Presentation Transcript

Slide 1)

Welcome to my research proposal. The title of my project is gender-specific predictive modeling for early detection of mental health issues. This proposal will cover the significance such a topic has in the current atmosphere, what this project aims to achieve, the key literature related to the project, the project's proposed methodology, the ethical considerations and risk assessment involved with the project, the description of the artefacts that will be created as a result of the project, and a timeline of the proposed activities.

Slide 2)

What is the contribution and the significance of this research? How does this project contribute to the discipline? First of all, it fills in a gap within the previous research. Men and women struggle with mental health differently, and the prevalence and presentation of mental health disorders can differ significantly between the two genders (Rosenfield & Smith, 2010). While reviewing the literature, there was a lack of predictive models that accounted for these gender-specific variations (Islam et al., 2024). This project aims to address this gap and improve the precision and fairness of early mental health detection models, allowing for more targeted and effective interventions. In addition, this project would help give other researchers new directions to conduct future research. Gender is not the only variable that it would be interesting to create specific models for. Age, for

example, or even gender-age as an intersection, may also have an impact on the accuracy of predictive models.

But why is this important to research? Individuals with mental disorders have a much higher mortality when compared with the general population (Walker et al., 2015). While it is tempting to think of suicide when considering mortality in mentally ill patients, many die from other health issues. In fact, Colton and Manderscheid (2006) discovered that individuals who struggle with mental health issues lose decades of potential life and die at younger ages from natural causes than their mentally healthy peers. If the mental illnesses can be discovered early on, then we give individuals the opportunity to access care much earlier than they may have. This ties into the next point on why this is important to research. There is potential here to create a digital platform where users answer a series of surveys that cover sociodemographic information as well as mental health data. With the answers, users would be told whether they are displaying no, low, medium, or high symptoms of a mental health disorder, for example stress or depression. This would enable a user to go and perhaps find a professional that can help them with a diagnosis, give them the care they may need, and more. However, this is just a potential idea, and nothing more.

Slide 3)

The research question that was developed after thinking about this topic is “Can gender-specific predictive models increase the accuracy of predicting the onset of mental illness?”. To verify whether the accuracy increases or decreases, the gender-

specific models would be compared to a general model that has both genders and all variables used in the gender-specific models included.

Slide 4)

The aims and objectives of the project are listed here. First of all, this project aims to develop not only gender-specific models, but also a general model using the survey data collected. Then, by identifying key predictors of mental health issues for the specific genders, we can develop the gender-specific models by seeing which variables are significant for one gender but not the other. In addition, identifying the key predictors of mental health issues may allow us to detect possible directions for new questions for future surveys. After developing the models, then the effectiveness of the models needs to be assessed. Not only will we assess the performance of the models by testing their accuracy, but we will also compare how they perform against each other, with the hypothesis that the gender-specific models will outperform the general model.

Slide 5)

Here is a list of some examples of key literature. One of these articles is a literature review, while another contains its own literature review. The literature review written by Islam et al. (2024) points out that there is a gap when it comes to studies done on gender-specific models for mental illness using machine learning algorithms in this field. This project aims to fill that gap and give future researchers a better understanding of the importance of creating models that focus on a specific variable, such as gender,

age, or other potential categorical variables. Verma et al. (2024) compare the effectiveness of a few different machine learning algorithms for mental health detection, as well as include a literature review that lists various studies that have undergone similar research. As this study already has done research on the optimal predictive models for mental health detection, choosing a model for this research project will be simpler. However, it is important to still test multiple models, as the accuracy of a model varies depending on the data it is trained on. The last study (Rosenfield & Smith, 2010) is a chapter in a book that examines the evidence that prove how and why men and women experience mental illness differently, as well as show different symptoms. This study will help in creating a survey that has questions targeted at either men or women as well as them both, or it will help in identifying existing surveys that fulfill the needs of this research project.

Slide 6)

For the methodology, the following overarching steps are shown on the slide. When it comes to data collection, quantitative sociodemographic data as well as quantitative mental health data will be collected via surveys or obtained through collaborating with mental health organizations. Exploring the available datasets on websites like Kaggle may be worth doing, but new data would be optimal. Controlling the sample so that there is an equal amount of male and female participants is necessary to split the data into 2 for the gender-specific models. After collecting the data, the data preprocessing step would occur. The necessary steps would be taken to ensure the data is ready for analysis, such as addressing missing responses, normalizing scores, and applying one-

hot encoding for any categorical variables. Exploratory data analysis comes next, where key predictors of mental health issues for each gender will be identified and extracted to use in the model creation, as well as any significant differences in variables between genders.

After these first three steps, the next three steps involve model creation, model evaluation, and model comparison. As mentioned, three models will be made in total. An all-genders model, a female-only model, and a male-only model. After reviewing the key literature, the model that performs the best appears to be logistic regression, however it would still be best to compare logistic regression, random forest, SVM, and other models before choosing a final model for each gender category. After optimizing the model performance, we then would evaluate the models using metrics such as ROC-AUC. These processes will likely occur simultaneously. For example, we may make a male-only logistic regression model, random forest model, and SVM. We would need to evaluate the models, optimize their performance, and then evaluate them again a few times, before finally choosing the one with the best performance. Only after doing this process for all the gender categories will we compare the final three models. By testing accuracy the way we did during the model evaluation test, we can determine whether the gender-specific models outperform the general model.

Slide 7)

When dealing with data, especially sensitive personal data such as mental health information, it is vital to undergo ethical considerations and assess the risk of the

project. The data of all participants will undergo anonymization to prevent the identification of individuals. The data will also be securely stored and encrypted to ensure the data is protected. If the data was to be breached it would lead to the exposure of a large amount of sensitive personal information, so only the researcher and the supervisor will ever have access to the data. The data will be deleted upon completion of the research project unless other guidelines from the university or ethical review boards determine the data should be kept for a different period of time.

Participants will be fully informed about the nature of the data being collected, how the data will be used, and the purpose of the research before data collection. Participants will also be given the option to withdraw their data at any stage of the research process without any consequences or penalties, up until the data has been anonymized. Once the data has been anonymized, there will be no way for the researcher to be able to know what data belongs to which individual. However, encryption will still be important to maintain, as no data can be fully anonymized. The participants will not be told the outcome of the research, as the impact of incorrect predictions may be harmful to the participants. Within the written research proposal, it will be important to emphasize that the models would be supplementary tools and not definitive diagnostic systems. Survey questions will undergo careful consideration as to prevent unconscious bias affecting the answers. The fairness of the models will also be evaluated by analyzing their performance across various demographics, ensuring that none of the models disproportionately affect any specific group.

Slide 8)

The only artefacts that will be created during this project are the three predictive models. A female-only predictive model, a male-only predictive model, and a general or all-genders predictive model. If it is necessary and data cannot be obtained from a mental health organization or other location, surveys will be made as well.

Slide 9)

Here is the proposed timeline that I developed after considering what needs to be done to ensure the research project is finished on time. The timeline is done as a Gantt table, to show how many things will occur simultaneously. The first two project activities – “clarifying the research question and the topic” and “securing consistency in design” - are things that I believe require the help of a supervisor. Although I very much like the research question and topic I have come up with and believe I have come up with a solid design, I do not have the same experience with research that someone experienced in academia has. I have therefore given these activities quite a bit of time, but I believe that 2-3 weeks should be enough. Then the ethical approval application must get approved before any data collection and take place. I hope that this is done as quickly as possible, so the data collection can occur as soon as possible. However, I have still given the activity 2 months of time, just in case.

I believe that data collection may take the longest time out of all the activities, outside of writing the report itself. I would like to collect data from as many different sources as possible to cover as many demographics and cultures as possible. As soon

as the data collection is finished, then I will begin the data analysis. I will be writing the report as I undergo data collection and data analysis, as there will be many parts of the report that can and should be written before any conclusions about the data can be drawn. Then, I hope to take one last month to review my dissertation, perhaps discuss it with my supervisor, and then submit it.

Slide 10)

Here are the references I used throughout the presentation. Thank you very much for your time. I hope that you found this presentation enjoyable.