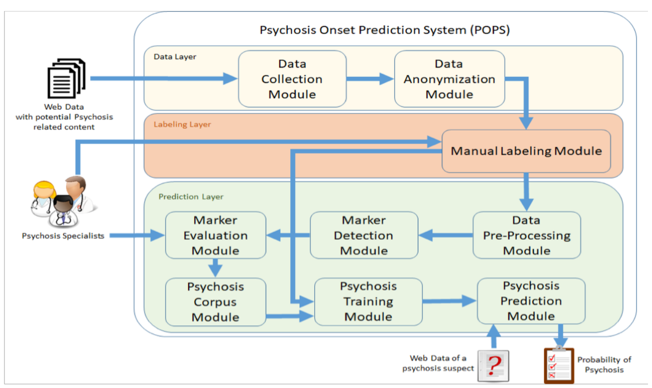
Anonymous Prediction of Psychosis in Social Media

Srikanth Tadisetty, Kambiz Ghazinour
Department of Computer Science, Kent State University

ABSTRACT

Psychosis is one of the mental illnesses that many people are struggling with and its early detection can result in better chances for successful treatment. Unfortunately, symptoms of psychosis are not easy to be discovered and that makes the diagnosis difficult. Some people are unaware that they are suffering from psychosis. In this work we propose a method to predict if someone is potentially dealing with psychosis through detection using posts history on social media. The framework applies NLP techniques with domain expert knowledge to create a custom psychosis corpus that consists of terms consistent with how individuals with psychosis post of social media, applied towards psychosis prediction. Examination through real-world samples from numerous patient posts on social media indicates that the model is an accurate predictor and validates the corpus effectiveness.

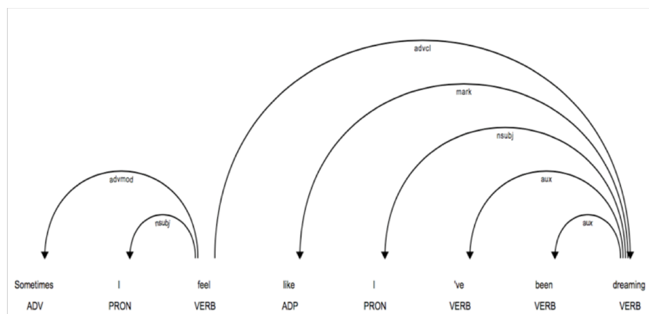
POPS FRAMEWORK



BACKGROUND

POPS aims at predicting psychosis presence using social media post history. Its pipeline consists of:

- Post extraction and preprocessing
- User anonymization
- Identification of personal vs general vs non-mental health related posts through lexical analysis parameters



- Corpus creation using posts consistent with mental illness rated by a domain knowledge expert.
- Applying corpus for prediction of psychosis within comments and of user based on post history

Rating	Description
1	Nothing about this comment sample raises any suspicion at all about the possibility that its author is experiencing psychosis
2	This comment is a little off. But suspicion for psychosis is mild.
3	There is very strong suspicion that the writer has psychosis
4	The writer unambiguously identifies self as someone with psychosis

RESULTS

Classifier	Post Prediction Vector	Actual Vector	Accuracy
NB	[0000000000]	[0001100000]	80%
SVM	[0000000000]	[0000000000]	100%
KNN	[0000000000]	[0000100100]	80%

Classifier	User Prediction Vector	Actual Vector	Accuracy
NB	[1001001100]	[1001111100]	80%
SVM	[0000000000]	[1001111100]	40%
KNN	[0001000100]	[1001111100]	60%

Anonymous Prediction of Psychosis in Social Media

Srikanth Tadisetty
Department of Computer Science
Kent State University
Kent, Ohio, United States
stadiset@kent.edu

Kambiz Ghazinour
Department of Computer Science
Kent State University
Kent, Ohio, United States
kghazino@kent.edu

Abstract—**Psychosis is one of the mental illnesses that many people are struggling with and its early detection can result in better chances for successful treatment. Unfortunately, symptoms of psychosis are not easy to be discovered and that makes the diagnosis difficult. Some people are unaware that they are suffering from psychosis. In this work we propose a method to predict if someone is potentially dealing with psychosis through detection using posts history on social media. The framework applies NLP techniques with domain expert knowledge to create a custom psychosis corpus that consists of terms consistent with how individuals with psychosis post of social media, applied towards psychosis prediction. Examination through real-world samples from numerous patient posts on social media indicates that the model is an accurate predictor and validates the corpus effectiveness.**

***Index Terms*—Psychosis, natural language processing, machine learning, classification, prediction.**

I. Introduction

Posting on the internet, including weblogs or social media, is one of the ways individuals seek for an outlet to express themselves or mental health concerns [2]. One of the major advantages the internet offers is the ability for the individual to stay anonymous [17]. 90% of US youth use social media daily, surpassing texting and email. [16]. Over 2 billion users engage with social media websites [9]. For many mental health issues such as psychosis, the timing of detection and treatment is critical; short and long-term outcomes are better when individuals begin treatment close to the onset of psychosis [5, 22]. The National Suicide Prevention Lifeline [3] for instance adopted a community to help other's suffering from mental illness. The Lifeline aims at providing help when an individual is in crisis often associated with mental disorders – mood disorders, schizophrenia, anxiety disorders, personality disorders and aggressive tendencies [3]. Such illnesses can be diagnosed well in advance without the need for crisis prevention and allow for early treatment plans. With the adoption of texting as a major communication standard, Crisis-Text-Line [2] specializes in crisis prevention through text messaging by connecting individuals to their trained social workers. People *feel* their identity/privacy is more preserved when they text. Law enforcement are involved in cases of imminent threat. Crisis prevention is not absolute, and these services are dependent on individuals to reach out for support. People considering suicide seek assistance and studies show that 64% of people who attempt suicide visit a doctor in the month before their attempt, while 38% visit in the week before [1]. This is much too late considering that 30 to 70% of suicide victims suffer from one of the aforementioned mental disorders [1]. Early identification and monitoring of schizophrenia can increase the chances of successful management to reduce the chance of psychotic episodes leading to a more comfortable life [18]. While the internet offers a positive medium for short term therapy, it is not a face to face therapy session, wherein a trained professional is better able to deduce the root of the problem [5]. The drawback of psychiatry is that it lacks objectified tests for mental illnesses that would otherwise be present in medicine. Current neuroscience has not yet found genetic markers that can characterize individual mental illnesses [13].

Blended-care treatment takes the infusion of face-to-face therapy and online therapy. Social masks become unnecessary, while still giving the opportunity to nullify misunderstanding during correspondence without the lack of perception that would otherwise be reassuring [5,17]. Developments in natural language processing present such an avenue to psychiatry as words may present a peek into the mind [6]. A thought disorder (ThD), which is a widely found symptom in people suffering from schizophrenia [14], is diagnosed from the level of coherence when the flow of ideas is muddled without word associations. Many people who have mental illness do not get the treatment that would alleviate their suffering because it often goes undiagnosed unlike illnesses such as asthma, diabetes or bronchitis or heart disease [7]. While there exist medical dictionaries such as SNOMED [11] and MedDRA [12], they are unable to assess the psychiatric status of individuals based of speech. To our knowledge, there are no linguistic markers for how individuals with mental illness speak on social media. The 2014 Marysville Pilchuck high school shooting [10] is considered to be one of the most brutal school shootings to date. The shooter had given signs of mental illness on Twitter, stating such comments as *"It breaks me… It actually does… It know it seems like I'm sweating it off… But I'm not… And I never will be able to…"* and *"It won't last….It'll never last"* [10].

Our focus in this paper will be classifying whether or not an individual exhibits signs of mental illness based on social media comment content. Often people suffering from mental illness have a diagnosis for more than one disorder – concomitance. The concomitance of schizophrenia and bipolar disorder is schizoaffective disorder [18] and we omit focus on subtypes. Since not every comment may be concerned with mental illness, our initial focus is to build a custom corpus of terms consist with mental illness discussion on social media. We focus on unsupervised grouping of words and see how well they are able to extract comments of individuals displaying mental illness. We propose a system capable of classifying individuals using these terms along with lexical dependency analysis acting as features. We find that using all of these features leads to a reasonable accuracy of 80% on a hand-tagged data set and therefore, reasonable for implementation in online surveillance systems.

II. Related Literature

There has been recent work done in identifying features in health and specific mental illnesses [2,15,18,19,20,21,22]. While predictions are based on association rule mining algorithms, language categorization is not as straight forward. It has been shown that 19 - 28% of all internet users participate in health focused groups and medical forums, while the rest may use abstract language from which we may derive a mental health attribute [15]. Crisis-Text-Line [2] is a text messaging emergency hotline exchanging over 72 million messages to date. They have created a corpus of high-profile markers that indicate different levels of urgency that is used to queue conversations based on urgency. Ghazinour et al. [15] explored the effectiveness of SNOMED and MedDRA dictionaries to mine personal health information on MySpace using the bag of words content-based model. They observe that traversing down the dictionary hierarchy, the less prevalent that terms appear in comments, while more general terms may be used out of context. Mitchell et al. [18] explore the topic probability based on word probability using LDA on Tweets collected between schizophrenics and non-schizophrenics. The method performed better when tested against brown clustering and CLM (5-gram) obtained features. Rose et al. [19] felt that the stigma against people with mental illness is a barrier to those seeking help for mental illness. Their study investigated what words students use to refer to those with mental illness in an elementary environment among five schools. A negative stigma undermines the need to seek help during early stages, though it may come from a lack of knowledge [19]. Sokolova et al. [20] used to a custom ontology of 500 terms comprised of commonly used terms by patients in clinical settings, which performed exceptionally greater than MedDRA and SNOMED in extracting personal health information on Twitter. Their terms consisted of general body and organ vocabulary. Wei and Singh [21] use network-based features along traditional content-based features to detect extremism on Twitter based from sentiment towards ISIS. They further this research by constructing weighted networks that models the information between publishers and mentioners of ISIS content. They then identify 50 features associated with negative/positive ISIS sentiment supporters to construct a weighted graph based on centrality to identify user extremists [22].

III. Implementation

POPS – Psychosis Onset Prediction System consists of various modules to predict if an individual is mentally ill using social media comment history.

To achieve this, a custom raw data scraper gathers thread, user, post and timestamp data from the website. Data is cleaned and organized with openpyxl. To keep in accordance with the United States Health Information Portability and Accountability Act (HIPPA), the comments were made untraceable to a particular user. A username that was an individual's actual name is kept hidden by replacing it instead with the retrieved userID that the website associates them with.

A custom regex parser is used to eliminate mistakes within sentences users might have written and creates tokens from each word. Stop-words removal and lemmatization then remove redundancy of terms. Tokens are stemmed using the NLTK snowballStemmer [8] and aggregated back into a sentence format. A term frequency distribution, collocations and TF-IDF are run to isolate frequent unigrams, bigrams, trigrams (The posts are from people with mental illness).

Posts are categorized into 3 categories using lexical analysis, keyword presence and spaCy POS tagger [4] applying the following subject-verb connectives: "nsubj", "nsubjpass", "poss".
Categorization is evaluated by psychosis expert.

We consider 50 random comments from 10 users; 5 users chosen from the personal mental health and general mental health categories were tagged by our psychosis expert without indicating what user it came from to prevent bias. Each comment is tagged using a custom evaluation schema. With each comment given a rating, the ratings are stripped and bunched based on what user they came from and returned to our psychosis expert for an unbiased overall user score following the same evaluation schema.

We apply bag of words for sentence vectorization. Scikit-learn classifier (Naïve Bayes, SVM, KNN) is trained with an 80% by 20% training/set split since the tagged data amount is scarce.

IV. Results

Our findings show that the SVM classifier performed most accurately (100%), while the NB and KNN (K=3) classifiers both capped at 80%.

This supports the fact that our corpus keywords consisting of unigrams and bigrams are effective.

References

[1] *Suicide*. Available: http://www.mentalhealthamerica.net/suicide
[2] *Crisis-Text-Line*. Available: https://www.crisistextline.org
[3] *National Suicide Prevention Lifeline*. Available: https://suicidepreventionlifeline.org
[4] *spaCy Linguistic Features*. Available: https://spacy.io/usage/linguistic-features
[5] K. D. Baker and M. Ray, "Online counseling: The good, the bad, and the possibilities," *Counselling Psychology Quarterly,* vol. 24, no. 4, pp. 341-346, 2011.
[6] G. Bedi *et al.*, "Automated analysis of free speech predicts psychosis onset in high-risk youths," vol. 1, p. 15030, 2015.
[7] B. (MD), *National Institute of Health (US); Biological Sciences Curriculum Study.* (NIH Curriculum Supplement Series [Internet]). National Institutes of Health (US), 2007.
[8] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python: O'Reilly, 2009. [Online]. Available: https://www.nltk.org/book/
[9] M. L. Birnbaum, S. K. Ernala, A. F. Rizvi, M. De Choudhury, and J. M. Kane, "A Collaborative Approach to Identifying Social Media Markers of Schizophrenia by Employing Machine Learning and Clinical Appraisals," *Journal of Medical Internet Research,* vol. 19, no. 8, p. e289, 2017.
[10] G. Botelho. (2014, July 23). *Jaylen Fryberg: From homecoming prince to school killer*. Available: https://www.cnn.com/2014/10/24/us/washington-school-shooter/index.html
[11] E. G. Brown, L. Wood, and S. Wood, "The Medical Dictionary for Regulatory Activities (MedDRA)," *Drug Safety,* journal article vol. 20, no. 2, pp. 109-117, February 01, 1999.
[12] College of American Pathologists. Committee on Nomenclature and Classification of Disease. and R. A. Côté, *Systematized nomenclature of medicine*, 2nd ed. Skokie, Ill.: CAP, 1979, p. v.
[13] M. Corcoran Cheryl *et al.*, "Prediction of psychosis across protocols and risk cohorts using automated language analysis," *World Psychiatry,* vol. 17, no. 1, pp. 67-75, 2018/07/19 2018.
[14] B. Elvevåg, P. W. Foltz, D. R. Weinberger, and T. E. Goldberg, "Quantifying incoherence in speech: An automated methodology and novel application to schizophrenia," *Schizophrenia research,* vol. 93, no. 1-3, pp. 304-316, 2007.
[15] K. Ghazinour, M. Sokolova, and S. Matwin, "Detecting Health-Related Privacy Leaks in Social Networks Using Text Mining Tools," Berlin, Heidelberg, 2013, pp. 25-39: Springer Berlin Heidelberg.
[16] E. Hawkes, D. Rose, B. Schulman J, and M. Ross. (1995). *Schizophrenia Forum*. Available: https://forum.schizophrenia.com
[17] G. M Innocente, *Client-Clinician Texting: An Expansion of the Clinical Holding Environment*. Pennsylvania: University of Pennsylvania Scholarly Commons, 2015.
[18] M. Mitchell, K. Hollingshead, and G. Coppersmith, "Quantifying the Language of Schizophrenia in Social Media," presented at the Proceedings of the 2nd workshop on Computational Linguistic and Clinical Psychology: From Linguistic Signal to Clinical Reality, Denver, Colorado,
[19] D. Rose, G. Thornicroft, V. Pinfold, and A. Kassam, "250 labels used to stigmatise people with mental illness," *BMC Health Services Research,* journal article vol. 7, no. 1, p. 97, June 28, 2007.
[20] M. Sokolova, S. Matwin, Y. Jafer, and d. schramm, "How Joe and Jane Tweet about Their Health: Mining for Personal Health Information on Twitter," in *International Conference Recent Advances in Natural Language Processing, RANLP*, Hissar, Bulgaria, 2013.
[21] Y. Wei and L. Singh, "Using Network Flows to Identify Users Sharing Extremist Content on Social Media," in *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, Cham, 2017, pp. 330-342: Springer International Publishing.
[22] Y. Wei, L. Singh, and S. Martin, "Identification of extremism on Twitter," in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2016, pp. 1251-1255.