McMaster University | ENGINEERING Computing & Software

Natural Language Processing for Mental Health Risk Prediction

Requirements Standard Plan

Michael Breau, Jessica Dawson, Benjamin Chinnery, Yaruo Tian, Matthew Curtis

Version 1, 2024-04-04

Table of Contents

Academic Integrity Disclaimer

We would like to acknowledge that as a dedicated students of McMaster University, we have thoroughly read and comprehended the Academic Integrity Policy published by the university. We are committed to upholding the principles of academic honesty and integrity in all aspects of our educational journey. We understand the importance of acknowledging the work and ideas of others, and we pledge to ensure that all our academic endeavors are conducted with the utmost originality and compliance with the university's policy.

We affirm that the content presented in this document is entirely our own, and any external sources used have been appropriately cited and referenced.

Michael Breau

As I submit my work, I, **Michael Breau**, take full responsibility for the integrity of my work and promise to avoid any form of plagiarism, cheating, or dishonest behavior. This acknowledgment serves as a testament to my dedication to academic excellence and the fostering of a trustworthy academic community at McMaster University.

Jessica Dawson

As I submit my work, I, **Jessica Dawson**, take full responsibility for the integrity of my work and promise to avoid any form of plagiarism, cheating, or dishonest behavior. This acknowledgment serves as a testament to my dedication to academic excellence and the fostering of a trustworthy academic community at McMaster University.

Benjamin Chinnery

As I submit my work, I, **Benjamin Chinnery**, take full responsibility for the integrity of my work and promise to avoid any form of plagiarism, cheating, or dishonest behavior. This acknowledgment serves as a testament to my dedication to academic excellence and the fostering of a trustworthy academic community at McMaster University.

Yaruo Tian

As I submit my work, I, **Yaruo Tian**, take full responsibility for the integrity of my work and promise to avoid any form of plagiarism, cheating, or dishonest behavior. This acknowledgment serves as a testament to my dedication to academic excellence and the fostering of a trustworthy academic community at McMaster University.

Matthew Curtis

As I submit my work, I, **Matthew Curtis**, take full responsibility for the integrity of my work and promise to avoid any form of plagiarism, cheating, or dishonest behavior. This acknowledgment serves as a testament to my dedication to academic excellence and the fostering of a trustworthy academic community at McMaster University.

Control Information

Version	Delivery		Feedback	
	Deadline	*Delivered*	*Received*	*Integrated*
V1	October 6, 2023	October 6, 2023		
V2				
V3				

(G) Goals

Control Information

Table 1. Natural Language Processing for Mental Health Risk Prediction — Versionning Information — Goal Book

Section	Version	Lead	Delivered	Reviewer	Approved
G.1	1.0	Matthew Curtis	2023-10-06	Jessica Dawson	October 6, 2023
G.2	1.0	Matthew Curtis	2023-10-06	Jessica Dawson	October 6, 2023
G.3	1.0	Michael Breau	October 2, 2023	Jessica Dawson	October 6, 2023
G.4	1.0	Michael Breau	October 2, 2023	Jessica Dawson	October 6, 2023
G.5	1.0	Jessica Dawson	October 6, 2023	Michael Breau	October 6, 2023
G.6	1.0	Jessica Dawson	October 6, 2023	Michael Breau	October 6, 2023
G.7	1.0	Matthew Curtis, Michael Breau	2023-10-06	Jessica Dawson	October 6, 2023
G.4	1.1	Michael Breau	November 2, 2023	Jessica Dawson	November 2, 2023
G.1	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
G.2	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
G.3	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
G.4	1.2	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
G.6	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024

(G.1) Context and Overall Objectives

Many people struggle with mental health issues that go unnoticed, and undiagnosed as well, and are isolated so they never get the diagnosis and the support they need. In response to those two points working in tandem, Dr. Mosser's past teams, along with support from UQAM and several psychologists from around the world, have developed methods and approaches to assist psychiatrists with diagnosing mental health issues.

As society advances so does the integration of technology into our lives. This social media use represents a body of information that could be used to more efficiently diagnose patients and improve psychologist workloads, in turn helping lessen barriers around seeking treatment.

Utilizing this information effectively requires systems that can process a large amount of data, much of it irrelevant to issue's of mental health, and provide predictions for specific disorders or direct a psychologist towards certain online interactions that may be relevant. This implies the usage of machine learning techniques, and specifically natural language processing techniques, to analyze all this data quickly. The objective of this project is to explore the applications of machine learning to this area of mental health identification.

In order to do this the project will need labelled data to develop models on. This is not easy to find, attempting to collect this ourselves would involve many consent and privacy concerns and be both difficult and time consuming. In order to have easy access to already developed data the team will be joining the "early risk prediction on the internet" competition, or eRisk for short. ERisk is a competition that presents a number of tasks relating to predicting mental health indicators from individuals' online post history with the goal of encouraging research into this area. The competition provides labelled data that models can be trained on as well as a structured environment through which the system developed for this project can be compared to other entries. This provides a golden opportunity for the team to meet their objectives and as such this project will largely be structured around this competition.

Our goal is to help eRisk expand the number of use cases that their NLP systems are applicable for and cover a wider range of scenarios and personal profiles. They have a strong base of NLP systems and data to work with, but the greater the breadth of the research the more people it can help. Early intervention is very important for mental health issues, and for issues such as suicidal tendencies and online predatory tendencies, it can be a matter of life and death. That is why this work is so important, and we will be working on expanding the number of situations in which it will be able to help people.

(G.2) Current situation

Currently, it can be a very difficult task to diagnose people with mental health conditions. If you are suffering from a mental health condition, it can often be hard to even originally reach out for help, and it is easy to feel isolated and alone. Due to this, many people never reach out to anyone and suffer alone especially in this post pandemic world where people still often feel very isolated as a hangover effect of the pandemic isolation. That is why it is so important to be able to identify the people who need help and offer them said help.

Currently there is no good system that will help you identify potential early warning signs of mental health issues based on a person’s online posting history. eRisk has been working on developing an interdisciplinary research field that aims to be applicable to a wide range of situations and to many different personal profiles. These can be used to help identify potential online predators, people with suicidal inclinations, or people that are susceptible to depression.

eRisk has been running since 2017 and has grown from having just one proposed task to now having three each year. Each year the competition releases labelled training data for each task in late December, the teams then have until April to develop their models before they run them on unlabelled test data eRisk releases. The results the models produce on this test data are submitted to the competition for evaluation and results are released in May. Each task is built around predicting mental health indicators from online posts but each explores a different facet of this process. The tasks for this year are as follows:

- **Task 1: Search for Symptoms of Depression** This task explores ranking sentences by relevance to a mental health issue so the most important information from a large body of data can be identified.

- **Task 2: Early Detection of Signs of Anorexia** This task explores detecting mental health issues quickly. It simulates monitoring an individual's posting activity in real time and encourages participants to come to an accurate conclusion in as few posts as possible.
- **Task 3: Measuring the Severity of the Signs of Eating Disorders** This task explores directly predicting the severity of different mental health symptoms from an individual's posting activity.

All three of these tasks have been run in two prior years of eRisk already. The first years for each task no labelled training data was provided; the second years the test data from the first year, now with labels, was provided as training data; and the third year, this year, the test data from the first and second years was labelled and provided as training data.

(G.3) Expected Benefits

This project aims to benefit this area of research by contributing to the literature surrounding the eRisk competition and machine learning techniques for detection of mental health issues. The team intends to build a rudimentary system, unlikely to be fit for actual clinical use, that helps advance this area of research and serve as a proof of concept for future endeavours.

The implementation of a natural language processing model that can detect early signs of mental health issues will have multiple benefits to its stakeholders. By participating in eRisk, the system aims to benefit eRisk and its participants by providing any information that could possibly provide useful for furthering this research topic. The system intends to benefit individuals who are affected by mental health issues by detecting early signs of these issues before they lead to a point that is irrecoverable for the individual.

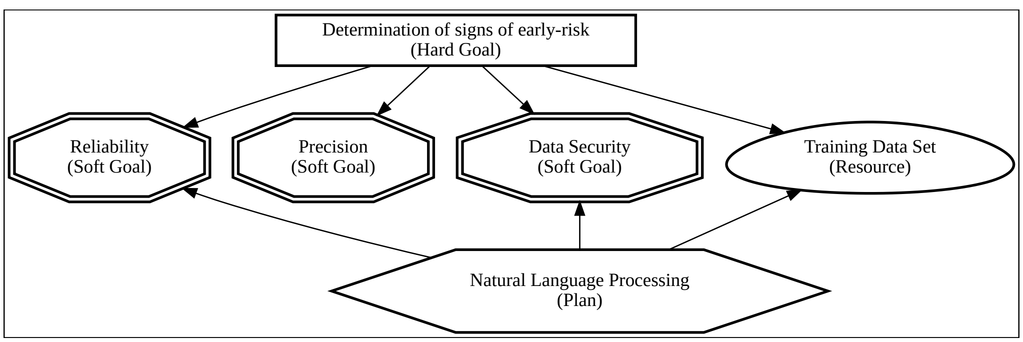

Figure 1. Goal Model Diagram for Natural Language Processing for Mental Health Risk Prediction

(G.4) Functionality overview

The requirements listed represent the core functionality of the Natural Language Processing system used to detect early signs of mental health issues. Detailed descriptions are available in the system section of this document.

Functional Requirements

The functionality of the system is built around each of eRisk's three tasks for the year:

- **Detecting Signs of Depression:** The system will be able to rank a set of written sentences based on how related they are to the 21 symptoms of depression laid out in the Beck Depression Inventory (the BDI) [3]. The system should produce a separate ranking for each of the 21 symptoms.
- **Early Detection of Anorexia:** The system will be able to predict whether an individual shows signs of anorexia based on their online posting history. The system will do this post by post, like it is monitoring an individuals online activity in real-time, and will attempt to make the prediction as quickly as possible.
- **Detecting Signs of Eating Disorders**: The system will be able to fill 22 of the 28 questions on the Eating Disorder Examination Questionnaire (the EDE-Q, for an overview see: https://www.corc.uk.net/outcome-experience-measures/eating-disorder-examination-questionnaire-ede-q/) [4][5] for an individual based on their online posting history. The six questions that were left out of the questionnaire, questions 13 to 18, do not follow the same answer format as the rest of the questions, a 0 to 6 rating of severity, and ask for an arbitrary number instead. These questions were likely omitted to make the task more focused.

Non-Functional Requirements

- **Security of Data**: The system should maintain privacy standards and not leak confidential information (see section E.3 "Data Privacy Regulations" for more information).

(G.5) High-level usage scenarios

Use Case 1: Search for symptoms of depression

Goal:	Rank sentences based on symptoms of depression
Actor:	Developer (Franklin, Matthew)
Precondition:	The developers have access to data
Main Pattern:	• The developers submits the raw data to the system • The system returns a sentence ranking

Use Case 2: Early detection of signs of anorexia

Goal:	Detect anorexia from a person’s posts
Actor:	Developer (Michael, Ben)
Precondition:	The developers have access to data
Main Pattern:	• The developers submits the raw data to the system • The system returns an anorexia prediction

Use Case 3: Measuring the severity of the signs of eating disorders

Goal:	To predict a person's eating disorder severity
Actor:	Developer (Jessica)
Precondition:	The developers have access to data
Main Pattern:	• The developers submits the raw data to the system • The system returns a list of predictions for an eating disorder questionnaire

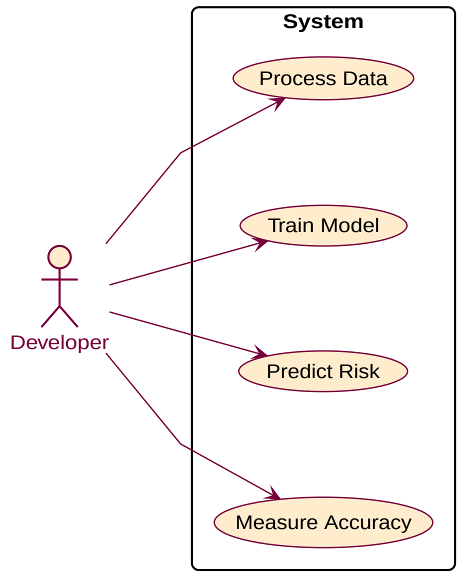

Figure 2. High Level use cases diagram

(G.6) Limitations and Exclusions

- The system need not address the acquisition of an individual's post history from external sites. The focus of the system is on developing and optimizing prediction models, implementing functionality to connect to a site and retrieve an individuals post history is outside the scope of this project and left for future development.
- The system need not provide an interface that psychologists can use in their everyday practice. Like above, the focus of the system is on developing prediction models, implementing these models into software that psychologists with limited technical knowledge can use is outside the scope of this project.

(G.7) Stakeholders and requirements sources

Table 2. Stakeholders

Stakeholder		Persona	Category
Direct Users	Psychiatrists	[Jessica]	Direct
	Police Officers	[Joe]	Direct
	Family Doctors	[Julia]	Direct
eRisk Competition Organizers		N/A	Direct
University of Montreal NLP Team		N/A	Direct
Mental Health Capstone Ops Team		N/A	Direct
Local Community			Indirect

Direct Stakeholders

Direct stakeholders have an interest in the implementation and can be used to elicit requirements. The groups that will be using the eRisk data and NLP architecture directly have been provided with personas to aid with requirement elicitation. They will be the primary users and will use the architecture in order to screen individuals to hopefully discover potential early signs of different mental health disorders. Although each group may use the NLP architecture to uncover different mental health issues and their motives will be different from one another.

Potential Stakeholders:

Jessica (Psychiatrist Persona):

Jessica is a 45-year-old Psychiatrist that works at her private practice. Her patient is a 14-year-old girl named Avery, and her mother has come to Jessica saying she is worried about Avery's mental health and thinks she may be beginning to suffer from depression and suicidal tendencies. When Avery comes into the doctor's office, she agrees to speak with Jessica about the topic but is very shy and reserved. Understandably so since the topic can be a hard and sensitive topic to approach. She ends up not really giving Jessica any information about how she is feeling and Jessica is left with no way of understanding if she is at risk or not. In moments like this Jessica loves to use eRisks NLP architecture to look at Avery social media accounts and determine if the model thinks she would be identified as someone that may be at risk for these issues. Without this model, Jessica is unsure of other ways to get more information about Avery's situation, but the NLP model can provide some useful insight for Jessica. As a result, she can bring up the topic of Avery's social media to her and create an open discussion about the feelings she may be having.

Joe (Police Officer Persona):

Joe is a 40-year-old police officer working in Hamilton, Ontario. He is on duty one night and gets a call from a distressed teenager. The teenage girl says that her friend's father has been acting very strange around her and is being creepy. This is all the information Joe is provided with and is stuck on what his next steps should be. Thankfully, he has the NLP architecture model from eRisk as a tool he uses when he gets calls like these or others about potential signs of mental health problems people are exhibiting. This is a key tool to see potential signs of early risk of mental health issues before drastic actions are taken and also helps find potential dangerous situations due to mental health issues exhibited online. When he runs the NLP model for this father, it comes back with a high risk of predatory behavior for him. This helps solidify to Joe that this is not just a one-off case and needs to be investigated swiftly before anything bad happens.

Julia (Family Doctor Persona):

Julia is a 35-year-old family doctor at her local doctors office. She knows the importance of mental health and how it can be hard for people to come forward and seek the help they need whether it is due to being shy, embarrassed, or any other number of reasons. As a result, she runs the eRisk NLP model on her patients the day before they come in. By doing this, she can find if there are any potential early signs of mental health issues and, if so, she can propose starting an open dialogue with her patient when they meet or ask her patient if they have been experiencing any feelings recently that would be related to said mental health issue. A lot of the time people just need an opportunity to speak about what they are going through, and this is a good way to get the dialogue started.

eRisk Competition Organizers:

The eRisk organizers will compare our systems performance in a variety of metrics in a competition format against other teams in order to further the research in the early risk prediction on the internet field.

University of Montreal NLP Team:

As previous competitors in the competion as well as mentors to the current project, the UoM NLP team has a direct interest in the implementation of the project.

Mental Health Capstone supervisors to the current project:

The projects supervisor's have a direct interest in the implementation as well as the outcome at the competition from the project.

Indirect Stakeholders

Local Community:

In the event that the project leads to the direct users, a change of an environment due to sucessful detection of early risk signs could provide a benefit to the local community.

(E) Environment

Control Information

Table 3. Natural Language Processing for Mental Health Risk Prediction — Versionning Information — Environment Book

Section	Version	Lead	Delivered	Reviewer	Approved
E.1	1.0	Benjamin Chinnery	October 2, 2023	Michael Breau	October 2, 2023
E.2	1.0	Yaruo Tian	October 2, 2023	Michael Breau	October 2, 2023
E.3	1.0	Yaruo Tian	October 2, 2023	Michael Breau	October 2, 2023
E.4	1.0	Yaruo Tian	October 2, 2023	Michael Breau	October 2, 2023
E.5	1.0	Benjamin Chinnery	October 2, 2023	Michael Breau	October 2, 2023
E.6	1.0	Benjamin Chinnery	October 2, 2023	Michael Breau	October 2, 2023
E.3	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024

(E.1) Glossary

- **Natural Language Processing (NLP)**: A subfield of computer science concerned with giving computers the ability to understand and manipulate human language.
- **Neural Networks**: Sometimes referred to as Artificial Neural Networks, are machine learning systems modeled on the neural pathways of the human brain. This type of system endeavors to recognize underlying relationships in a set of data.
- **Mental Health**: Mental health encompasses emotional, psychological, and social well-being, influencing cognition, perception, and behavior
- **Training Models**: Data provided to machine learning systems to help teach it correct assumptions about a given topic
- **Early Risk Detection Error (ERDE)**: An Evaluation Metric that evaluates the number of writings processed by the algorithm before producing a positive prediction
- **eRisk**: eRisk organizes a competition every year to compete various Artificial Intelligences against each other to help identify mental health issues. We will be entering our NLP into their contest this year, and as a result are tuning our algorithm to their specifications.

(E.2) Components

- **Message Servers**: The system must utilize servers provided by ERisk for when we validate our system during the testing stage to interatively get user writings and send responses (This is the scope of the operations teams project).

- **Text Data and Profiles**: The data sources and profile information mainly come from various platforms such as Reddit. Using their APIs could provide real data (out of scope).
- **Data Visualization Tools** Incorporating data visualization and analysis APIs and libraries such as Matplotlib and Pandas will aid in testing phase to generate an more accurate result (out of scope).

(E.3) Constraints

- **Data Privacy Regulations**: The system must adhere to the regulations of the Personal Information Protection and Electronic Documents Act [8] of Canada which governs how personal information can be used by organizations in Canada, and the Personal Health Information Protection Act [9] of Ontario which governs how personal health information can be used by organizations in Ontario.
- **eRisk Privacy Agreement**: The system must conform to the terms laid out in the eRisk data usage agreement, which are as follows:
  1. The information may only be used for research purposes. Portions of the data maybe copyrighted, and may also have commercial value as data, so you must be careful to use it only for research purposes.
  2. Summaries, analyses and interpretations of the linguistic properties of the information may be derived and published, provided it is not possible to reconstruct the information from these summaries.
  3. You may not try to identify the individuals whose texts are included into this dataset. You may not cross-reference individuals with the dataset against any other dataset or collection of data. You may not try to establish any kind of contact with the individuals of this dataset.
  4. You are not permitted to publish any portion of the dataset (e.g. example post) other than summary statistics, or share it with anyone else.
  5. We grant you the right to access the collection's content in the manner described in this agreement. You may not otherwise make unauthorized commercial use of, reproduce, prepare derivative works, distribute copies, perform, or publicly display the collection or parts of it.
  6. You may present research findings concerning knowledge obtained using the collection provided that the aforementioned presentation is within the limits of this agreement. Any scientific publication derived from the use of this collection should explicitly refer to: D. Losada, F. Crestani. A Test Collection for Research on Depression and Language Use. Conference and Labs of the Evaluation Forum (CLEF), Évora, Portugal, 2016.
  7. You shall not use results obtained through the use of the collection for profitable purposes including advertisement and/or defamatory or slanderous purposes.
  8. If we or the copyright holders request you to discontinue the use of the collection, or your use of the collection is deemed to be in violation of this agreement, you shall immediately discontinue use of the collection and promptly delete the collection and all data obtained by processing it from any computer or media onto which it has been copied.
- **eRisk Specification** The system must follow ERisk's output specification in order to be able to submit to the competition.

(E.4) Assumptions

- **Availability Servers**: It is assumed that message testing and deployment servers are active and available at all times during the testing and deployment phase.
- **Accuracy of User Profiles**: The assumption is made that all profile data provided for testing are formatted correctly and will be able to be processed without any errors, disruptions or further development procedures beyond the scope.
- **Connection Stability**: It is assumed that the connection between the system and message servers is smooth and stable, without connectivity disruptions from internet service providers.

(E.5) Effects

- **Workload of Mental Health Specialists**: The system could help ease up on the workload of Mental Health Professionals by being an additional tool they could use which could help provide support with diagnosis.
- **Barriers to Accessing Mental Health Services**: The online and virtual nature of the system was designed to help ease the barriers to support that people in remote communities often face.
- **Improved Detection**: The system may result in better detection of symptoms at varying stages by altering mental health professionals of signs that they might have missed.
- **Improved Treatment Plans**: Improved detection may result in mental health professionals being able to create more efficient treatment plans.
- **Cost of Mental Health Services**: The assistance and accessibility of these tools may reduce the overall cost of certain Mental Health Diagnostic Services.

(E.6) Invariants

- **User Data Privacy**: The system must maintain the privacy and anonymity of the individuals involved in either the training set or in any writings it evaluates.
- **Inputted Data Privacy**: The system must maintain privacy for any evaluations done for a user, by only sharing that information to the appropriate mental health assessor of the user.
- **Authority of Mental Health Professionals**: The system must maintain the current status quo for Mental Health Professionals being the only ones qualified to give a definitive diagnosis. This system is meant to aid the process, it is not meant to be the sole authority.
- **Ethical Guideline Accountability**: The system must act in accordance with the proper ethical standards and practises provided to us from the correspondent mental health associates.
- **Data Retention Guidelines**: The system must act in accordance with the nesscessary data retention guidelines of the appropriate environment, only saving and storing what is allowed.

(S) System

Control Information

Table 4. Natural Language Processing for Mental Health Risk Prediction — Versionning Information — System Book

Section	Version	Lead	Delivered	Reviewer	Approved
S.1	1.0	Michael Breau	October 5, 2023	Jessica Dawson	October 6, 2023
S.2	1.0	Michael Breau	October 5, 2023	Jessica Dawson	October 6, 2023
S.3	1.0	Jessica Dawson	October 6, 2023	Michael Breau	October 6, 2023
S.4	1.0	Jessica Dawson	October 6, 2023	Michael Breau	October 6, 2023
S.5	1.0	Michael Breau	October 5, 2023	Jessica Dawson	October 6, 2023
S.6	1.0	Jessica Dawson	October 6, 2023	Michael Breau	October 6, 2023
S.2	1.1	Michael Breau	November 2, 2023	Jessica Dawson	November 2, 2023
S.2	1.2	Jessica Dawson	March 6, 2024	Michael Breau	March 6, 2024
S.1	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
S.2	1.3	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
S.3	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
S.5	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
S.6	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024

(S.1) Components

- **Depression Component**: Handles task 1 of the eRisk competition. Ranks sentences based on their relevancy to the 21 depression symptoms of the Beck Depression Index [3], takes in a database of sentences and returns a ranking of the sentences for each symptom.
- **Anorexia Component**: Handles task 2 of the eRisk competition. Takes in an individuals posting history one post at a time and returns a warning when it is confident it has identified signs of anorexia.
- **Eating Disorders Component**: Handles task 3 of the eRisk competition. Fills out a copy of the EDE-Q [4][5] for an individual based on their posting history, takes in an individuals posting history and returns a filled questionnaire.

(S.2) Functionality

This section describes in further detail the components listed in the previous section by describing their functional and non-functional requirements.

General and Security

Non-Functional Requirements:

- GR1: The system should be compatible on MacOS, Windows, and Linux.
- GR2: Copyrighted content should not be included in the output without reference.
- SR1: The system will be tested periodically to detect a potential lack of processing power.
- SR2: Sensitive user data must not be present within the results generated.

Each task will have their own accuracy metrics that the quality of the models predictions will be evaluated on.

Task 1: Search for Symptoms of Depression Component

Functional Requirements:

- T1FR-1: The system will be able to parse both xml and jsonl files.
- T1FR-2: The system will rank input sentences from users based on their correlation to 21 depression symptoms from the beck depression index.
- T1FR-3: The system will send the predictions to a txt file in the format "{symptom_number}, Q0, {sentence-id}, {position_in_ranking}, {score}, {system_name}".

Task 2: Early Detection of Signs of Anorexia Component

Functional Requirements:

- T2FR-1: The system will be able to parse both xml and jsonl files.
- T2FR-2: The system will make predictions given a set of a user's posts.
- T2FR-3: The system will allow posts to be added to a user's set of posts and the system will update the prediction for that user.
- T2FR-4: The system will send the predictions to a txt file in the format "{username} {prediction (0 or 1)}" with each entry being on a separate line.

Task 3: Measuring the Severity of the Signs of Eating Disorders Component

General Functional Requirements:

- T3FR-1: The system will be able to parse jsonl files of an individual's post history.
- T3FR-2: The system will make predictions given a set of a user's posts.
- T3FR-3: The system will send the predictions to a txt file in the format "{username} {prediction (0-6) for Q1} ... {prediction (0-6) for Q28}" with each individual being on a separate line.

Accuracy Functional Requirements:

These functional requirements cover how the accuracy of the model will be evaluated. These requirements involve evaluating the system's performance against a set of three baselines: guess zero for every question, guess six for every question, and guess the average for every question. If the system fails to outperform these baselines it does not represent an improvement over guessing and will be considered a failure.

Performance will be evaluated on a number of metrics, the following is taken from [7] and discusses the metrics that will be used to evaluate accuracy for task 3:

Evaluation is based on the following effectiveness metrics:

- Mean Zero-One Error (*MZOE*) between the questionnaire filled by the real user and the questionnaire filled by the system (i.e. fraction of incorrect predictions).

$$MZOE(f,Q) = \frac{|\{q_i \in Q : R(q_i) \neq f(q_i)\}|}{|Q|}$$

  where f denotes the classification done by an automatic system, Q is the set of questions of each questionnaire, q_i is the i-th question, $R(q_i)$ is the real user's answer for the i-th question and $f(q_i)$ is the predicted answer of the system for the i-th question. Each user produces a single *MZOE* score and the reported *MZOE* is the average over all *MZOE* values (mean *MZOE* over all users).

- Mean Absolute Error (*MAE*) between the questionnaire filled by the real user and the questionnaire filled by the system (i.e. average deviation of the predicted response from the true response).

$$MAE(f,Q) = \frac{\Sigma_{q_i \in Q}|R(q_i) - f(q_i)|}{|Q|}$$

  Again, each user produces a single *MAE* score and the reported *MAE* is the average over all *MAE* values (mean *MAE* over all users).

The following measures are based on aggregated scores obtained from the EDE-Q questionnaires.

- Restraint Subscale (RS): Given a questionnaire, its restraint score is obtained as the mean response to the first five questions. This measure computes the RMSE between the restraint ED score obtained from the questionnaire filled by the real user and the restraint ED score obtained from the questionnaire filled by the system. Each user u_i is associated with a real subscale ED score (referred to as $R_{RS}(u_i)$) and an estimated subscale ED score (referred to as $f_{RS}(u_i)$). This metric computes the RMSE between the real and an estimated subscale ED scores as follows:

$$RMSE(f,U) = \sqrt{\frac{\Sigma_{u_i \in U}(R_{RS}(u_i) - f_{RS}(u_i))^2}{|U|}}$$

where *U* is the user set.

- Eating Concern Subscale (ECS): Given a questionnaire, its eating concern score is obtained as the mean response to the following questions (7, 9, 19, 21, 20). This metric computes the RMSE between the eating concern ED score obtained from the questionnaire filled by the real user and the eating concern ED score obtained from the questionnaire filled by the system.

$$RMSE(f,U) = \sqrt{\frac{\Sigma_{u_i \in U}(R_{ECS}(u_i) - f_{ECS}(u_i))^2}{|U|}}$$

- Shape Concern Subscale (SCS): Given a questionnaire, its shape concern score is obtained as the mean response to the following questions (6, 8, 23, 10, 26, 27, 28, 11). This metric computes the RMSE between the shape concern ED score obtained from the questionnaire filled by the real user and the shape concern ED score obtained from the questionnaire filled by the system.

$$RMSE(f,U) = \sqrt{\frac{\Sigma_{u_i \in U}(R_{SCS}(u_i) - f_{SCS}(u_i))^2}{|U|}}$$

- Weight Concern Subscale (WCS): Given a questionnaire, its weight concern score is obtained as the mean response to the following questions (22, 24, 8, 25, 12). This metric computes the RMSE between the weight concern ED score obtained from the questionnaire filled by the real user and the weight concern ED score obtained from the questionnaire filled by the system.

$$RMSE(f,U) = \sqrt{\frac{\Sigma_{u_i \in U}(R_{WCS}(u_i) - f_{WCS}(u_i))^2}{|U|}}$$

- Global ED (GED): To obtain an overall or 'global' score, the four subscales scores are summed and the resulting total divided by the number of subscales (i.e. four). This metric computes the RMSE between the real and an estimated global ED scores as follows:

$$RMSE(f,U) = \sqrt{\frac{\Sigma_{u_i \in U}(R_{GED}(u_i) - f_{GED}(u_i))^2}{|U|}}$$

MAE is the most important metric out of all of these as it gives the clearest picture of how well the system is performing overall. MZOE is less valuable as it only shows how many answers the system got exactly right, having a good MZOE but a bad MAE isn't a good result as getting a solid amount of answers exactly right but all the others very wrong is not desirable behaviour. The measures based on aggregated scores are useful for seeing what parts of the questionnaire the system is good at predicting for, but again, if any of these metrics are good while MAE is poor this does not represent desirable behaviour. The system should outperform the three baselines on MAE or at least tie them on MAE and outperform on another metric.

This leads to a set of three accuracy requirements:

- T3FR-4: The model will outperform guessing all zeros on MAE OR tie MAE and outperform on at least one of the other metrics.
- T3FR-5: The model will outperform guessing all sixes on MAE OR tie MAE and outperform on at least one of the other metrics.
- T3FR-6: The model will outperform guessing the average on MAE OR tie MAE and outperform on at least one of the other metrics.

(S.3) Interfaces

All three components of the system will use a command line interface that allows the user to point to a file of writings and will then print the results of the respective model into a location specified by the same command line.

(S.4) Detailed usage scenarios

Use Case 1: Search for symptoms of depression

1. User provides a series of sentences.
2. Model ranks sentences based on the 21 symptoms of the beck depression index questionaire.

Use Case 2: Early detection of signs of anorexia

1. User provides an series of posts belonging to a single user.
2. Model provides a 0 if the user is predicted to not have anorexia and a 1 if the user is predicted to have anorexia.

Use Case 3: Measuring the severity of the signs of eating disorders

1. User provides an series of posts belonging to a single user.
2. Model provides a 0-6 score predicting the user's answers to the eating disorder examination questionnaire.

(S.5) Prioritization

Depression Component	must-have	should-have	could-have	won't-have

The system will divide text into individual units of words and or sub-words (Tokenization).	x			
The system will convert generated tokens into lowercase to preserve consistency.		x		
The system will remove stopword tokens (Common words that do not commonly have an effect on the meaning).		x		
The system will reduce tokens to their root forms (Ex: "moving" to "move").		x		
The system will produce an output that is useable by the Vectorization Component.	x			
Anorexia Component	must-have	should-have	could-have	won’t-have
The system will convert tokens into numerical vectors.	x			
The system will produce an output that is useable by the Prediction Component.	x			
Eating Disorders Component	must-have	should-have	could-have	won’t-have

The system will be able to parse jsonl files of an individual’s post history.	**x**			
The system will make predictions given a set of a user’s posts.	**x**			
The system will send the predictions to a txt file in the format "{username} {prediction (0-6) for Q1} … {prediction (0-6) for Q28}" with each individual being on a separate line.	**x**			
The model will outperform guessing all zeros on MAE OR tie MAE and outperform on at least one of the other metrics.	**x**			
The model will outperform guessing all sixes on MAE OR tie MAE and outperform on at least one of the other metrics.	**x**			
The model will outperform guessing the average on MAE OR tie MAE and outperform on at least one of the other metrics.	**x**			

(S.6) Verification and acceptance criteria

- The system should meet the accuracy requirements for each component laid out in S.2. This will be determined by taking the labelled data from eRisk, splitting it into a train set and a test set, training the prediction models on the train set, and then predicting answers for the test set. These predictions will be evaluated against the known values of the test set on the accuracy measures laid out in S.2 and if they outperform the baselines described in S.2 the implementation will be deemed satisfactory.
- The system should perform in the top 50% of systems submitted to eRisk this year for each of the tasks. Given these results will be released by eRisk sometime after the capstone course wraps up the result of this criteria will not be included in the final capstone documentation.
- The predictions models used in the system should be generalizable beyond the eRisk data we are developing on. This is a difficult and vague thing to prove, that the system's models don't just perform well on the data they are developed on but that they will perform well on any instance of the problem the model is built to solve. For the purposes of the project this criteria will be considered met if a reasonable justification of generalizablity is made for each model, including an analysis of any factors that negatively impact generalizability.

(P) Project

Control Information

Table 5. Natural Language Processing for Mental Health Risk Prediction — Versionning Information — Project Book

Section	Version	Lead	Delivered	Reviewer	Approved
P.1	1.0	Matthew Curtis	2023-10-06	Jessica Dawson	October 6, 2023
P.2	1.0	Matthew Curtis	2023-10-06	Jessica Dawson	October 6, 2023
P.3	1.0	Yaruo Tian	October 5, 2023	Michael Breau	October 5, 2023
P.4	1.0	Yaruo Tian	October 5, 2023	Michael Breau	October 5, 2023
P.5	1.0	Yaruo Tian	October 5, 2023	Michael Breau	October 5, 2023
P.6	1.0	Benjamin Chinnery	October 4, 2023	Michael Breau	October 5, 2023
P.7	1.0	Benjamin Chinnery	October 4, 2023	Michael Breau	October 5, 2023
P.2	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024
P.5	1.1	Jessica Dawson	April 3, 2024	Michael Breau	April 3, 2024

(P.1) Roles and personnel

- **Developers**: Matthew Curtis, Jessica Dawson, Michael Breau, Benjamin Chinnery, Yaruo Tian

We will be working with professors and many different other teams throughout our project with our project as one specific task (Developing a NLP architecture). We are unsure of the exact specifics for our projects as of right now due to the fact eRisk has not released the topics that we will be choosing from for the 2024 competition we will be taking part in. As a result, each of our main roles at the moment are just general developers.

- **GitHub Issue Tracker**: Michael Breau

Michael will be responsible for organizing and managing all of our github issues. This includes creating issues, creating the respective tags and categories and also making sure they are closed when they are finished.

- **Team Liaison**: Benjamin Chinnery

Benjamin will be our team liaison and as a result he will be the primary person reasonable for organizing our meeting as a group with all of our input put into consideration. He will also be our primary contact between our team and Professor Mosser along with the other teams involved as well. We all will of course have access to talk with Professor Mosser and the other teams and will do so. Benjamin will just act as our primary point of contact.

(P.2) Imposed technical choices

From our research and the advice of our supervision team it was determined that the majority of libraries and common tools for natural language processing are written for python. As such the system will be written in python to maximize the number of available techniques that can be explored. This will also allow for easy interoperability with the ops team who have made the choice to use python as well.

(P.3) Schedule and milestones

- Task 1: Learn required knowledge for project development such as Natural Processing Techniques and its corresponding libraries.
- Task 2: Initial setup and install required tools.
- Task 3: Use the Anorexia Model from past year as reference for initial prototype.
- Task 4: Develop version 1 (functional prototype) for proof of concept.
- Task 5: Split into 3 teams and focus on the three different models for the three different types of early risk detection.
- Task 6: Analyze results of the three tasks and focus on potential improvements.
- Task 7: Develop version 2 (final product) and test for accuracy.

(P.4) Tasks and deliverables

Iteration #1: Weeks 1-4

- Task 1: Learn required knowledge for project development such as Natural Processing Techniques and its corresponding libraries.
- Task 2: Initial setup and install required tools.

Iteration #2: Weeks 5-8

- Task 3: Use the Pathological Gambling Project as reference for initial prototype.
- Task 4: Develop version 1 (functional prototype) for proof of concept.

Iteration #3: Weeks 9-12

- Task 5: Split into 3 teams and focus on the three different models for the three different types of early risk detection.
- Task 6: Analyze results of the three tasks and focus on potential improvements.

Iteration #4: Weeks 13-16

- Task 7: Develop version 2 (final product) and test for accuracy.

(P.5) Required technology elements

- The project will likely require the use of standard data manipulation libraries like numpy and pandas to manage and organize data.
- The project will require the standard python libraries for parsing json and xml as the data from eRisk that the systems' models will be developed on comes in a mixture of these formats.

(P.6) Risk and mitigation analysis

NLP Learning Curve

When the original description for the project was released to our group, it was clearly stated that this project would not require any prior experience with AI work, and would instead be a gradual learning process with a ramp up period. To help mitigate the risk of team members falling behind schedule, regular progress status check-ins in our team meetings could help mitigate difficulty. Solutions to help out teammates could come from other members of the group sharing their understanding, or which resources they used if they ran into similar problems. If team member support isn't sufficient, it is recommended for team members to reach out to project facilitators such as Dr. Mosser if further clarification is required.

Difficulty regarding eRisk Specification

Due to the ambiguity of the project specification in the early states, team members may feel overwhelmed with attempting to stay ahead of the curve in terms of preparing elements of the project before details regarding this year's competition are revealed. Team members are encouraged to keep up to date on all released eRisk specifications and familiarize oneself with the provided past documentation and specification from the prior teams. Additionally, risk mitigation can be attained by team members practising the exercises provided by Dr. Mosser and Dr. Meurs.

Difficulty Balancing Project Deliverables and Competition Deliverables

Due to this project having two independent timelines, one for the course deliverables and one for the contest deliverables, team members may experience difficulty balancing both responsibilities. The best way to manage this is to stick to clearly defined schedules and have fair divisions of work in order to ensure that all important project aspects are assigned to someone at all times.

(P.7) Requirements process and report

-eRisk Competition Organizers: Arrange internal team meetings and discussions to stay informed and up to date regarding any specifications and important deadlines for the eRisk Competition, in order to keep our project on track and to the right specifications. The competition organizers will release details such as what mental health issues will be focused on, timing and processing constraints, and general milestone deadlines. This information will be found on the website as well as communicated through their corresponding e-mail services.

-University of Montreal NLP Team: The team of researchers and experts working at the University of

Montreal offer invaluable insight towards creating the project, as well as being major stakeholders for our team, as they have a major say in quality assessment of our project. Our team will have regularly scheduled meetings with the core Natural Language Processing Team as well as with any other collaborative experts that can provide feedback and instruction for us.

-**Mental Health Capstone Ops Team**: As this project is not solely the responsibility of 1 Capstone team, it is important that we as the Natural Language Processing team work closely with the Software Architecture Operations team to ensure team synergy, which is vital due to the direct reliance that both teams have with each other. Our primary means of accomplishing this is through regular inter-team communications and meetings to discuss any challenges or highlight key aspects of implementation to further each others understanding, as well as make sure that both projects are on the right track.

-**Mental Health Professionals**: Due to the scope of this project being so closely intertwined with the work of mental health professionals, it is imperative that we arrange to receive insight from people in that field. The overall prior team associated with this project is a widely multidisciplinary effort, and luckily for us, that team includes some therapists and other mental health workers from around the world who can provide us insight and feedback on our progress. It is important to put a strong emphasis on their outlook, as they are most closely related to the codes and standards that this project should follow in order to not violate any ethical standards.

-**Users of Online Mental Health Services**: It is important to conduct some user interviews with people who have either used virtual mental health services in the past or are interested in using them. Getting user feedback and input can further our understanding to give us a better idea of the virtual mental health landscape.

References

- [1] Bertrand Meyer. *Handbook of Requirements and Business Analysis*. Springer. 2022.
- [2] Ian Sommerville and Peter Sawyer. *Requirements Engineering: A good Practice Guide*. Wiley. 1997.
- [3] Beck, A. T., Ward, C. H., Mendelson, M., Mock, J., & ERBAUGH, J. *An inventory for measuring depression*. Archives of general psychiatry. 1961. https://arc.psych.wisc.edu/self-report/beck-depression-inventory-bdi/
- [4] Fairburn CG, Beglin SJ. *Assessment of eating disorder psychopathology: Interview or self-report questionnaire? International Journal of Eating Disorders*. 1994. https://doi.org/10.1002/1098-108X(199412)16:4%3C363::AID-EAT2260160405%3E3.0.CO;2-%23
- [5] Fairburn CG, Beglin S. *Eating disorder examination questionnaire*. In: Fairburn CG, editor. *Cognitive behavior therapy and eating disorders*. New York, NY: Guilford Press; 2008.
- [6] Fairburn CG, Cooper Z. *The eating disorder examination*. In: Fairburn C, Wilson GT, editors. *Binge eating: nature assessment and treatment*. 12. New York, NY: Guilford Press; 1993.
- [7] Javier Parapar, Patricia Martín-Rodilla, David E. Losada, and Fabio Crestan. *Overview of eRisk at CLEF 2022: Early Risk Prediction on the Internet (Extended Overview)*. Springer. 2022. https://ceur-ws.org/Vol-3180/paper-66.pdf
- [8] *Personal Information Protection and Electronic Documents Act*. S.C. 2000, c. 5. https://laws-lois.justice.gc.ca/eng/acts/P-8.6/index.html
- [9] *Personal Health Information Protection Act*. S.O. 2004, c. 3. https://www.ontario.ca/laws/statute/04p03
- [10] D. Losada, F. Crestani. *A Test Collection for Research on Depression and Language Use. Conference and Labs of the Evaluation Forum (CLEF)*. Évora, Portugal, 2016.

Reflections

1. What knowledge and skills will the team collectively need to acquire to successfully complete this capstone project? Examples of possible knowledge to acquire include domain specific knowledge from the domain of your application, or software engineering knowledge, mechatronics knowledge or computer science knowledge. Skills may be related to technology, or writing, or presentation, or team management, etc. You should look to identify at least one item for each team member.

The knowledge and skills required include: NLP Architecture (Multiple different methods), File Parsing, NLP evaluation skills, Mental Health Knowledge, etc…

1. For each of the knowledge areas and skills identified in the previous question, what are at least two approaches to acquiring the knowledge or mastering the skill? Of the identified approaches, which will each team member pursue, and why did they make this choice?

   Matthew Curtis

   A knowledge area that I thought was imperative to acquire for success was a deeper understanding of vectors, how to manipulate them, and how they can be used in the NLP process. Task 1's main approach we were looking at involved treating each symptom as a vector and each sentence as a vector and comparing them but for that I thought we needed a deeper knowledge of vectors and their formation. For this the main two methods I used were watching youtube videos and reading articles on vectors and nlp vectorization. To understand the basics I read the article "Understanding NLP Word Embeddings - Text Vectorization" on towardsdatascience.com which was very helpful in gaining a basic technical knowledge about vectors and vectors in the NLP space. Then I wanted a different method of learning so I wanted some youtube videos first about vectors themselves, then about vectorization and how it is used in NLP. I also watched "Word Embedding - Natural Language Processing| Deep Learning" by Krish Naik on youtube which gave me a lot of insight into how the whole process comes together and how vectors are used for word embeddings. Lastly, I looked at the article "Vectorization Techniques in NLP [Guide]" by neptune.ai to see specific examples of how vectors can be used and manipulated for NLP purposes. This was very important for me to actually see the possible use cases.

   Jessica Dawson

   I will be working to learn what the current state of techniques are in NLP and the eRisk competition so I can better orient myself and my work. I will do this by reading past research papers and search for anything I do not understand through google.

   Michael Breau

   Since I will be working on the binary classification for anorexia I will be looking specifically into natural language processing methods that are more focused on classification. One way to gain this knowledge is from the BERTopic documentation page which is a python library used for topic modeling which can be used in a classification pipeline. The other method is from the sckit-learn documentation page which is another python library but used specifically for training and classifying documents which is directly related to the binary classification problem.

Benjamin Chinnery

A skill that I found necessary in regards to project management, is trying to learn more about the process required for two distinct teams operating in unison. For the final project to be efficient, it is imperative that both the NLP team and the OPS team have good synergy and understanding of each other's goals. Difficulties arise in the sense that although each group member has a level of control over their own group, it's important to also have some sort of central management that can guide the progress of the other team, without delving too far from your personal responsibilities. It is important to not extend oneself too far both because the other team would not want to take orders from someone who isn't as intimately familiar with their task, and because it is important to not overencomber oneself and lose sight of keeping their own project on track. To improve my inter-team project management skills, I could refer to past year guides for the competition and attempt to translate their process into something that's more applicable. Alternatively, while the prior approach holds some merit, I believe a more effective strategy would be to opt for more open communication with the teams through a liaison and use the existing capstone management guidelines as a basis.

Yaruo Tian

I needed a compounding of domain specific knowledge and technical skills. Firstly, acquiring general knowledge and having a general idea in mental health and wellness is crucial. Understanding the symptoms, risk factors associated with depression will provide the basis for designing an effective NLP system that will accurately predict depression symptoms. Additionally, knowledge of existing research and literature on depression prediction models and NLP techniques applied in healthcare contexts will be valuable as this will provide a platform to work towards and on top of. This can be achieved by finishing self study of academic papers, online courses, or consultations with experts in the field. Furthermore, gaining insights into right considerations and regulations surrounding mental health information will also be beneficial. On the technical side, understanding where and how to use natural language processing techniques and machine learning algorithms is crucial to the success of this capstone project. Acquiring skills in information preprocessing and developing models is essential to build accurate prediction models. Online courses could be found on YouTube and active projects in NLP and machine learning could aid these skills. Additionally, gaining experience with applicative programming languages and frameworks such as Python, Numpy, Scikit-Learn and TensorFlow will also streamline the development of the capstone project.