CA4016 Application Domains

Iya Chivileva Student № 19348896

Personal Data Assignment

Mood Analysis and Predictions.

Introduction

We live in a very dynamic and rapidly developing world. Our rhythm of life has become so fast and stressful that sometimes we do not notice how fleeting our life is.
In my project, I decided to try to answer one of the eternal questions. 'What is happiness?' 'What makes us happy?' In my opinion, this is more of a philosophical question, but I chose to approach this from a more scientific point of view.
I made a decision to tackle this question from the opposite direction and study this problem from a different perspective. What makes us unhappy? Or in other words 'What contributes to the deterioration of our mood?'. My search led me to such terms as depression and anxiety.
One of the most fashionable and popular terms nowadays is depression. It is considered to be a disease of the 21st century. But is it just a trendy word?
Today approximately 300 million people worldwide have been diagnosed with depression. Depression is distinct from common mood fluctuations and short-lived emotional responses to challenges in everyday life. Especially when recurrent and with moderate or severe intensity, depression may become a serious health condition. The affected person may experience severe suffering and function poorly in his daily routine. Suicide can result from depression at its worst. Every year, over 700 000 people die due to suicide. For people aged 15 to 29, suicide is the fourth most common cause of death.
According to experts (World Health Organization, 13 September 2021) over the next fifty years, depression is projected to come out on top in terms of prevalence, ahead of cardiovascular disease.
It turned out that my chosen topic is more important than I initially anticipated. Since depression is a very common condition that can affect any of us, irrespective of gender, age or background I would like to discover this topic more. I believe that depressive state is often underestimated and many people do not even consider it as a disease. In my project, I contribute to find out what depression is, from a scientific point of view and what factors from our everyday life can positively contribute to the onset of depression.
After discovering the definition of depression, I wanted to learn more about its symptoms. Depression is a serious medical condition that can negatively affect your thoughts, motivation, emotions, and behaviour. It is a mental health condition.
I found following the most common symptoms of depression (Support line - depression support (2021) Aware).

1. Depressed mood most of the day, nearly every day.
2. Significant (≥5%), unintentional weight loss or weight gain or loss of appetite.
3. Insomnia or hypersomnia.
4. Fatigue or loss of energy.
5. Feelings of worthlessness or excessive or inappropriate guilt.
6. Decreased ability to concentrate or think.

7. Decreased interest or pleasure in all or most activities.
8. Psychomotor agitation or retardation.
9. Recurrent thoughts of death, recurrent suicidal ideation and/or suicidal plans.

Early detection and consistent care are essential to a successful outcome in the case of recovery. Treatment of profound and chronic depression is often very difficult, requiring in most cases inte nsive psychotherapy.

Personal data sources

Talking about myself, I have never been diagnosed with depression. At the same time, like everyone else, I sometimes experience an anxiety or oppressed state. In this project, I would like to identify factors and trends that can affect the emotional condition based on my own example. So, let's start with a discussion on what can affect the deterioration of our mood.
These are key factors that can lead to bad mood:

- Stress
- Lack of sleep (tiredness and overwork)
- The news
- The weather
- Hormonal changes (menstruation, pregnancy, puberty and menopause)
- Drugs and alcohol
- Poor nutrition (any changes in diet)
- Medication side effect
- Physical illness or chronic pain
- Lack of physical activity

After reviewing the above list, I analysed the various applications and resources that I could use to collect my personal data.

1.At first, I decided to use an application with which I track my menstrual cycle. It is called Clue. (Clue: *Period and Ovulation Tracker for iPhone and Android. Clue Period & Ovulation Tracker with Ovulation Calendar for iOS, Android, and watchOS. 2012)*
I decided to use data from the past three months. My start day is 01.09.2022 and my last day is 15.11.2022. Thus, in total, I have 76 days to work with.
With this application, you can track any hormonal changes, including period, pain, sexual activity, energy levels and emotions.
I obtained my target attribute from this application. I was tracking my emotional condition on daily basis. There are 4 categorical features representing my mood state: “Happy”, “Sensitive”, “Sad” and “PMS”.
In the case of happiness and sadness, everything is clear. I was tracking ‘Sad’ days when I felt a bit down or was just tired. I noted the sensitive state of the days when I felt especially vulnerable. The state of PMS is a very unique emotional state and females usually experience it a few days before their following menstrual cycle. During PMS days your mood is quite changeable during the day.
If I had to encode these features in an ordinal way, they would look like this:
[0, 1, 2, 3] = [“Happy”, “PMS, ”Sad”, “Sensitive”].
Thus, in this assignment, I say that the state of sensitivity is the closest to depression.

The exported data from the application was completely unprepared for processing. I found an open-source Python code on the Internet to convert it into an appropriate format.

2. The second app I decided to use was FatSecret (Your key to success. FatSecret 2007).
I am taking care of my health and nutrition. I am quite selective in my diet and the products that I consume. With this application, I monitor the number of calories I eat per day and the percentage/grams of fats, proteins and carbohydrates.
Exporting data from this application was as fast and convenient as possible. Every month I had to export separately. All the spreadsheets were immediately in CSV format. Moreover, the data was quite extensive and contained 11 columns: Cals (kcal), Fat, Sat, Carbs, Fiber, Sugar, Prot (all in grams), Sod, Chol and Potassium (in mg). Since the last three columns didn't have data for all days I decided to exclude them straightaway.

3. The third application that I was using as my data source is Apple Health. (*Ios – Health. Apple December 6 2022)* I consider myself addicted to the phone and I am practically inseparable from it. It is always by my side. That's why I expected my Apple Health data to be the most informative and valuable.
Unfortunately, Apple's health data turned out to be the most difficult to process and export. I again had to find Python code to convert it into a readable format.
After I processed it, I found out that the only attribute that I have there is "Distance walked" and 'Active Energy Burned'. The second attribute did not have data for every day. It was gathered from another application that is connected with my Apple Health. It is called 'Goals-Fitness'. I use this application to track my jogging workouts progress. I am not using this application in this project as I only have a couple of days of worthy data within that application. I will be considering these two columns from Apple Health as my physical activity.
I also found that I have sleep analysis on my phone and my screen time data. But to my disappointment, I was not able to easily export this data. I researched this question and apparently there is no tool for screen time or sleep data extraction from an iPhone. Moreover, screen time data is available on an iPhone only for the past 3 weeks, so I had to store it manually for the project. Thus, I had to fill in sleep and screen time attributes manually. If my data was bigger it won't be a realistic method to gather these kinds of data.
4. The fourth data source became my smart scale. As I mentioned before, I adhere to a proper lifestyle and nutrition, respectively, I carefully monitor my weight. For this, I use Repho Health (*Renpho: Smart healthy living* 2016). application. This app supports very simple data export. It was straight away in a clean csv spreadsheet format. Data was presented by 13 columns. Weight, BMI, Body Fat (%), Free-fat body weight (kg), Subcutaneous Fat (%), Visceral Fat, Body water (%), Skeletal Muscle (%), Muscle Mass (kg), Bone Mass (kg), Protein (%), BMR (kcal), Metabolic Age and Date. As "Metabolic Age" was a constant number for all days I excluded it from my data.

5. Since Stress is one of the key factors that can affect our state of mood, I decided to use my Google calendar to manually create a Stress attribute. I decided to make a binary "Stress" column (1;0). The 1 represents days when I had any work or college assignment deadlines. As I am looking at the data for the past 2,5 months only I remember certain events and days when I experienced high levels of stress. They were associated with news and quarrels with my closest people.

6. Another attribute that I also made up manually was “Holiday / night out”. It is also a binary data column. I created it also using my Google calendar and my gallery data. I represented days when I had time out with ones in my data.

I realize that if I had to work with data that included a larger number of days, this would be quite problematic, so for a future commercial implementation, another approach is required to collect these types of features (5 and 6).

7. The last data source that I used was Met Eireann's (*Historical data Met Éireann)* official website to collect weather data. I was exporting historical weather data from Dublin Airport station, as it is the closest to me. It was a very big dataset as starting day was 1st January of 1942. It has more than 20 columns. Rain (mm), Maxtp and Mintp (C), gmin, soil, wdsp, hm, ddhm, hg, cbl, sun, g_rad, pe, evap, smd_wd, smd_md, smd_pd and ind. All of these attributes are daily weather data. I explained each of them in more detail in my Jupyter Notebook.

Data cleaning and engineering

A detailed description of the data as well as my code and more results of the experiments can be found in the Jupyter notebook **https://github.com/Chiviya01/Mood-Analysis-/blob/main/Mood_Analysis.ipynb**

I merged all the exported data on the day attribute. I started with the general analysis of my data. My merged dataset contained 46 columns and only 76 rows. I had 5 categorical columns and the rest were numerical.

The first step I took was encoding my categorical attributes into numerical. I used Label Encoder to do it.
I also had 7 columns that I obtained from my FatSecret up, that had missing values since I wasn't consistent in filling in my food consumption. Since I forgot to fill in my food consumption for a couple days. My diet is quite of the same type. So, I decided to fill in missing values through mean, standard deviation and then data normalisation.
I had some missing days in 'Screen time' column, so I filled them in in the same way as my missing food consumption data.
One of the columns was 'Active Energy Burned', I got it from my Apple Health data, but it was generated through another application that I am using to track my jogging workouts. This application is called Goals-Fitness and it is connected to my Apple Health data. So, the day when I performed a workout I had Active Energy Burned represented by the number of calories burned in my Apple Health. I did not have this kind of data for every day, so I decided to encode this column into a binary one. It became a [0,1] column where 1 stands for 'went for a jog'.
The last change that I made to my dataset was adding a new feature, which was 'Week day'. I obtained this feature with help of the daytime python library. I did it because I believe, that our mood might also be dependent on the current day of the week, as many people feel more stressed and gloomy on Monday when by the end of the week our mood usually improves.

As previously mentioned, my dataset contained data only for 76 days. It is impossible to train an accurate model on such a small amount of data. I decided to synthesise more data. I used Synthetic Data Vault (SVD) python library to generate more data instances. Surprisingly, when I was running predictions on the extended dataset I got a worse result than on my original dataset. I concluded synthesising more data approach as inefficient and therefore I concluded to work only with my original data.

For choosing the most appropriate model I decided to use PyCaret. It is an AutoML python package. It is very handy as it automatically pre-processes and transforms data for you. It also runs multiple ML models at the same time and returns the table with their performance results. I am providing the results obtained from PyCaret. The best model turned out to be Gradient Boosting Classifier (GBC). Since my dataset is very imbalanced the best metric to use for model evaluation is F1 score. From the table below it can be seen that the initial model F1 score is about 50%, which I was not happy with.

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
gbc	Gradient Boosting Classifier	0.5467	0.2281	0.4944	0.5039	0.5021	0.2862	0.2997	0.364
lr	Logistic Regression	0.5600	0.2703	0.5500	0.4781	0.4882	0.3622	0.4111	0.026
ridge	Ridge Classifier	0.4400	0.0000	0.4292	0.4103	0.3935	0.2036	0.2271	0.014

Figure 1: The best models according to Pycaret

This is my initial confusion matrix. It can be seen that '1' (PMS) and '2' (Sad) classes were fully misclassified. I wanted to improve my results and I completed multiple experiments.

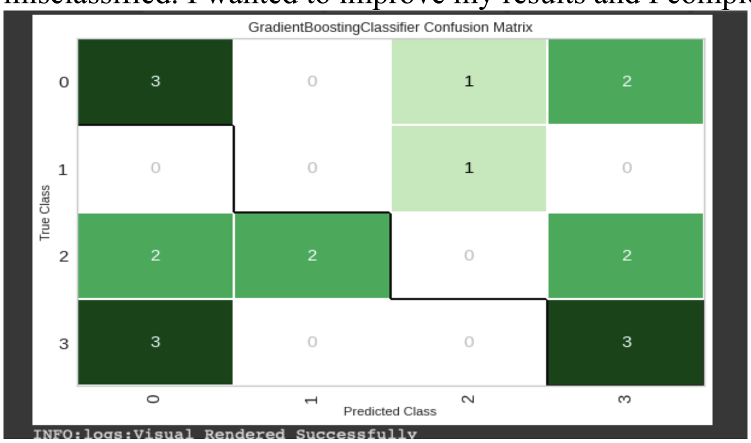

Figure 2: Gradient Boost Classifier confusion matrix – PyCaret

Due to the low accuracy of the model, I decided to research more into Gradient Boosting models. I found out that eXtreme Gradient Boosting classifier usually shows more accurate results than Gradient Boosting Classifier.
For comparison, after running these two models on my initial dataset my F1 score for GBC was 0.210 and for XGBoost it was 0.315.
XGBoost applies advanced regularization which improves model generalization capabilities. I concluded to use XGB model in my further research. To maximise the results of my XGBoost model I needed to improve my data quality.
One of the biggest challenges that I came across is the imbalance of my dataset. My target column – 'mood', had only 6 instances of PMS. Below I am providing the value count of my target column before resampling.

```
happy          33
sensitive      19
sad            18
pms             6
```

Figure 3: Value counts for ‘Mood’ – target column. Imbalance data challenge.

In order to solve this problem, I used the Nearest Neighbour method to resample my dataset. In the end, I Normalised my dataset and applied Principal Component Analysis (PCA). Another challenge was that many of my features were correlated between themselves. Like ‘Body fat %’ and ‘Weight (kg)’ and ‘Muscle Mass (kg)’. Because of this, my data had a lot of multicollinearities. I decided to reduce the number of used features. I wanted to verify the least important features with two approaches. The first one was Variance Inflation Factor (VIF) and the second one was Feature Importance obtained from my model.

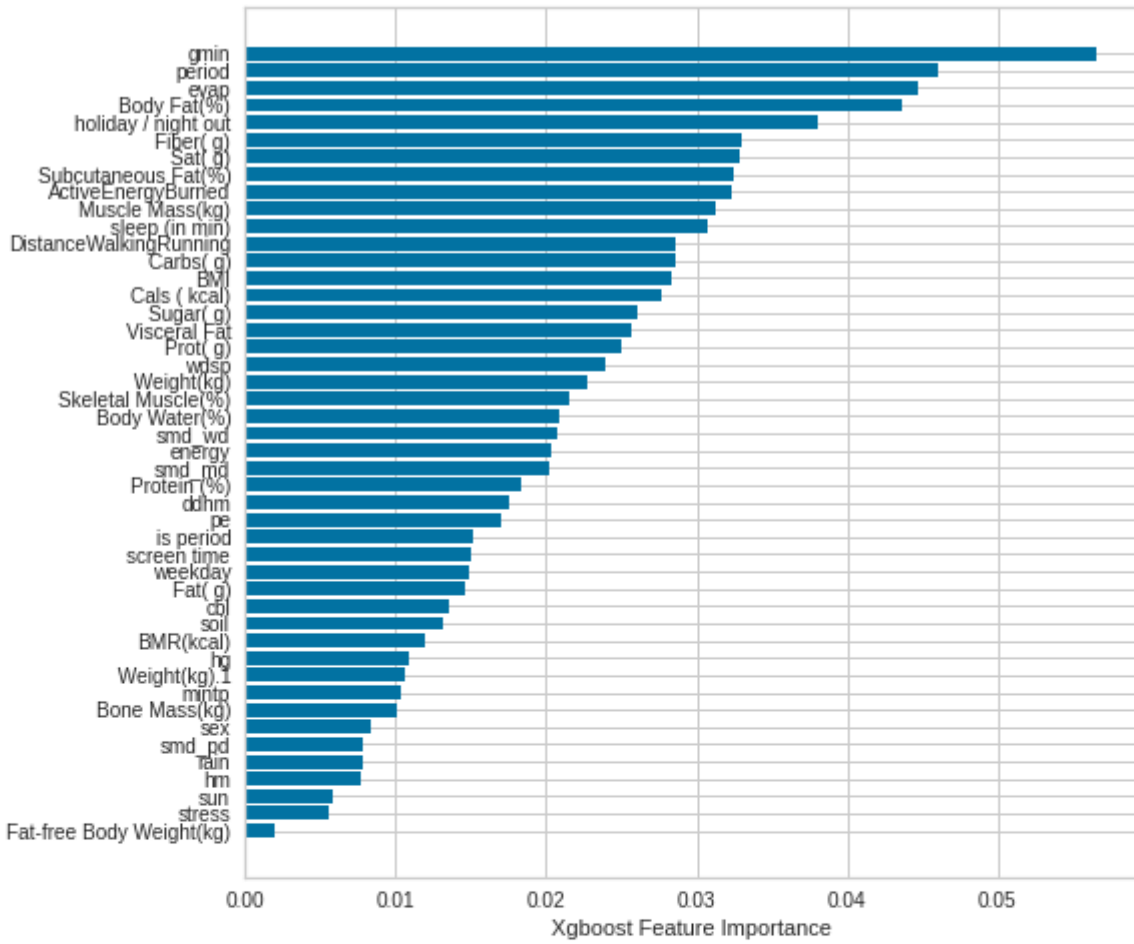

Figure 4: Feature importance according to Xgboost classifier

I had a high interest in the above graph. It basically displays the most important factors that affect my mood.
From the above plot, it can be seen that the top attributes that affect my mood are: grass minimum temperature (gmin); my menstrual cycle (period); holidays or nights out; the food that I consume (Fiber and Sat). I was not very surprised by the results. Things like the weather and

my menstrual cycle, which, are out of my control. I decided to analyse more deeply factors that are manageable.
At this stage, I got really curious about how my nourishment affects my mood.

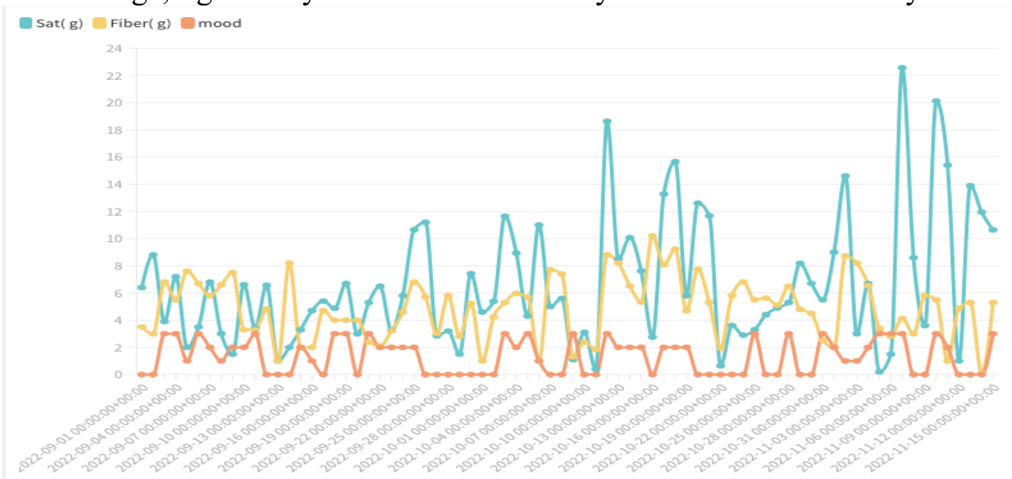

Figure 5: Daily trends of Mood state vs consumed Fiber (in grams) and Saturated Fat (Sat) (in grams).

From the above plot, I found out that my emotional state tends to worsen when I consume less Fiber and more Saturation fat. I conducted small research and the key findings are

1. According to the study (Ph.D a.Takeshi Kochi, M.D. cKeusuke Kuwahara, Ph.D adRei Ito 2015) performed by The Furukawa Nutrition and Health Study in Japan. Increased consumption of dietary fiber derived from fruits and vegetables was associated with significantly reduced risk of depression.
2. According to a Professor in the Department of Nutrition of Université de Montréal's Faculty of Medicine Stéphanie Fulton (*Stephanie Fulton 2018)* – the anxious, depressive and compulsive behaviours as well as metabolic changes were observed with a diet rich in sugar and saturated fat.

Thus, scientific articles and theories only confirmed my own results obtained from the graph of the influence of the amount of consumed Fiber and Sat as one of the key factors affecting my emotional health.
Combining the results obtained from VIF and the above feature importance plot I removed the least important features and retrained my model on my normalised and balanced dataset. I also tuned the model's hyperparameters with Grid Search. The results got slightly better. Now my resampled confusion matrix looked like this.

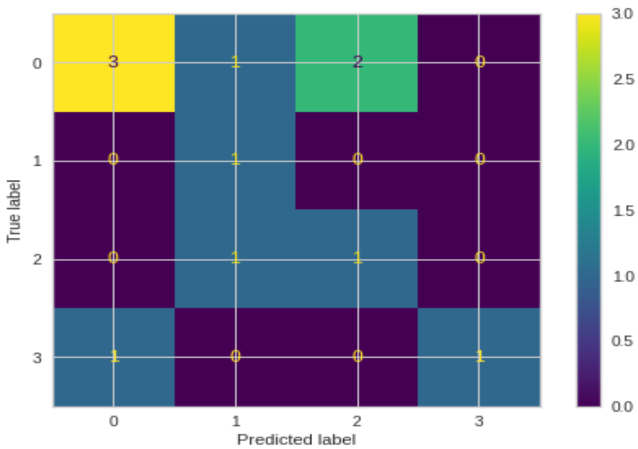

Figure 6: XBG confusion matrix on improved dataset and tuned model.

It can be seen now that even having only one sample example of class ‘1’ (PMS) this time it was predicted correctly. Classes ‘0’ (Happy), ‘2’ (Sad) and ‘3’(Sensitive) have 50 % accuracy. These results are better because now each class has at least half of the correct predictions. The final F1 score now is equal to 54.25%.
The summary of final results.

```
0.5454545454545454
              precision    recall  f1-score   support

           0       0.75      0.50      0.60         6
           1       0.33      1.00      0.50         1
           2       0.33      0.50      0.40         2
           3       1.00      0.50      0.67         2

    accuracy                           0.55        11
   macro avg       0.60      0.62      0.54        11
weighted avg       0.68      0.55      0.57        11
```

Figure 7: XBG results on improved dataset and tuned model.

From Figure 7 it can be seen that comparing my initial results with all approaches that I took to improve my data quality and my model performance I had succeed in obtaining a better result. I believe with bigger data I would be able to improve my results even more.

Visual data analysis

After the first discovery, I made as a result of visual data analysis, see Figure 5, I made a decision to compile more visual analysis of my data.
Due to the fact that my data contained too many columns, the visual representation of the correlation of all of them with each other looks very overloaded and was not very informative. I have it provided in my Jupyter notebook. I decided to consider only the most interesting features for me.

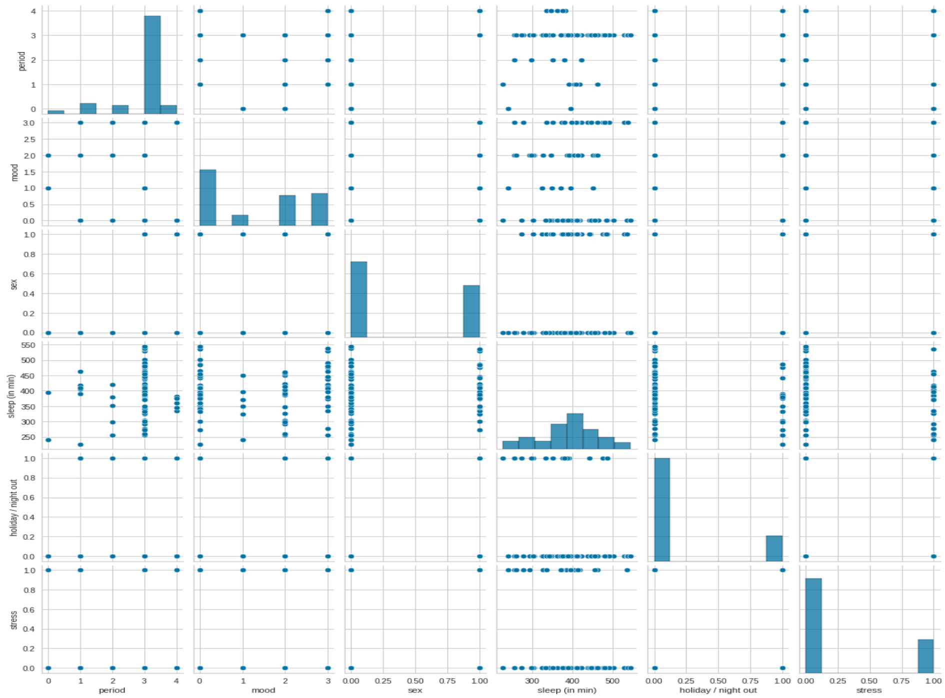

Figure 8: Features correlation.

It can be seen from the plot above that all the distributions are linear but it is difficult to identify any trends or dependencies from it. I decided to map smaller, several visualizations that would include the most related attributes.
From Figure 4 (the feature importance plot) the most important feature that affects my mood the most was gmin.
In the following graphs, I will be treating any mood that is not equal to 0 ('happy') as abnormal.

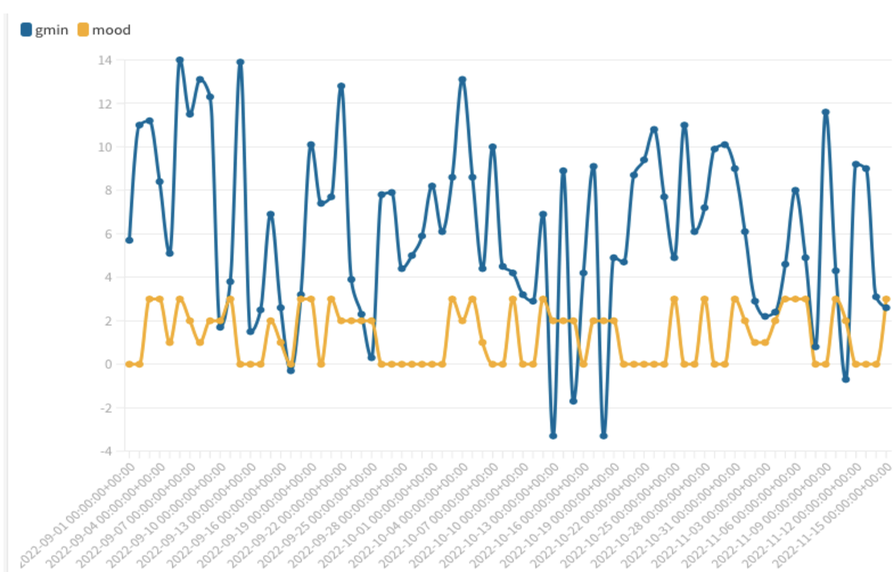

Figure 9: Daily trends of Mood state vs Grass Minimum temperature.

One thing that can be noted from the plot is that there is a tendency when the changes in the weather are substantial my mood raises from 0 to 2 or 3. Moreover, when the weather was at its lowest, my mood was equal to 2. I researched this question in more detail.

1. A recent paper (M.J.T.W. December 2019) called "Temperature and mental health: Evidence from the spectrum of mental health outcomes" revealed that lower temperatures reduced negative mental health outcomes while higher temperatures increased them. Hotter temperatures were associated with visits increase in emergency room for mental illness and also higher rate of suicides.
2. Seasonal Affective Disorder (SAD) is a major depressive disorder with seasonal pattern. This is one of the forms of depression that generally starts during the fall season as light diminishes. SAD worsens in the winter periods and trends to occur again at the same time annually. It was proved (Melrose, S. 2015) that SAD is mostly due to the lack of sunlight. While Sad appears more often during winter seasons it is not due to the cold temperatures. Research has found that people with SAD symptoms often have low vitamin D levels.

Since it turned out that the temperature has a direct effect on my mood state I chose also to plot a graph for my mood states compared to sun and rain. Sun represents here the sunshine duration (hours) and rain represents precipitation amount (mm).

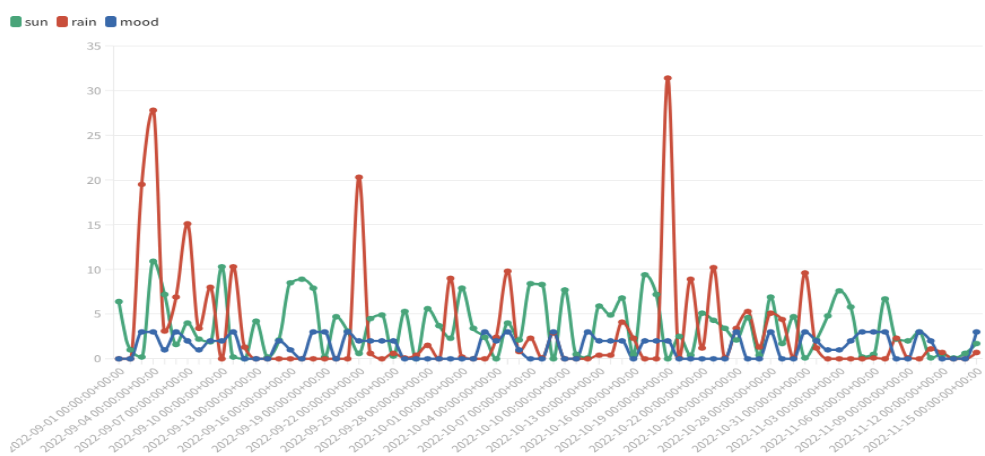

Figure 10: Daily trends of Mood state vs the amount of Sun and Rain.

Unfortunately, from the above plot, I wasn't able to highlight anything.
Very often doctors prescribe to their patients to spend more time outdoors in fresh air. That is why I came up with idea to plot a graph to establish any dependencies along my mood state and the quantity of Distance Walking / Running. I also included Active Energy Burned, my jogging activity that mentioned previously.

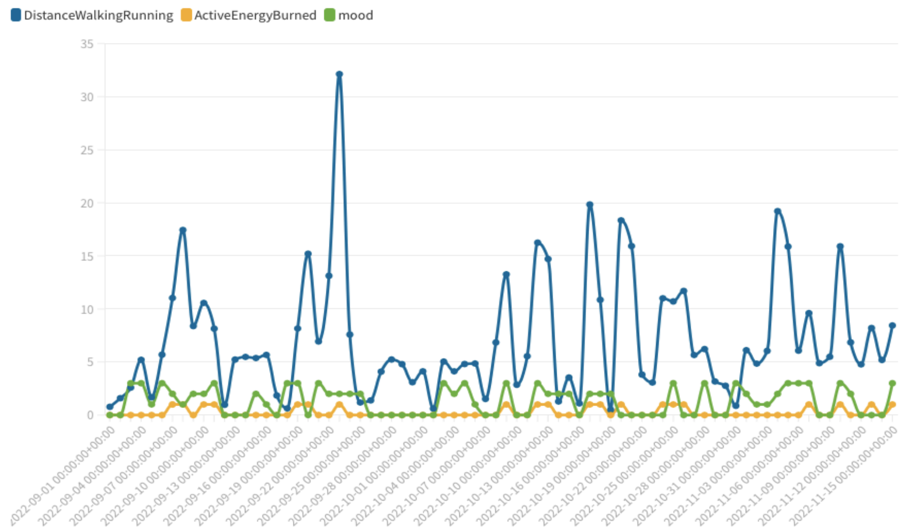

Figure 11: Daily trends of Mood state vs Distance Walked (in km) and Outdoor workouts.

From the graph above, you can see the correlation between my mood and the number of kilometres I have walked. Moreover, it looks like the mood line repeats the amplitude of the line of the distance walked. Upon further review, I can easily explain it by the fact that when I feel down I like to go for walks or jogs. I'm not sure what insight I could get from this for myself, but it is definitely a good aspect for the ML model to train from.
And again, I decided to research this question from a more scientific standpoint. As one of the studies (Matthew Pearce, P.D. 2022) claims that adults who did activities equivalent to a minimum of 1.25 hours of walking per week had an 18% lower risk of depression compared with those who did not exercise. Moving up to an “activity volume equivalent to 2.5 hours of walking per week was associated with 25% lower risk of depression,” the study authors said.
Talking about myself I always feel way better after a walk. Given my search results, I use the right method to improve my mood.

To make the charts less overloaded, I decided to turn my data range from daily into weekly. To reshape the columns, I used mean or mode. I got 11 weeks and the following charts.

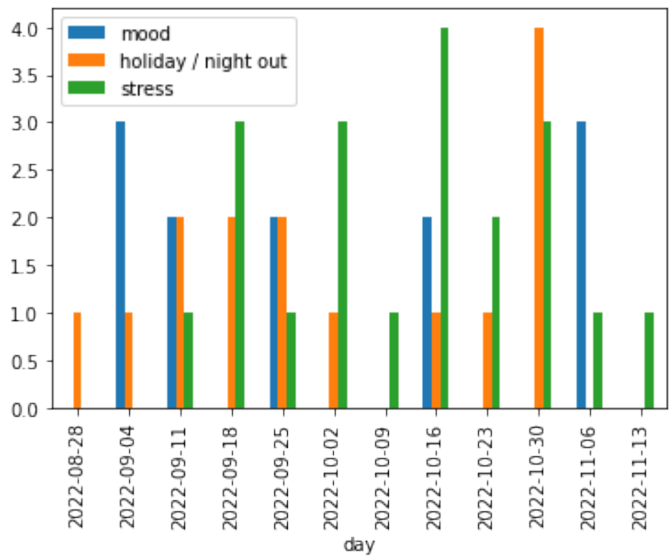

Figure 12: Weekly trends of Mood state vs Stress level and Times out.

I started by plotting the chart from the most obvious factors that affect our usual emotional state: stress and time out. I've noticed an interesting thing that on the weeks where my time spent out was higher my mood is usually considered as 'happy'. This can be explained by the fact that socialization reduces the symptoms of depression and makes us happier.

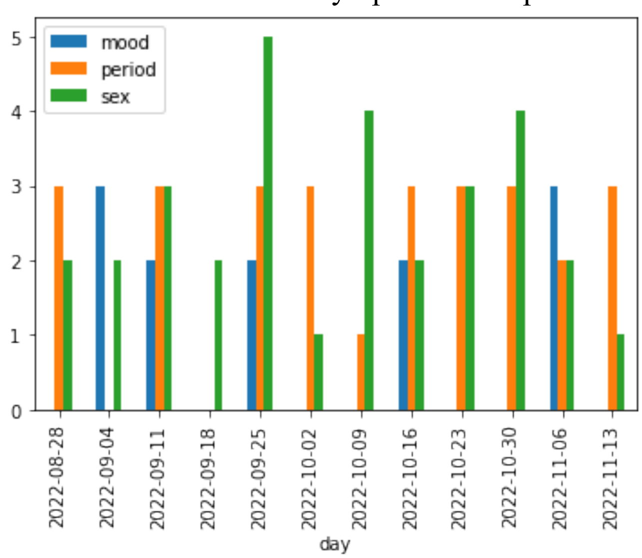

Figure 13: Weekly trends of Mood state vs Menstrual cycle and Sexual Activity.

The above graph represents the relationships between my state of mood with my menstrual cycle and sexual activity. From the above chart, I wasn't able to spot any correlations.

Commercial potential

The initial idea of my project was to find out for myself what factors can affect my mood. Can I change anything in my lifestyle to become happier? After researching the problem in more detail, I came to the conclusion that a similar application could be very useful to many people. In particular potential customers would be people who like using tracking apps and devices and who suffer from poor mental health often. I believe it would be utilized mostly by the younger generation in the range of 15-35 years old. The application will also let its potential customers choose and give access to their data from other tracking applications and devices that are being used, such as Health, Weather, smart watches and others.

As mentioned in the introductory part of this report, most of us have experienced symptoms of depression, which can range from simple mopes to serious mental illnesses. Our mood affects our productivity at work or school, our communication with others, and of course our health. Such an application could be useful for everyone, however, let's not forget that each person is unique and everyone's mood can be influenced by completely different factors. In my example, I tried to cover the most popular factors that affect our mood as much as possible. Due to the limitations of my data, since it was only 76 days, I was not able to achieve highly accurate results, my accuracy is only about 60 percent. However, I think with a bigger dataset obtained from multiple users, if also considering some biological characteristics and previous medical history, I think it would be possible to train a more accurate model.

The main focus of such an application would be to predict and monitor the current and future emotional state of its users. The application will only process the data that the user provides it with. The application will then send certain in-app messages to the user after personalizing their data. For example, individual insights, reminders or suggestions regarding information the user might find useful in their particular example. The main goal of such application would be to provide some insight in order to improve the user's mood. For example, for one user it might be good to exclude or, on the contrary, add something to their diet, for someone else it might be advisable to change their sleep schedule, for another the recommendation could be spending more time outdoors.

This application should be primarily aimed at identifying the user's potential depression. This could be done if the number of days with a negative mood increased. In this case, I would suggest conducting and offering the user more serious options to improve their health or prevent its worsening. For example, provide telephone numbers or mental health services of local care centres. Thus, the necessary assistance could be provided right there and then, preventing possible deterioration and future consequences. Seeking professional help early can save life.

I believe that even living in this digital and informative century, many people are still misinformed. Sometimes when you feel depressed, you just don't know where to seek help. Moreover, many of us ignore and neglect our mental state and health until the situation becomes critical. Living in the age of technology, maybe something as simple as a recommender application could solve some of these problems. However, it is important to remember that depression cannot be cured through mood tracking/managing because it is a neurobiological illness that can sometimes only be helped with proper medication.

I investigate the market and I managed to find out that such an application does not exist yet. There are only self-care journals provided for mood tracking. All mood prediction experiments that I looked at are usually associated with only one attribute. I failed to find an integrated approach for mood prediction based on multiple features. I believe that this app can be the first in the market and be a breakthrough.

I would also like to empower users to learn more about mental illness. I would include a content tap in my application where all users can get a possibility to educate themselves. A short material about various mental diseases, their symptoms, consequences and methods of treatment would be provided there.

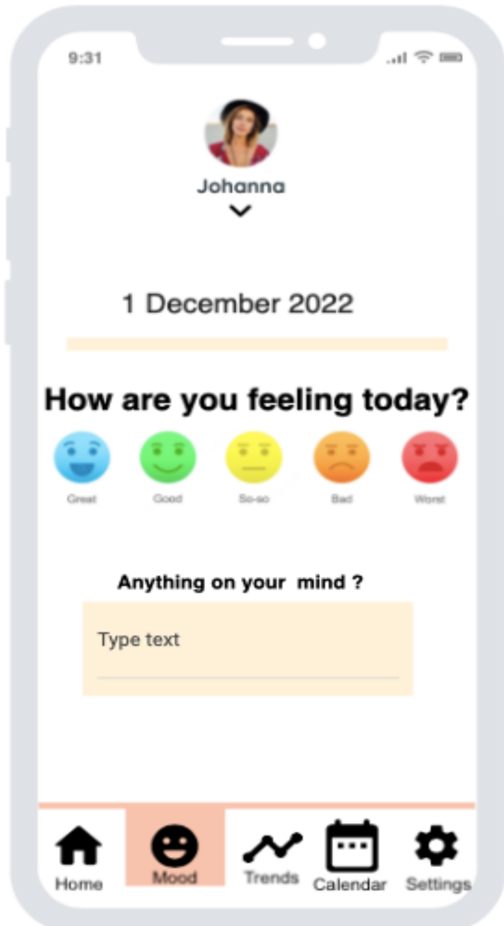

Figure 14

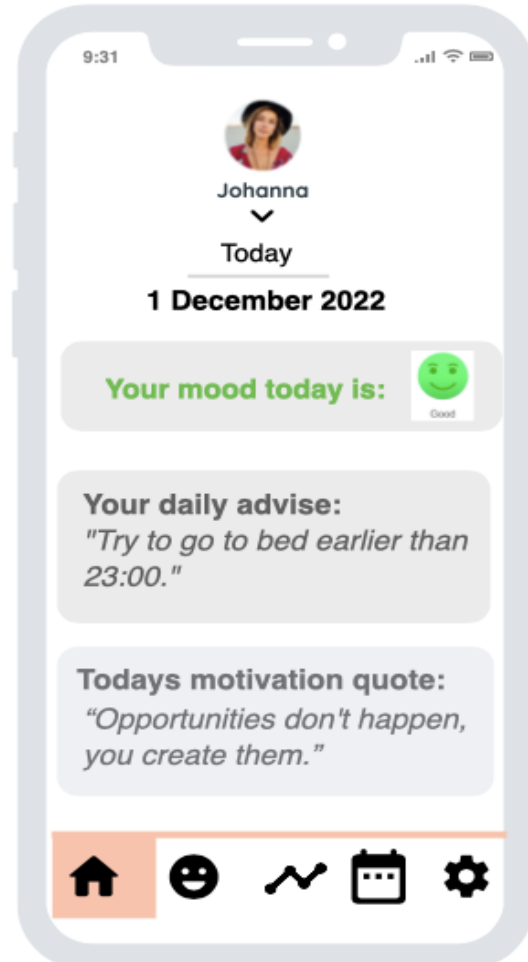

Figure 15

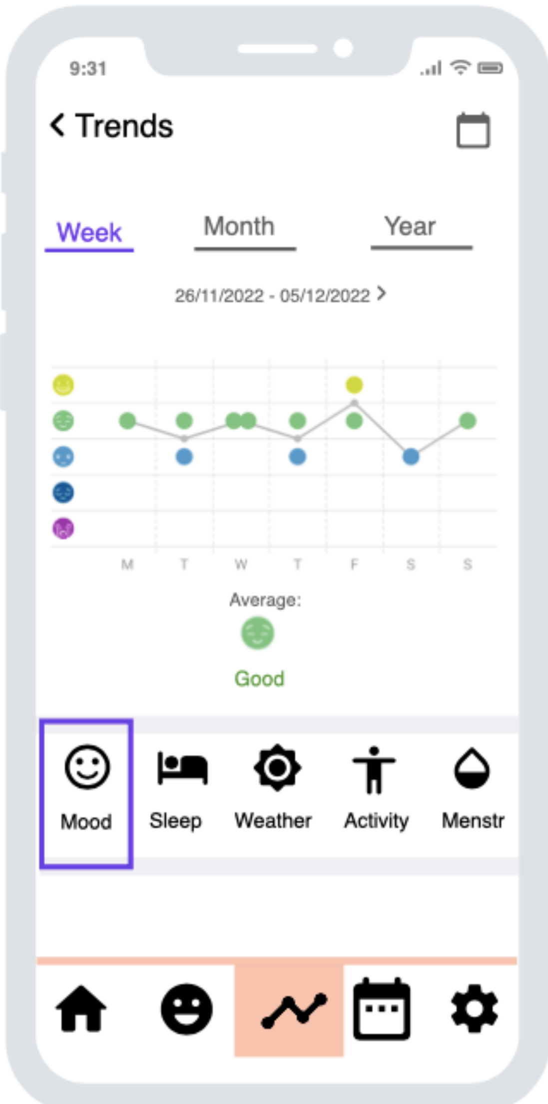

Figure 16

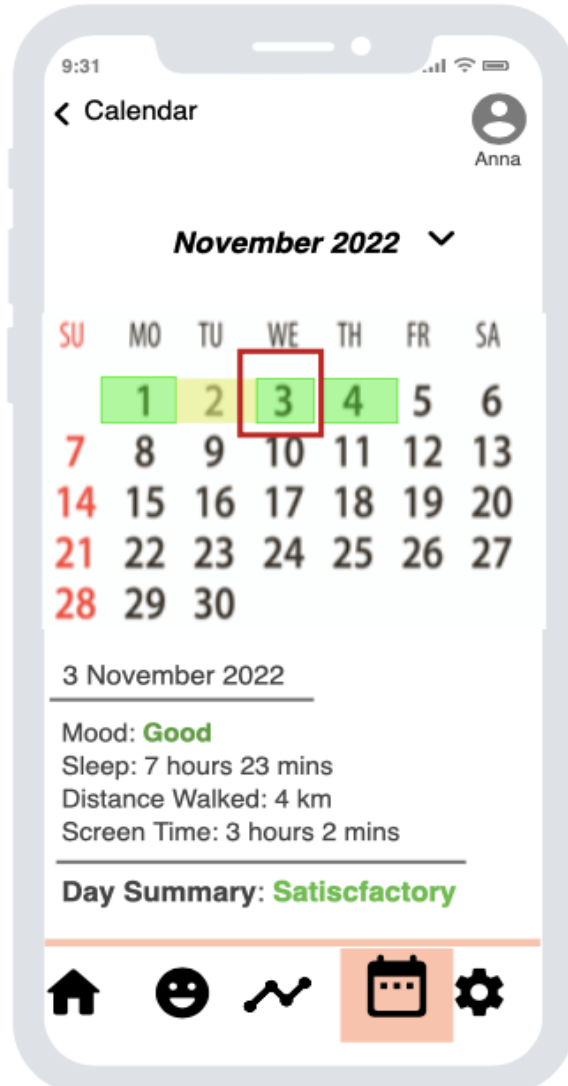

Figure 17

The images above represent a possible design of my proposed app. Initially, the user will have to create a profile. They will need to provide their biological characteristics such as age, gender and complete a survey regarding their mental state. The users will also get asked which apps they want to connect and transfer data from. The user has the opportunity to also track their energy levels, pain levels, appetite and others by adding it in the setting tab. Figure 14 shows the 'How are you feeling today?' question that the user has to answer every day so that the app can make accurate predictions. During the day, the user might change their mood. There is also 'Anything on your mind?' question, that is optional to answer. The user can provide additional information about their emotional state which can be processed with NLP to get more training data and make predictions more individual. The user can always change the mood over the course of the day. Figure 15 shows the home screen that displays your current mood, your daily advice and motivational quote. Figure 16 shows the Trends tab. Here the user should be able to find all the trends and graph visualizations of all the data that they provided to the app. Figure 17 displays the calendar which contains all past data for each of the days. It also returns a daily summary, so the user can have an overview of their past records.

Ethical Challenges and Issues

The technology and data presented in the potential application are an opportunity to empower people to take control of their health.

Any application that uses personal data would increase its user's footprint. Storing and processing personal or especially health data is a big responsibility. Medical confidentiality is one of the most important principles in professional medical ethics, in addition, it is protected by law. The highest standards of security and privacy should be met.

The legal basis for the processing of user's health data should be done in accordance with GDPR European General Data Protection Regulation (General Data Protection Regulation (GDPR). 2018).

Application and data that are presented in my project may encounter the following ethical challenges:

1. Protection of confidential data

AI provides the potential to qualitatively enhance diagnostics, personalize treatment, radically change medical decision- making, expand the possibilities of early detection and prevention of diseases. For the most effective work of AI, the most comprehensive patient data is required. This includes both medical and social data.

If access to sensitive information about the state of physical, mental health, suicidal tendencies is made available, this can lead to discrimination in hiring, inequality in obtaining health insurance.

Another big problem is that medical-related personal information is very valuable for cybercriminals. Such data can be used to obtain loans to a third party, for tax fraud, issuing fictitious invoices to insurance companies, obtaining strictly accountable drugs, selling personal data bases to drug distributors, and for other illegal action.

Considering all the above there is a high chance that patients can be harmed due to the leakage of their medical and personal data.

2. The presence of any biases in data

Any application that is related to our physical or mental health must be closely monitored.

One of the most important challenges is a problem of bias in AI systems is one of the most critical in the application of AI. If the system is trained on any bias, the objectivity of the system, the decisions made by it can be erroneous and biased.

For an AI system to be fair, it is necessary to eliminate bias in the input data on which the AI learns. As practice shows, even with careful data preparation, this is not always possible.

In healthcare, we are talking not only about patient's discrimination, where machine learning (ML) algorithms can amplify biases around race, gender, political leanings, despite the intention to do good and improve a given system, but also about not so obvious bias that can lead to incorrect training and, as a result, incorrect diagnoses.

For an effective algorithm, transparent monitoring of human values and moral considerations must also take place. As well as all diagnoses, treatments and recommendations that the AI system offers to its patients must be checked and agreed upon with the medical worker.

Predicting the emotional state is an absolutely individualistic and difficult task, so the risk of bias is extremely high. My application is just a type of recommender service. Users should remember that this is only a recommendation and not a doctor's prescription.

To overcome these challenges, it needs to establish patients' control over their confidential data. The application cannot store any personal data without its users' explicit permission. Moreover, all patients must be informed how their data is stored, processed and used. The privacy policy should be very clear and transparent to its users. The application will be using its users' personal data:

- to provide great service, understand their needs and communicate with them.
- to deliver personalised insights by processing the health data entered by users.
- to improve the application services.

Data protection laws are one means of safeguarding individual rights and place obligations on data controllers and processors. The users must have the following rights:

- to make sure that their data is not being retained in an identifiable format for longer than necessary.
- to request the complete deletion of their personal data.
- withdraw their consent from ongoing data processing at any time.
- lodge a complaint with the relevant supervising authority if they believe that their personal data is processing in violation of applicable data protection regulations.

Application should use strong security measures to protect its users' data against misuse, loss or alteration of personal information. The industry standards must be followed while processing and storing the data.

Conclusion and my personal findings

Each person interprets the term happiness in their own way. In my project, I tried to define happiness and depression from a more scientific point of view. I completed this research project by predicting mood states by using various ML models. These models were trained on my personal data which included the most important factors influencing our mood.

We are all absolutely unique and for an effective and non-bias model, it is necessary to train it on the data obtained from different individuals. But in the course of my experiment, I managed to identify some aspects that concern me exclusively.

I cannot say that my final model is accurate enough, but it did well with half of the predictions, given the limitations and high imbalance of my data.

First of all, I found out that regardless of the emotional state the predictions are hard to make. I can say that considering several experiments I conducted, it was the class of happiness that was always predicted by about 50-60. This was a discovery for me since this class had the largest number of instances. While the classes that represented a negative mood had several times fewer instances returned the same accuracy or even slightly better. I think that this is due to the fact that the factors that I selected initially corresponded to the negative impact.

Thanks to my research of all referenced scientific articles and visual analysis of my own data, I was able to come up with several insights. My first insight is, proper and balanced nutrition as well as maintaining an active lifestyle significantly reduces the risk of developing depression. I always thought that sugar is responsible for the level of serotonin in our bodies. I fallaciously believed that to improve your mood you need to eat a dessert. Moreover, typically when people already in a bad mood they preferably would resort to consuming junk foods that can be prepared with less cooking time and energy spent. From my own findings, I had a breakthrough that food that contains high levels of sugar and saturated fat, in other word junk food will only make our mood worse. The key point here is that I need to keep a high level of Fiber and a low level of Saturated fat. For the past two weeks, I have been following this diet and I feel physically better, moreover, I have also noticed that there are a lot more 'happy' days on my calendar now. I plan to continue to follow this principle.

I got acquainted with the new term SAD. After I studied the papers related to this term and also took into account the results obtained from my data, I found out that I am prone to suffer from the seasonal affective disorder (SAD). I often feel sluggish and depressed during the autumn months mostly from October onwards when the days get shorter. My future approach would be the next time when these symptoms appear again, I will try to increase my vitamin D levels.

I also found out that socialisation and days off have a positive effect on me. In the future, I will try to organise my calendar so that there are no long stressful periods without rest.

My project started with a hypothesis about whether it is possible to make a person happier with the help of AI. I found this question quite childish and was relatively sceptical about the results. Nevertheless, the results of my research showed that this is a rather serious problem that affects millions of people, because even a prolonged state of apathy can lead to more severe mental illnesses, including depression.

In conclusion, I can say that mood prediction is a very complex and at the same time subjective problem. But despite this, my experiments have shown that this is a feasible task. I would like to have more data so that I can conduct more efficient experiments. I also believe that professional experts in psychology and mental health are needed to solve this problem. After all, inaccurate AI predictions can only harm human health.

References

1. Department of Health & Human Services. (2017). *Monitoring-your-mood*, *Better Health Channel*. Department of Health & Human Services. https://www.betterhealth.vic.gov.au/health/healthyliving/monitoring-your-mood
2. World Health Organization. (2021, 13 September). *Depression*. https://www.who.int/en/news-room/fact-sheets/detail/depression
3. Edwards, M.K. and Loprinzi, P.D. (2018, 7 July) *Experimental effects of brief, single bouts of walking and meditation on mood profile in Young Adults*, *Health promotion perspectives*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6064756/
4. *Clue: Period and Ovulation Tracker for iPhone and Android*. *Clue Period & Ovulation Tracker with Ovulation Calendar for iOS, Android, and watchOS*. https://helloclue.com/
5. *Your key to success*. *FatSecret. (2007)*. https://www.fatsecret.com/
6. *Ios – Health. Apple (December 6 2022) (United Kingdom).* https://www.apple.com/uk/ios/health/
7. *Renpho: Smart healthy living.* (2016). *Renpho EU*. https://renpho.eu/?gclid=Cj0KCQiA1ZGcBhCoARIsAGQ0kkobgDG79yDfabhTVc3uGDrJ5FUQdsgZy-AFIoVprS7gaaIulcOQK_0aAj38EALw_wcB
8. *Historical data Met Éireann - The Irish Meteorological Service*. https://www.met.ie/climate/available-data/historical-data

9. Ph.D a.Takeshi Kochi, M.D. cKeusuke Kuwahara, Ph.D adRei Ito. (May 2016) *Dietary fiber intake and depressive symptoms in Japanese employees. The Furukawa nutrition and Health Study, Nutrition. Elsevier.* https://www.sciencedirect.com/science/article/pii/S0899900715005237
10. Stephanie Fulton (April 2018) *Nucleus accumbens inflammation mediates anxio-depressive behaviour and compulsive sucrose seeking elicited by saturated dietary fat.* Molecular Metabolism. https://www.sciencedirect.com/science/article/pii/S2212877817309389?via%3Dihub
11. M.J.T.W. (December 2019) *Temperature and mental health: Evidence from the spectrum of mental health outcomes*, *Journal of health economics*. U.S. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/31590065/
12. *EU Data Protection Rules* (2020) *European Commission - European Commission*. https://ec.europa.eu/info/law/law-topic/data-protection/eu-data-protection-rules_en
13. General Data Protection Regulation (GDPR). (2018). *General Data Protection Regulation (GDPR)* https://gdpr-info.eu/
14. Melrose, S. (2015 November 25) *Seasonal affective disorder: An overview of assessment and treatment approaches*, *Depression research and treatment*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4673349/
15. *The General Data Protection Regulation* (2022) *Consilium*. Available at: https://www.consilium.europa.eu/en/policies/data-protection/data-protection-regulation/
16. Khandelwal, N. (2020) *A brief introduction to xgboost*, *Medium.* Towards Data Science. https://towardsdatascience.com/a-brief-introduction-to-xgboost-3eaee2e3e5d6
17. Matthew Pearce, P.D. (2022) *Association between physical activity and risk of depression*, *JAMA Psychiatry*. JAMA Network. https://jamanetwork.com/journals/jamapsychiatry/fullarticle/2790780?guestAccessKey=67cf8fd3-e6b0-49af-be4f-d08f5219fc7b&utm_source=For_The_Media&utm_medium=referral&utm_campaign=ftm_links&utm_content=tfl&utm_term=041322
18. Newman, T. ZOE. (2022 October 15) *Can eating more fiber improve mental health? Can Eating More Fiber Improve Mental Health?* https://joinzoe.com/learn/dietary-fiber-benefit-mental-health
19. Smith G. New Food Magazine (2018 February 15) *Saturated fat and sugar-rich diets can lead to depression and anxiety*. New Food Magazine. https://www.newfoodmagazine.com/news/64695/saturated-fat-sugar-rich-depression-anxiety/
20. Schimelpfening, N. (2021 June 28) *What is seasonal affective disorder? Verywell Mind*. Verywell Mind. https://www.verywellmind.com/what-is-seasonal-affective-disorder-1065408
21. *Support line - depression support* (2021) *Aware*. https://www.aware.ie/support/support-line/
22. UCD News (2012 June 13) *Socialising helps to alleviate symptoms of depression*. https://www.ucd.ie/news/2012/06JUN12/130612-Socialising-helps-to-alleviate-symptoms-of-depression.html
23. Asmelash, L. (2019, August 14). *Social media use may harm teens' mental health by disrupting positive activities, study says*.

CNN. https://www.cnn.com/2019/08/13/health/social-media-mental-health-trnd/index.html