ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

Facebook Vs Linkedin: A Text Mining Approach for Job Preference

[1] Ms. Sarita. Byagar, [2] Dr. Anil Jadhav
[1] Ph.D Scholar, Symbiosis International University,
[2] Assistant Professor, Symbiosis Centre of Information Technology

***Abstract* - Social media has become the catchword as every individual is directly or indirectly hooked to these sites. The topical growth of online social networks has enabled job seekers to stay connected and find anticipated jobs through these sites. The amount of textual data generated on these sites is massive which can be mined to find what information they focus on and also discover which site is ideal to the user for job search or job related bustle. The paper is organized as follows: Section 1 is the introduction of social media, social media for jobs and data mining of social media. Section 2 gives a brief review of text mining, Section 3 explains the research questions explored in this paper, the context of the study, Section 4 details its methodological approach i.e procedures. Section 5 discusses the results. Section 6 concludes discussion and conclusion.**

I. INTRODUCTION

The focus of Social Networking Websites is to build online groups or communities of people sharing interests and activities, or who are interested in exploring others interests and activities. These websites provide many ways for users to network or interact, such as e-mail and instant messaging services. Social Networking Sites (SNS) are intended to connect users to each other and to visually display each individual's network of friends, display their personal happenings in their life, showcase their professional skills etc. Social media potentially offer a place for job seekers to tap into their networks for finding work, alert their friends when they get to know of jobs availabilities, or endorse their own skills in a way that is publicly visible to prospective employers (Smith, 2015). The boost in Internet access has led to a rapid rise in use of online social networking sites like Facebook. However, most SNS have unique characteristics and thus all are not used for job search. SNS like LinkedIn have grabbed the lion's share in professional networking space. A recent cover story article in Fortune magazine (Hempel 2010) suggested that connecting on LinkedIn is more useful than exchanging business cards or churning resumes. Along with serving as a prime location for finding and researching jobs, potential job seekers can also utilize their social media presence to highlight relevant skills to prospective employers (Smith, 2015).

"How to effectively search for jobs?" is an extremely significant question for individuals, firms and policy makers. Over the past decades job seekers have tailored their search efforts as the technology has shaped this process. A vital element in the process has been the role of job seekers' social connections. There is significant literature that suggests that "who you know" plays a very important role in finding a job. (Granovetter 2005) argues that social networks are valuable because they affect the flow and quality of information and improve the trust and confidence in the information. These factors are especially important as online platforms have enabled every job post to be accessible to every job seeker across the globe. Increase in Internet penetration has led to a meteoric rise in use of online social networking sites like Facebook. However, most SNS have unique characteristics and thus all are not used for job search. SNS like LinkedIn have grabbed the lion's share in professional networking space. According to (Social Recruiting Survey Results 2014) 73% of recruiters have hired a candidate through social media. The tactics which are generally used for recruiting on social networks like searching for candidates, contact candidates, keeping tab on potential candidates, vet candidates pre-interview and posting jobs are comparatively carried out more on Linkedin than facebook. Job seekers always keep a tab on social networking sites to find out about job openings or opportunities and Linkedin has comparatively more posts (92% to be precise) related to jobs than facebook(48%) or twitter(39%). The daily traffic created by these web sites and the number of users are overwhelming. (Kluemper, D. H., & Rosen, P. A, 2009). Social media has generated a wealth of textual data, which include hidden knowledge that can be explored for variety of research related tasks. In particular, many businesses can mine into the huge amount of social media data to discover new facts, knowledge and interesting patterns and use the findings and enhanced understanding to generate valuable information (Dey, Haque, Khurdiya, & Shroff, 2011).

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

As manual coding of social media data is too time-consuming, in the recent years application of text mining and data mining techniques to analyze social media has gained a lot of consideration.(Barbier& Liu, 2011). For a business to survive it is important to collect their own social media data as well as that of their competitors and then mine the large amount of textual content in order to reveal hidden relationships, insights, patterns, and trends. A recent trend is to conduct opinion mining on social media data in order to identify consumer feelings, opinions, and sentiments on certain subjects/issues and to detect possible changes of opinion(Cheng, Ke, &Shiue, 2011; Pang & Lee, 2008).

II. A BRIEF REVIEW OF TEXT MINING

Text mining also referred as text analytics involves obtaining quality information from various text sources. Text mining typically involves the procedure of structuring the input text; extract patterns within the structured data, and finally evaluation and interpretation of the output. Text mining tasks comprise of text clustering, text categorization, text summarization, sentiment analysis and many more. There are wide ranges of sources from where text data arrive. This data may incorporate documents, emails, tweets, blog posts, posts to name but a few, and is typically defined as 'unstructured data'. The essential challenge of text mining is the precise analysis of both structured and unstructured data in order to extract, find meaningful associations, trend and pattern in large mass of text. The escalating volume and accessibility of digital data online in social media environments like Twitter, Facebook, Pinterest, Youtube, GooglePlus, LinkedIn and collaborative online environments like Wikipedia, offer innovative opportunities for researchers to investigate social, cultural, economic and political behavior.(Ampofo, Collister, O'Loughlin, Chadwick, 2015). Various text mining techniques are used by researchers to analyze huge volume of text data related to health care, businesses, education domain, social media and many more. According to the literature, for mining and analyzing social media business data there is no explicit standard method. In this paper, an open source approach for text mining using a set of R packages for mining facebook and linkedin data is presented [22], which is applicable for other social media sites too. The analysis in this paper utilizes data mining algorithms, such as association rules, classification, clustering, and many more in exploring and discovering new information and relationships in textual sources. It is an inter-disciplinary research field combining information retrieval, data mining and statistics.

III. RESEARCH QUESTIONS

User generated contents are considered as very imperative in the study of social media. This paper focuses on the two popular social media websites facebook and linkedin and observed the posts updated and shared by users.
The goal of this paper can be formulated in form of research questions as follows:

- Is it possible to exploit information coming from the social networks, Facebook and LinkedIn, for finding which site is preferable and focuses more on professional activities and endeavors?
- Which websites keywords demonstrate job related activities?

IV. METHODOLOGY AND PROCEDURES

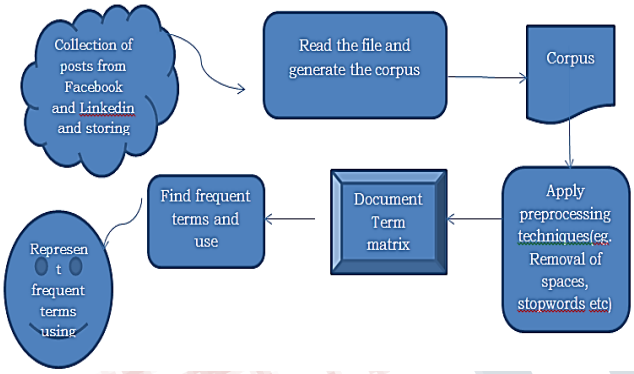
To answer the research questions, we conducted a social media comparative analysis for the Facebook and LinkedIn sites. First, we collected quantitative data manually from their linkedin site and facebbok data from an online website. These text posts were posted by different users. Secondly, we applied text mining to analyze the text messages posted on their Facebook and LinkedIn sites in order to discover which are the frequently used terms on these websites and to find which website is more related to job, interview or new opportunities related activities. This study used the posts collected between April 2015 and November 2016 as the sample for text mining. The posts were saved into Excel Spreadsheets in CSV format for analysis.
The process followed for mining Facebook and LinkedIn data is shown below in Figure 1. The steps involved in the methodology are described as follows:
1. Data Access: Manual data collection of posts posted and shared by users from facebook and linkedin sites and stored in CSV format.
2. Data Cleaning: Using some additional tm package to get the text, then, clean the data by removing stop words, special characters, spaces, punctuation, URLs and performing stemming (get the root of the words). This step produces a structured representation of data called Term-Document Matrix which is a mathematical matrix that describes the frequency of terms that occur in a collection of documents. In a document-term matrix, rows correspond to documents in the collection and columns correspond to terms.

3. Data Analysis: The structured representation produced in the previous step enables performing Mining tasks such as finding association rules and finding frequent terms.
4. Visualization: The wordcloud package and bar plots in R has been used to show the frequency of words in the posts.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

A user can add information to Facebook and LinkedIn site in many ways, such as posting messages to the wall and uploading photos. A popular communication feature on both these sites is the wall post (Bender, Jimenez-Marroquin, & Jadad, 2011). A wall post is a blurb that can be posted by any user who visits the site (McCorkindale, 2010). The wall post can be publicly viewed by anyone who visits the site too (McCorkindale, 2010). Thus, our analysis mainly focused on the wall posts. A total of 400 wall posts from LinkedIn was collected. Facebook data was obtained from an online website(approximately 400 posts). Specifically, for a particular user, Facebook has around 25 wall posts on average per day and LinkedIn has around 10 wall posts on average per day.



V. RESULTS

R package with version 3.3.2 has been used for analysis of the collected data. The R package enables authentication and access to twitter messages by using keyword search queries but not for facebook and linkedin messages due to some privacy policy with effect from May 2016. After getting the data, the extraction of text and the cleaning has been done using tm package which is a framework for text mining applications within R. The output obtained from the previous step is a structured representation of text, data-term matrix. This structured representation can be used to perform text mining using tm package. One of the outputs of this is the word cloud representation of posts. We have the 'wordcloud' representation and the frequent terms used in posts from both the sites. The size of each term in the cloud indicates the number of mentions of that term in the posts, reflecting its importance. It is also possible to obtain the most frequent terms with a specific occurrence threshold.

The result of analysis for LinkedIn

After applying the function to find frequent terms, below are the frequent terms used in the dataset using the function findFreqTerms()

"best" "business" "developer" "developers" "digital" "experience" "first" "get" "good" "hiring" "india" "interested" "interview" "job" "like" "looking" "most" "new" "one" "openings" "opportunity" "people" "sales" "sap" "time" "what" "why" "work" "years"

The frequent words can also be plotted in a histogram(bar plot) to show a clear pictorial representation of the frequency of words. The below figure 2 depicts top 26 frequent words and figure 3 depicts top 10 frequent words.

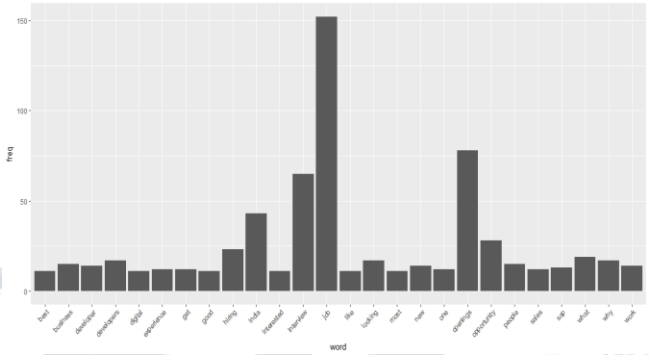


Figure 2 LinkedIn : Bar Plot representing frequency of words (Top 26 words)

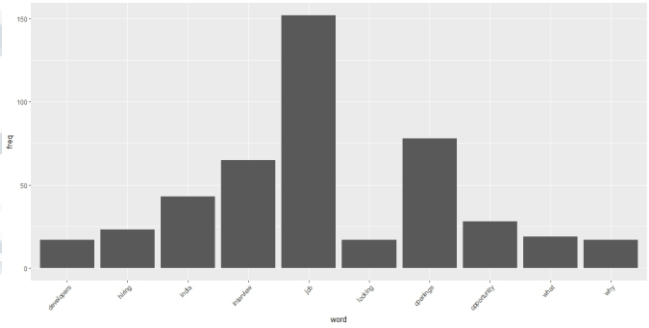


Figure 3 LinkedIn : Bar Plot representing frequency of words (Top 10 words)

And finally the data visualization in Figure 4 with 'wordcloud' representation of the terms used in the data. Wordcloud give superior prominence to words that appear more frequently in the source text. It offers businesses a valuable way to communicate and convey important information at a glance. Using a word cloud can make data sizzle and immediately express crucial information.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 5, May 2018

Figure 4 LinkedIn Wordcloud

The result of analysis for Facebook

The same approach was applied for mining data collected from facebook. For facebook these are the list of frequent terms(105 terms) listed below after carrying out data cleaning . The number of terms can be found by choice i.e by giving the number.

"and" "ani" "anyon" "back" "becaus" "befor" "best" "better" "birthday" "call" "can" "class" "come" "day" "done" "even" "ever" "everyon" "excit" "facebook" "feel" "final" "find" "first" "friend" "fun" "get" "god" "good" "got" "great" "happi" "hate" "head" "home" "hope" "hour" "just" "keep" "know" "last" "let" "life" "like" "littl" "live" "lol" "long" "look" "love" "made" "make" "miss" "much" "need" "never" "new" "next" "night" "now" "one" "onli" "peopl" "play" "propname" "read" "realli" "right" "say" "school" "see" "sick" "sleep" "someon" "someth" "start" "still" "take" "thank" "the" "thing" "think" "time" "today" "tomorrow" "tonight" "tri" "two" "use" "veri" "wait" "want" "watch" "way" "week" "weekend" "well" "what" "whi" "will" "wish" "wonder" "work" "world" "yay" "year" "you"

The below histogram (bar plot) figure 5 shows top 24 frequent words and figure 6 show top 11 frequent words.

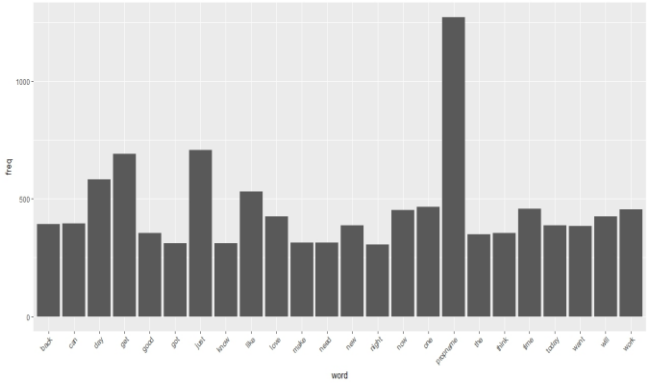


Figure 5 Facebook: Bar Plot representing frequency of words (Top 24 words)

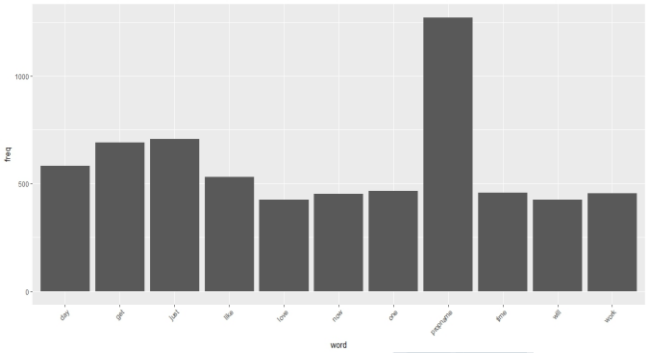


Figure 6 Facebook: Bar Plot representing frequency of words (Top 11 words)

Figure 7 Facebook wordcloud

VI. DISCUSSION AND CONCLUSION

As social media has become a subject of interest for research, it is imperative to understand how social media data can be harvested for finding which sites are more preferable for jobs, new trends in technologies and hiring related undertakings. It is viable to apply text mining tasks for facebook and linkedin data to analyse user contributed and shared posts and the same can be done for other social networking sites also. This paper investigates the role of online social networks for job search and job related activities, new trends in market etc. Through result of mining functions(i.e bar plot and wordcloud) applied it is evident that linkedin is more inclined towards job related happenings as the frequently used terms viz; job, opportunity, hiring, opportunity, openings etc direct towards the same. Facebook is generally used to express ones views, and display social actions and events rather than professional endeavors. The word cloud illustrates clearly

the recurrent terms used in the source text data and evidently linkedin focuses on professional activities. Results from the text mining of social media data analysis show that job seekers would be actively engaged to linkedin as it focuses on new job opportunities, current hiring, interviews etc. If you're in the process of a job search or networking then LinkedIn is far superior. While LinkedIn doesn't have reviews about companies but one can look up employees and see if they have high turnover, what the profiles are like for people they've hired for roles you're looking for, what the management team is like, where they came from, etc whereas facebook gives more of social aspect of a specific employee like his family, friends and acquaintances and happenings in his personal life. Hence linkedin is the most used and chosen social networking site for placements, opportunities, hiring, interviews and most important job search.

Future research in this area can focus on prediction of employee's attrition or whether an employee is in process of new job or opportunities and such prediction can help companies to retain their valuable employee(s).

REFERENCES

1. Ampofo, L., Collister, S., O'Loughlin, B., & Chadwick, A. (2015). Text Mining and Social Media: When Quantitative Meets Qualitative and Software Meets People. Innovations in Digital Research Methods, 161-91.

2. Thiel, K., Kötter, T., Berthold, M., Silipo, R., &Winters, P. (2012). Creating usable customer intelligence from social media data: Network analytics meets text mining. KNIME. com Report.

3. Mostafa, M. M. (2013). More than words: Social networks' text mining for consumer brand sentiments. Expert Systems with Applications, 40(10), 4241-4251.

4. He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. International Journal of Information Management, 33(3), 464-472.

5. Dai, K., Nespereira, C. G., Vilas, A. F., & Redondo, R. P. D. (2015). Scraping and Clustering Techniques for the Characterization of Linkedin Profiles. arXiv preprint arXiv:1505.00989.

6. Younis, E. M. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. International Journal of Computer Applications, 112(5).

7. Job Search on Online, Rajiv Garg, Rahul Telang, David Bowman & Murray Johannsen, TTG Consultants, Carnegie Mellon university

8. Case, T., Gardiner, A., Rutner, P., & Dyer, J. (2013). A linkedin analysis of career paths of information systems alumni. Journal of the Southern Association for Information Systems, 1(1).

9. Robinson, Winter &Sinar, Social Media as a Tool for Research: A Turnover Application Using LinkedIn, The Industrial Organizational Psychologist, July 2014 Volume 52 Issue 1

10. Skeels, M. M., &Grudin, J. (2009, May). When social networks cross boundaries: a case study of workplace use of facebook and linkedin. InProceedings of the ACM 2009 international conference on Supporting group work (pp. 95-104). ACM.

11. Kluemper, D. H., & Rosen, P. A. (2009). Future employment selection methods: evaluating social networking web sites. Journal of managerial Psychology, 24(6), 567-580.

12. Garg, Rajiv, and Rahul Telang. "To be or not to be linked on LinkedIn: Job search using online social networks." Available at SSRN 1813532 (2011).

13. http://theundercoverrecruiter.com/linkedin-vs-facebook-battle-professional-network-infographic/

14. Treem, J. W., &Leonardi, P. M. (2013). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. Annals of the International Communication Association, 36(1), 143-189.

15. Broughton, A., Foley, B., Ledermaier, S., & Cox, A. (2013). The use of social media in the recruitment process. Retrieved May, 2, 2014.

16. Huberman, B. A., Romero, D. M., & Wu, F. (2008). Social networks that matter: Twitter under the microscope. Available at SSRN 1313405.

17. Betances, D., Solarczyk, R., & Bellows, C. (2012). Personal social networking: Affects on companies hiring decisions.

18. Jobvite, Social Recruiting Survey Results 2014 http://www.jobvite.com/blog/2014-social-recruiting-survey-infographic/

19. Martensen, M., Börgmann, K., & Bick, M. (2011). The Impact of social networking sites on the employer-employee relationship. In Proceedings of BLED Conference. BLED, Bled.

20. Ahmad, M., &Lodhi, S. (2015). Impact of Social Media on Job Search a case of Karachi. International Journal of Scientific and Research Publications, 628.

21. Stoughton, J. W., Thompson, L. F., & Meade, A. W. (2013). Big five personality traits reflected in job applicants' social media postings. Cyberpsychology, Behavior, and Social Networking, 16(11), 800-805.

22. http://cran.rproject.org

23. Dey, L., Haque, S. M., Khurdiya, A., & Shroff, G. (2011, September). Acquiring competitive intelligence from social media. In Proceedings of the 2011 joint workshop on multilingual OCR and analytics for noisy unstructured text data (p. 3). ACM.

24. Smith, A. (2015). Searching for work in the digital era.