

## MENTAL HEALTH AND SOCIAL MEDIA

Mental health conditions affect a large percentage of individuals each year. Traditional mental health studies have relied on information collected through contact with the mental health practitioner. There has been research on the utility of social media for depression, but there have been limited evaluations of other mental health conditions (Jan-Are, Jan & Deede, 2002). First, we will examine specific techniques that have previously been used to analyze forum data, define behavioral health and public health issues, and lastly, we will explore the implications that this research has for big data analytics.

### Analysis of Social Media

In this part of the paper, we explore the various techniques that have been previously used to analyze the data found in social media sites. The rise of social media sites, forums, blogs, and other communications tools has created an online community of individuals who are able to socialize and express their thoughts through various applications (Georgios & Mike, 2012). Microblogging has become a very popular tool for communication among users. The individuals who write these messages blog about their lives share opinions, and discuss current events. As more individuals participate in these microblogging services, more information about their messages becomes available. The massive amount of data in user updates creates the need for accurate and efficient clustering of short messages on a larger scale (Chen & Liu, 2014). Certain research areas have chosen to focus on the opinions and sentiments of these messages (Si et al., 2014), community detection (Newman, 2004), politics (Tumasjan, Sprenger, Sandner, & Welpe, 2010), and user interests (Li et al., 2014). Techniques for clustering this data have included document clustering, topic modeling sentiment analysis, and text mining.

### Topic modeling

Recent years have seen a surge in information that is both digitized and stored. As this trend continues, it has become increasingly difficult for users to find what they are looking for. Novel computational tools are needed to help organize, search, and comprehend these large amounts of data (Chen & Liu, 2014). Currently, we are able to type keywords into a search and find documents that are related to them. However, there is a crucial element that is missing from this process. Specifically, it is important to utilize themes to explore specific topics. A thematic structure could serve as a portal through which users could explore and obtain knowledge about various topics. Topic modeling algorithms are statistical methods that analyze the words of the original documents and discover themes that occur. Furthermore, topic modeling analyzes how these themes relate to one another, and how they differ over time (Blei, 2012). These algorithms do not need any previous annotations or labeling of the documents, these topics surface automatically from the analysis of the original texts. Blei (2012) describes latent Dirichlet allocation (LDA), which is the simplest type of topic model. LDA is a statistical model of a collection of documents that tries to validate the intuition that documents exhibit multiple topics. The simple LDA model provides an effective and powerful way to discover and exploit the hidden thematic structures

found in large amounts of text data.

### Sentiment analysis

Microblogging websites have developed into a source for varied types of information. Individuals post messages about their opinions, current events, complaints, and sentiments about products they use in their daily lives (Liu, 2012). It is very often that companies study these user reactions on microblogging sites. The challenge then becomes how to build a technology that can detect and summarize an overall sentiment. A large amount of social media contains sentences that are sentiment-based. Sentiment is defined as a personal belief or judgment that is not founded on proof or certainty (Davidov, Tsur & Rappoport, 2010). Sentiment involves the use of Natural Language Processing (NLP), statistics, or machine learning methods to extract, identify, or characterize the sentiment content of a text source (Liu, 2012). The automated identification of sentiment types can be beneficial for many NLP systems.

### Text mining

Text mining is the discovery of new information by automatically extracting information from a large amount of various unstructured textual resources (Aggarwal & Zhai, 2012). Text mining can help an organization gain valuable insights from text-based content such as word documents, email, and postings on social media sites like Facebook, Twitter and LinkedIn (Rossi, Malliaros & Vazirgiannis, 2015). Mining unstructured data with natural language processing (NLP), statistical modeling and machine learning techniques can be challenging because natural language text is usually inconsistent. It contains ambiguities caused by inconsistent syntactic structures (Liu, 2012).



a different theme. We describe the themes corresponding to the top five partitions of the graph shown in Figure 1. For example, the first partition contains personal experience and recommendations such as the Linden method for dealing with anxiety, panic attacks. The second partition was focused on other drugs such as Xanax and Benzodiazepine. The third partition contained clinical issues such