

ABSTRACT

Schizophrenia is a mental issue that typically shows up in late immaturity or early adulthood. Described by dreams, mental trips, and other subjective troubles, schizophrenia can regularly be a long lasting battle. Schizophrenia is a mental issue wherein practical and basic cerebrum systems are upset. Old style organize investigation has been utilized by numerous scientists to measure mind systems and to examine the system changes in schizophrenia, however lamentably measurements utilized in this old style technique exceptionally rely upon the systems' thickness and weight; the examinations made by this strategy are one-sided. Customary techniques either need enough memorable information or to keep the customary checking on persistent exercises for recognizable proof of a patient related with a mental ailment infection. Strategy: In request to address this issue, we propose a philosophy to characterize the patients related with constant mental sickness maladies (for example Uneasiness, Depression, Bipolar, and Consideration Deficit Hyperactivity Disorder) in view of the information removed from the Reddit, an outstanding arrange network stage. The proposed technique is utilized through Co-training technique by joining the discriminative intensity of generally utilized classifiers.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF FIGURES	viii
	LIST OF ABBREVIATIONS	ix
1	INTRODUCTION	1
	1.1 OVERVIEW	3
2	LITERATURE SURVEY	4
3	METHODOLOGY	
	3.1 AIM OF THE PROJECT	7
	3.2 HARDWARE REQUIREMENTS	8
	3.3 SOFTWARE REQUIREMENTS	8
	3.4 OVERVIEW OF PYTHON	9
	3.5 SYSTEM DESIGN	10
	3.5.1 System Architecture	10
	3.5.2 Module overview	11
	3.5.2.1 Loading Live Data	12
	3.5.2.2 Data Preprocessing	13
	3.5.2.3 Building the Model	14
	3.5.2.4 Training the Model	15
	3.6 RANDOM FOREST ALGORITHM	16
	3.7 DECISION TREE ALGORITHM CLASSIFICATION	19
	3.8 RAD	25
	3.9 BENEFITS	28

4.	RESULT AND DISCUSSION	
	4.1 DECISION TREE BOUNDARIES	30
	4.2 LOGISTIC REGRESSION	30
	4.3 DATA PREPROCESSING	31
	4.4 TEXT TOKENIZATION	32
	4.5 CLASSIFICATION	33
5.	SUMMARY AND CONCLUSION	
	5.1 SUMMARY	34
	5.2 CONCLUSION	35
	5.3 FUTURE WORK	35
	REFERENCES	37
	APPENDIX	
	A.SOURCE CODE	39
	B.PUBLICATION	46

LIST OF FIGURES

FIGURE NO.	NAME OF THE FIGURE	PAGE NO.
3.1	ARCHITECTURE DIAGRAM	11
3.2	RANDOM FOREST	17
3.3	DECISION TREE	18
3.4	ARCHITECTURE OF DECISION TREE	20
3.5	TREE DIAGRAM	22
3.6	RAD MODEL	25
4.1	PREPROCESS DATA	31
4.2	KEYWORD VECTORS	32
4.3	PREDCTION	33
4.4	MODEL ACCURACY	33

LIST OF ABBREVIATIONS

RAD	Rapid Application Development
SDLC	Software Development Life Cycle
NLP	Natural Language programming
NE	Named Entity
RNN	Recurrent Neural Network

CHAPTER 1

INTRODUCTION

Schizophrenia is a dangerous neurological sickness, which seems in roughly 1% of the adult population around the globe. It affects remembrance, awareness, linguistic and executive functions of the brain. According to World Health Organization (WHO) stated that the worldwide incidence of global Schizophrenia is just about 20 million people. Among that 69% percentage of Schizophrenia disorder patients are not obtaining proper meditations. There are various causes for Schizophrenia disorders.

Any research does not distinguish one particular aspect. Hereditary and environmental influences perform a major role in its diagnosis. It is normally developed in the adulthood under the age of 26. The patients which are affected by this disease will have intellectual deficiencies, including reading disorders. They will have difficulty to select suitable phrases while they are talking and composing. It modifies patients' observation, points of view, and conduct as confirm by fantasies, dreams, muddled discourse or conduct, social withdrawal, and differed intellectual deficiencies. Fig 1.

Schizophrenia Word Cloud Schizophrenia Discourse (SD) is a mental ailment, which is the highly predominant neurodegenerative disease in this globe. It is a radical and dense mental illness influencing susceptible peoples, whose pervasiveness is projected to twice by 2030. Diagnosis and medication of neurological Discourse are emphases in brain study. Despite the fact that meds have been created to treat schizophrenia, they don't work for all patients, and there is proof that a few types of schizophrenia seem, by all accounts, to be treatment safe. This just underlines the significance of leading fruitful schizophrenia clinical preliminaries.

According to the artificial intelligence study, social media text analysis has been a leading area of study in the past two to three decades that has been commonly applied in wide array of research such as healthcare, smart home, surveillance systems, human-computer interaction, gaming and so on. With an exceptional development of web-based social networking, huge number of individuals deliberately share a lot of information by communicating their states of mind, sentiments, feelings,

and day by day battles with emotional well-being issues via web-based networking media stages like Twitter.

This offers open doors for new comprehension of these networks. With the improvement of Internet utilization, individuals have begun to impart their encounters and difficulties to emotional wellness issue through online discussions, small scale web journals or tweets. Their online exercises roused numerous specialists to present new types of potential human services arrangements and techniques for early melancholy recognition frameworks. Utilizing distinctive Natural Language Processing (NLP) methods and content characterization draws near, they attempted to prevail in a better improvement.

The singular aim of this application is to develop a machine learning especially deep learning approach, named Long-Short Term Memory (LSTM) plus Convolutional Neural Network (CNN), for detecting Schizophrenia disorders using features obtained from a different dataset from kaggle.com, namely the Schizophrenia Discourse dataset. The proposed LSTM+CNN is competent to handle high-ranking dimensional data by expanding the model interpretability and discovering the highly appropriate features while dispensing with the typical inequitable class division. Natural language processing (NLP) has gained significantly from the revival of neural networks (NNs), expected to their superior execution with fewer necessary of contrived features.

Convolutional neural network (CNN) and recurrent neural network (RNN) are the two major types Deep neural network architecture. Gating systems have been established to improve certain constraints of the basic recurrent neural network, causing in two fundamental recurrent neural network types: long short-term memory (LSTM) and gated recurrent unit (GRU). This application is therefore dedicated on approaches to categorize symptoms of Schizophrenia Discourse in posted tweets as early as feasible established on deep learning methodology.

This paper content is organized into six divisions. The first division contains the Introduction section. The second division specifies about social network. The third division illuminates related work. Our proposed system is deliberated in the fourth division which covers the datasets, data analysis procedure, the proposed LSTM+CNN approach. The experiment results are presented in the fifth division and

conclusions are illustrated in the final segment. The final segment also contains the discussion as well as the future work in that direction.

1.1 OVERVIEW

The social media network has become immeasurable and reputation along with all the service industries today. Today most of the people around the world are using the social media network to share their individual, qualified interest and condition consciousness. Social media networks are the collection of collaborative web-based technology where interested communities produce and share expertise, concepts, intelligence, specialized interests and other information through virtual communities using online communication channels. There are various types of social medias like wikis, Social Networks, Media Sharing Networks, Discussion Forums, Bookmarking & Content Curation Networks, Social Review Websites, Community Blogging & Publishing Websites, Image & Video Hosting Networks, Sharing Economy Networks, Social Shopping Networks, Interest-Based Networks, Anonymous Social Networks and many more. The popular social networks are Facebook, Twitter, LinkedIn etc. The social networks are web based social media applications which are constantly evolving. The communities uses the social network to connect people and share knowledge through human-to-human interaction. It is also called as relationship networks where likeminded people can share ideas and knowledge. The popular image and video hosting websites are Flickr, photobucket, Instagram, Snapchat, YouTube. Vimeo etc. It is also called as Media Sharing Websites. These media sharing networks can be used to share photos, images, videos, live videos, audios and other medias which can be accessed from anywhere and anytime around the world. The image and video hosting websites can be used to promote brand recognition, lead creation etc. These networks mostly promote user generated content. The widespread Discussion

CHAPTER 2

LITERATURE SURVEY

Joseph Kambeitz, Stefan Leucht, Lana Marija Kambeitz-Illankovic, Stephen James Wood Detecting Neuroimaging Biomarkers for Schizophrenia In the year 2015. Schizophrenia is a mental issue that typically shows up in late immaturity or early adulthood. Described by dreams, mental trips, and other subjective troubles, schizophrenia can regularly be a long lasting battle. Schizophrenia is a mental issue wherein practical and basic cerebrum systems are upset. Old style organize investigation has been utilized by numerous scientists to measure mind systems and to examine the system changes in schizophrenia, however lamentably measurements utilized in this old style technique exceptionally rely upon the systems' thickness and weight; the examinations made by this strategy are one-sided.

Kimberly McManus, Emily K. Mallory, Rachel L. Goldfeder, MS· Winston A. Haynes BA Mining Twitter Data to Improve Detection of Schizophrenia In the year 2016. Customary techniques either need enough memorable information or to keep the customary checking on persistent exercises for recognizable proof of a patient related with a mental ailment infection. Strategy: In request to address this issue, They proposed a philosophy to characterize the patients related with constant mental sickness maladies (for example Uneasiness, Depression, Bipolar, and Consideration Deficit Hyperactivity Disorder) in view of the information removed from the Reddit, an outstanding arrange network stage. The proposed technique is utilized through Co-training technique by joining the discriminative intensity of generally utilized classifiers.

Jonathan D. Tatum, IStanford University, Stanford, CA Multivariate Pattern Recognition Studies In the year 2015 Individuals who suffer from schizophrenia comprise 1 percent of the United States Population and are four times more likely 10 die of suicide than the general US population. Identification of at-risk individuals with schizophrenia is challenging when they do not seek treatment. Microblogging platforms allow users to share their thoughts and emotions with the world in short snippets of text. In this work. They leveraged a large corpus of Twitter posts and machine-learning methodologies to detect individuals with schizophrenia. Using features from tweets.

Wolfram Hinzen and Joana Rossell The linguistics of schizophrenia thought disturbance as language pathology across positive symptoms. In the Year 2015 Emoticon use, posting time of day, and dictionary terms. They trained, built, and validated several machine learning models. The Support Vector machine model achieved the best performance with 92% precision and 71% recall on the held test set. Additionally, They built a web application that dynamically displays summary Statistics between cohorts. This enables outreach to undiagnosed individuals. Improved physician diagnoses, and destigmatization of schizophrenia. This Method hypothesize that linguistic organization in the schizophrenic brain plays a more central role in the pathogenesis of this disease than commonly supposed. Against the standard view, that schizophrenia is a disturbance of thought or selfhood, They argue that the origins of the relevant forms of thought and self hood at least partially depend on language. The view that they do not is premised by a theoretical conception of language that They identified as 'Cartesian' and contrast with a recent 'un Cartesian' model. This linguistic model empirically argues for both (i) a one-to-one correlation between human specific thought or meaning and forms of grammatical organization, and (ii) an integrative and codependent view of linguistic cognition and its sensory motor dimensions. Core dimensions of meaning mediated by grammar on this model specifically concern forms of referential and propositional meaning. A breakdown of these is virtually definitional of core symptoms. Within this model the three main positive symptoms of schizophrenia falling to place as failures in language-mediated forms of meaning, manifest either as a disorder of speech perception (Auditory Verbal Hallucinations), abnormal speech production running without feedback control (Formal Thought Disorder), or production of abnormal linguistic content (Delusions). The hypothesis make settable predictions for the language profile of schizophrenia across symptoms; it simplifies the cognitive neuro psychology

Deanna Greenstein , James D. Malley , Brian Weisinger , Liv Clasen and Nitin Gogtay Multivariate machine learning methods can be used to classify groups of schizophrenia patients and controls using structural magnetic resonance imaging (MRI) In the year 2014. However, machine learning methods to date have not been extended beyond classification and contemporaneously applied in a meaningful way to clinical measures. They hypothesized that brain measures would classify groups, and that increased likelihood of being classified as a patient using regional

CHAPTER 3

METHODOLOGY

3.1 AIM OF THE PROJECT

To implement an end-to-end extensible framework for extracting the information about Schizophrenia **symptoms** delusions. The hallucinations, catatonic or disorganized .behavior, apathy, reduced thought fluidity, disperse and unproductive language Using Machine Learning.

One of the primary choices we face for our e-voting project implementations is -Which development methodology should we use?! There are various software development methodologies are available in the ICT industry. They are

- Agile Software Development
- Crystal Methods
- Dynamic Systems Development Model (DSDM)
- Extreme Programming (XP)
- Feature Driven Development (FDD)
- Joint Application Development (JAD)
- Lean Development (LD)
- Rapid Application Development (RAD)
- Rational Unified Process (RUP)
- Scrum
- Spiral
- Systems Development Life Cycle (SDLC)
- Waterfall (a.k.a. Traditional)

Among all these methodologies, we have decided to compare the two most popular and widely used methodologies: