Final Project Report

Project management

1. *The contributions/roles of each member of the group:*

Ye: I took the role of technical lead and mainly contributed to developing the data science pipeline. I performed data analysis, preprocessing, prediction model training and testing, model optimization, and hyperparameter tuning. I discussed with my teammates in each step for better solutions. I participated in meetings and in preparing the project charter, especially business needs, public health, and business impact and strategic alignment. I also took the responsibility of project manager for 1 week.
Shweta: In the initial phases I was involved in finding a relevant problem and the appropriate dataset. After discussion with the team, I drafted the project requirements and the scope of the project. This helped us to decide the data processing and analysis pipeline when Ye and I took up the role of technical leads. I performed data preprocessing, analysis, model training and visualization and organized all of this for the final Jupyter notebook. I made the diagnostic, predictive and prescriptive analytics tab of the dashboard. I was the project manager during the last week and coordinated with team mates on various deliverables.
Sidra: I was involved in different aspects of the project. Initially we all contributed to each assignment equally, the tasks became more defined as we moved along. I was the project manager of week 39 in which we had to prepare a preliminary draft to show in the seminar. I have mainly been responsible for the dashboard with some of my other team members but we have decided as a team about what goes on the dashboard and how it should look like.
Mathilda: My role during the project changed depending on the different tasks. I was the project manager for the second week of the project, and apart from working on different parts of the project, like the project charter and so on where we divided tasks equally, I have mainly been responsible for the dashboard, both in the early stages and final stages. I have discussed with my teammates the requirements and checked in with them on the different solutions during the progress. This was done both to keep them up to date but also to get feedback, making everyone involved in a way. Since I was responsible for the Dashboard, I also asked for help when needed and divided up the tasks to some of my group members when they also worked on the dashboard.
Moa: I started by compiling information for the project charter, but as roles evolved, I transitioned to the dashboard team, and later I became primarily responsible for usability testing. Although I didn’t contribute as much to the technical development of the dashboard as some of my teammates, my role was to ensure the dashboard's user-friendliness through testing. I was project manager for one week but as the project progressed, roles shifted based on individual strengths. For example, while other team members took on more of the technical tasks, I focused on usability testing and supported the team with what was needed for the time being.
Amanda: I started as the project manager and created documents for the tasks we would work on. Subsequently, I have adapted my roles according to the needs that have arisen in order to support the group in the best way. I researched and compiled the necessary studies to back our main objective, ensuring we had a solid foundation. Following that, I took on the

task of organizing our PowerPoint presentation. Additionally, I worked on drafting the goals for our user testing to establish clear guidelines for us to follow. I also gathered the group's feedback on our project, which allowed me to write the project management part in this final project report.

2. *During our project, we experienced the different development phases of Wheelan's model [1] as follows:*

1. Dependency and Inclusion

At the beginning of the project, we were all cautious and tried to understand each other's working styles and strengths. We searched for different datasets and jointly answered assignment 1. We also decided which model to follow, and it was a relatively unanimous process where everyone was keen to avoid conflicts. An event that reflected this phase was the selection of datasets. Although we had different proposals, those who presented their datasets were able to argue their choices without major contradictions, which helped us reach a common decision.

2. Counterdependency and Fight

After we split into two teams – a dashboard team and a data science team – we realized that working on each sub-project could become challenging if all three in each group were working on the same thing. This led to some team members starting the tasks themselves, while others found it difficult to find their role and contribute directly. Within the data science team, one of the members, feeling left out as the others had already started dataset cleaning, wrote to the other two on the team questioning their role. This was later brought up in a meeting, where we decided that person would instead act as a bridge between the data science team and the dashboard team. In this way, we got a clearer division of responsibilities that was more in line with the needs of the project.

3. Trust and Structure

When we realized that it wasn't necessary to have so many people in each team, we started to clarify our roles. After discussing the group's needs, we distributed tasks based on our strengths: two people with coding experience focused on developing algorithms and data processing for the data science part, while another, with design and usability experience, worked on the dashboard layout. Each took on responsibilities that were more aligned to the needs of the project, which helped us create a clearer structure and division of labor.

4. Work and productivity

In the final stages of the project, we worked very effectively together. Our roles were clarified and everyone knew their area of responsibility. Thanks to our weekly meetings and the ongoing communication in the group chat, it was easy to follow up and support each other when needed. This phase was characterized by being able to focus on our respective tasks without distractions, which allowed us to deliver a final result that we were all happy with.

5. Termination

In our group, we had a short meeting after the last class where we discussed what we had learned and expressed our gratitude for each other's contributions. This phase gave us the opportunity to formally end the collaboration and feel satisfied with the work we had accomplished.

3. *Whether the members of the team felt comfortable or challenged in the roles they performed:*

Ye: The project has some challenges like difficulty in achieving good performance of the ML model. I am also an introverted person who likes to work on my own. However, I tried to discuss in the group with other members, talked about the problems, and tried to get a solution together. Everyone in our group is positive-minded, open, helpful, and supportive of each other. Thus, finally, I can overcome the challenges and I feel more comfortable working with others.

Shweta: Since this project had a steep learning curve, I felt challenged and have gained confidence in handling the technical responsibilities. I am more open to using new programs and collaborating with my classmates. I am more oriented towards working alone, therefore, having to communicate with others and adapt to their working styles may have felt challenging at times. As a team, we had a good balance of personalities and basic work ethics were followed by all, which I greatly appreciate. I was comfortable in the group as we all adapted well to each other and there was freedom to choose roles and responsibilities.

Sidra: I did feel a little out of my depth during the week I was project manager but was quickly made to feel comfortable by my team members. I also started getting more courageous and comfortable with working on the dashboard. As an introverted person, I excel in an environment where I work alone and am responsible for myself, but this experience has taught me that there are many benefits to working in a team.

Mathilda: I felt mostly comfortable in the roles that I had. I was comfortable being project leader for the week that I was as I felt like I had a good overview of the project and it was clear what I had to do and I felt comfortable communicating with my group. Sometimes I felt less comfortable, like when the structure of the week was less defined. For example some weeks there were some confusions in who was the project leader and when the next meeting was. This part was slightly challenging for me as I easily take a leading role in structuring this, so it was a good exercise to take a step back. I felt challenged in my role as responsible for the Dashboard but in a way where I learned a lot. It was nice to be responsible for this task, but it was also hard because it felt like quite new territory for me. What I felt comfortable with and uncomfortable with can be related to my personality leadership type which is "Steadiness" as someone who likes to pursue a defined course of action and is also good at handling technical assignments.

Moa: In our project, I felt comfortable throughout the process, primarily due to the open communication and supportive environment within the team. Everyone was respectful and encouraged one another, which helped us gain a positive group dynamic. However, when it came to specific tasks, such as developing the prediction model and dashboard, I felt more

challenged. These areas required more technical knowledge, and while I contributed where I could, I found myself learning from the experience. I identify with the "influencing" type, which tends to be sociable, enthusiastic, and supportive. This might explain why I felt comfortable in the collaborative aspects of the project, as I enjoy team interaction and discussion. However, the technical tasks might have been more challenging for me as they don't align with the typical strengths of an "influencer" personality, which focuses more on creativity and relationships than on analytical tasks.

Amanda: In the role of project manager, I generally felt comfortable and motivated. Creating documents and structuring our work were tasks I enjoyed, and I appreciated the opportunity to drive the project forward. This sense of comfort and motivation can be partly linked to my dominant personality type. At the same time, I was flexible and adapted my roles to the needs of the group, which meant that I also felt challenged to learn new aspects of the project. Researching and compiling the necessary studies to support our project was a stimulating task that I really enjoyed. However, related to my dominant personality, I can sometimes find it difficult to handle situations where I am given specific tasks without a clear understanding of the goals. When I felt uncertain about the goals, it could affect my ability to be effective, which was challenging. In conclusion, this experience has given me valuable insights into both my strengths and the challenges that my personality type brings.

4. *Each member of the groups most suitable leadership style:*

Ye: My personality test result was steadiness, and so I determined a course of action to achieve objectives. As I am analytical, I analyze everyone's task, condition and achievement. Then, I tried to support what they needed, and do some administrative tasks like submitting assignments and updating the Jira board so that they could focus fully on their assigned tasks. I also tried to communicate with others to know what they need and how to help them. By this way, we could achieve the objectives of the project.

Shweta: The personality test result of being Compliant was a little surprising for me. I usually take time to adjust to new people. The detailed description was accurate, and because of the need to be structured, goal oriented and perfect, I am aware I can be difficult to work with. I am observant which helps me to communicate differently with each person and that I think usually helps me in group work. The experiences gained by this collaboration have helped me to become more self aware and enriched my communication skills.

Sidra: According to the personality test, my leadership style is Complaint, which is characterized by focus on following rules, maintaining structure, discipline and ensuring accuracy in tasks. While working on this project, my personality traits of being detailed oriented and precise have come in handy. I do believe that sometimes I can get a bit nit-picky with things which can take valuable time but I have learned to manage that during the course of this project.

Mathilda: From taking the personality test in the beginning of the course I found out I had a Steadiness leadership style, which means that I like to follow a specific course of action to achieve my goals. For me it is easier to do tasks when I follow a clear structure that is logical. During this project I feel like that mostly that was actually beneficial for me especially since we were applying the waterfall method. I found that the weeks where I was team leader or

where someone who also likes clear structure were the easiest for me.
Moa: As an influencing person, my leadership style focuses on promoting relationships and encouraging collaboration within the team. However, due to personal challenges, I haven't always been able to perform at my best, which I've noticed affecting my contributions. While I enjoy bringing energy and enthusiasm to the team, I sometimes prioritize people and ideas over details and structure. That said, I felt the group had a strong balance, providing effective support in staying organized and handling the more technical aspects of the project.
Amanda: As a person with a dominant personality type, I am often driven and decisive, but I work best when I get to choose tasks that I really understand and find interesting. It is important for me to work on projects where I have a clear picture of what is expected of me. Then I can really be efficient and fast in my work. I am also flexible and willing to adapt to the wishes of others, which allows me to cooperate well with my colleagues. However, I can have difficulty handling situations where I am given specific tasks without having a clear understanding of the goals. This mix of strengths and weaknesses affects how I interact with others and my approach to leadership, where I value both independence and collaboration.

We felt that everyone took on the leadership role well during the week that each person was project manager. Each member listened carefully to that person and gave them space to schedule meetings and make important decisions. There was a clear respect for the project manager, and no one tried to take over that role. This contributed to a harmonious work environment where we could focus on our tasks and collaborate effectively.

5. *The things that worked well, what did not and what could have been done differently:*

What worked well
One of the strengths of our work was the clear communication and spirit of cooperation that characterized the group. By using tools like WhatsApp to quickly book meetings and give feedback, as well as Google Docs to work together on documents, we created a structure that facilitated our collaboration. In addition, the flexibility to adapt roles to the needs of the group was a great asset, allowing us all to contribute in the best possible way.

What didn't work well
Despite the positive aspects, the group encountered some challenges. Sometimes it was difficult to find meeting times that suited everyone, especially since we had another course with other meetings to adjust to. This made us sometimes feel stressed before deadlines. We also experienced on certain occasions uncertainty about how we would move forward in the project, which affected our efficiency and the group's progress. In addition, there were situations where some members felt unsure of their roles, which made it difficult to get everyone involved, especially in the early phases of the project.

What could have been done differently and how
To improve our work process, we could have been more proactive in communicating about our availability. Since we had a different course with uncertain meeting times, instead of scheduling specific meetings, it would have been good to have shorter, more frequent "meetings" via WhatsApp. By continuously agreeing when we could all be available for

quick discussions, we could have reduced the stress of deadlines and created a better workflow. We could also have introduced a more structured approach to clarify roles and responsibilities already at the beginning of the project. By clearly defining who was responsible for which tasks, we would have reduced uncertainty and ensured that everyone felt included and valued in the team.

6. *The project development approach we followed:*

The waterfall development methodology worked well for our OCD severity prediction model project because it provided a clear and structured approach, ensuring each phase was completed before moving on to the next. This was particularly effective when handling the dataset and developing a dashboard connected to it, as the linear timeline allowed for a step-by-step progression. It also facilitated a well-organized requirements gathering and analysis phase, minimizing the risk of unexpected changes later in the process, thereby keeping the project on track and well-aligned with its goals.

7. *If we could restart the project:*

If we were to start another project, the Agile approach might be more suitable for our needs because it offers greater flexibility and adaptability. With the experience gained from this project, we now have a better understanding of the strengths and weaknesses of the development process. This would allow us to more effectively organize tasks and divide the team into specialized groups, such as a dashboard development group and a model development group, which could work in parallel. This division of labor would enable faster progress.

8. *Lessons learned about the chosen approach:*

The Waterfall approach worked well for our project because it provided a clear, linear structure that helped bridge our initial knowledge gap. It allowed us to collaborate closely in the early stages, ensuring everyone was aligned before dividing tasks based on individual strengths. We could have divided tasks, like model development and dashboard creation, earlier to work more efficiently in parallel. Additionally, incorporating more flexibility would have allowed for adjustments during development. However, given our lack of experience with these types of projects, a more flexible approach might have been less effective. Overall, the Waterfall methodology suited our team's knowledge gap and inexperience well.

9. *General feeling among the group members about the way the conflicts or disagreements were being solved:*

The general feeling in the group was that we were mostly cooperative, but there were variations in how members experienced conflict and disagreement. Some often felt they had to adapt and be more accommodating, while others felt redundant or unsure of their roles. For example, when we realized that the dataset was synthetic (as mentioned in the answer to question 10), frustration arose among some members, which led to a discussion about how to adapt our strategy. Despite differing opinions, we showed a willingness to cooperate and

decided to proceed with the synthetic dataset. In conclusion, we can say that although the group generally had a cooperative spirit, there were times when some felt more forced to adapt, while others felt they needed to stand on their own to achieve their goals.

10. An event that significantly influenced the development of the project:

During the course of the project, more precisely when we were almost done with our project charter, we encountered a significant challenge that affected our strategy. We discovered that the data set we used was synthetic. This led to a discussion about how to proceed, as we had originally planned to use real medical data to create a more accurate prediction model.

To resolve the situation, we decided to adapt our approach and continue working with the synthetic dataset instead of trying to find a dataset that contained real medical data. We justified this choice with several arguments. First, access to real-world medical data is highly regulated and poses significant barriers, including the need for special permits due to privacy and security concerns. Second, the synthetic dataset mirrors real data in a way that allows similar techniques for analysis and interpretation to be used, without the ethical and privacy concerns that come with handling real data. Continuing with the synthetic data set allowed us to focus on developing a pilot model that can be used in the future when real data becomes available, which we considered a practical and strategic solution.

11. The scheduling table:

Our team was generally able to follow the schedule and perform the tasks described in the scheduling table but we encountered some delays. Most deadlines were met according to plan, but we fell behind with the tasks Complete First Version and Demonstrated and Test on Experts in IT and Healthcare. Completing the first version of the dashboard took significantly longer than expected, which in turn delayed user testing. This delay affected our timeline for the following activities, but we were nevertheless able to complete the project within the deadline.

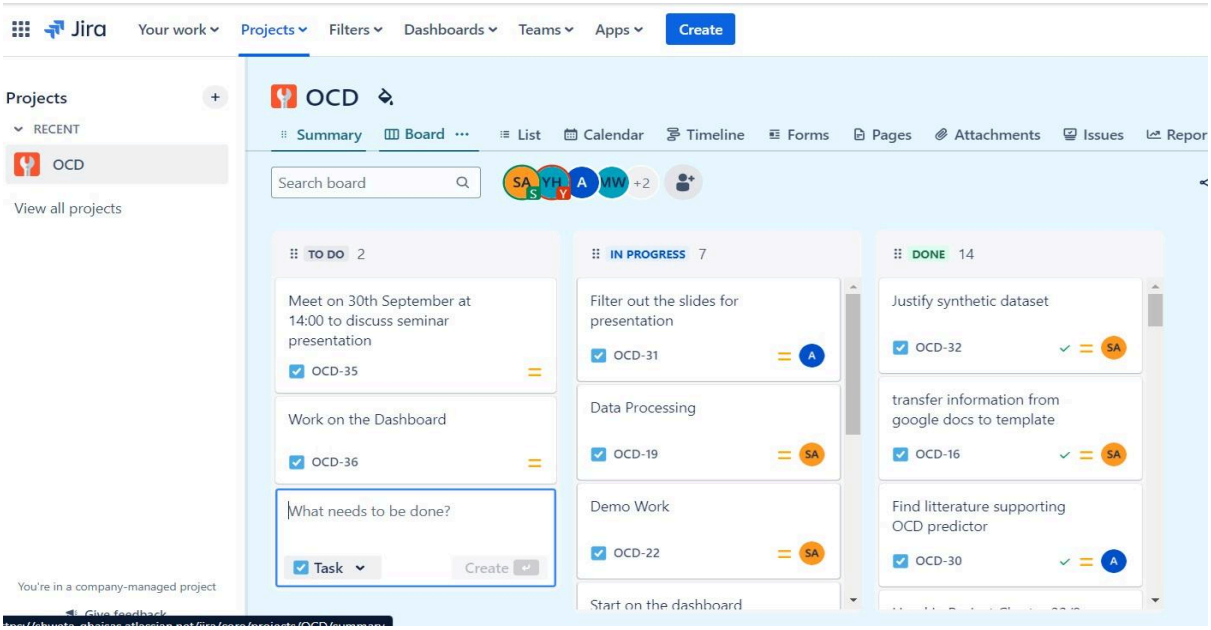
12. A significant deviation:

During the course of the project, a deviation occurred when we received feedback that suggested we change our classification of OCD severity from three categories (high, medium, low) to only two categories (high and low). This change required us to adapt our methods and analyzes to ensure that our model would be both effective and useful. To support our work, we used several development tools that proved to be very useful. We used GitHub for version control, which facilitated collaboration and tracking changes to the code. Python and VSCode were central to our data analysis and modeling, enabling efficient handling of the adjustments we needed to make. Jira was used to organize and track tasks according to our waterfall approach, giving us a clear structure and timeline for the project. These tools were critical to navigating the changes we faced.

13. Tools used for project management:

Jira Board

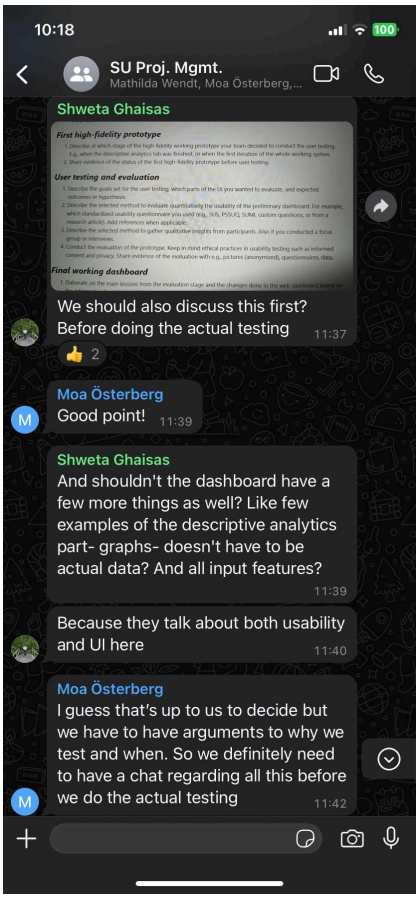
We used Jira Board to manage project tasks, track progress and organize workflow. After each meeting, the project manager was responsible for assigning tasks (“To do”) to each of the group members based on what we agreed on at the meeting. Each team member was then responsible for updating the Jira Board by moving their assigned task to "In Progress" and when the task was complete to "Done". In this way, we could all keep track of where in the process one's group members were.



WhatsApp

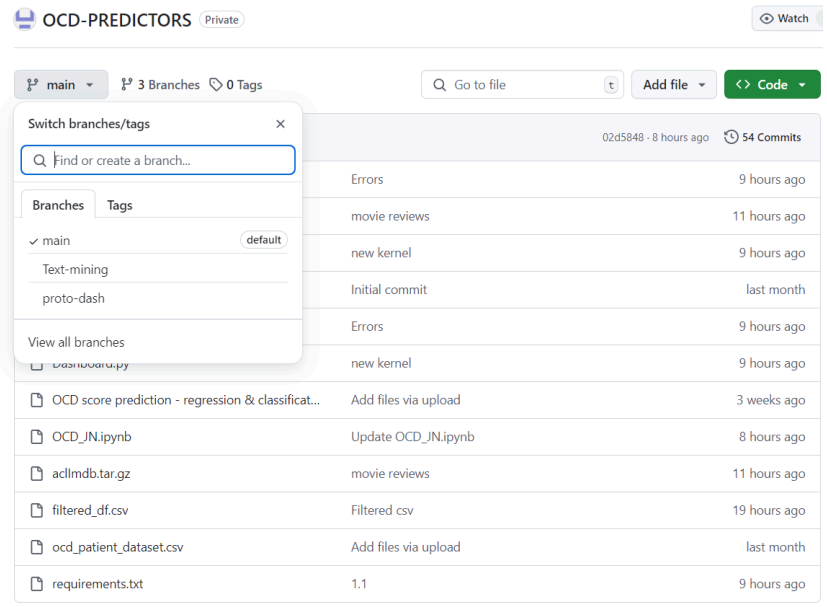
WhatsApp was a tool used to facilitate communication with others in our project group. It made it easier to schedule appointments and verify everybody's free times to set dates that fit into the group. Also, WhatsApp allowed us to provide and receive immediate feedback, which is of great relevance for making all the necessary adjustments as soon as possible.

We also shared important project updates, reminders, and other important information in our group chats, ensuring everyone was always informed throughout the project. File management in the app was also very easy, enabling us to share documents and other files among ourselves, thereby giving all team members quick and easy access to materials any time.



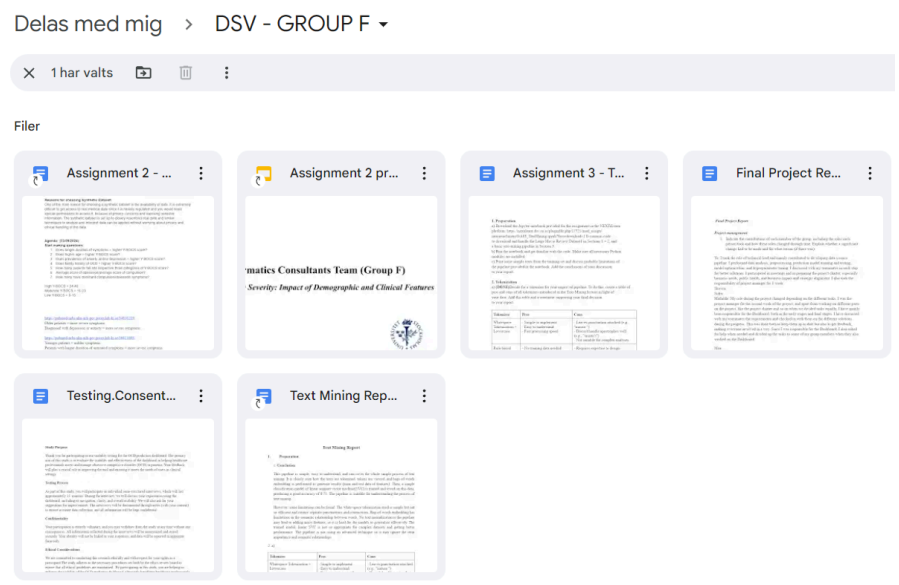
GitHub

In our project, we used GitHub for version control, collaborative coding, and developing the project dashboard. By creating a central repository, all team members could contribute code while keeping track of changes and ensuring the integrity of our work. The branching feature enabled us to work on different tasks simultaneously without disrupting the main codebase.



Google Docs

In our project, we created a shared folder in Google Drive where we collected all our files, including text documents and PowerPoint presentations. This setup allowed us to work collaboratively on the same documents and see what each team member had contributed. The ability to comment and suggest edits made it easy to provide feedback and refine our work together, ensuring that we maintained a clear and organized record of our project throughout its duration.



Data science pipeline

1. *The selected dataset, the problem domain, and intended target users:*

The selected synthetic dataset includes 1500 patients with obsessive compulsive disorder (OCD) and includes demographic information such as age, gender and education level, as well as clinical details such as duration of symptoms and date of diagnosis. The problem area focuses on predicting the severity of OCD symptoms by identifying patterns between some of these demographic factors. Designed for healthcare professionals and researchers, the dashboard aims to analyze and visualize factors that influence the severity of OCD symptoms. By using the dataset, users can identify patterns and relationships between patient characteristics and symptom severity, which can lead to better treatment planning and early interventions.

2. *The data processing pipeline that we followed:*

For data processing, data types of all columns are checked first. Then, null values in the dataset are also checked and 248 null values are found in the 'previous diagnosis' column. Then, these values are filled with 'None', meaning no previous diagnosis. Null values are checked again and it shows no null.

After that, the feature engineering process is conducted. A new column, named 'Total_Score' is calculated by adding Y-BOCS scores for compulsion and obsession. Then, the rows with a total score below or equal to 40, and the scores for compulsion and obsession, each below or equal to 20 are filtered, as they are possible meaningful values. A new required column, 'Score_Category' is formed by the value, 'low' or 'high' if the total score is between 0 and 20, 21 and 40 respectively. This binary column will be used as the target for model training.

Then, necessary features only are selected and extracted into a dataframe, which will be used in descriptive, diagnostic, and predictive analysis. This simplifies the dataset and model training.

For diagnostic and predictive analytics, data types are corrected and boolean values are converted to integers to avoid computational errors, improving accuracy. The numerical values are normalized using min-max normalization, ensuring that all features equally contribute to the model.

Then, categorical features undergo one hot encoding required for model training, and the score category experiences label encoding to be able to be used as a target.

3. *6 analytical questions for the descriptive analytics tab will be as follows:*

1. What is the distribution of OCD severity (Total_Score) across different age groups?

This question helps to determine if certain age groups tend to experience more severe OCD symptoms. The age has been split into 5 age groups and we evaluate how our dataset is

represented across the various age bins. All age groups have a similar distribution, and the total scores mostly lie between over 10 and below 30. Only age groups over 19 reach up to nearly 40 (maximum value). The youngest group (0-18) has the highest score, nearly 30.

2. Is there a significant difference in OCD severity based on gender?

The literature does not indicate a relation between gender and OCD. Gender-based analysis can uncover any significant differences in the severity between males and females, useful in targeting gender-specific interventions. It is seen that females have a distribution of higher total scores than males. The total scores of most females range from 15 - 29, but those of males have only a lower range from 12 -24.

3. How does a family history of OCD influence the Total_Score?

We try to examine whether individuals with a family history of OCD are more likely to have severe symptoms, potentially highlighting genetic or environmental influences. It shows that most patients without family history have a higher score range (from around 10 to nearly 30) while those having family history have a range up to 25.

4. How does the diagnosis of Anxiety or Depression affect Total_Score?

It interprets the presence of anxiety or depression or both affecting OCD severity. Patients having only one condition: anxiety or depression, have a higher score range from over 10 to nearly 30, while those having both conditions have a lowest range of up to 24. Those without both conditions can have a slightly higher range up to 26.

5. Are the Obsession types represented equally in this dataset? What is the average Total_Score for each type?

It detects the occurrence of different types of obsessions. Among the 5 types, contamination and harm-related have the highest number of instances over 88, while hoarding and symmetry have a slightly lower number, around 83 but the religious type has only 77. Regarding average total score, the harm-related type has the highest (20.7), and contamination and hoarding, around 20 while symmetry and religious types have the lowest score, just above 19.

6. Are the Compulsion types represented equally in this dataset? What is the average Total_Score for each type?

It examines the occurrence of 5 compulsion types. Checking and washing have the highest number of instances, about 90, counting has a lower number, 83 while ordering and praying account for only just above 78. However, praying has the highest average total score over 22. Checking and ordering have lower scores just above 20. Counting accounts for the score, about 19 and washing has the lowest, over 17.

4. 6 analytical questions for the diagnostic tab will be described as follows:

1. Is there a relationship between the diagnosis of anxiety and the severity of symptoms?

By conducting a chi-square test between anxiety diagnosis and severity score category, it results in a chi-square value of 4.3 and a p-value of 0.03. The p-value is lower than the common significance level of p-value, 0.05, so there is a statistically significant relationship between them. Therefore, a person's diagnosis of anxiety is related to the severity of the symptoms.

2. Is there a relationship between a family history of OCD and the severity of symptoms?

By conducting a chi-square test between family history and severity score category, it results in a chi-square value of 1.23 and a p-value of 0.2. The chi-square is relatively low, showing weak association but the p-value is higher than the common value for significance, 0.05, so there is no statistically significant relationship between them. We cannot conclude that a family history of OCD has a relationship with the severity of the symptoms.

3. Is there a relationship between the diagnosis of depression and the severity of symptoms?

The chi-square test statistic is run between the diagnosis of depression and the severity of symptoms. It produces a chi-square of 0.067 and a p-value of 0.79. The p-value is higher than the common value for significance, 0.05, so there is no statistically significant relationship between them. Thus, we cannot conclude that the presence or absence of depression in OCD is not related to the severity of the symptoms.

4. What is the correlation between Age and Total_Score?

We calculate the Pearson correlation coefficient between age and total score, and the result is 0.07. This indicates a very weak positive correlation between them, which would be almost negligible. Age and total score can be assumed to be nearly uncorrelated.

5. What is the correlation between the duration of symptoms and OCD severity?

The Pearson correlation coefficient between the duration of symptoms and OCD severity(total score) is 0.03, which is an extremely weak positive correlation. Thus, it can suggest that there is no relationship between the duration of symptoms and OCD severity.

6. How are Age and Duration of Symptoms related?

The Pearson correlation coefficient between the duration of symptoms and age is also 0.07. So, there is a weak positive association, meaning that if age increases, the duration of symptoms may increase, but it is a small relationship and can be negligible.

5. *The one classification problem addressed through machine learning:*

In our project, we have a problem of predicting the high or low severity of symptoms in OCD patients, which can be addressed by a machine learning solution. For this purpose, we will develop a model by using the features: age, gender, duration of symptoms (months), family history of OCD, depression diagnosis, anxiety diagnosis, obsession type, and compulsion type to predict the target: 'high' or 'low' score category meaning high or low severity. We

chose the random forest model, which has the best performance among the models trained on the dataset, to use for the prediction, integrated into the dashboard. Regarding the dataset, it is highly relevant to this classification of OCD severity. It contains necessary features, that are found to have an impact on disease severity in previous articles, and a Y-BOCS score, that is used in clinical practice to indicate the high or low disease severity and is modified into high or low score category as the target for the model development.

If you did the optional extra task with Prescriptive Analytics: Briefly describe which SHAP visualization method you chose, whether you chose to explain a single prediction or multiple predictions, and why.

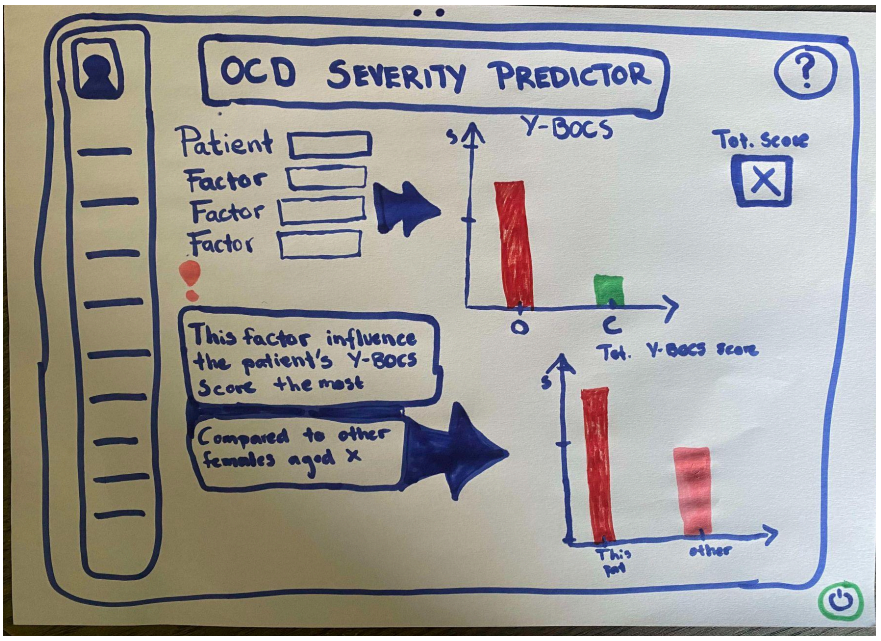
For prescriptive analytics, we used SHAP visualization method, especially the SHAP summary plot. In the Jupyter Notebook, we use it for multiple predictions and can provide a comprehensive overview of how each feature in all instances has an impact on the random forest model predictions in both magnitude and direction(positive or negative). However, the dashboard uses a single prediction to visualize the impact of the features of a particular instance on the model. Thus, the dashboard users, especially clinicians, can understand how each feature contributes to the model prediction result (high or low severity) of the patient examined. Therefore, it is an effective tool for understanding overall feature importance in the model's performance and for individual contribution of all instances.

Early-stage prototyping

1. *The decisions and activities related to the early-stage prototyping of the web dashboard:*

The early stages of the prototyping of the Dashboard involved the group discussing what our end-users actually want and need so that we all shared a common vision. After we had established our common vision for the Dashboard, one of our group members created a low fidelity sketch on paper to make our vision visible.

2. *Evidence on the design process of the low-fidelity prototype:*



3. *How the preliminary layout of the dashboard and the chosen UI inputs can help the target users get answers for the medical problem domain:*

The preliminary layout of the Dashboard would easily help the target users insert inputs as key factors and then visually see the predicted total Y-BOCS score for their patient. The First prototype is not what we actually created later, where graphs are visual representations we show under the diagnostic and descriptive tabs. However, it was a first idea and it helped us organize what we wanted to show.

4. *The advantages and lessons learned by the implementation of early-stage low-fidelity prototyping in the project:*

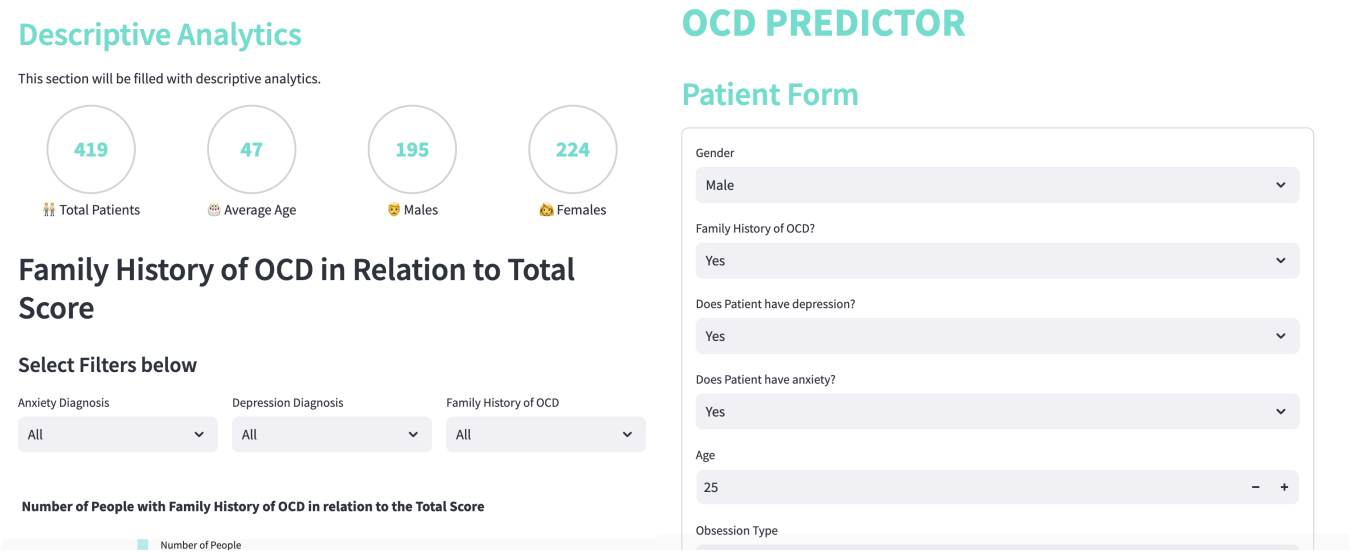
By implementing a low-fidelity prototype the group's common vision of the dashboard got visual. It also enables us to later see what we wanted to change and what requirements to add where. It helped us structure the Dashboard and communicate efficiently regarding the dashboard.

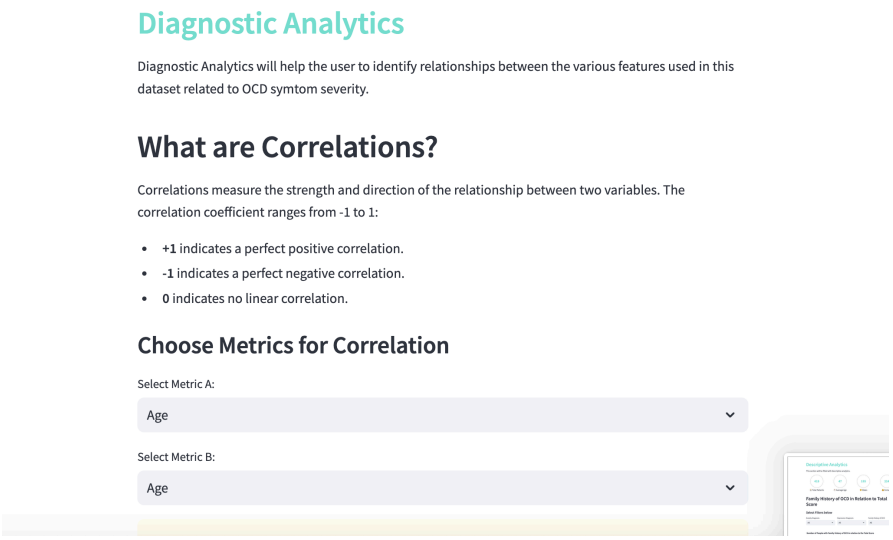
First high-fidelity prototype

1. *At which stage of the high-fidelity working prototype our team decided to conduct the user testing:*

We decided to conduct user testing quite late in the project, when the 'About' and 'Diagnostic' page was finished, most of the descriptive and the predictive page was working but not with the appropriate prediction model, just a prototype.

2. *Evidence of the status of the first high-fidelity prototype before user testing.*





User testing and evaluation

1. *Goals for the user testing:*

1. We want to evaluate whether the users can enter the patient's features easily and intuitively. This includes ensuring that it is clear what each option represents and that the menus are easy to navigate and are accessible.

2. We want to investigate whether users feel safe when using the system to enter patient information and submit responses. The system diagnoses the severity of OCD symptoms, and

the testing aims to ensure that the user receives a clear presentation of the result along with an explanation the user can comprehend.

3. It is important for us to evaluate whether the explanation associated with the assessment of the severity of the patient's OCD symptoms was clear and easy to follow. We wanted to know whether this explanation informed the users why the patient was classified in a particular way, increasing trust in the system and its usefulness.

4. We want to investigate how easily users can navigate between the four tabs: "About", "Predictive Analytics", "Descriptive Analytics" and "Diagnostic Analytics". The testing aims to ensure that each tab is intuitive and that users understand what the different tabs offer.

5. We want to evaluate whether users consider the information presented under the "Descriptive Analytics" and "Diagnostic Analytics" tabs to be clear and informative. It is important that users can understand and trust the analyses offered, which increases their confidence in the dashboard's usefulness and relevance in patient assessment.

2. *Parts of the UI to evaluate:*

Drop-down menus: We evaluate the user's interaction with the drop-down menus to see how intuitive it is to select different patient features.

Result presentation: We review how clearly the result and explanation are presented after the user submits patient information in the "Predictor" tab.

The Tabs: We focus on the navigation between the "About", "Predictor", "Descriptive Analytics" and "Diagnostic Analytics" tabs to ensure that users can easily understand the purpose of each tab.

Clarity of information: We evaluate how clear and informative the content is under the "Descriptive Analytics" and "Diagnostic Analytics" tabs, and whether it provides users with the information they need.

3. *Expected outcomes:*

We expect users to be able to navigate easily and intuitively between the tabs and feel confident entering patient information. We believe that the explanation of the severity of the OCD symptoms is perceived as comprehensible and that it increases confidence in the system's diagnosis.

We also expect that the information under the tabs "Descriptive Analytics" and "Diagnostic Analytics" is satisfactory and informative, and that this strengthens users' confidence in the dashboard's usefulness in clinical contexts.

Quantitative Usability Testing

The System Usability Scale (SUS) was selected as the quantitative method for usability testing of the OCD prediction dashboard because of its simplicity, reliability, and strong reputation in usability research [2]. The SUS-questionnaire is quick and straightforward for participants to complete, and its broad applicability enables us to effectively assess overall usability. Furthermore, the SUS yields a clear, quantifiable score that helps measure user satisfaction and identify specific areas for improvement. Participants rate topics on a scale from 0 (strongly disagree) to 5 (strongly agree), allowing us to measure how user-friendly the dashboard is and ensure it meets users' needs effectively.

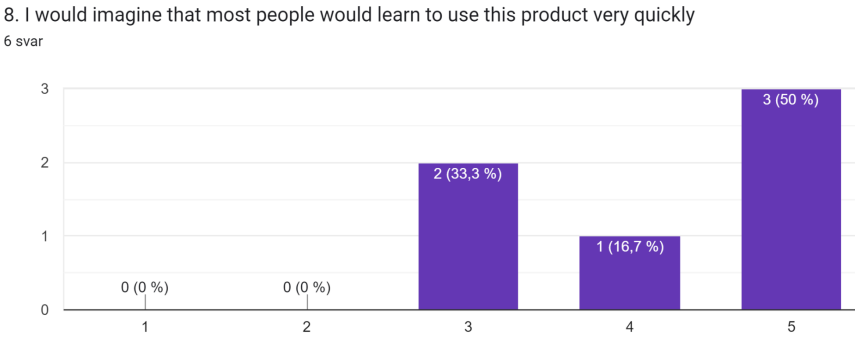
Qualitative Usability Testing

Individual usability testing will be conducted to gather nuanced feedback from participants on the OCD prediction dashboard. We will employ the “Think Aloud” method, which encourages participants to verbalize their thoughts while completing tasks, offering valuable insights into their experiences [3]. This approach helps us identify challenges with the design and areas for improvement. Additionally, semi-structured interviews will be utilized to facilitate in-depth discussions, allowing interviewers to explore participants' experiences, challenges, and suggestions in a flexible dialogue [4]. This method encourages a comfortable environment for participants to share their thoughts openly, leading to honest and detailed responses. The semi-structured format also enables us to focus on specific topics while considering each participant's unique perspective, ultimately informing the ongoing development and enhancement of the dashboard [4].

Usability testing Results

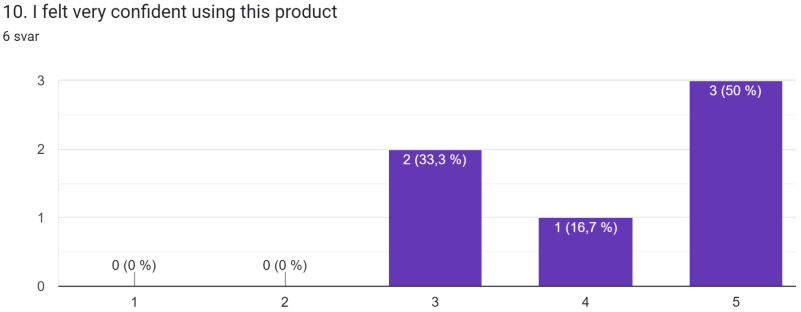
The System Usability Scale (SUS) [2] questionnaire was conducted on six participants who have a health informatics background. Scores ranged from 62.5 to 100, indicating varying levels of user satisfaction and perceived usability. The overall average SUS score of approximately 79.4 suggests that the system is perceived as above average in usability. However, the significant variance in individual scores highlights that while some users are highly satisfied, others experience challenges. To gain deeper insights into user experiences, we also conducted interviews with two participants during usability testing. Their feedback complements the SUS results and provides further context for areas needing improvement. The results can be structured into areas for improvements based on our usability testing goals:

1. Ease of Entering Patient Features: Most users found the data entry process straightforward, but some struggled with unclear options in the drop-down menus. Enhancing labels and adding tooltips could improve clarity. One participant stated, “The features list is good,” but suggested that clearer explanations for each option would aid decision-making.



Picture: Example of question and results from questionnaire

2. User Safety and Confidence: Feedback on safety while entering patient information was mixed. While some users felt secure, others desired more reassurance regarding data privacy. Clear security indicators could help improve user confidence.



Picture: Example of question and results from questionnaire

3. Clarity of Diagnostic Results: The presentation of diagnostic results received varied feedback. Some users found it hard to understand the severity and what to expect. Simplifying language and using clearer explanations could enhance understanding and trust in the system's assessments. One participant also mentioned, “I got a high risk, and that was very obvious because of the color choice—red. It was very informative.” This feedback emphasizes the importance of using color effectively for clear communication.
4. Navigation Between Tabs: Users generally found tab navigation intuitive but suggested clearer visual prompts for each tab's purpose. Brief descriptions or tooltips could improve usability. One participant noted that the navigation was “fast” and the questions were “very direct and easy to understand”. However, they recommended changing titles like "Diagnostic Analytics" for clarity, stating that the term “diagnostic” could imply it refers to information about the diagnosis rather than the dataset being used.
5. Clarity of Information in Analytics Tabs: Users found the information in the "Descriptive" and "Diagnostic Analytics" tabs informative but recommended clearer labeling and better organization to enhance comprehension.

Overall, the usability testing revealed a strong foundation in user satisfaction but also identified areas for improvement, particularly in clarity and user safety. The qualitative

feedback from the interviews reinforces these points, and implementing the suggested changes will increase trust in the system and improve overall usability.

Final working dashboard

1. *The main lessons from the evaluation stage and the changes done in the web dashboard based on the information from the usability tests:*

The main lessons from the evaluation phase based on System Usability Scale (SUS) was to focus on the points mentioned by the user that gave a low score while maintaining the features that received high score. For instance, some users struggled with unclear options in the dropdown menu of the patient entry form, therefore it was rectified and made simpler and easier to understand. Moreover, some users found it difficult to understand the result that was achieved through the prediction form, therefore, SHAP analysis for each result is shown to give the user more insight.

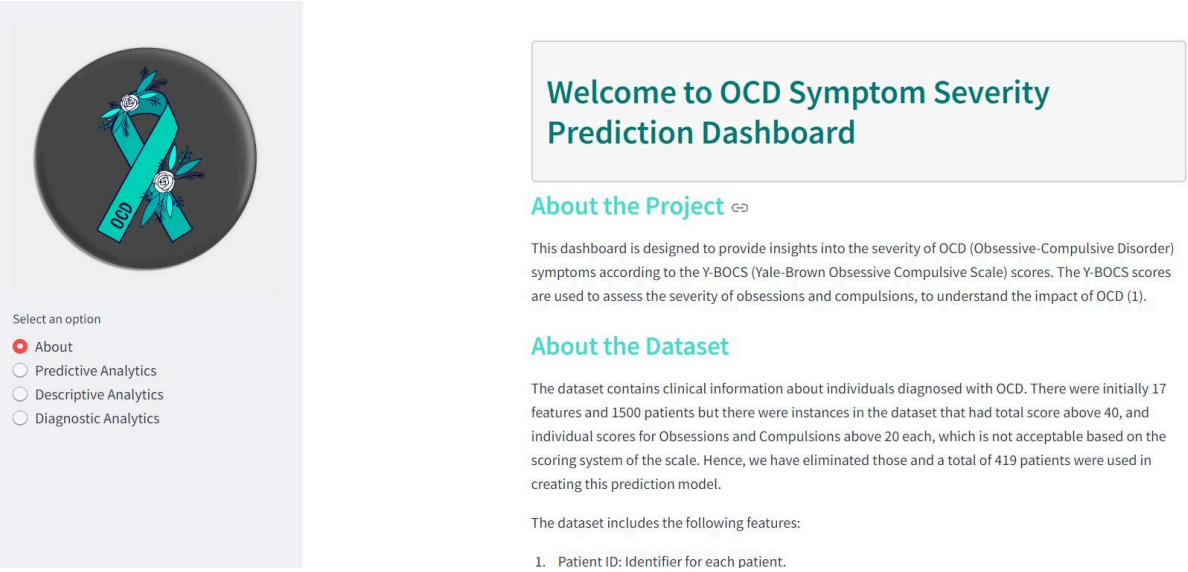
Furthermore, through some user feedback, the “Diagnostic” tab is changed to “Diagnostic Analytics” for more clarity. Another change that we have made is adding proper labels, headings and more information in the Descriptive Analytics and Diagnostic Analytics tab.

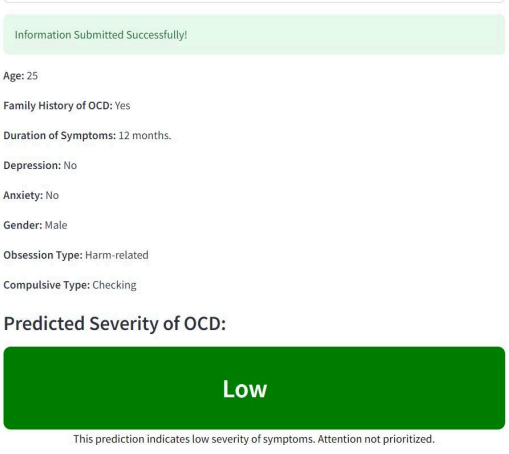
2. *Evidence of the final dashboard. E.g., screenshots, links to online material:*

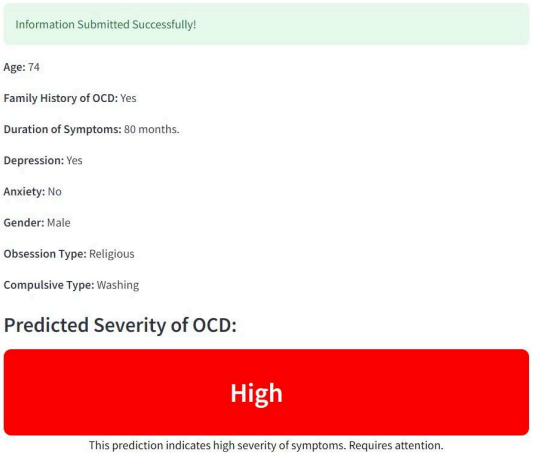
Link to the Dashboard:

https://ocd-predictors.streamlit.app/

Some Screenshots of the final dashboard:







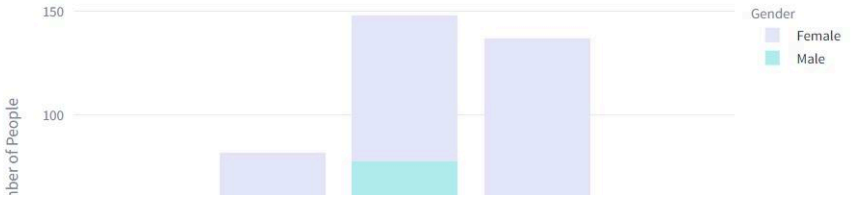
Descriptive Analytics

This tab provided key summary statistic and descriptive insights about the dataset used to train the prediction model.

419	47	195	224
Total Patients	Average Age	Males	Females

Age and Gender Distribution

Number of People by Age Group and Gender



Analysis of Type of Symptom with Total Score

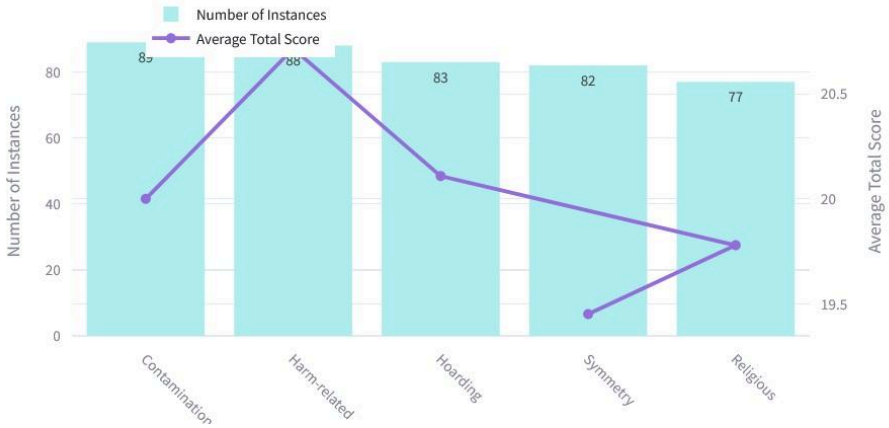
Choose the type of Symptom to analyze:

Obsession Type

Select Obsession Types to Filter

Harm-related × Contamination × Religious × Hoarding × Symmetry ×

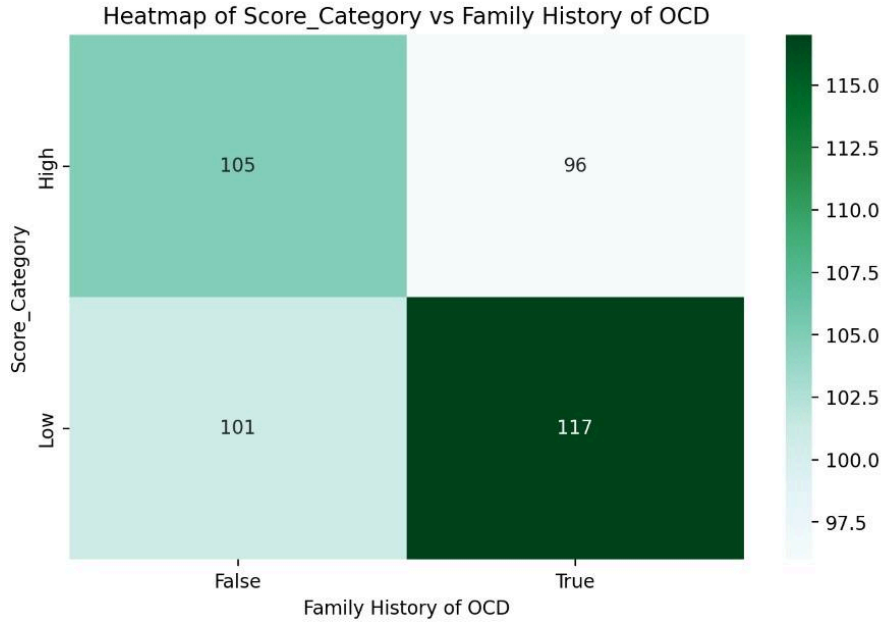
Number of Instances and Average Total Score by Obsession Type



Chi-square Test Results

Chi-square Statistic: 1.23

P-value: 0.2666



Prescriptive Analytics

This section helps you understand which patient factors, like demographics and clinical details, influenced the algorithm's prediction of OCD severity. Prescriptive analytics can guide you in setting treatment goals for the patient. The analytics presented here are specific to the patient for whom you entered the above data.

We have used SHAP (SHapley Additive exPlanations) to explain the output of machine learning model. It provides insights into how much each feature contributed to a specific prediction, making complex models more interpretable.

References

1. Hyper Island. Creating effective teams: the detailed curation. Hyper Island [published 2023-10-03; accessed 2024-10-15]. Available from: https://hyperisland.com/en/blog/thought-leadership/creating-effective-teams-the-detailed-curation

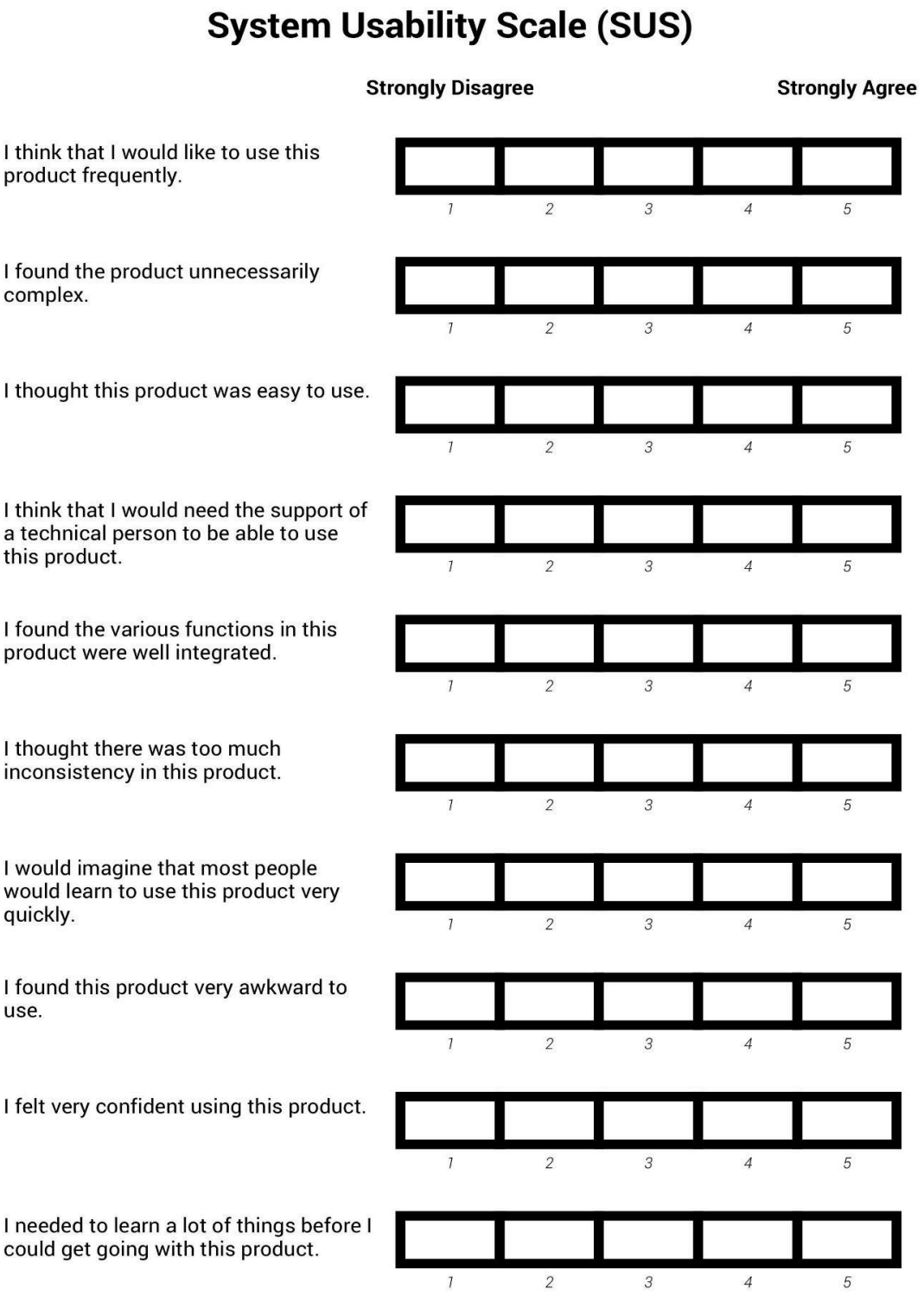
2. Laubheimer P. Beyond the NPS: Measuring Perceived Usability with the SUS, NASA-TLX, and the Single Ease Question After Tasks and Usability Tests [Internet]. Nielsen Norman Group. 2018. Available from: https://www.nngroup.com/articles/measuring-perceived-usability/

3. Maze. 8 Top Usability Testing Methods Explained [Internet]. Maze. Available from: https://maze.co/guides/usability-testing/methods/

4. Davis R, Gardner J, Schnall R. A Review of Usability Evaluation Methods and Their Use for Testing eHealth HIV Interventions. Curr HIV/AIDS Rep. 2020;17(3):203-218. doi:10.1007/s11904-020-00493-3

Attachment 1. The SUS Questionnaire

Evaluation was performed via a form:
https://docs.google.com/forms/d/e/1FAIpQLSdLaA7nz_Q3HqYj9_2jimQs0u8Av4qw5MTDvMLa3Qgwogiqcw/viewform?usp=sf_link



Attachment 2. Semi-structured interview questions.

Interview questions

1. Is the navigation smooth and clear? If not, why?
2. Does the dashboard so far provide all the information you need on the topic? If not, what is missing?
3. Can you describe your experience interpreting the severity predictions for OCD from the dashboard?
4. Generally speaking, do figures seem correct to you? If not, please list the potential errors you have spotted.
5. Did the dashboard provide enough context and information to support clinical decision-making?
6. What features would you like to see added or improved?

Attachment 3. Informed consent for participants of Usability Testing.

Study Purpose

Thank you for participating in our usability testing for the OCD prediction dashboard. The primary aim of this study is to evaluate the usability and effectiveness of the dashboard in helping healthcare professionals assess and manage obsessive-compulsive disorder (OCD) in patients. Your feedback will play a crucial role in improving the tool and ensuring it meets the needs of users in clinical settings.

Testing Process

As part of this study, you will participate in individual semi-structured interviews, which will last approximately 15 minutes. During the interview, we will discuss your experiences using the dashboard, including its navigation, clarity, and overall usability. We will also ask for your suggestions for improvements. The interviews will be documented through notes (with your consent) to ensure accurate data collection, and all information will be kept confidential.

Confidentiality

Your participation is entirely voluntary, and you may withdraw from the study at any time without any consequences. All information collected during the interviews will be anonymized and stored securely. Your identity will not be linked to your responses, and data will be reported in aggregate form only.

Ethical Considerations

We are committed to conducting this research ethically and with respect for your rights as a participant.The study adheres to the necessary procedures set forth by the ethics review board to ensure that all ethical guidelines are maintained.. By participating in this study, you are helping us enhance the usability of the OCD prediction dashboard, ultimately benefiting healthcare professionals and patients.

Consent

Before proceeding, you will be asked to provide informed consent, indicating that you understand the study's purpose, procedures, and your rights as a participant. If you have any questions or concerns about the study or your involvement, please feel free to ask.