive
- TN = True Negative
- FP = False Positive
- FN = False Negative

Another common tool used with binary classifiers is the receiver operating characteristic (ROC) curve. Similar to the precision/recall curve, the ROC graph plots the true positive rate (another name for recall) versus the false positive rate (FPR), instead of precision versus recall. The FPR measures the proportion of negative events that are misclassified as positive [12]. In the ROC space, the best prediction would result in a point in the upper left corner (0,1) representing 100% sensitivity (no false negatives) and 100% specificity (no false positives) [7]. In addition to ROC curves, classifiers can also be compared using the area under the curve (AUC), which measures how well they distinguish between classes. In a perfect classifier, the AUC will be 1, while in a purely random classifier, it will be 0.5 [3].

CHAPTER 4

Results and discussion

In this chapter, the results of executing the machine learning model on several patterns of the data with 2 different hyperparameter settings are presented. The performance of the model is explained using a variety of evaluation approaches. Moreover, the results of implementing feature importance methods are visualised and discussed.

4.1 Result of the model

As aforementioned in the methodology chapter, the machine learning model used in this study was the random forest model. Begin with a full model that contains all features, inclusive of all triggers and migraine patient information (Gender, Age, and BMI). In this model, the default hyperparameters defined by Scikit-learn were installed. This model will be called 'Model 1' for the rest of this paper. The result of executing the confusion matrix for this model on the test dataset is shown in Figure 4.1. Out of 842 migraine attack cases, the model correctly predicted 799, while 43 cases were mislabeled, while for non-migraine attacks, 794 correct predictions were made, while 48 were incorrect. This means this model provides the accuracy, precision, recall, and F1 score on the test data of 94.6%, 94.86%, 94.3%, and 94.58%, respectively. According to its performance, the model is superior to the untrained predictions when comparing its results to the class imbalance levels (92.66%). The ROC curve and AUC score are also calculated to visualize the performance of this model, as presented in Figure A.5 in the appendix. The ROC curve directing near to perfect point and 0.98 AUC scores exhibit that this model provides good performance and it is able to differentiate classes of the output (migraine and non-migraine). The performance of the model, however, needs further investigation to see if it can be improved.

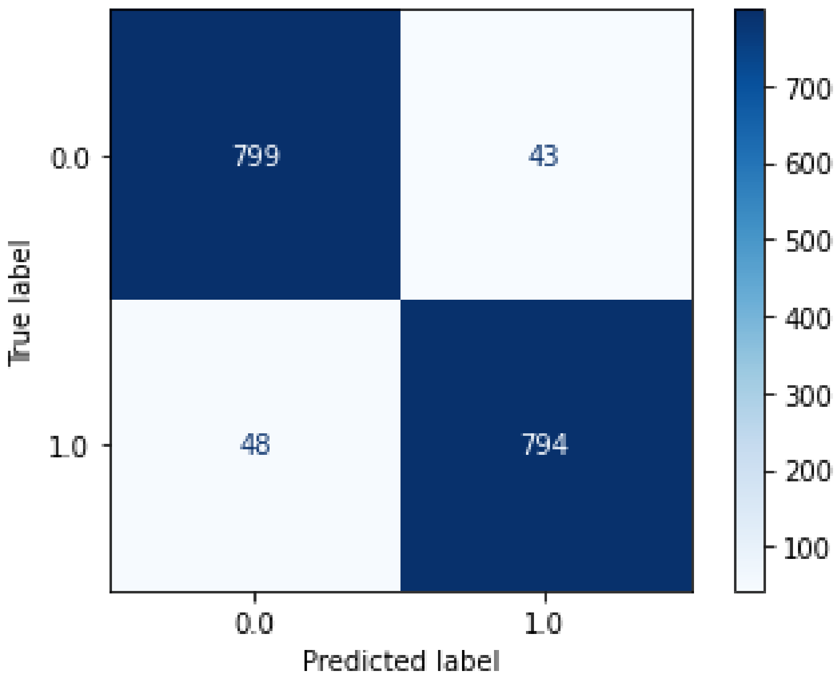

Figure 4.1: Confusion matrix of Model 1

According to the result from the EDA section, features that do not give a significant statistical result are removed, and only those features with a P-value less than 0.05 are included. Several features, inclusive of 'Caffeine', 'Smoking', 'CheeseChocolateConsumption', 'Overillumination', 'Ovulation', 'ExcessiveSleep', and 'Exercise' were excluded from the dataset used in this model (named Model 2). And the default hyperparameters are applied. Compared with Model 1, the correct prediction rate for migraine cases is similar, but for the non-migraine cases, the accurate predictions reduce from 794 to 778 for non-migraine cases. On the other hand, the incorrect rate for the non-migraine class increase. Consequently, this model 2 provides lower performance in comparison with model 1. The accuracy decreases to 93.65%, while the precision, recall, and F1-score reduce to 94.76%, 92.40%, and 93.57%, respectively. Figure A.6 in appendix visualises performances of this model.

As for the Model 3, the triggers from the same group are merged into one column, as explained in the EDA area. This model provides an accuracy of 94.30%, which is higher than Model 2 but lower than Model 1. Similarly, the precision of 93.78% and the F1-score of 94.33% are lower than those of Model 1. However, this model has a better recall score in comparison with Model 1. Since the target of this study is to predict migraine attack cases as much as possible, so the patients will be able to prevent them or reduce their effects, Model 1 remains the best model among all 3 models. The confusion matrix of this model are attached in the appendix (Figure A.7). Although this model is not selected for further study in this paper, when benchmarking model performance versus data collection cost, it may be the most useful model in actual practice, since it can help improve dataset complexity.

Even though feature selection was insufficient to improve the model's performance, as explained in the methodology section, in the random forest model, the model's performance, both in terms of the correct prediction rate and the training speed, can be improved by modifying hyperparameters. Model 1 exhibits the best performance among the two models above, so it is selected for hyperparameter adjustments to improve performance and named Model 4. Model 4 is then defined by a full model (including all migraine triggers and patient information) with adjusted hyperparameters. The best-tuned hyperparameters compared with the default hyperparameters are shown in Table 4.1.

Table 4.1: Hyperparameter table

	n_ estimators	**bootstrap**	**max_ features**	**criterion**	**max_ depth**	**min_ samples**	**oob_ score**
Default	100	True	auto	gini	None	5	True
Adjusted	500	False	auto	gini	None	7	False

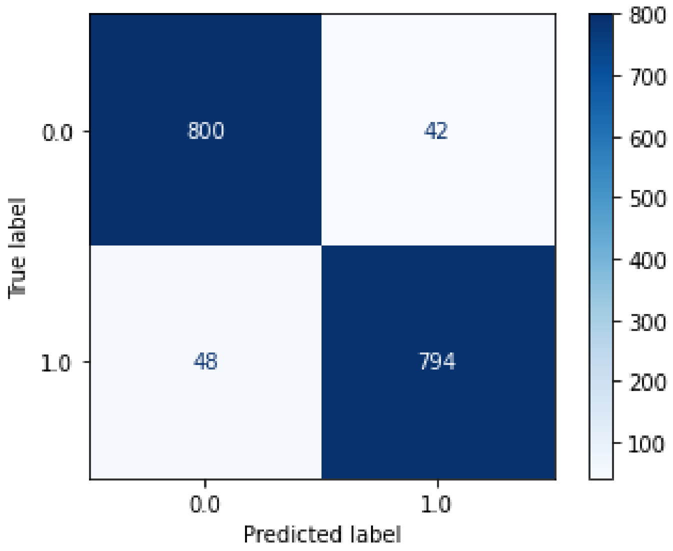

Figure 4.2: Confusion matrix of Model 4

In Figure 4.2, the confusion matrix of this model is exhibited. The improvement is not significantly displayed. For migraine cases, only 1 more correct prediction increases, whereas the result for non-migraine cases remains the same. The accuracy, precision, recall score, and F1 score of this model are 94.66%, 94.98%, 94.30%, and 94.64%, respectively. Even though the improvement is not significantly exhibited here, as expected, this model provides a better result than model 1 and makes it the best model in this study. In addition to the confusion matrix, the ROC-AUC curve is also studied for this model. This ROC curve of Model 4 is closer to the perfect point than the one of Model 1, and the 0.9845 AUC scores indicate that this model provides more performance and better differentiation between output classes than Model 1.

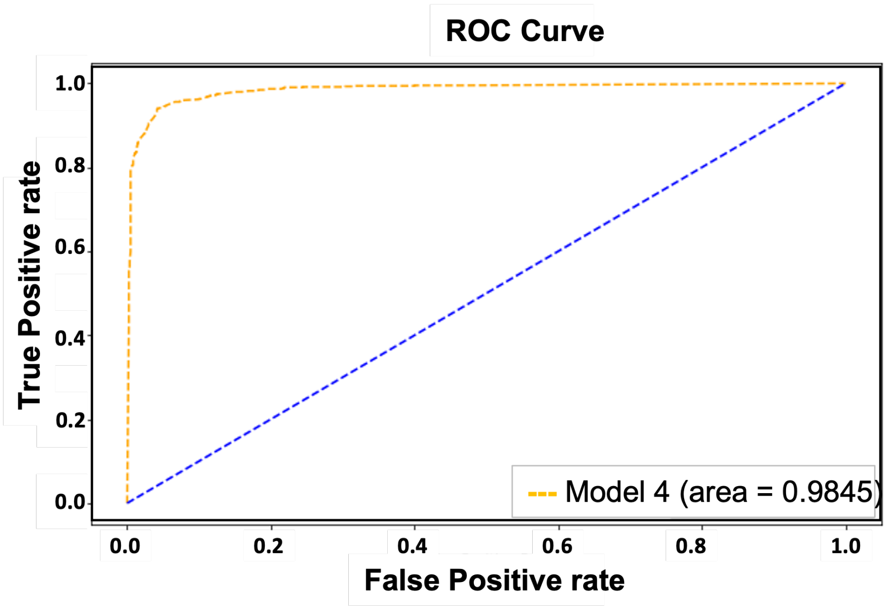

Figure 4.3: ROC curve of Model 4

There are several cases when acquiring patient personal information, such as gender, age, and BMI, is difficult. The example data that was provided by the business side at the beginning of this project also excluded patient information. So in this study, we also investigated the model to support this case. Following that, Model 5 is trained using only migraine triggers (without participant information). This model demonstrates the lowest performance scores with an accuracy score of 90.02%, a precision score of 93.43%, a recall score of 86.10%, and an F1 score of 89.62%.

Table 4.2: Model performance comparison table

Model	Accuracy	Recall	Precision	F1-score
Model 1	94.95%	94.89%	95.01%	94.95%
Model 2	93.82%	92.40%	95.11%	93.73%
Model 3	94.06%	94.30%	93.85%	94.08%
Model 4	95.13%	94.66%	95.56%	95.11%
Model 5	89.73%	84.92%	93.96%	89.21%

Table 4.2 demonstrates the summary of all model performance, which apparently explains why Model 4 is the best model among the 5 studied models for this data set. ROC with AUC scores in Figure 4.4 gives a similar result to Table 4.2 but in a more visual way. The ROC curve

nearest to the perfect classifier point (the top-left corner) explains that Model 4 provides the best prediction performance, and the highest AUC score confirms that Model 4 is the best model to distinguish between migraine attacks and non-migraine attack cases.

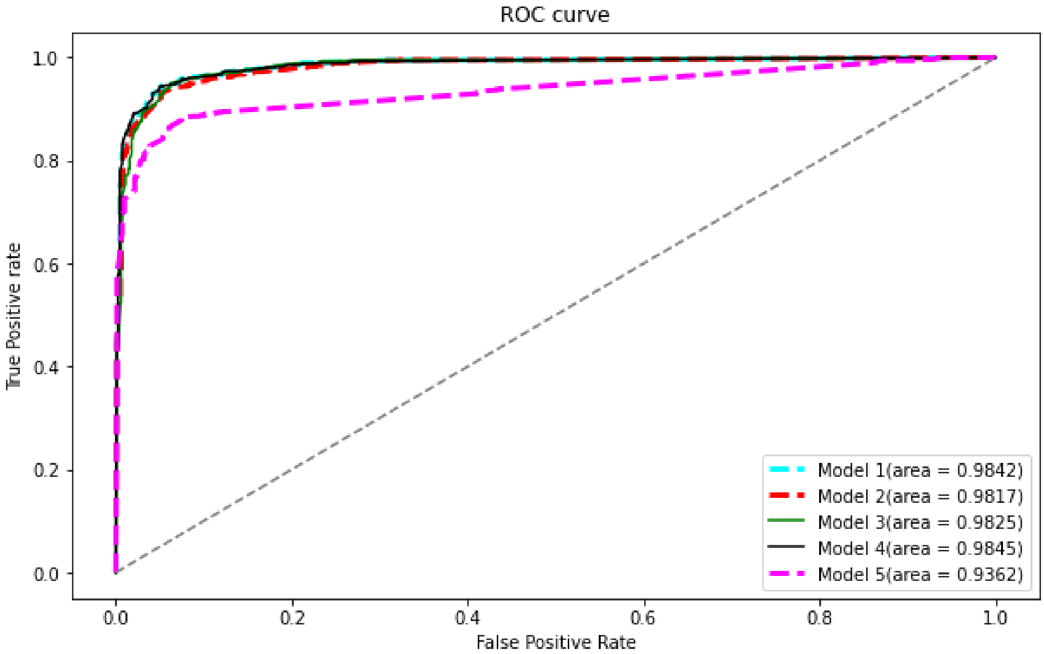

Figure 4.4: ROC-AUC comparison

4.2 Result of feature importance

As mentioned in the introduction section, one of the targets of this paper is to study how considerably each trigger influences migraine attacks.

Before starting to analyse the feature importance for the whole model, an individual prediction is conducted in this study, and the waterfall plot using SHAP value is conducted in order to predict one observation of the dataset. An analysis of waterfall plots provides vital insight into why a case is predicted given its variable values. A waterfall plot shows the bottom value, which is augmented (red) or subtracted (blue) to get the final value [17]. Based on the first observation in X_test, the Figure 4.5 below shows the prediction. In this case, there is a base value of 0.5 shown at the bottom, and as indicated by the model output, there is a prediction that Observation 1 is a non-migraine case (F(x) =1) illustrated at the top. With a 0 value for 'Stress' adds 0.08 to the probability of the model for non-migraine cases, 22.9 of 'BMI' affects the model in the same direction (for non-migraine cases) with an additional

0.07 to the probability. All of the triggers in this record have similar effects and all result in non-migraine cases. And as a result, this record is predicted as a non-migraine case with a probability of 1 (100% non-migraine for this case).

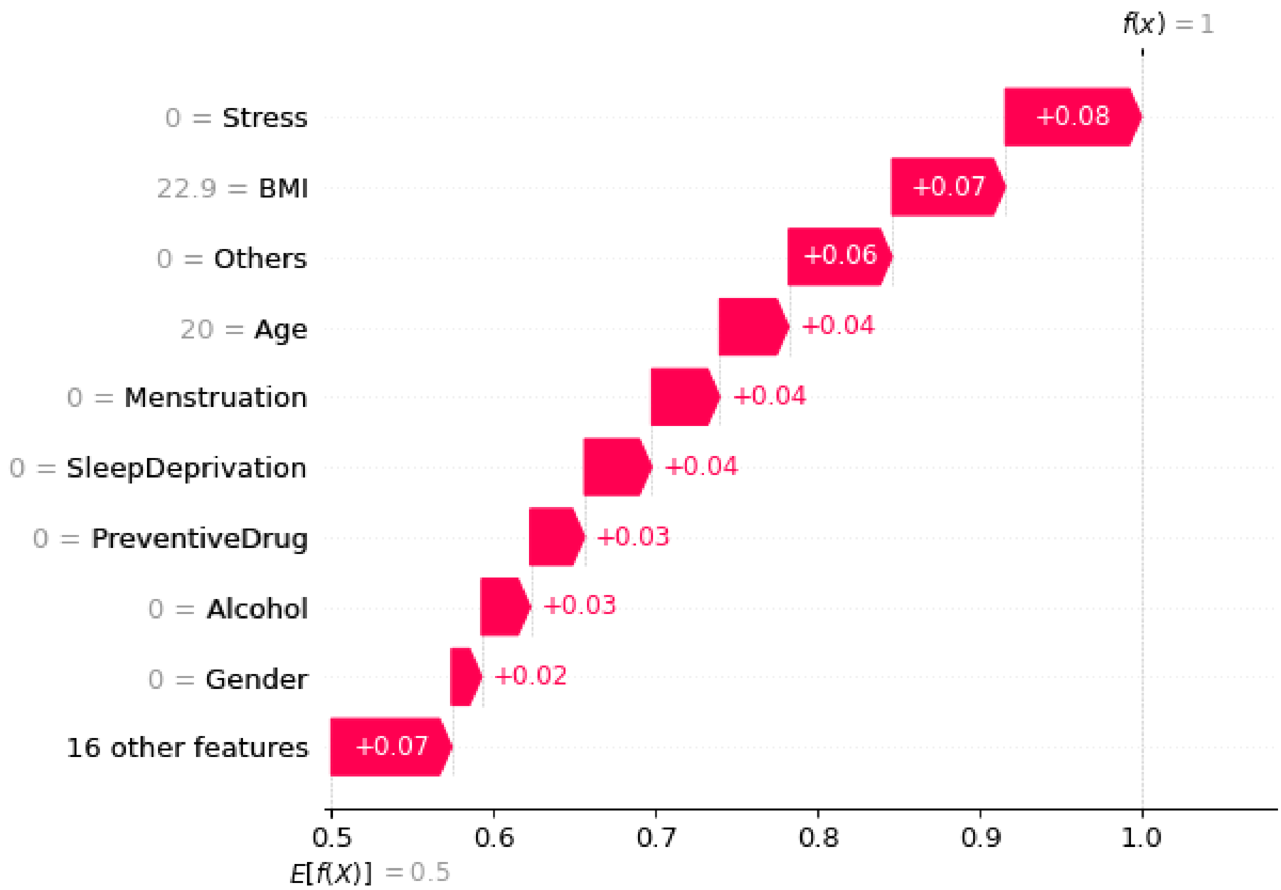

Figure 4.5: SHAP waterfall plot of Model 4

Besides the waterfall plot, the force plot can also be used to visualize the prediction of an individual instance using the SHAP method. This analysis was also performed for instance number 3333, and the results are shown in Figure 4.6 of the appendix section. According to the study, a patient with an 18.4 BMI who is having menstruation has a 0.97 probability of not experiencing migraines, or else she has a 0.03 chance of experiencing migraines.

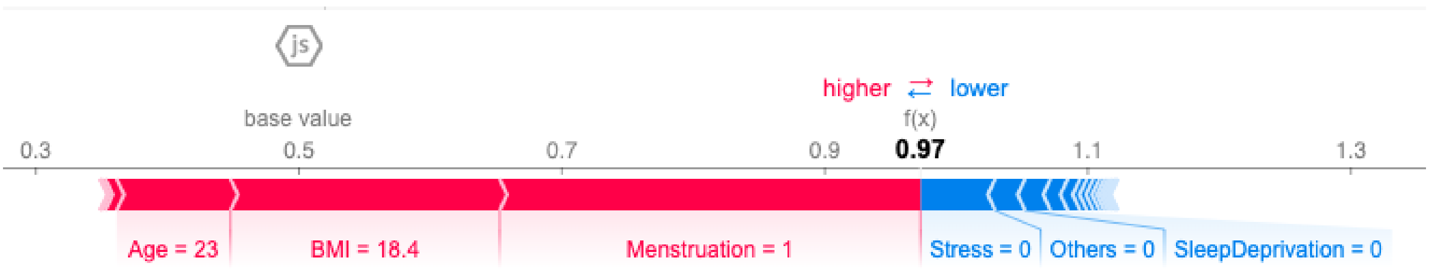

Figure 4.6: SHAP force plot of Model 4

The MDI algorithm, which is the built-in algorithm in a random forest, is used in the first analysis of feature importance. According to MDI results, 'BMI' has the greatest influence on migraine attacks. Among the factors that correlate with migraine attacks, 'Age' is the second most important factor. In this figure, stress is the first binary feature revealed and represents the third most important feature. Smoking, on the other hand, represents the least impactful migraine trigger. It may be that MDI behaviour is biased so BMI and age have the greatest influence on migraine attacks. As mentioned in [16] that 'it has a tendency to inflate the importance of continuous features' [16].

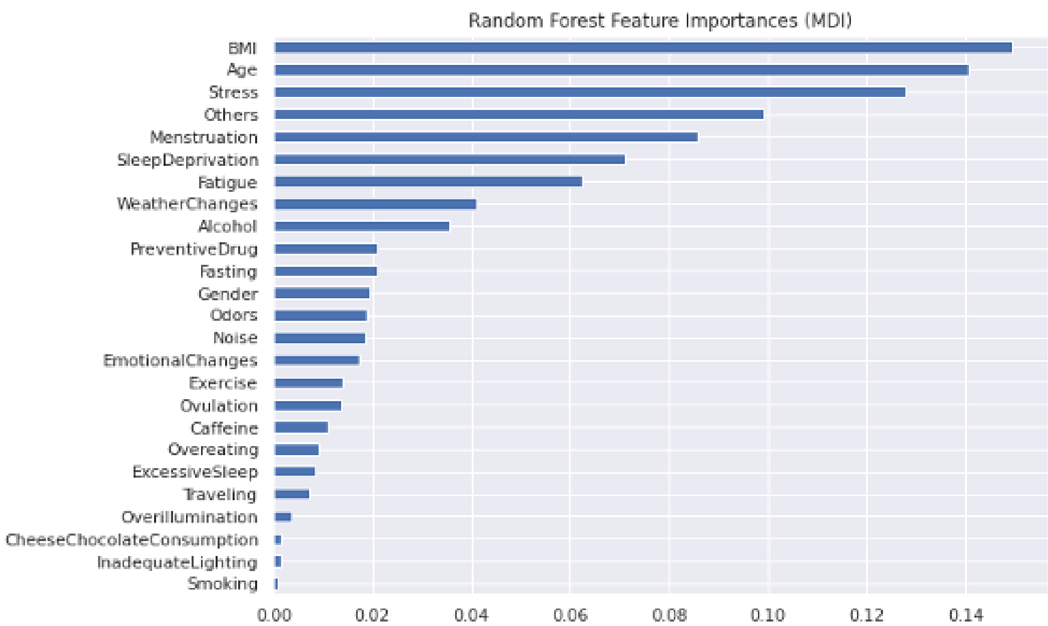

Figure 4.7: Model 4 feature importance by MDI algorith

In addition to this algorithm, another more reliable algorithm called SHAP was applied and compared with it. Compared with the MDI algorithm, the SHAP approach makes 'Stress' (which is the third most important factor in the MDI) the most important factor here, followed by Age, BMI, and 'Menstruation', respectively, in the top four. According to the individual prediction, this gives a similar result, 'Stress' is the most prominent feature that affects migraine attacks. Figure 4.8, it only displays the 10 most important features, the triggers with the lowest importance are not displayed. However, as illustrated in Figure A.8 the appendix, two features that correlates to migraine attacks the least are 'CheeseChocolateConsumption' and 'Smoking'. The SHAP result seems more reasonable than the MDI result compared to the correlation studied in the EDA session.

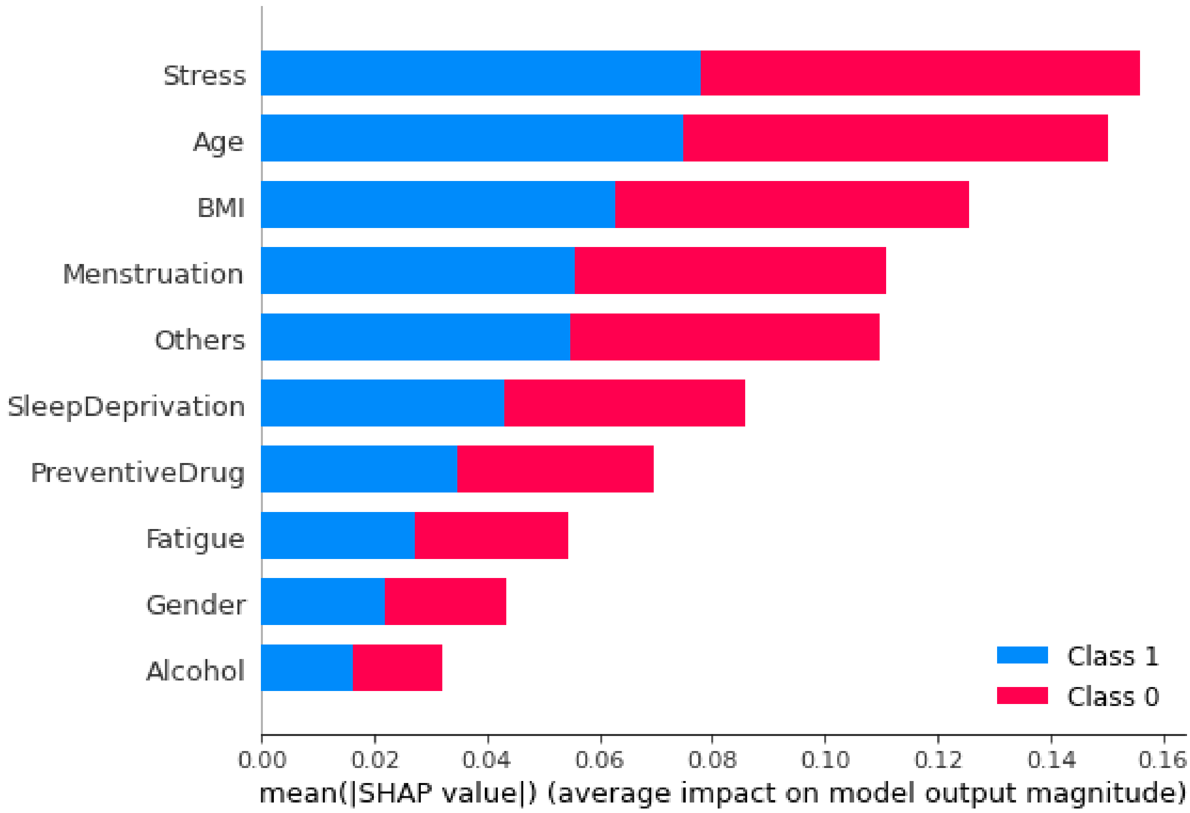

Figure 4.8: Model 4 feature importance by SHAP algorithm

In this study, SHAP is not just used to explain the importance of each feature in general but also to understand each feature's behaviour within the model in more detail. Model 4, which is the best model in this study, is selected and investigated in more detail with SHAP. In addition to the feature importance shown in the Figure 4.8, the relationships between the predictors and the target variable can be depicted both positively and negatively on the SHAP value plot.

In the summary plot (Figure 4.9), feature importance is combined with feature effects.

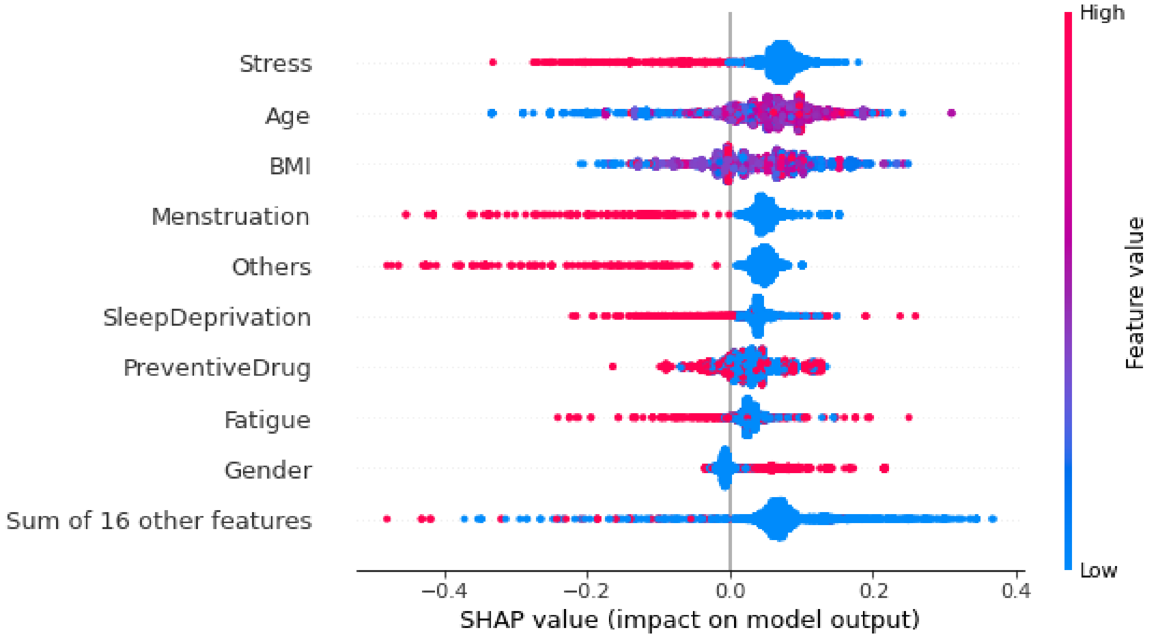

Figure 4.9: Model 4 summary plot by SHAP algorithm

Feature and instance Shapley values are displayed on each point of the summary plot. Features are located on the y-axis while Shapley values are on the x-axis. Feature values range from low to high according to colour (blue to red). The distribution of Shapley values per feature is exhibited by the scatter of overlapping points along the y-axis. The features are listed in ascending order of importance [17]. 'Stress' is the most corresponding feature to migraines, with a high value of 'Stress' clearly directing to migraine cases, while a low value of stress obviously leads to non-migraine cases. Similar to 'Stress', 'Menstruation' apparently differentiates between migraine and non-migraine cases, it has a positive impact on migraine attacks (a high value of 'Menstruation' causes migraine, while a low value of it leads to non-migraine cases). However, 'Age' and 'BMI' reveal themselves as the second and third most important features, but their values are not distinctly explained as to how they contribute to migraine attacks.

The feature importance information can be used while the data collecting process is executed. It allows us to understand which information is important and has significant impacts on the model. And in case there is a limitation to the data acquisition process where we cannot collect all triggers, this also shows which triggers we can exclude from our model.

CHAPTER 5

Conclusions

We can conclude from all the analyses above that SHD is an effective dataset to analyse migraine triggers. Random Forest, one of the most powerful machine learning models, is applied in 5 different ways. And we found that Model 4, the full model, which includes all features with hyperparameter tuning, provides the best result for the dataset used in this study. The model's performance is evaluated using accuracy, precision, recall score, and F1-score. And Model 4 demonstrates its performance at 95.13% of accuracy, 95.56% of precision, 94.66% of recall, and 95.11% of the F1-score. A 0.98 AUC demonstrates that the model can effectively differentiate between migraine and non-migraine attack cases. In addition to the migraine prediction model, this study investigates which triggers affect migraine attacks the most and how their impacts differ. As a result, we found that stress is the most important feature that corresponds to migraine, followed by patient age and BMI in the case that these information are provided.

For further study, there is only the Random Forest machine learning algorithm used in this study. Other machine learning models are available for comparison. In addition, deep learning algorithms can be used to perform this type of study if a bigger and more appropriate dataset is available. Furthermore, there are several approaches that can be used to deal with data imbalance issues. In this study, we only use the SMOTE approach, other approaches could be implemented to see if they produce significantly different results. In terms of the dataset, there are several patterns to combine or extract the features, and we can examine and benchmark the results to get the most accurate model.

APPENDIX A

Appendix

As mentioned in EDA for univariate descriptive statistics of patient information, in addition to the boxplot in Figure 3.1, the describe table and histogram are also applied, and the results are shown here.

Table A.1: Migraine patients information

	Age	BMI
Mean	37.71	22.41
Standard deviation	8.64	2.76
Min	20	16.4
Max	56	32.5

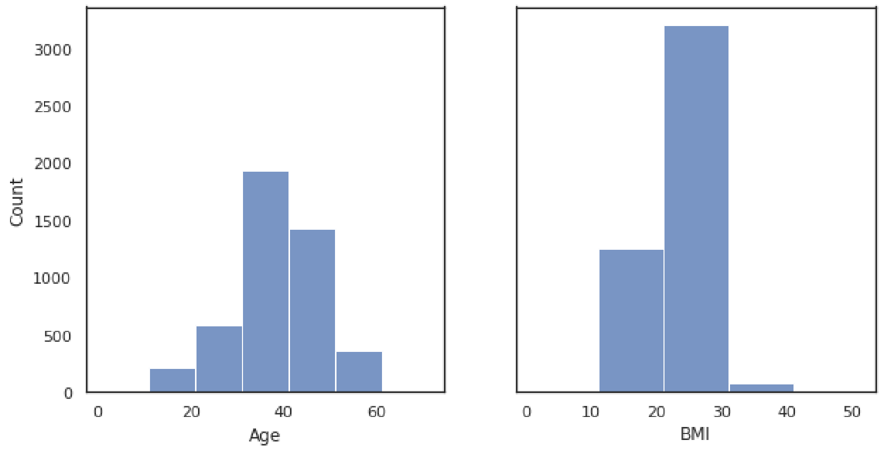

Figure A.1: Univariate descriptive statistics for Age and BMI

In addition to Figure 3.3 in EDA, the proportion between each level of each trigger and the migraine attacks are examined, and this Figure A.2 shows the example of the study.

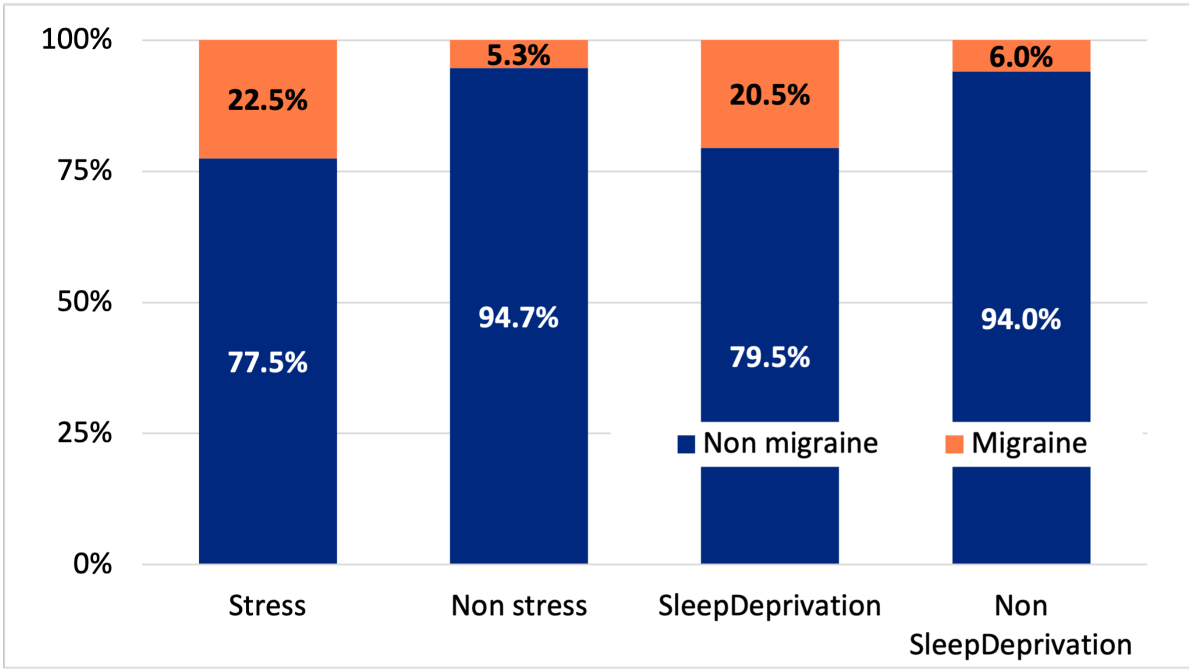

Figure A.2: Proportion of migraine attack by Stress and SleepDeprivation

The completed version of Table 3.1 illustrates here.

Table A.2: Completed table of migraine vs trigger association

Trigger	Migraine		Non migraine		Chi square	DOF	P_value
	W/T	W/O	W/T	W/O			
Stress	121	215	417	3826	203.35	1	$3.89e^{-46}$
Others	55	281	99	4144	184.43	1	$5.23e^{-42}$
Menstruation	62	274	139	4104	167.27	1	$2.93e^{-38}$
SleepDeprivation	82	254	318	3925	109.56	1	$1.22e^{-25}$
Odors	31	305	54	4189	103.78	1	$2.26e^{-24}$
Alcohol	21	315	29	4214	84.25	1	$4.37e^{-20}$
Fatigue	75	261	387	3856	58.36	1	$2.18e^{-14}$
EmotionalChanges	25	311	62	4181	56.55	1	$5.47e^{-14}$
Noise	27	309	88	4155	42.79	1	$6.08e^{-11}$
Traveling	11	325	21	4222	30.76	1	$2.92e^{-08}$
PreventiveDrug	174	162	1679	2564	18.78	1	$1.47e^{-05}$
InadequateLighting	4	332	3	4240	18.77	1	$1.48e^{-05}$
Fasting	30	306	171	4072	16.65	1	$4.49e^{-05}$
Overeating	15	321	74	4169	10.7	1	$1.07e^{-03}$
WeatherChanges	27	309	209	4034	5.54	1	$1.86e^{-02}$
Caffeine	8	328	199	4044	3.33	1	$6.80e^{-02}$
CheeseChocolateCons	5	331	24	4219	2.87	1	$9.02e^{-02}$
Overillumination	6	330	32	4211	2.87	1	$9.03e^{-02}$
Exercise	5	331	124	4119	1.85	1	$1.74e^{-01}$
Ovulation	7	329	53	4190	1.09	1	$2.96e^{-01}$
ExcessiveSleep	8	328	64	4179	1.02	1	$3.13e^{-01}$
Smoking	2	334	44	4199	0.25	1	$6.19e^{-01}$

This Figure A.3 is a full version of Figure 3.5. It demonstrates the correlation matrix of all migraine triggers.

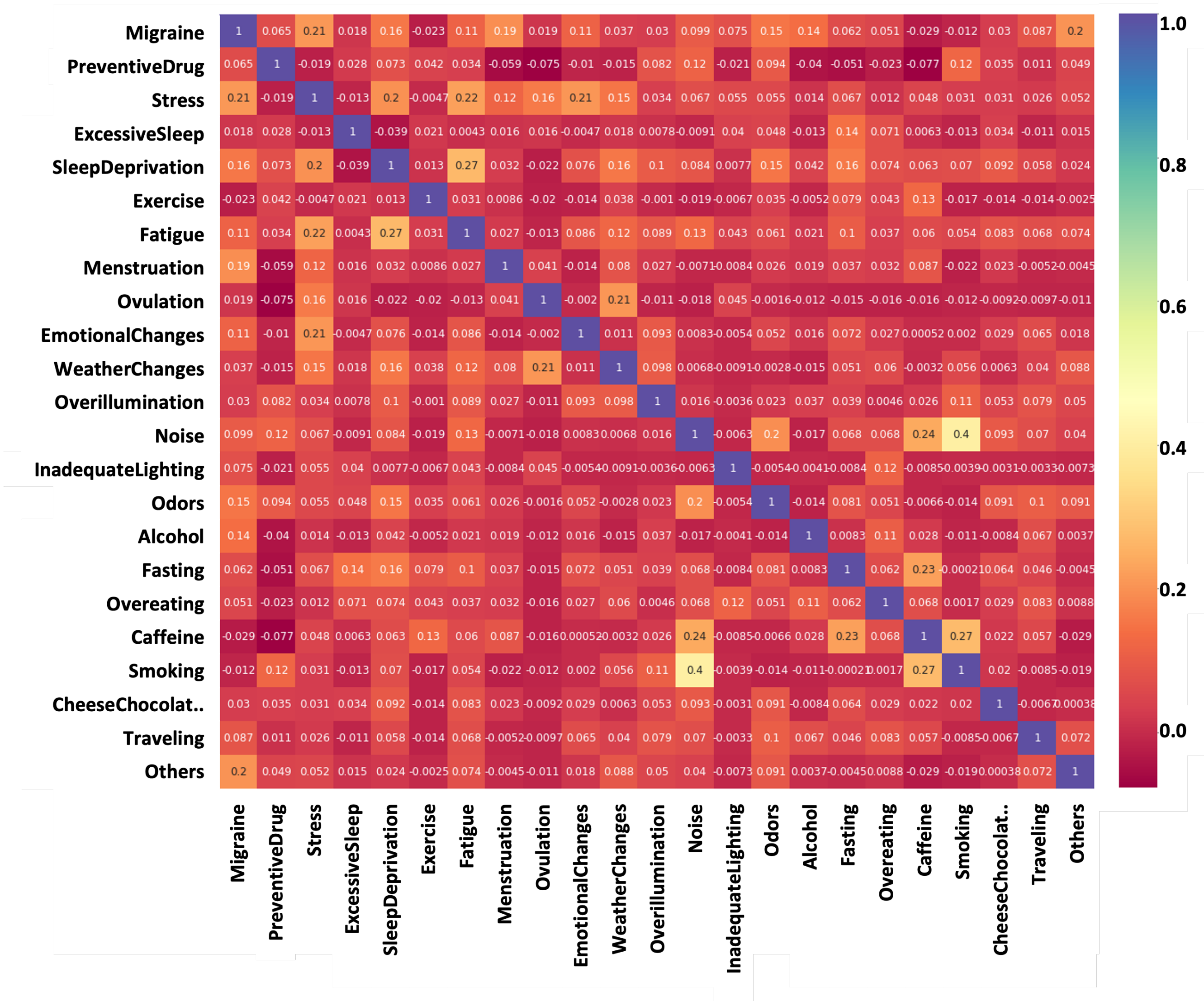

Figure A.3: Correlation matrix for all triggers

As cited in EDA area, this visualises the correlation matrix of combined features.

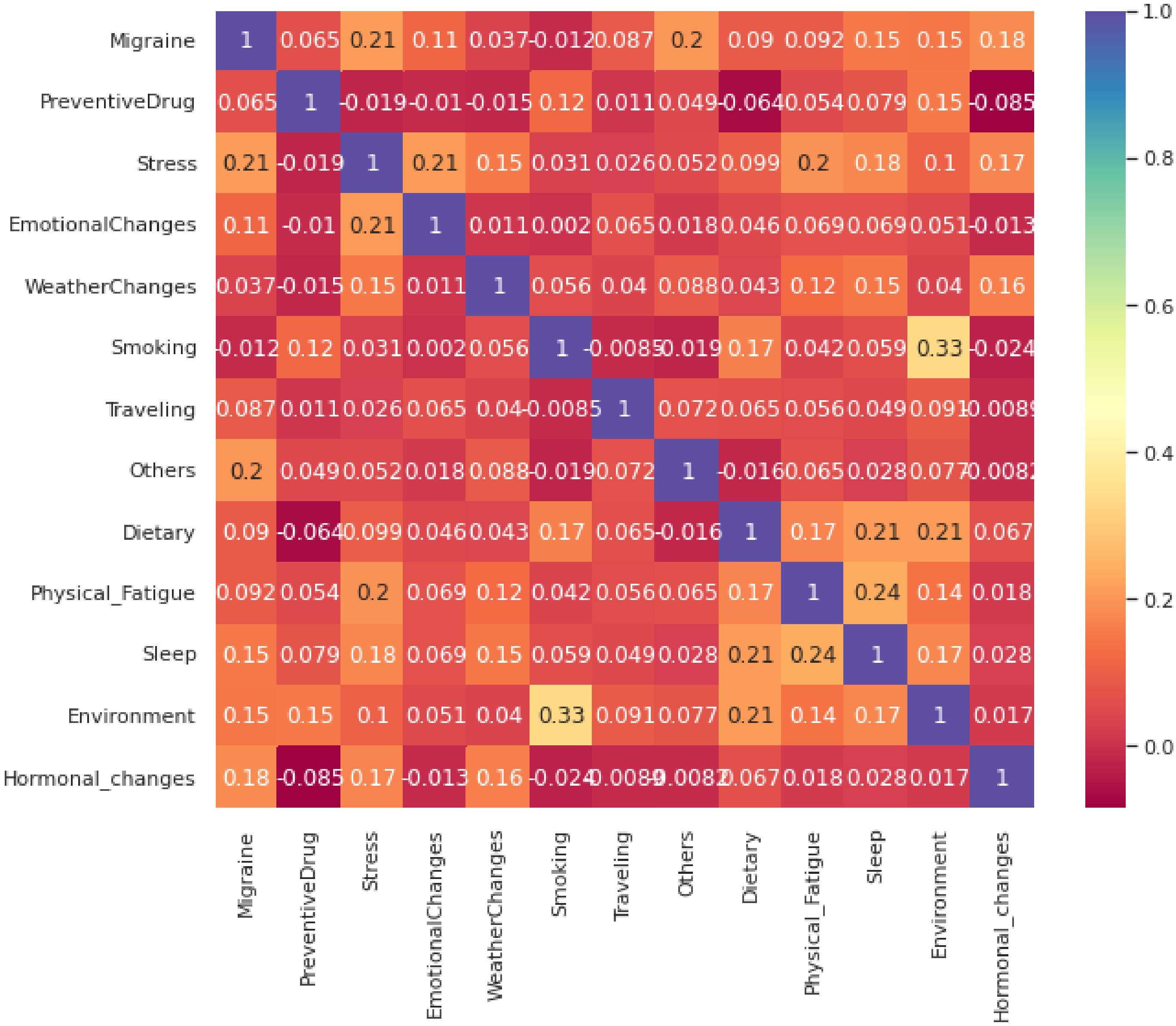

Figure A.4: Correlation matrix for combined triggers

The figure below shows the ROC-AUC result of the model 1.

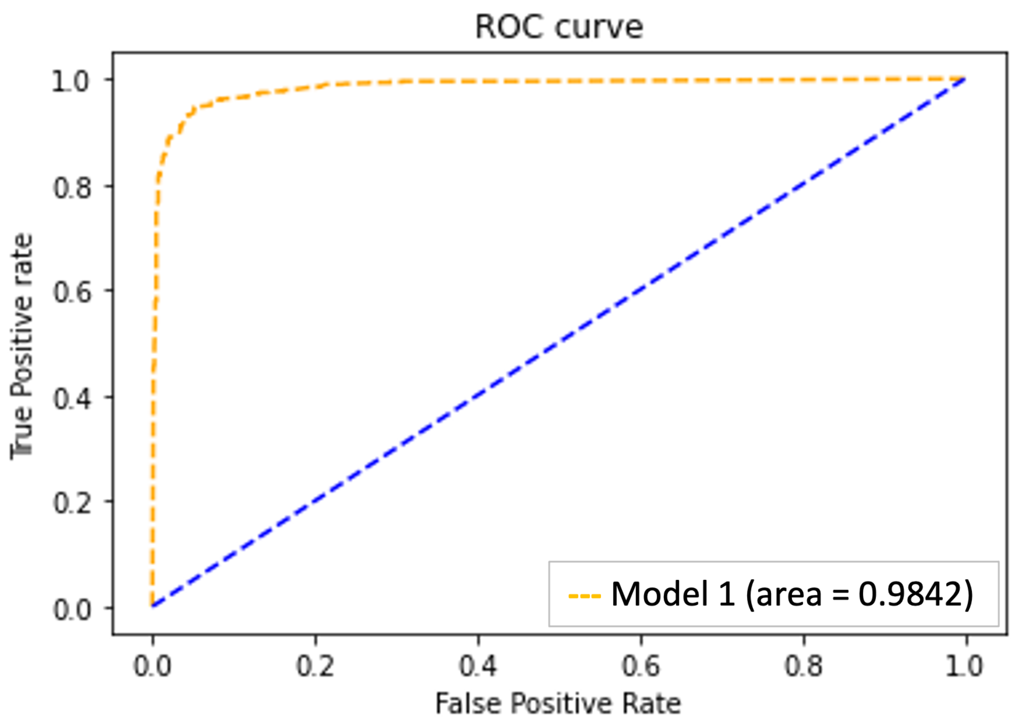

Figure A.5: ROC curve of Model 1

The the confusion matrix of the model 2 is presented below.

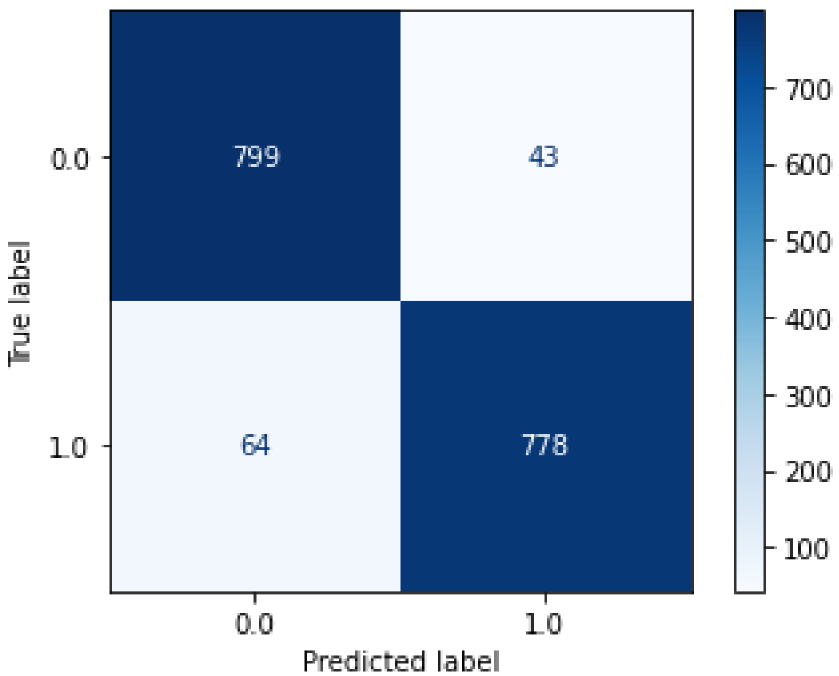

Figure A.6: Confusion matrix of Model 2

The figure below exhibits the confusion matrix result of the model 3.

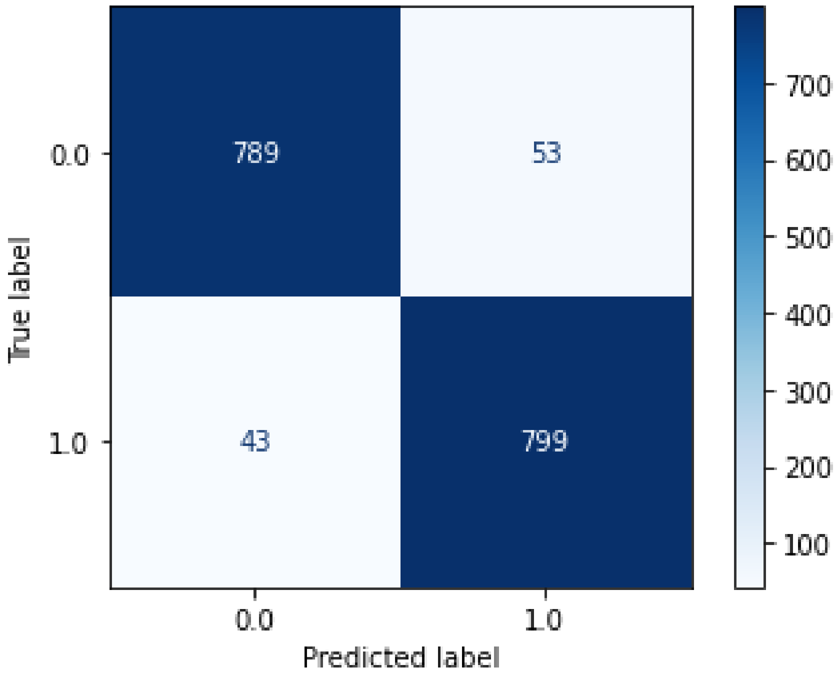

Figure A.7: Confusion matrix of Model 3

As mentioned in Results section, this figure illustrates the full version of Figure 4.8

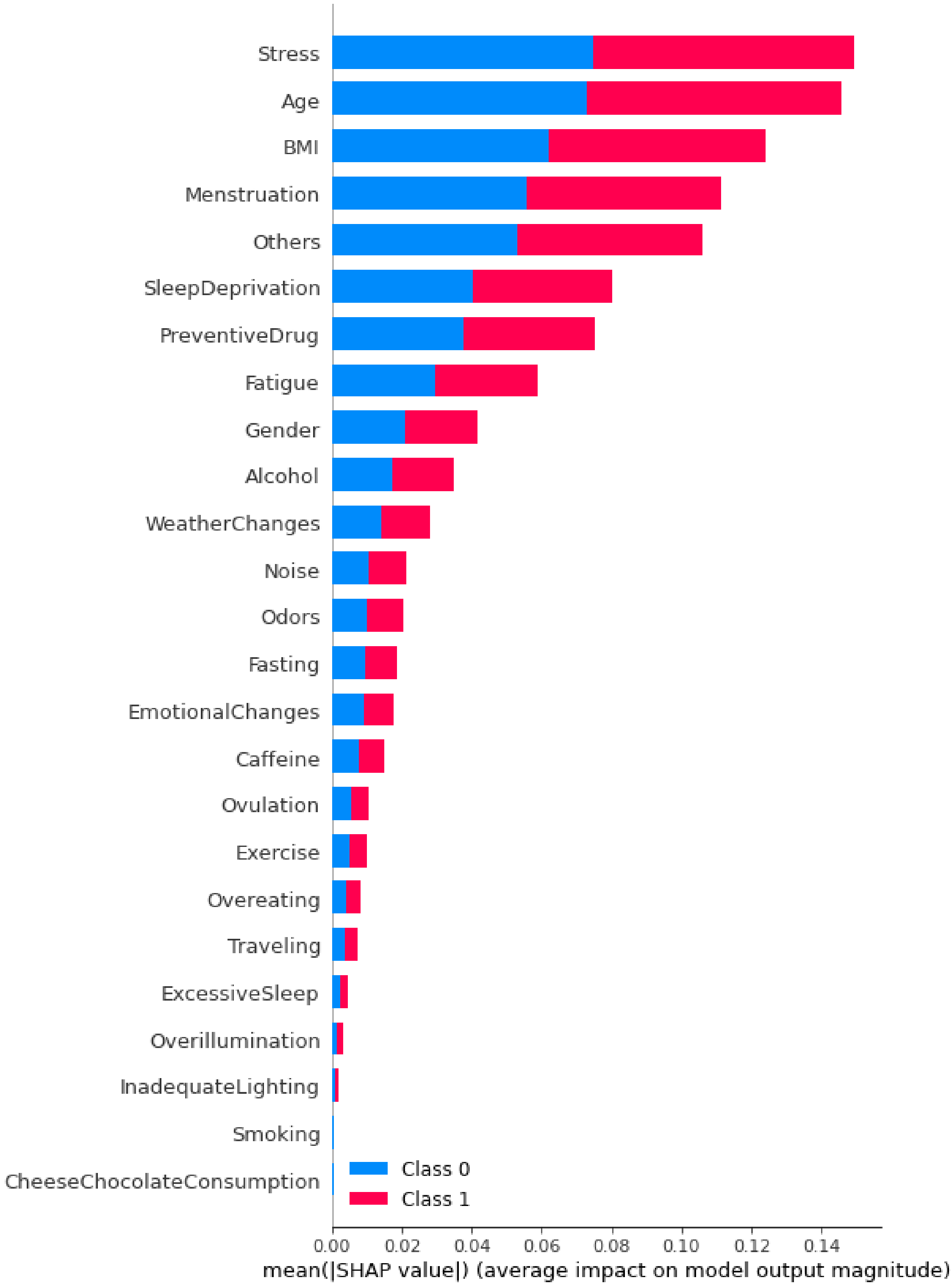

Figure A.8: Model 4 Feature importance by SHAP (full)

Bibliography

[1] The European Migraine Headache Alliance. Headache migraine - european migraine headache alliance, 05 2022.

[2] Pablo Aznar. Decision trees: Gini vs entropy quantdare, 12 2020.

[3] Aniruddha Bhandari. Auc-roc curve in machine learning clearly explained - analytics vidhya, 06 2020.

[4] Jason Brownlee. How to calculate precision, recall, and f-measure for imbalanced classification - machinelearningmastery.com, 01 2020.

[5] NHS Choices. Causes - migraine, 2022.

[6] Wikipedia Contributors. Precision and recall, 11 2022.

[7] Wikipedia Contributors. Receiver operating characteristic, 10 2022.

[8] Alberto Fernández, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of artificial intelligence research*, 61:863–905, 2018.

[9] Todd Michael Franke, Timothy Ho, and Christina A. Christie. The chi-square test: Often used and more often misinterpreted. *American Journal of Evaluation*, 33(3):448–458, 2012.

[10] Patrícia Timy Fukui, Tessa Rachel Tranquillini Gonçalves, Cintia Giunchetti Strabelli, Natalia Maria Fernandes Lucchino, Fernanda Cunha Matos, Juliana Pinto Moreira dos Santos, Eliova Zukerman, Vera Zukerman-Guendler, Juliane Prieto Mercante, Marcelo Rodrigues Masruha, et al. Trigger factors in migraine patients. *Arquivos de neuro-psiquiatria*, 66:494–499, 2008.

[11] GeeksforGeeks. Decision tree introduction with example - geeksforgeeks, 12 2017.

[12] Aurelien Geron. *Hands-on machine learning with Scikit-learn and Tensorflow : Concepts, tools, and techniques to build intelligent systems*. O'reilly Media, 2017.

[13] AW Hauge, M Kirchmann, and J Olesen. Trigger factors in migraine with aura. *Cephalalgia*, 30(3):346–353, 2010. PMID: 19614703.

[14] NHS inform. Migraine - symptoms, triggers and treatments, 2020.

[15] Syed Irfan. Hyperparameter tuning in random forest | by syed irfan | sep, 2021 | medium | geek culture, 09 2021.

[16] Eryk Lewinson. Explaining feature importance by example of a random forest, 02 2019.

[17] Christoph Molnar. *Interpretable machine learning : a guide for making Black Box Models interpretable*. Lulu, 2019.

[18] The World Health Organization. Headache disorders, 04 2016.

[19] Alessandro Panconesi. Alcohol and migraine: trigger factor, consumption, mechanisms. a review. *The Journal of Headache and Pain*, 9:19–27, 01 2008.

[20] Jeong-Wook Park, Min Kyung Chu, Jae-Moon Kim, Sang-Gue Park, and Soo-Jin Cho. Analysis of trigger factors in episodic migraineurs using a smartphone headache diary applications. *PLOS ONE*, 11(2):1–13, 02 2016.

[21] Jeong-Wook Park, Min Kyung Chu, Jae-Moon Kim, Sang-Gue Park, and Soo-Jin Cho. Analysis of trigger factors in episodic migraineurs using a smartphone headache diary applications. *PLOS ONE*, 11:e0149577, 02 2016.

[22] Piotr Płoński. Random forest feature importance computed in 3 ways with python, 06 2020.

[23] Sruthi E R. Random forest | introduction to random forest algorithm, 06 2021.

[24] Pekka Siirtola, Heli Koskimäki, Henna Mönttinen, and Juha Röning. Using sleep time data from wearable sensors for early detection of migraine attacks. *Sensors*, 18:1374, 04 2018.

[25] Jahnavi Thekkada. Linkedin, 2022.

[26] The Migraine Trust. Impact of migraine - the migraine trust, 02 2022.

[27] The Migraine Trust. What is migraine? - the migraine trust, 08 2022.

[28] Shaun Turney. Chi-square test of independence | formula, guide examples, 05 2022.

[29] Shaun Turney. Pearson correlation coefficient (r) | guide examples, 05 2022.

[30] Rosie Yuyan. The random forest algorithm for statistical learning - matthias schonlau, rosie yuyan zou, 2020.

[31] Rebecca Zhu and Yan Luo. Predicting migraines with machine learning and feature selection.

[32] Rebecca S. Zhu and Rucha Dave. Mygraine: Predicting migraines through various machine learning models utilizing user-inputted data. *International Journal of High School Research*, 2:65–71, 12 2020.