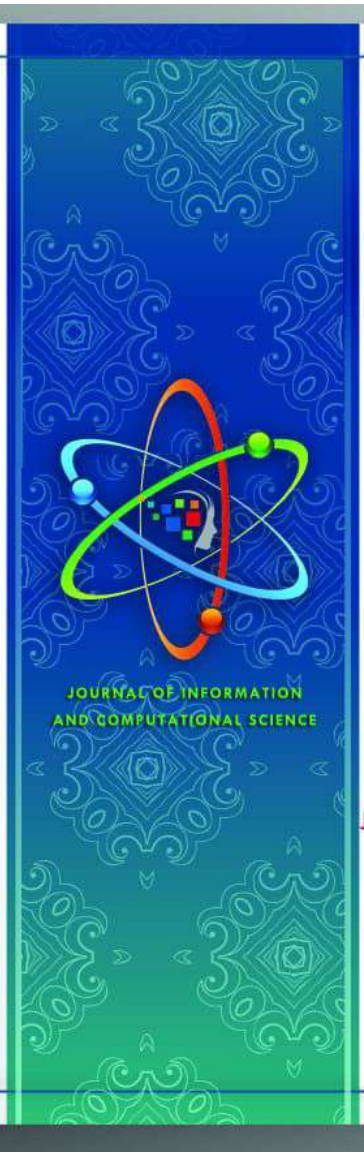


Journal of Information and Computational Science

UGC - Care Group - II Certified Journal

ISSN NO: 1548-7741 / web : www.joics.org / E-mail : submitjoics@gmail.com

Certificate of Publication

This is to certify that the paper entitled

Detecting Safest Cities for Women in India

Authored by :

DURGA LAVANYA BOLLA

From

SVKP & DR K S RAJU ARTS & SCIENCE COLLEGE(A), PENUGONDA, 666.W.G.DT., A.P, INDIA

Has been published in

JOURNAL OF INFORMATION AND COMPUTATIONAL SCIENCE,VOLUME 12 ISSUE 9 ,SEPTEMBER -2022

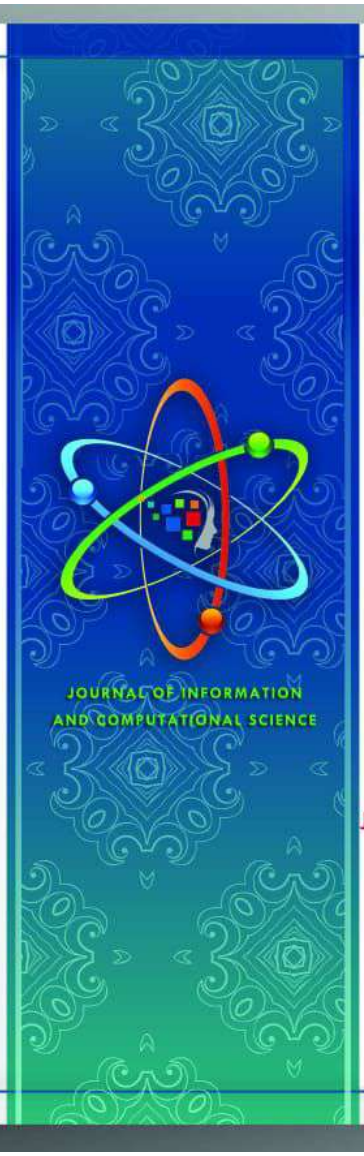
ज्ञान-विज्ञान विमुक्तये
An Emblem Symbolising The Future

S. Joseph

Joseph Sung
Editor-In-Chief
JOICS

6.2 IMPACT FACTOR

ISO International Organization for Standardization 7021-2008



Journal of Information and Computational Science

UGC - Care Group - II Certified Journal

ISSN NO: 1548-7741 / web : www.joics.org / E-mail : submitjoics@gmail.com

Certificate of Publication

This is to certify that the paper entitled

Detecting Safest Cities for Women in India

Authored by :

CH. SRINIVASA RAO, Associate Professor

From

SVKP & DR K S RAJU ARTS & SCIENCE COLLEGE(A), PENUGONDA, .W.G.DT., A.P, INDIA

Has been published in

JOURNAL OF INFORMATION AND COMPUTATIONAL SCIENCE,VOLUME 12 ISSUE 9 ,SEPTEMBER -2022

ज्ञान-विज्ञान विमुक्तये
An Emblem Symbolising The Future

S. Joseph

Joseph Sung
Editor-In-Chief
JOICS

6.2 IMPACT FACTOR

ISO International Organization for Standardization 7021-2008

Detecting Safest Cities for Women in India

DURGA LAVANYA BOLLA
PG Scholar, Department of Computer Science,
SVKP & Dr K S Raju Arts & Science College(A),
Penugonda, W.G.Dt., A.P, India
EMAIL-ID : lavanyasrinivasbolla@gmail.com

CH. SRINIVASA RAO*
Associate Professor in Computer Science,
SVKP &Dr K S Raju Arts & Science College(A),
Penugonda, W.G.Dt., A.P, India
EMAIL-ID : chiraparapu@gmail.com

Abstract—
Women and women are experiencing tons of violence and harassment publicly places in numerous cities ranging from stalking and resulting in abuse harassment or abuse assault. This analysis paper essentially focuses on the role of social media in promoting the protection of girls in Indian cities with special relevancy the role of social media websites and applications as well as Twitter platform Facebook and Instagram. This paper conjointly focuses on however a way of responsibility on a part of Indian society is developed the common Indian folks so we should always specialize in the protection of girls close them. Tweets on Twitter that sometimes contains pictures and text and conjointly written messages and quotes that specialize in the protection of girls in Indian cities is accustomed scan a message amongst the Indian subculture and educate folks to require strict action and penalize those that harass the ladies. Twitter and different Twitter handles that embrace hash tag messages that area unit wide unfold across the complete globe sir as a platform girls forladies for girls} to specific their views regarding however they feel whereas we have a tendency to depart for work or travel in an exceedingly transport and what's the state of their mind once they area unit enclosed by unknown men and whether or not these women feel safe or not?

Index Terms: **Social Medica,Tweet,N-Gram,Preprocess.**

I. Introduction

There are certain types of harassment and Violence that are very aggressive including staring and passing comments and these unacceptable practices are usually seen as a normal part of the urban life. There have been several studies that have been conducted in cities across India and women report similar type of sexual harassment and passing off comments by other unknown people. The study that was conducted across most popular Metropolitan cities of India including Delhi, Mumbai and Pune, it was shown that 60 % of the women feel unsafe while going out to work or while travelling in public transport. Women have the right to the city which means that they can go freely whenever they want whether it be too an Educational Institute, or any other place women want to go. But women feel that they are unsafe in places like malls, shopping malls on their way to their job location because of the several unknown Eyes body shaming and harassing these women point Safety or lack of concrete consequences in the life of women is the main reason of harassment of girls. There are instances when the harassment of girls was done by their neighbors while they were on the way to school or there was a lack of safety that created a sense of fear in the minds of small girls who throughout their lifetime suffer due to that one instance that happened in their

lives where they were forced to do something unacceptable or was sexually harassed by oneof their own neighbor or any other unknown person.

Safest cities approach women safety from a perspective of women rights to the affect the city without fear of violence or sexual harassment. Rather than imposing restrictions on women that society usually imposes it is the duty of society to imprecise the need of protection of women and also recognizes that women and girls also have a right same as men have to be safe in the City.Analysis of twitter texts collection also includes the name of people and name of women who stand up against sexual harassment and unethical behavior of men in Indian cities which make them uncomfortable to walk freely. The data set that was obtained through Twitter about the status of women safety in Indian society was for the processed through machine learning algorithms for the purpose of smoothening the data by removing zero values and using Laplace and porter's theory is to developer method of analyzation of data and remove retweet and redundant data from the data set that is obtained so that a clear and original view of safety status of women in Indian society is obtained.

II. LITERATURE SURVEY

People often express their views freely on social media about what they feel about the Indian society and the politicians that claim that Indian cities are safe for women . On social media websites people can freely Express their view point and women can share their experiences where they have faced sexual harassment or where we would have fight back against the sexual harassment that was imposed on the . The tweets about safety of women and stories of standing up against sexual harassment further motivates other women data on the same social media website or application like Twitter. Other women share these messages and tweets which further motivates other 5 men or 10 women to stand up and raise a voice against people who have made Indian cities and unsafe place for women. In the recent years a large number of people have been attracted towards social media platforms like Facebook, Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society. There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning. Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches. It is a common practice to extract the information from the data that is available on social networking through procedures of data extraction, data analysis and data interpretation methods. The accuracy of the Twitter analysis and prediction can be obtained by the use of behavioral analysis on the basis of social networks. Opinion mining involves several important tasks, including sentiment polarity and intensity assignment. Polarity assignment is concerned with determining whether a text has a positive, negative, or neutral semantic orientation. Sentiment intensity assignment looks at whether the positive/negative sentiments are mild or strong. Given the two phrases "I don't like you" and "I hate you," both would be assigned a negative semantic orientation but the latter would be considered more intense. Effectively classifying sentiment polarities and intensities entails the use of classification methods applied to linguistic features. While several classification methods have been employed for opinion mining, Support Vector Machine (SVM) has outperformed various techniques including Naive Bayes, Decision Trees, Winnow, etc. The most popular class of features used for opinion mining is n-grams . Various n-gram categories have attained state-of-the-art

results. Larger n-gram feature sets require the use of feature selection methods to extract appropriate attribute subsets. Next, we discuss these two areas: n-gram features and feature selection techniques used for Author profiling.

N-GRAM FEATURES FOR AUTHOR PROFILING

N-gram features can be classified into two categories: fixed and variable. Fixed n-grams are exact sequences occurring at either the character or token level. Variable n-grams are extraction patterns capable of representing more sophisticated linguistic phenomena. A plethora of fixed and variable n-grams have been used for opinion mining, including word, part-of-speech (POS), character, legomena, syntactic, and semantic n-grams. Word n-grams include bag-of-words (BOWs) and higher order word n-grams (e.g., bigrams, trigrams). Word n-grams have been used effectively in several studies .Typically, unigrams to trigrams are used, though 4-grams have also been employed. Word n-grams often provide a feature set foundation, with additional feature categories added to them. Given the pervasiveness of adjectives and adverbs in opinion-rich text, POS tag, n-grams are very useful for sentiment classification. Additionally, some studies have employed word plus part-of-speech (POSWord) n-grams. These n-grams consider a word along with its POS tag in order to overcome word-sense disambiguation in situations where a word may otherwise have several sensesfor example, the phrase "quality of the" can be represented with the POSWord trigram "quality-noun of prep the-det." Character n-grams are letter sequences. For example, the word "like" can be represented with the following two and three letter sequences "li, ik, ke, lik, ike." While character n-grams were previously used mostly for style classification, they have recently been shown to be useful in related affect classification research attempting to identify emotions in tex. Legomena n-grams are collocations that replace once (hapax legomena) and twice occurring words (dis legomena) with "HAPAX" and "DIS" tags.

FEATURE SELECTION FOR AUTHOR PROFILING

Prior sentiment classification studies have placed limited emphasis on feature selection techniques, despite their benefits. Feature selection can potentially improve classification accuracy, narrow in on a key feature subset of sentiment discriminators, and provide greater insight into important class attributes. There are two categories of feature selection methods, both of which have been used in prior Author profiling work: univariate and multivariate.

Univariate methods consider attributes individually. Examples include information gain, chi-squared, log likelihood, and occurrence frequency . Although univariate methods are computationally efficient, evaluating individual attributes can also be disadvantageous since important attribute interactions are not considered. It is also easier to interpret the contribution of individual attributes using univariate methods. Most opinion mining studies have used univariate feature selection methods such as minimum frequency thresholds and the log-likelihood ratio]. Information gain (IGhas also been shown to work well for various text categorization tasks, including Author profilingTsutsui et al. used the Chi-Squared test to select features for text sentiment classification.

OTHER FEATURE SELECTION METHODS

In addition to prior sentiment feature selection methods, it is important to briefly discuss multivariate and hybrid methods used in related tasks. Principal component analysis

(PCA) has been used considerably for dimensionalityreduction in text style classification problems. Recently, many powerful dimensionality reduction techniques have also been applied to nontext feature selection problems.

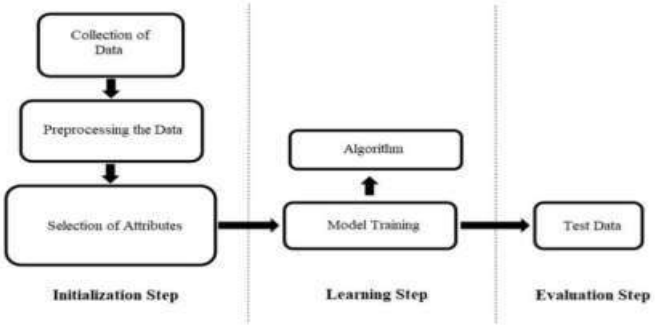
III. PROBLEM STATEMENT

We present a classifier to predict contextual polarity of subjective phrases in a sentence. Our approach features lexical scoring derived from the Dictionary of Affect in Language (DAL) and extended through WordNet, allowing us to automatically score the vast majority of words in our input avoiding the need for manual labeling. We augment lexical scoring with n-gram analysis to capture the effect of context. We combine DAL scores with syntactic constituents and then extract n-grams of constituents from all sentences. We also use the polarity of all syntactic constituents within the sentence as features. Our results show significant improvement over a majority class baseline as well as a more difficult baseline consisting of lexical n-grams.

Disadvantages:

- Twitter and Instagram point and most of the people are using it to express their emotions and also their opinions about what they think about the Indian cities and Indian society.
- There are several method of sentiment that can be categorized like machine learning hybrid and lexicon-based learning.
- Also there are another categorization Janta presented with categories of statistical, knowledge-based and age wise differentiation approaches.

IV. SYSTEM ARCHITECTURE



MODULES

Tweet Server:

Collect all from register users. Clean tweets by removing the stop words and noise like repetitive letters. Tokenize each word and allot strength to the words in the dataset and feed it to the program.For each word, compare it with positive sentiments and negative sentiments word dictionary and then increment positive count or negative count of the overall phrase. Finally, based on the positive count & negative count, we can get result percentage about sentiment to decide the polarity which is categorized in Positive, Negative and Neutral.

User

1. Create new user and login
2. View all tweet details and Likes, Dislikes and Make reviews
3. View all tweets reviews like user name, tweet name, review, sentiment analysis, review date and suggestion

V. PROCESS MODEL

Data Pre-Processing

The data were pre-processed in the first phase of the detection model. Pre-processing includes the transformation and normalization of data. To clean the dataset, URLs, retweets, mentions, and stop-words were removed from the dataset. Each row of the dataset wasthen tokenized by breaking the text into tokens or words. After that, the tokenized wordswere subjected to stemming and lemmatization. The stemmed input text was processed through the One-Hot procedure to extract features from these input words. The featureswere binary patterns that could be employed in the machine learning prediction modelto predict depression (1 represents depressed, and 0 represents other words). The list ofdepressive sentences and words is represented, extracted from the dataset.Represents some depressive tweets taken from the dataset.

Features Visualization Using Principal Component Analysis.

We embraced a dimensionality reduction technique called PCA to visualize features.PCA performs decomposition using the covariance matrix to generate eigenvaluesto minimize the inner scattering of samples and maximize inter-class scattering.

Modeling LSTM-RNN for Emotional State Analysis

While conversing with others, sentiments can be expressed as time-sequential sentences in textual data. Machine-learning models are pretty capable of encoding time sequential data. We implemented RNN because it is the most optimal method for working with sequential data. RNN comprises recurrent connections with past and present states and hidden states. Memory plays an essential function in neural networks and sometimes faces gradient disappearing problems. LSTM is the solution to memory processing-related problems.

VI. RESULTS

Home Page:

Login :

Tweet Server :

View All Tweets :

VI. CONCLUSION

In this paper, we tried to identify the presence of depression in Reddit social media; and searched for affective performance increase solutions of depression detection. We characterized a closer connection between depression and a language usage by applying NLP and text classification techniques. We identified a lexicon of words more common among the depressed accounts. According to our findings, the language predictors of depression contained the words related to preoccupation with themselves, feelings of sadness, anxiety, anger, hostility or suicidal thoughts, with a greater emphasis on the present and future.

VII.REFRENCES:

[1] Agarwal, Apoorv, Fadi Biadsy, and Kathleen R. McKeown. "Contextual phrase level polarity analysis using lexical affect scoring and syntactic n-grams." Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics. Association for Computational Linguistics, 2009.

[2] Barbosa, uciano, and Junlan Feng. "Robust sentiment detection on twitter from biased and noisy data." Proceedings of the 23rd international conference on computational linguistics: posters. Association for Computational Linguistics, 2010.

[3] Bermingham, Adam, and Alan F. Smeaton. "Classifyingnn sentiment in microblogs: is brevity an advantage?." Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010.

[4] Gamon, Michael. "Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[5] Kim, Soo-Min, and EduHovy. "Determining the sentiment of opinions." Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics, 2004.

[6] Klein, Dan, and Christopher D. Manning. "Accurate unlexicalized parsing." Proceedings of the 41st Annual Meeting on Association for Computational Linguistics-Volume 1. Association for Computational Linguistics, 2003..

[7] Charniak, Eugene, and Mark Johnson. "Coarse-to-fine n-best parsing and MaxEnt discriminative reranking." Proceedings of the 43rd annual meeting on association for computational linguistics. Association for Computational Linguistics, 2005.

[8] Gupta, B., Negi, M., Vishwakarma, K., Rawat, G., &Badhani, P. (2017). Study of Twitter sentiment analysis using machine learning algorithms on Python. International Journal of Computer Applications, 165(9), 0975-8887.

[9] Sahayak, V., Shete, V., & Pathan, A. (2015). Sentiment analysis on twitter data. International Journal of Innovative Research in Advanced Engineering (IJIRAE), 2(1), 178-183.

[10] Mamgain, N., Mehta, E., Mittal, A., & Bhatt, G. (2016, March). Sentiment analysis of top colleges in India using Twitter data. In Computational Techniques in Information and Communication Technologies

(ICCTICT), 2016 International Conference on (pp. 525-530). IEEE.

ABOUT AUTHORS:

DURGA LAVANYA BOLLAis currently pursuing MCA in SVKP & Dr K S Raju Arts & Science College, affiliated to Adikavi Nannaya University, Rajamahendravaram. Her research interests include Machine Learning, Data science.

Ch.Srinivasa Rao is a Research Scholar in the Department of Computer Science & Engineering at Acharya Nagarjuna University, Guntur, A.P, India. He is working as Associate Professor in SVKP & Dr K S Raju Arts&Science College, Penugonda, A.P. He received Master's degree in Computer Applications from Andhra University and Computer Science & Engineering from Jawaharlal Nehru Technological University, Kakinada, India. He Qualified in UGC NET and AP SET. His research interests include Data Mining and Data Science.& Machine Learning.