Collaborative Filtering Recommender System for Treatment of Depression

Abigail Orlando, Keenan Venuti, and Matthew Tesfalul
CS108 Final Project

The process of medical diagnoses, physician treatment, and pharmaceutical prescriptions has changed immensely in the last 50 years alone (Altman). With efficacy testing and improvements in education, patients are consistently receiving the best, and constantly improving, treatments for their conditions. Yet, the success of modern western medical systems does not translate to all types of diseases or disorders. With rarer or newer diseases, doctors often improvise and extrapolate from prior knowledge. However, mental health disorders are not rare, new or unknown. The application of the traditional observation, diagnosis, treatment, rehabilitation process does not coordinate as effectively with mental health disorders, particularly depression. In 2014 in the U.S., 15.7 million adults or 6.7% of the U.S. population experienced one major depressive episode (ADAA). The lifetime risk of exhibiting Major Depressive Disorder (MDD) symptoms, qualifying a diagnosis, is 17%. Despite the large portion of people affected by depression, there are very little standard diagnosis technologies in hospitals. Diagnoses procedures range from physical exams, to lab tests, to psychological evaluations using the DSM-V (Mayo Clinic). The DSM-V serves as an important manual to depression, but the generality of symptom descriptions and seemingly non-corporal causes of mental health disorders, compared to physical ailments, creates huge diagnosis and treatment limitations. These limitations can cause harmful outcomes. Depression is overdiagnosed; research found that 26%-45% of patients referred for depression do not meet diagnostic criteria (Bostwick). This inevitably harms treatment efficacy ratings as physicians may believe a treatment is ineffective because it does not alleviate symptoms when it may be treating the wrong disorder. It also shows that there are inherent flaws in methods of diagnosis that psychiatrists are currently using. Due to the diagnostic nature of mental health disorders, shared symptoms across disorders, and

comorbidity of disorders, current applications of processes will naturally be relatively ineffective (DSM).

When a patient experiencing symptoms of MDD visits their doctor, they typically fill out surveys accumulating information regarding themselves (demographical data), behavioral data, and any treatments they have tried. With this patient-specific information, an extensive medical education, and experience with depression diagnoses and prescriptions, the doctor makes a treatment recommendation focusing on severity of symptoms and perceived efficacy. We saw similarities in the process to how a recommender system, like Amazon's product recommendations chooses what to display to users. Both recommendations are based on user demographics, user behavior, and user ratings of products/treatments. A doctor's knowledge capacity is naturally limited; we see this as an opportunity to incorporate the extensive memory storage and computing capacity into mental health diagnoses and prescriptions. As such, our aim is to apply a modified diagnosis process by utilizing artificial intelligence to interpret, evaluate, and estimate treatment efficacy for individual patients based on similarities to patients that have been treated before.

We implemented a collaborative filtering (CF) recommendation system, utilizing patient demographic information and patient ratings of treatment efficacy to recommend treatments to new patients. Starting with a general demographic and behavioral dataset, segmented to focus on people with mental health issues, we created user clusters based on demographic and behavioral similarities. With an appropriate distribution of mental health afflicted users, our model needed efficacy of treatment ratings by user, to make recommendations post filtering. We utilized trends in the most popular depression-centric treatments to simulate a dataset of probable efficacy

ratings for each user. As our system is used, new patients will submit their demographic data and ratings of treatments to the continually growing dataset. By implementing both an item-item and a user-user hybrid learning model, our model can look at patient to treatment reactions objectively. Our goal is to replace the innate "psychiatrist's hunch" of prescription that can lead to misdiagnoses and ineffective prescriptions. By comparing and grouping users on predetermined demographic similarities and comparing treatments based on efficacy rating similarities across users, our model can recommend a treatment for an entirely new patient or one that has tried depression treatments before. In time, with re-entry of patient data, our system will learn from its own recommendations, figuring out what works and does not work. We focused on finding the most effective treatment for an individual user given information on them, rather than a blanket treatment applied to all users with the same disorder. This methodology accounts for some of the discrepancies in mental health treatment, primarily an unfair assumption that depression affects and influences every person the same.

Given the nature of medical diagnoses and the importance of patient safety and confidentiality, we considered several ethical concerns when designing our project. We decided to implement within a consequentialist, specifically utilitarian, framework with deontological constraints. We believe our system will make the best decision and only be implementationally functional if it optimizes treatment for the entire population of those suffering from MDD, rather than considering individual autonomies. However, we feel there are certain unalienable rights of an individual and a few general limitations, already implemented within the mental health community, that we must be sensitive too. While we did not focus on optimizing personal freedoms, we found a set of constraints to properly restrain a sometimes seemingly immoral

utilitarian framework. Our model inherently requires current patient data to make predictions and assessments for future patients. We felt that a pure deontological system may be limited when weighing current people's autonomy against future people's autonomy. Similarly, our system could not function within a deontological framework given the amount of user and treatment cold-starts it accounts for. When a patient enters the system with just demographic data or limited treatment efficacy ratings, our system decides as to the most effective treatment for them based on similar users or treatments. Our system bases this consensus on data, probabilities and how this will help the overall knowledge of depression treatments. While it considers side effects of treatment or possible harm to each patient, if benefit exceeds those costs, within our constraints, the system will recommend based on this utility calculation. Some deontologists may consider the possible harm to a patient as a sacrifice of their autonomy; our model considers the overall benefit to depression research and its convictions as to when a benefit exceeds associated costs.

Recommender systems use a variety of data mining techniques and algorithms to generate personalized recommendations for users. Recommender systems have applications in a variety of fields and in the past decade have become extremely popular due to their usefulness in ecommerce (Konstan et. al.). Users today are inundated with options, and recommender systems help to filter through these options by using algorithms to predict which items a user will like. Some examples of recommender systems are Google search, Spotify music recommendation, Amazon product recommendation, and Netflix movie recommendation. Recommender systems make recommendations based on several factors like user ratings of items, user demographic information, and context specific user behavior. The more accurate a recommendation is, the

more likely a user is to use or purchase the product. As such, a lot of research has gone into recommender system algorithms; Netflix even hosted a 1 million dollar competition to find a movie recommender system with the highest prediction accuracy (Netflix Prize).

There are many types of recommender systems, but most are fundamentally filtering algorithms. For recommending treatment for depression, a collaborative filtering model is the most appropriate as it makes predictions about the preference of a user by collecting preferences from many other users -- a technique that doctors naturally use all the time. The high level assumption is that similar users will have similar item ratings. Spotify utilizes this technique to make music recommendations for individual users. First, the system finds a group of users who are most similar to an individual user based on song ratings. Then, the system recommends songs to the individual user that were highly rated by the individual user's most similar users. While the predictions are personalized for the user, they come from the collaboration of many users. In other words, a large population collectively listen to all the songs and find the best ones over time. A new user benefits from all this hard work by only listening to the filtered best songs.

One big problem with a recommender system for depression treatment is that a lot of our users will have no previous treatment ratings, known as cold-start users. Collaborative filtering models rely on sufficient user rating history to find the most similar users and then to make predictions of user preferences based off those most similar users. Here we present a new hybrid collaborative filtering recommender system framework for treatment of depression, inspired by Tiwari and colleagues, that solves the cold-start user problem by taking advantage of user demographic and behavioral information to help find similarities between a new user and already existing users in the system (Tiwari et. al.). Demographic information is composed of various

user characteristics like age, gender, state, income, etc. Behavioral information is comprised of self-observed symptoms within the last 30 days; these symptoms range from lack of eating/sleeping to trouble concentrating and a total disinterest in daily activities. The system calculates recommendations for a new user by predicting preferences within a much smaller cluster of most similar demographic users rather than from the entire user base.

A collaborative filtering system predicts a user's rating of items not yet rated with two basic approaches. First, user-user CF, which recommends items that were rated highly by users who are similar to the new user. Second, item-item CF, which recommends items that are similar to the items that a new user has previously rated highly. Our system combines both approaches.

User-user CF relies on the assumption that people who have similar ratings in the past are likely to have similar ratings again. If so, we can predict a user's rating of an item, by utilizing similar users' ratings of that item. Similarities between each pair of users is calculated by finding the difference in ratings for each item and then taking the average of these differences. Then, prediction of a user's item rating is calculated by taking a weighted sum of the item's ratings from similar users (Ekstrand et. al.).

User-user CF has two main problems: user cold-start problem and sparsity. First, since user-user CF relies on finding similarities in item ratings between users to make a recommendation, a cold-start user, who has no item ratings, cannot be given a recommendation. Second, if the data is sparse, meaning users have rated only a few items, similarity measures between users are unreliable. A solution to both problems is to use demographic and behavioral data to find similarity between users. While it is hard for Spotify and Amazon to collect personal

demographic/behavioral information from their users, doctors are privy to this information and in many cases it heavily weights their treatment recommendation.

Item-item CF relies on the assumption that a user is likely to have the same rating of similar items. If so, we can predict a user's rating of an item by utilizing the user's ratings of similar items. Similarity between each pair of items is calculated by taking the average of the difference in item ratings across users. To predict a user's rating of an item, find the items that are most similar to the item of interest and take a weighted average of the user's ratings of those items (Ekstrand et. al.).

Item-item CF has two main problems: user cold-start and item cold-start. First, since item ratings come from a user's ratings of similar items, item-item CF also faces the user cold-start problem. Second, item cold-start is when a new treatment is created, but no users have rated the item. For our system this is a huge problem, because if nobody uses and rates a new treatment, then our system will never recommend it to a user. Unlike the user demographic solution to the user cold-start problem, there is not a clear solution here. Ultimately, new treatments will have to be recommended to new users based on limited data from clinical trials. Since this type of recommendation would have low statistical confidence, it raises several ethical concerns that we will address later on.

For our hybrid system, we combine user-user and item-item CF techniques (see Appendix 1 for diagram of system). For our user-user implementation we first cluster all users with a k-means clustering algorithm based on user demographic and behavioral data. All user-user filtering will come from users in this cluster, not from the entire dataset. If a user is a cold-start user, then treatment rating predictions are calculated by taking the average of treatment

rating predictions of all users in the same demographic cluster. If a user is not a cold-start user, then treatment rating predictions for unrated treatments are determined using traditional user-user CF. There are many techniques that can be used for traditional user-user CF, but the most common, and the one we use in our system, is the k-Nearest Neighbors technique, which returns the k most similar users to the user of the system. For each unrated treatment, predictions are made by taking an average of the k most similar users' ratings of the treatment. For our item-item implementation, treatment rating predictions for unrated treatments are determined using traditional item-item CF. Here, we also use the k-Nearest Neighbors (kNN) technique, but to find the k most similar treatments to the treatment of interest. For each unrated treatment, predictions are made by taking an average of the user's ratings of the k most similar treatments. Since the item-item CF relies on treatment rating data, it does not handle cold-start users. Lastly, our system takes a weighted sum of the rating predictions returned from item-item and user-user filtering to provide the top three recommendations to the user. In the case of cold-start users, the recommendation is exclusively user-user CF, because our item-item system does not handle cold-start users.

As mentioned in the design above, the recommender system operates by using two different collaborative filtering techniques, user-user and item-item filtering, to cluster users by both demographic and treatment information to recommend a treatment. For our hybrid system, we need two datasets to perform these techniques. These datasets include a demographic and a treatment efficacy rating dataset. The two are linked in that a user provides information and answers questions, and next an entry is created in each of the datasets with the appropriate

information is inputted. In a perfect scenario, these datasets would include only users dealing with mental health and depression.

First, we considered the design of the ideal demographic dataset to conduct the user-user filtering portion of the system on the similarities in demographic information and behaviors between users. In the design of the dataset and its interaction in our user-user model, we consider relevant information that was detailed enough to cluster the users and predict depression symptoms. This dataset describes all the user specific biographical information such as age, race, gender, weight, height, etc., and also includes identifiable characteristics that define the user's behavior patterns such as drinking habits, loneliness, sleeping history, and medical prescription history. In addition to the content, our design of the dataset also considers the types of responses whether it be quantitative or qualitative, and ensures that they are measurable given that similarity between users is quantified by k-means in order to create clusters and later by kNN to predict treatment recommendations. This meant transforming certain qualitative/categorical data responses into binaries in our dataset and framing certain demographic and behavioral questions into numerical values to capture differences in responses. This first dataset must contain limited missing or incomplete data in order to generate valid conclusions, recommendations, and for the clustering techniques to work properly.

To implement user-user and item-item CF, a treatment efficacy rating dataset with the same users as the demographic dataset is required. Ratings are defined on a scale from 1 to 5 with five having the highest efficacy rating. A user rates all the treatments that he or she has tried. Ideally, the dataset will include all used and or tested treatments and would grow as more treatments are discovered and users rate them. The treatment efficacy rating dataset contains

ratings of specific treatments such as the prescription of drugs like SSRIs or SNRIS, exercise, alter diet, and electroconvulsive therapy (ECT). The treatments range behavior, talking, pharmaceutical, holistic/spiritual, and direct-treatment therapies. Ideally, the users in our dataset should have tried and rated a few treatments in order to know exactly which treatments work and are preferred for an individual, however, as mentioned earlier, our system can handle cold-start users given their demographic information.

An issue arose when modeling these datasets for our hybrid system regarding the timeline of treatments and the demographic data. Given that the treatment dataset contains all the treatment ratings the user has ever tried in his life, there is a concern that changes in behavioral-demographic data will not be accurately reflected by changes in treatment data. Certain demographic questions such as behavioral and time sensitive experiences in the past 30 days will not always align with the timeline of the treatment efficacy ratings for when the user tried the treatment. One solution to this issue we considered was to have the entry in demographic and treatment rating dataset also factor the timing of when the user tried the treatments. This would allow the flexibility for the system to understand if patient-A rated a certain medicine working 2 years ago is different than if the patient-B rated the medicine not being effective in past 30 days. Our system should take this into account when predicting a recommendation for a patient with similar demographic information to patient-B and patient-A.

While collecting data from public datasets would be difficult, we do think we could obtain the necessary data for the two datasets by partnering or reaching out to psychiatrists to pilot our system. After the psychiatrist receives the patient's consent, the doctor would provide the patient with a questionnaire to fill the relevant demographic, user specific information, and

treatment ratings of past treatments tried. The psychiatrist would diagnose or recommend treatments for patient to try and the cycle continues. Note that on each occasion the user fills out the demographic information and treatment ratings. Thus, we could track the historical timing of the treatments and demographic ratings by adding a row for each visit.

Since mental health datasets are private and unregularized, for the purpose of implementing our version of the model, we improvised and made tradeoffs by using real demographic data and simulated treatment data. For the demographic dataset, we researched, cleaned, and segmented a dataset from CDC's 2015 Behavioral Risk Factor Surveillance System (BRFSS) phone survey data to represent the dataset. The dataset initially consisted of (~500,000 people) and contained responses to survey questions in modules spanning categories such as anxiety and depression, cardiovascular health, diabetes, and cognitive decline (CDC-1). Since the survey was conducted in several states, some states asked questions from only a few modules (CDC-2). In order to compensate for this we segmented down the samples to mostly the respondents (~5,000 people) that were asked and answered the mental health module (see appendix 1 for Cohort diagram). Next, we utilized the DSM V as a source for important depression symptoms and characteristics to extract the relevant questions from the CDC's BFSS survey that provided information on user's behavior trends, insights regarding rates of depressive symptoms, and additional biographical information (DSM V).

For the treatment efficacy rating dataset, we listed several specific treatments like in the ideal design described in the earlier section and created a short questionnaire that a new user could fill to inform if he or she tried any of the treatments (see appendix 1 for Treatment Dataset Documentation). For the simulation of the treatment efficacy rating, we used small assumptions

from a user's response to the demographic and behavior questions to generate a rating for each of the treatments. We also ensured random propagation of N/A for treatments to simulate a typical user's treatment history that we would expect in an ideal system. For instance, a user who reported "yes" to a question asking, "Are you now taking medicine or receiving treatment from a doctor or other health professional for any type of mental health condition or emotional problem" would most likely have tried and rated an SSRI drug. Therefore, they were given a higher probability in simulation to rate the treatment in that genre. However, the treatment's exact rating on the scale was randomized since it would be difficult to draw trends to make a consistent simulation for all treatments. Because this simulation draws on assumptions and small probabilistic sampling, no conclusions should be drawn from the results of our system that justify recommendations for treatments. Rather the system should be evaluated for the effectiveness of the design and implementation of the model and datasets for the goal of generating predictions for a new user using techniques to cluster the user based off of demographic information and treatment ratings.

Given the nature of input of our system, the large amount of people it makes decisions on, and the potentiality of misuse, we chose a utilitarian framework with deontological constraints to help guide decisions and practices. Despite our system's ability to address and potentially benefit everyone suffering from Major Depressive Disorder, we feel that there are certain individual rights that cannot be ignored. In pursuit of a system that makes decisions most people would agree with and see as necessary, we implemented the following constraints to an otherwise utility driven decision making process. First, and perhaps most important, our model requires the consent of usage by a patient. This consent carries certain implications. A patient

must consent to submitting both treatment and demographic data for the system to evaluate and use. Despite our system's ability to possibly make beneficial recommendations based on demographic or treatment data regarding the entire depressed population, we believe a patient should know when a psychiatrist's decision was in part from a recommender system's recommendation. All patients must consent to the possibility of being patient zero. In other words, patients must tolerate being the first patient to receive a new to the market treatment that might not have been prescribed by psychiatrists yet. Most patients may prefer a treatment that has documentation and years of implementation behind its usage. However, our system attempts to make recommendations that benefit the entire MDD population. If no one is willing to try or guinea pig a new treatment, our system will not be able to implement that treatment as a recommendation effectively. There is a low chance of this occurring patient to patient, but we felt that it required consent as it may cause slight harm or unknown side effects. Lastly in our requirements of consent, every patient must submit honest demographical/behavioral and treatment efficacy data. For our system to work based on patient subjectivity, it must utilize as accurate of data as possible.

Beyond necessary consent, we considered two other important constraints. In accordance with current psychiatric practices, our system will not recommend children or the elderly "black-banded" or tier level 3 medications even when the data recommends an extreme prescription. Prescribing these types of drugs to children or the elderly is an unnecessary risk; in their current form they may cause more harm than good. Finally, our system will never go as far as to sacrifice a life to benefit others. Given the relative safety of depression treatments, we do not foresee this as a problem. However, we are sensitive to the power effects certain treatments

and mistreatments can have on those afflicted.Despite this rule being relatively anti-utilitarian, we do not believe it is appropriate for our system to make a trolley-esque decision, sacrificing one to save five.

We also considered the importance of accuracy and confidence ratings when our system makes a recommendation. Given the nature of misdiagnosis and incorrect initial prescription within the mental health world, we considered when it would be appropriate or necessary to implement our system based on accuracy and confidence. For example: if our system could predict with X accuracy and Y confidence that is below a doctor's ability is it morally permissible to use our system? Further, if our system can predict with X accuracy and Y confidence that is equal to a doctor's ability, is it morally permissible to use our system or a doctor's recommendations interchangeably? Lastly, if our system could predict with X accuracy and Y confidence that is above a doctor's ability, is it morally necessary to use our system instead of a doctor's recommendations? However, looking through the modern lens of human/computer AI system interactions, a combination of human and artificial intelligence usually leads to the most effective overall system. We predict, the greatest possible recommender system is one that utilizes both a psychiatrist's trained and experienced knowledge and our algorithm. Yet, we cannot prove or disprove this claim until post implementation; it is instead based on similar situations utilizing human and artificial intelligence.

Despite inspiring advances in the field of mental health diagnosis and treatment, prescribed treatment efficacy and accurate diagnosis in its current form needs improvement. Due to innate obstacles such as, cross-disorder applicable symptoms, comorbidity, and subjective symptoms, psychiatric professionals must innovate and create new processes regarding mental

health. Our aim, to describe, create, and implement a CF recommender system, used in tandem with a doctor, will provide fresh methodology to the current psychiatric diagnosis process. Starting with a demographic dataset paired with simulated treatment data, our hybrid user-user, item-item CF model can cluster users based on demographic similarities and compare associated treatment efficacy ratings. Foreseeing possible ethical contradictions regarding our model's goal in providing the best treatment to the most amount of people and the personal freedom to choose who makes individualized healthcare recommendations, we constrained our naturally utilitarian system with appropriate consent-related deontological constraints. With the success of our clustering algorithm, we can focus on expanding its input, output, and applicability.

Our initial research and implementation yielded promising functionality, but this paper serves as a baseline framework for future projects. Down the line, we would like to further document and analyze user demographic data by creating a more comprehensive survey and reporting system. This report would also include how often a patient misses a treatment and account for time specificity of treatment to behavioral symptoms. We would also like our treatment dataset to eventually contain treatment interactions to measure efficacy of when treatment A and B are taken in tandem. Finally, given the similar processes of mental health diagnosis across disorders, our model can be scaled to function similarly on other psychiatric disorders. This paper outlines ideas and proposes a path to creating a recommender system for Major Depressive Disorder treatments. This system, when implemented correctly will complement current diagnosis and prescription psychiatric practices.

References

Altman, Lawrence K.. “So Many Advances in Medicine, So Many Yet to Come”. The New York Times. Dec. 26 2006. Nov. 20th 2016.

American Psychiatric Association. DSM 5. American Psychiatric Association, 2013.

Bostwick, Michael, et al. “Recognizing mimics of depression: The ‘8 Ds’”. Current Psychiatry, MDedge. June 11th 2012. Nov. 20th 2016.

Centers for Disease Control and Prevention (CDC-1). *Behavioral Risk Factor Surveillance System Survey Data*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2015. Nov. 20th 2016.

Centers for Disease Control and Prevention (CDC-2). *Behavioral Risk Factor Surveillance System Survey Questionnaire*. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, 2015. Nov. 20th 2016.

Ekstrand, Michael, et. al. “Collaborative Filtering Recommender Systems”. Foundations and Trends in Human Computer Interaction. 2012.

"Facts & Statistics. Anxiety and Depression Association of America”. ADAA. . Web. Nov. 20th 2016.

Kontan, Joseph, et. al. “Recommender systems: from algorithms to user experience”. User Modelling and User Adapted Interaction. March, 2012.

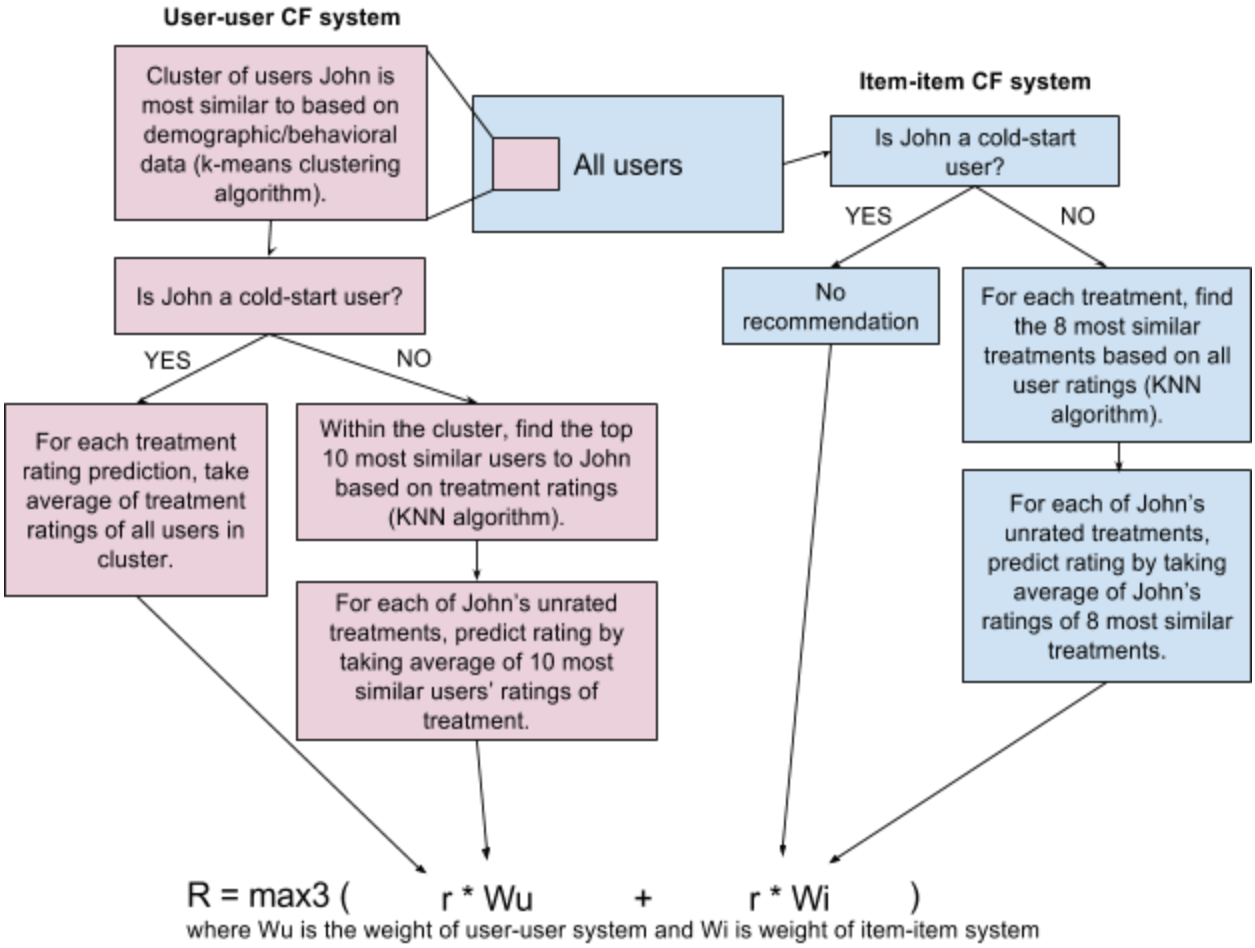
Mayo Clinic Staff. “Depression (major depressive disorder)”. Mayo Clinic. July 7th 2016. Web. Nov. 20th 2016.

Netflix Prize. Retrieved from http://www.netflixprize.com

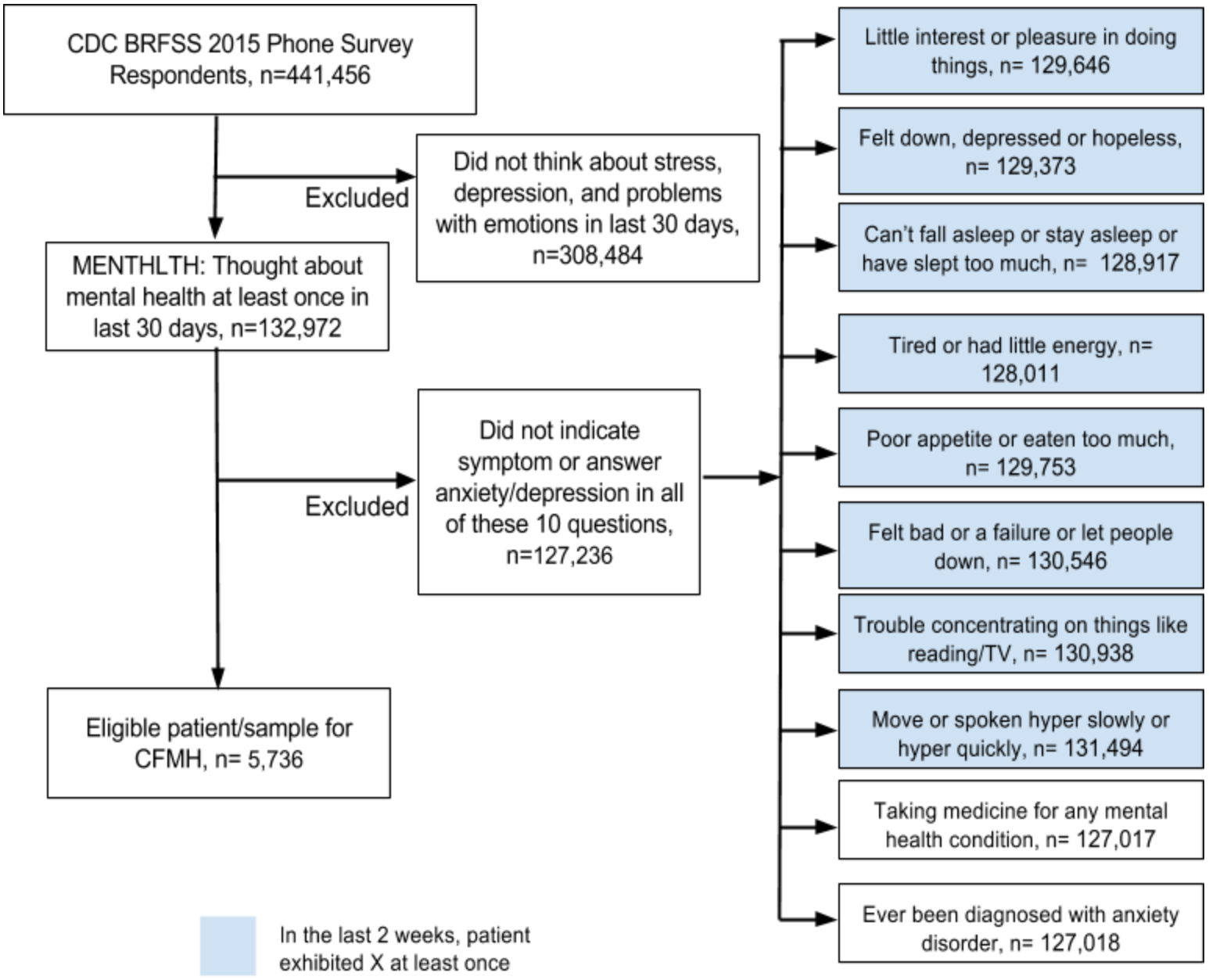
Tiwari, Saurabh, et. al. “An approach for Recommender System by Combining Collaborative Filtering and User Demographics and Item Genres”. International Journal of Computer Applications. October, 2015.

Appendix 1

1. Diagram of system



2. Cohort diagram



3. Treatment Dataset Documentation and Questionnaire

Treatment input: This dataset will be accumulated through detailed patient questionnaire responses within the MDD diagnosed population

Talking Therapies:

- Subset - Psychoanalytic Approach
- Subset - Cognitive Approach
- Subset – Clinical Behavior Analysis

Pharmaceutical therapies

- SSRIs
  - Celexa
  - Lexapro
  - Luvox
  - Paxil
  - Prozac
  - Zoloft
- SNRIs
  - Cymbalta
  - Effexor
  - Pristiq
- NDRIs
  - Wellbutrin
- Trycyclics
  - Elavil
  - Norpramin
  - Pamelor
  - Tofranil
  - Ludiomil
  - Asendin
- SARIs
  - Nefazodone
  - Trazodone
- MAOIs
  - Marplan
  - Nardil
  - Parnate

Behavioral therapies

- Exercise
- Water
- Changed Diet
- Decrease in Alcohol
- Meditation
- Abstention of recreationally used drugs

Holistic/Spiritual Therapies

- Prayer
- Spiritual Advice
- Aroma Therapy

Direct-Treatment Therapies

- Transcranial Magnetic Stimulation
- Electroconvulsive Therapy
- Psychosurgery
- Deep Brain Surgery (implementation of electrodes into brain)

Questionnaire for Treatment Efficacy Ratings:

Talking Therapies

- PSYDAN: Have you participated in any form of psychodynamic therapy in the last 30 days? (This includes psychoanalysis) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- COGAPR: Have you participated in any form of cognitive therapy in the last 30 days? (This includes CBT) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- CBA: Have you participated in any form of clinical behavior analysis therapy in the last 30 days? (This includes RFT) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)

Pharmaceutical Therapies

- USESSRI: Have you taken a SSRI as a part of your medical prescription in the last 30 days? (This includes Celexa, Prozac, Zoloft, etc.) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- USESNRI: Have you taken a SNRI as a part of your medical prescription in the last 30 days? (This includes Cymbalta, Effexor, Pristiq, etc.) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- USENDRI: Have you taken a NDRI as a part of your medical prescription in the last 30 days? (This includes Wellbutrin, etc.) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- USETRY: Have you taken a Trycyclic as a part of your medical prescription in the last 30 days? (This includes Elavil, Tofranil, Asendin, etc.) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- USESARI: Have you taken a SARI as a part of your medical prescription in the last 30 days? (This includes Nefazodone, Trazodone, etc.) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- USEMAOI: Have you taken a MAOI as a part of your medical prescription in the last 30 days? (This includes Marplan, Nardil, Parnate, etc.) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)

Behavioral Therapies

- EXER: Have you made any positive (increase in effort, time, or intensity) changes towards your exercise in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- DIETPOS: Have you made any positive changes towards your diet in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- ALCRED: Have you reduced your alcohol intake in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- DRUGRED: Have you reduced your recreational drug intake in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- MEDINC: Have you increased time spent performing mindfulness actives such as (mediation) in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)

Holistic/Spiritual Therapies

- PRAYINC: Have you increased time spent praying in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- SPIRITAD: Have you sought and received spiritual advice in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- AROM: Have you practiced aroma therapy in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)

Direct-Treatment Therapies

- TMS: Have you received transcranial magnetic stimulation in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- ECT: Have you received electroconvulsive therapy in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- PSYSUR: Have you received any form of psychosurgery in the last 30 days? (This includes) If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)
- DEEPBRAIN: Have you received deep brain surgery (implantation of electrodes) in the last 30 days? If yes, please rate its effectiveness as a treatment on a scale of 1-5. (N/A, 1-5)

4. **CDC Behavioral Risk Factor Surveillance System (BRFSS) 2015 Land-Line and Cell-Phone Data**

Link: http://www.cdc.gov/brfss/annual_data/2015/pdf/codebook15_llcp.pdf

5. **Link to ipython notebook on github**

Link: https://github.com/abigailorlando/CS108-FinalProject

Appendix 2

1. **Work Division**

Abigail: Took the lead on the research, cleaning, and segmentation of demographic dataset. Led design of hybrid collaborative system and the implementation of the system using python, numpy, scikit, and pandas. Led the implementation of the algorithms and presented the design of the model in ipython notebook.

Keenan: Took the lead on the implementations of treatment dataset design, research on MDD in DSM V, discussion and framework for ethical focus of our project. Analyzed strengths and weaknesses of model for solving this problem using psychological and health research. Provided techniques to measure similarities and organize the dataset responses and questions to best fit our collaborative filtering model.

Matthew: Helped Abigail with technical implementation of the model, assisted in the design and interactions of the two datasets, and simulation of treatment efficacy ratings. Researched clustering techniques in order to find best fit for our problem's context and available data. Also led discussions and work on front end interactions and design of system between patient and doctor.

Together: We met weekly to discuss direction of the design of our technical and ethical focus. While each of us led phases in the development of our project, we all contributed research, ideas, and implementation before and during meetings. We utilized our strengths and diverse backgrounds to bring expertise to the group and teach other. Our shared personal goals of learning new computational and statistical methodologies, and engaging ethical challenges that our system our system created helped facilitate this collaboration smoothly.