

Social Communications using Big Data Applications

S.Prakashkumar, I.S.Beschi

Abstract— *Social computing is interactive and collaborative behavior between computer users. Personal computing is an individual user activity in that one user generally commands computing. In social computing, the Internet allows users to interact through many mediums, including: Social computing is basically the use of a computer for social purposes. Before the Internet, computers were largely used as tools for increasing productivity. The Internet introduced a social element where users could network, share interests, publish personal insights and use their computers for more than just doing a job faster. Social computing can still benefit businesses synergistically if it is used for business purposes. Social computing can be used to market products and promote customer relations. Online marketing and viral marketing are two types of promotional advertising that have grown out of social computing.*

Keywords: *Big data applications, Data mining, Data visualization, Machine learning algorithms, Natural language processing, Sentiment Analysis, Social computing, Social Intelligence.*

I. INTRODUCTION

Social computing can still profit businesses synergistically if it's used for business functions. Social computing area unit typically accustomed market merchandise and promote shopper relations. on-line marketing and infectious agent marketing unit of measurement a pair of sorts of promotional advertising that have massive out of social computing. Social Computing is one in all the foremost necessary marketing platforms for any business at intervals the globe. With faster conversion at intervals the banking sector of Asian nation, the Social Computing has become one in all the come well-liked mechanism for banks to push their merchandise, offers and services to the purchasers. for purchasers the Social Computing has become an easy and quick due to share their feedback and comments from anywhere at intervals the globe. Although the mechanism to interconnect between banking sector and shopper has become easy, to extract the sentiment of the purchasers and their perspective towards the merchandise continues to be a challenge to the banking businesses. usually at intervals the business, the offerings and responses to the competition organization is extraordinarily crucial for added planning. This thinking paper proposes a system that features a Social Computing based approach towards understanding the baking

Revised Manuscript Received on July 10, 2019.

S.Prakashkumar, Assistant Professor/Department of Computer Science Marudupandiyar College(Arts & Science), Trichy main Road, Pillayaratti, Vallam Post, Thanjavur

I.S.Beschi, Research Scholar/Department of Computer Science Marudupandiyar College(Arts & Science), Trichy main Road, Pillayaratti, Vallam Post, Thanjavur

market from customers browse purpose and improve the bank's complete, vogue future marketing ways that and campaigns format.

- Abbreviations and Acronyms
API – Application Programming Interface
DB – Database
NLP – Natural Language Processing

II. RELATED WORK

Strategic use of social media data not exclusively impacts the approach throughout that the money institutions market their product and services, but jointly on but they conduct competitive analysis for product and repair vogue. Banks have established their presence on social media like Face book, Social Computing r, and LinkedIn. Social computing may well be a large social media that allows micro blogging through tweets that unit of measurement public. Every word, photo, video and follower can have an effect. there is vast volume and kind of data on Social Computing which can be analyzed mistreatment the huge data approaches fitted to cash sector.[1,2].

In addition, Social Computing API's unit of measurement offered for publically practitioners and researchers which can aid in data analysis of Social Computing data. [1]. Each account on Social Computing is said to a singular id and a singular Social Computing handle which can be accustomed retrieve the profile and tweets for data analysis.

Social Computing data analytics was researched among the past on a ramification of domain like exchange [4], provide Chain [5], welcome [6] etc. Tweets extraction was in strait a span variable from 3 to 12 months by altogether completely different authors. Tweet is associate unstructured data, that has to be filtered by mistreatment varied tongue method techniques like stemming, stop word and noun removal [7] to urge the useful data.

The most important a part of this analysis lies in mining {the data the knowledge the data} for extracting information for gaining deep insights into the tweets for consumer behavior, feedback on product and complaints.

Sentiment analysis will establish the emotions of the arments designer to be positive, negative or neutral. Most of the literature uses lexicon based totally approach, but this desires an honest and powerful reference that may not invariably offered. linguistics of the text jointly plays a major role once performing sentiment analysis, that's usually unnoticed.

Researchers have found that accumulated accuracy is achieved if linguistics unit of measurement incorporated. [8]. A bunch of study area unit done on prediction of the exchange data supported the sentiment analysis and various algorithms like - SOFNN (Self Organizing Fuzzy Neural Networks) [10].

A variety of information science algorithms unit of measurement used for agglomeration like K-mean, hash tagging, TF-IDF [11] then context analysis have to be compelled to be applied on the tweets for understanding the correlation, aggregation and association kind the tweets. Recent focus is on the topic Spatio-temporal agglomeration of social media data [12].

Units

This Section presents the various modules of our system. Overall, the system has 5 main method elements (refer figure 1) via – knowledge extraction, tweets pre-processing, data processing, knowledge benchmarking and knowledge image as represented in figure one.



Fig. 1: System Components



Fig. 2: The process follow is depicted

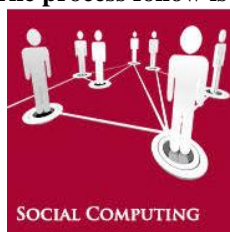


Fig 3: Process Flow

3.1. Input data :

The supply for input file is Social Computing to make sure numerous coverage, Tweets for the four banks – 2personal –ICICI, HDFC and 2 nationalized banks – SBI, PNB ar thought-about in scope for the analysis. to accumulatea enough volume of knowledge tweets for a year are going to be into account. any to uncover the association trends within the product bought by customers, multiple Social Computinghandles for a similar banking establishments are going to be used for extracting the information.

3.2. Data Acquisition and Filtering

This module extracts the domain specific information from Social Computing for the four banks in scope, victimisation the Social computing API – Tweepy. The domain specific tweets area unit store in an exceedinglyNoSQL info – MongoDB for additional process. NoSQL info was

sensible|an honest|a decent} option to store the unstructured tweets and conjointly features a good vary of aggregation functions for querying unstructured information. 3. 3. Tweets Pre-processing and Extraction

This tweets dataset composed for brief messages (tweets) need a series of custom preprocessing to extract the keyword For quality analysis.

As a primary step, all correct nouns area unit removed, this make sure that person and place names don't seem to be inferred as keywords. in contrast to this analysis wherever correct nouns haven't any weightage, there might become situations wherever correct nouns area unit necessary keywords like for election campaign analysis person name has got to be classified as per the organization, then the right noun elimination ought to be skipped. morewithin the tokenizing step the computer address and special characters area unit discarded.

Tweets area unit normalized by correcting common English words used on social media like plisses is replaced with please, between with between. This method uses wordbook for common social media representations, Stop words like is, an, and, that etc that don't add price to the domain specific words area unit removed. The stop words reference wordbook is customizable. Next stage words area unit stemmed, works like “complain” and “complaint” have a similar root. thus within the next part of filtering the foundation is preserved and remaining words area unit discarded. Length primarily based filtering is finished on the domain specific words. Words with length but three and quite fifteen area unit removed.

3.4. Knowledge extraction using data mining :

Until this step, the input file is mere individual text entities, with none categorization, association and no abstract thought drawn from the info. information extrctation module will the subsequent victimization the tongue processalgorithms:

cluster the topics by victimization the Hashtags utilized in the tweets, replies and retweets, that aids in understanding the wide mentioned topic by the purchasers and may be a part of focus.

Sentiment analysis to see customer's perception regarding the product that is be a significant input for the long run strategy style. Sentiment analysis is well-known and therefore the wide used technique, that helps in distinctive public sentiment from underlying text. A lexicon based mostly rule is employed, that is proved to the most effective accuracy, with support of a well-constructed wordbook.

K-means cluster rule is any wont to categorise the info supported the product that area unit into account. Association within the product are often uncovered victimization the Mongolian monetary unit sound unit association and aggregation techniques. The association input are of nice importance to deciding the offers on the product.

3.5. Data Benchmarking:

Social media analytics of your account against the business peers' social media presence would reveal key metrics on positioning your merchandise.

3.6. Data Visualization:

In addition to the higher than steps of process and data processing, it's equally necessary to presents the insights drawn through numerous visual techniques in a very perceivable manner. Usage of acceptable visualisation techniques aid in simple understanding and thus will increase the likelihood of usage of this metrics into actions.

III. TOOLS AND TECHNOLOGY TO BE USED

The following open supply languages, tools and libraries would be used for implementing the projected system:

4.1 Python :

Python is Associate in Nursing ASCII text file and object-oriented artificial language.. because it is open supply, there's Associate in Nursing handiness of the many libraries and genus Apis to perform one perform. Extraction of Social Computing APIs is well expedited by the usage of Python. It are often utilised for a good vary of applications like scripting, developing and testing. it's majorly most popular over the opposite scripting languages, as a result of its magnificence and ease.

4.2 Social Computing API (Tweepy&TextBlob) :

Tweepy may be a library in Python, for accessing the Social Computing genus Apis. These genus Apis square measure accustomed extract and transfer the messages and tweets in real time. High volume of tweets and making a live feed is expedited solely mistreatment this library.

TextBlob is another library provided by Python, that is employed to hold out process on matter knowledge, extracted through the Tweepy library. It provides a straightforward API to hold out natural language processing tasks. It will with efficiency perform natural language processing tasks like phrase extraction, translation and analysis.

4.3 MongoDB

Mongo dB is Associate in Nursing ASCII text file, platform freelance document-oriented information . It are often simply paired with Python mistreatment the official connectors provided by MongoDB. MongoDB is especially used with Python as a result of the support in field, vary question, and regular expression searches. Besides simply a information program, MongoDB can even be used as a filing system.

4.4 PyQT

PyQt is Associate in Nursing open supply plugin in Python to bind Python and interface toolkit Qt for visual illustration of the findings.

IV. CONCLUSION

With exponential increase in social media usage in recent years by numerous banks primarily to promote their product. This analysis work can exploit social media knowledge

analytics mistreatment knowledge to providesome bespoke metrics, which might be inputs to the bank for style of latest product, editing the social media strategy etc. the info analysis would cowl multiple aspects – sentimental analysis, domain central approach and customised image. additionally resolving some descriptive context analysis with the clustered tweet knowledge, can aid in understanding the precise client perspective on the merchandise at the side of the feeling.

REFERENCES

1. Shamanth Kumar, Fred Morstatter, HuanLiu, "Social Computing Data Analytics" in Springer, Aug 2013, Pages – 1-5
2. Jennifer. Q. Trelewicz, Big Data and Big Money: The Role of Data in the Financial Sector , Accession Number: 16948585 , DOI: 10.1109/MITP.2017.45, Publisher: IEEE
3. Anber, H., A. Salah, and A. A. Abd El-Aziz, "A Literature Review on Social Computing Data Analysis", International Journal of Computer and Electrical Engineering, vol 8(3), 2016: Journal of Computers, 2016..
4. Spin-offs in Indian Stock Market owing to Social Computing Sentiments, Commodity Prices and Analyst Recommendations
5. Voice of Customers: Text Analysis of Hotel Customer Reviews (Cleanliness, Overall Environment & Value for Money)
6. Bongsug (Kevin)Chae, Insights from hashtag #supplychain and Social Computing Analytics: Considering Social Computing and social computing data for supply chain practice and research, International Journal of Production Economics Volume 165, July 2015.
7. Anber, H., A. Salah, and A. A. Abd El-Aziz, "A Literature Review on social computing Data Analysis", International Journal of Computer and Electrical Engineering, vol 8(3), 2016: Journal of Computers, 2016.
8. Crockett, KA and Mclean, D and Latham, A and Alnajran, N (2017) Cluster Analysis of Social computing Data: A Review of Algorithms. In: 9th International Conference on Agents and Artificial Intelligence (ICAART), 24 February 2017 - 26 February 2017, Portugal.
9. Anshul Mittal , ArpitGoel , "Stock Prediction Using Social computing Sentiment Analysis", Stanford University.
10. Zhichao Han, DATA AND TEXT MINING OF FINANCIAL MARKETS USING NEWS AND SOCIAL MEDIA, A Dissertation submitted to the University of Manchester
11. M. Budde, J. De Melo Borges, S. Tomov, T. Riedel, and M. Beigl. Leveraging spatio-temporal clustering for participatory urban infrastructure monitoring. In Proceedings of the First International Conference on IoT in Urban Space, pages 32{37. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 2014.