

**PROPOSING MENTAL-HEALTH INTERVENTION STRATEGIES
BASED ON CLIENT INTAKE FORM DATA (COGNIXR)**

St. Clair College, Center for the Arts



CAPSTONE REPORT

Submitted to the

MR. UMAIR DURRANI

(Professor, St. Clair College Center for the Arts)

SUBMITTED BY

Priyanka Ghetiya (0779258)

Ravinder Singh (0782184)

Digvijaysinh Madansinh Desai (0776134)

Parth Bhupendrabhai Sheth (0781634)

AUGUST 2022

INDEX

CHAPTER 1 : INTRODUCTION	4
CHAPTER 2 : DATA DESCRIPTION	5
CHAPTER 3 : METHODOLOGY	6
3.1 NLP (Natural Language Processing).....	6
3.2 Machine Learning Models... ..	7
CHAPTER 4 : RESULTS.....	8
CHAPTER 5 : CONCLUSION AND FUTURE WORK	9
ACKNOWLEDGEMENTS.....	10
REFERENCES... ..	11

TABLE OF FIGURE

Figure:1	Different NLP techniques which used in our project.....	6
Table:1	Training and Test Accuracy of Machine Learning Models.....	8

CHAPTER 1 : INTRODUCTION

To bridge the gap between a patient and a mental health professional, CogniXR has started a platform for those willing to take the help of a professional who can help cure them in the best manner possible.

Mental Health is one of the biggest issues in today's era. It has been said that *by age 40, about 50% of the population will have or have had a mental illness.* There is an exceptionally low ratio of patients to therapists in this field. So, to increase the

In this project we are aiming to achieve a methodology where we can reduce the time that takes place when a patient visits any mental health professional for the very first time. When any patient visits the therapist for the first time, the patient needs to fill in an intake form which consists of some of their personal information as well as some of the information regarding the life of a patient. When this intake form is being sent to the therapist, he analyses the information provided by the patient in the intake form and tries to find out the best possible treatment for that patient. We are trying to ease that process with the help of the knowledge we have gained during Data Analytics for Business from the year 2021.

The time taken by the professional to analyze the form and get the best treatment for the patient, this project may help a professional in terms of primary analysis and thinking procedure to predict the treatment strategies with the help of a machine learning model we create. The more accurately the model predicts, the better the therapy can be provided. This may save time and make the process of the therapy faster.

This will channelize the two procedures of the current treatment phases:

1. Analyzing the whole form, understanding the conditions of the patient, and suggesting the therapy.
2. Developing an Intake Form for the therapist according to the needs of an organization.

CHAPTER 2 : DATA DESCRIPTION

The dataset we are using in this project is being collected by us with the help of an intake form we have designed for the data collection. The data is all about the mental health of an individual survey to determine the mental health condition and suggest the mental health intervention strategies depending on the data filled by the user when they want to visit the mental health professional at the first time.

There are total of thirty-four questions which are open ended to be asked to the patient to get the primary information regarding the mental condition of them. So that will help the Professional who is going to treat them in terms of they know the primary information of patient and they can suggest better treatment to the patient.

Dataset consists of long text and short text data mostly as all the answers from the intake form are in the same format. So, it will become easy for us to get it analyzed and reach a conclusion with different techniques. The data storage will be done as per the regulations designed by Federal Canadian Government under *PIPEDA Act, 2000 (Personal Information Protection and Electronic Documents Act, 2000)*.

The questions we are targeting for our analysis are as follows:

- ✓ In your opinion, what is the cause of that stress and how does it affect you ?
- ✓ Can you briefly describe your main symptom(s)?
- ✓ Can you tell us about how you feel today?
- ✓ What makes you happy ?

The questions and answers do not consist of any more personal information, so we are in a good state to store this information for our project analysis purposes. We are taking care of whatever information is being provided to us by users in terms of their privacy.

The techniques which we are planning to use are Natural Language Processing, Text Analysis and Text Mining. Natural Language Processing will help us in analyzing the text data very well.

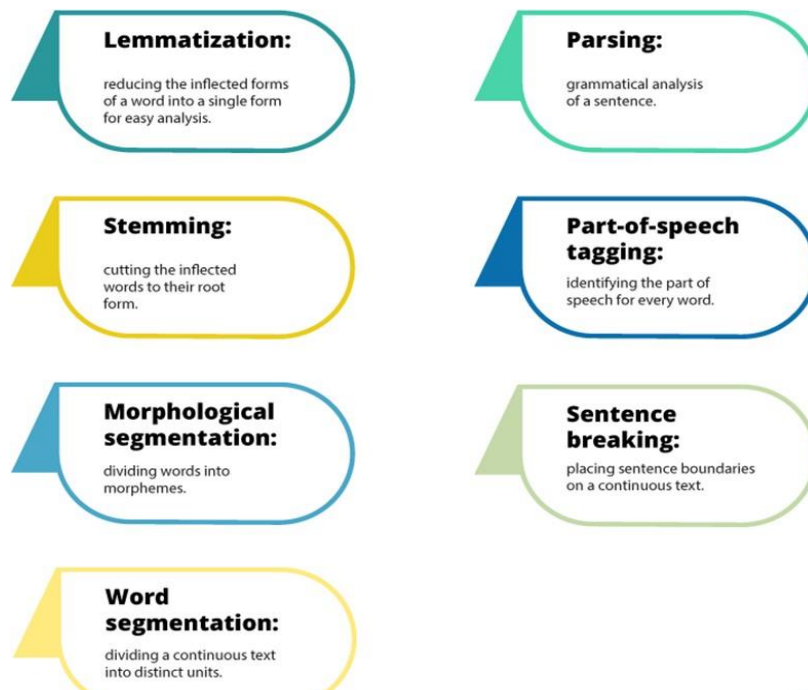
CHAPTER 3 : METHODOLOGY

We did not have any data for our project at first, and the firm had a basic intake form that required several modifications in accordance with our study. As a result, under the direction of our professor and the founder of the CogniXR firm, we created a Google intake form for patients. This form includes the thirty-five open-ended, text-based questions that we have discussed in total. We utilized NLP and text mining techniques because the majority of the data on this form is text data.

3.1 NLP (Natural Language Processing):

Creating computers that understand and react to text or voice data—and answer with text or speech of their own—much like humans do is the goal of natural language processing. NLP combines statistical, machine learning, and deep learning models with computational linguistics—rule-based modelling of human language. With the use of these technologies, computers are now able to interpret human language in the form of text or audio data and fully "understand" what is being said or written, including the speaker's or writer's intentions and mood.

We used many NLP techniques in our project which you can see in the given figure:



(Figure:1 Different NLP techniques which used in our project)

3.2 Machine Learning Models:

- **SVC and Linear SVC Classifier:** Linear SVC's (Support Vector Classifier) goal is to split or categorize the data you provide by returning a "best fit" support vector. You may then feed some characteristics to your classifier to get the "predicted" class after collecting the support vectors.
- **Random Forest Classifier:** Leo Breiman and Adele Cutler are the creators of the widely used machine learning technique known as random forest, which mixes the output of several decision trees to produce a single outcome. Its widespread use is motivated by its adaptability and usability since it can solve classification and regression issues.
- **Decision Tree Classifier:** A supervised learning approach called a decision tree may be applied to classification and regression issues. The method generates rules that may be represented by a tree structure using training data. Both categorical and numerical data may be used with decision trees. In contrast, some machine learning techniques need encoding to numeric values and are unable to oversee categorical input.

CHAPTER 4 : RESULT

Due to its numerous potentials uses in marketing, political science, psychology, human-computer interaction, artificial intelligence, etc., mental health detection in text data has recently gained increased popularity. A significant part of drawing attention to this topic was the availability of a vast volume of textual information, particularly opinionated and self-expressed writing. We have thoroughly examined several feature engineering, text mining, and modeling techniques. In the following table, we can demonstrate that the LinearSVC model provides the greatest accuracy when compared to other models based on the performance of all models. LinearSVC and the Random Forest model helped us reach an accuracy of 99%, while SVC and the Decision Tree model helped us achieve an accuracy of 90% and 92% percent, respectively.

Classifier/Models	Training Accuracy	Test Accuracy
SVC	90%	45%
Linear SVC	99%	57%
RandomForest Classifier	99%	55%
DecisionTree Classifier	92%	46%

(Table:1 Training and Test Accuracy of Machine Learning Models)

CHAPTER 5 : CONCLUSION AND FUTURE WORK

Since there are several machine learning approaches accessible, it is important to compare them all and then choose the one that best fits the target domain. Today, there are several specialized programs in the medical field that can forecast mental health disease quite precisely in advance, allowing for effective and quick therapy. One significant area of research in human-computer interaction is mental health prediction. Research on predicting mental health concerns from face and audio data has advanced sufficiently, whereas predicting mental health from patient's intake form data is still a new research area. In this study, text data were used to predict a mental health intervention method, and some relevant analysis or discussion were also conducted. We utilized a mental health survey form for training and testing purposes. It has thirty-five total textual open-ended questions concerning various mental health conditions. On our dataset, we assessed many models, and LinearSVC had the greatest accuracy. In the future, we will focus on usability, accuracy, and efficiency. In future we can develop an API which can directly get input from the intake form data and can directly predict the diseases. We can enhance this project with Audio and Video as well Speech to Text can also be added so that one does not need to fill out the whole form and we can get the voice modulated answers.

ACKNOWLEDGEMENTS

We would like to convey our deep gratitude to our supervisor, Mr. Umair Durrani, Professor of St. Clair College, for his unwavering support and guidance during the capstone project's implementation.

We would also like to thank **Ms. Emmanuella Michel**, founder of **Cognicorp + Therapie Inc.**, for giving us the opportunity to work in the organization. We are grateful to all the organization's mentors and personnel for their invaluable collaboration, advice, and direction throughout the internship time.

We also offer our best regards and gratitude to all our colleagues, fellow workers and others who have supported us in any manner during the duration of the project.

Priyanka Ghetiya (0779258)

Ravinder Singh (0782184)

Digvijaysinh Madansinh Desai (0776134)

Parth Bhupendrabhai Sheth (0781634)

REFERENCES

1. GitHub Repository Link:
https://github.com/0779258/DAB402_Group_4
2. Google Intake Form Link :
<https://docs.google.com/forms/d/e/1FAIpQLSfaNItiZufVISf9Oci61B5lul-C8x3zrl3PbEm1h3P01KnQQ/viewform>
3. <https://psychiatry.org/Patients-Families/What-Is-Mental-Illness#:~:text=Mental%20illnesses%20are%20health%20conditions,Nothing%20to%20be%20ashamed%20of>
4. <https://www.ukessays.com/essays/social-work/mental-health-policy-thesis-statement-social-work-essay.php>
5. <https://cmha.ca/brochure/fast-facts-about-mental-illness/>
6. Zhang, T. (2022, April 8). Natural language processing applied to mental illness detection: a narrative review. Nature. Retrieved May 25, 2022, from <https://www.nature.com/articles/s41746-022-00589-7>