



## A REVIEW ON SOCIAL MEDIA ANALYSIS: CHALLENGES AND APPLICATION

Shachi Shukla  
PG Scholar, Dept. of CSE  
RITS, Bhopal, INDIA

Seema Narvare  
Asst. Prof., Dept. of CSE  
RITS, Bhopal, INDIA

Chetan Agrawal  
HOD, Dept. of CSE  
RITS, Bhopal, INDIA

**Abstract:** Social networks (SNs), such as Facebook, Twitter and WeChat have emerged and tightly connected web clients all over the world. By analyzing and mining social networks, we could assemble information on the comments made by people with respect to a particular product. Analysis of such comments shows its value for the design of marketing and advertising campaigns. The typical examples are viral marketing, influential bloggers finding, social advertising, social healthcare, expert finding, personalized commendation, citation networks, and so on. Social media includes interactive applications and proposals for creation, sharing and replacing client-produced matters. The earlier period of few years have brought vast escalation in social media, particularly social networking services, and it's varying our systems to systematize and correspond. It aggregates judgments and sentiments of various clusters of group at low price. Mining the characteristics and matters of social media provides us a prospect to find out social organization attributes, to evaluate action prototypes qualitatively and quantitatively, and rarely calculate upcoming human correlated occasions. In this research paper, we firstly review the areas which could be calculated with existing social media platforms, then draw general idea about obtainable predictors and systems of prediction, and to end with discuss problems and probable potential research scopes.

**Keywords:** Social media, Sentiment analysis, Social network, prediction, influence analysis, Text mining

### 1. INTRODUCTION

A huge organization of conventional time-series models of calculating potential auctions of a product or service have relied on the earlier period chronological and seasonal auctions and often provided untrustworthy calculation conclusions. Significantly, since these analytical models depend on only the earlier period information of auctions, they lean to overlook the dynamic collision of modern events that might have a significant persuade on auctions. Additionally, even though clients' existing judgment and opinion about companies and products (e.g., qualified evaluates of products, individual biased private estimations) would influence buying actions and potential auctions by words-of-mouth causes, the conventional models don't have any source of inputs to judge these sorts of socio-emotional issues.

Social networking (e.g., news gatherings, item audit locales, web journals, twitter and facebook) may fill in as an intermediary of individuals' assessments, for example, their prior period encounters and current assumptions about items or administrations. Consequently, joining new key elements from the investigation of related social networking matter could enable organizations to add another layer to their current prescient models and lift the expectation exactness.

The mix of social networking expectations with existing systematic estimating models is probably going to be superior to both of the two in confinement because of the way that each model spotlights on only one part of the earth which is basic to deciding potential results. By and large, social factors alone are available a deficient photo of the present situation and are in this manner not sufficiently intense to anticipate correct request. Be that as it may, without social networking input, current request figures can't consider troublesome get-togethers and huger societal

patterns that will without a doubt impact request too. Along these edges, this mix is probably going to demonstrate advantageous in any space where customer conduct is a vital segment what is to be anticipated. What's more, since existing prescient models of interest are extremely mind boggling and hard to alter, it's desirable over make changes in view of social networking without using specifics of the first request demonstrate detailing.

Social networking is a platform that enables normal people to make and distribute matters. Two overall well known social networking sites, Twitter and Facebook, show its hazardous development and significant impact. Both Twitter and Facebook are in the best 15 most-visited sites on the planet as per Alexa positioning [1]. Facebook has in excess of 1400 million dynamic customers [2] by 31 December 2017, and by December 2017, on Twitter, there were around 140 million data pieces made and exchanged day by day [3]. There is other specific social networking that is centered on stimulation, games, fund and legislative issues.

Since there are numerous customers imparting their insights and encounters by means of social networking, there is accumulation of individual intelligence and distinctive perspectives. Such collection has confinements as perspectives are liable to change with time. It might be said the social networking predict issue is paralleled by expectation of budgetary time arrangement in light of prior period history, which has its uses in exchanging. By and large, if extricated and examined appropriately, the information via social networking media could prompt helpful predicts of certain human related occasions. Such expectation has extraordinary advantages in numerous domains, for example, fund, item advertising and governmental issues, which have pulled in expanding number of specialists to this subject. Investigation of social networking additionally gives bits of knowledge on social

elements and general wellbeing. A review gives us point of view and is useful for completing further research.

Rest of the paper is organized as following: in section 2 we briefly describes about Social Networks, in section 3 overview of social media search and its application is described, different social networking services are presented in section 4, in section 5 offline and real time social media analytics is explained, how prediction is done is explained in section 6, section 7 gives idea about the major metrics about social media used in prediction, key techniques to analyze textual data is given in section 8, research challenges and scopes are studies in section 9 and 10, finally we summarized our paper with facts and findings in section 11.

## 2. SOCIAL NETWORK

A social media is a social structure including people or associations, which more often than not are spoken to as hubs, together with social relations, which compare to the connections among hubs. The social connection could express both, for example, family relationship and cohorts, and understood, for instance fellowship and normal intrigue. For example, fig. 1 is a case of undirected informal community in an organization, from open source programming GUESS [4]. In fig. 1, every hub speaks to a worker. The edge between two hubs implies these two representatives have a few interchanges in work and the heaviness of each edge is the correspondence recurrence.

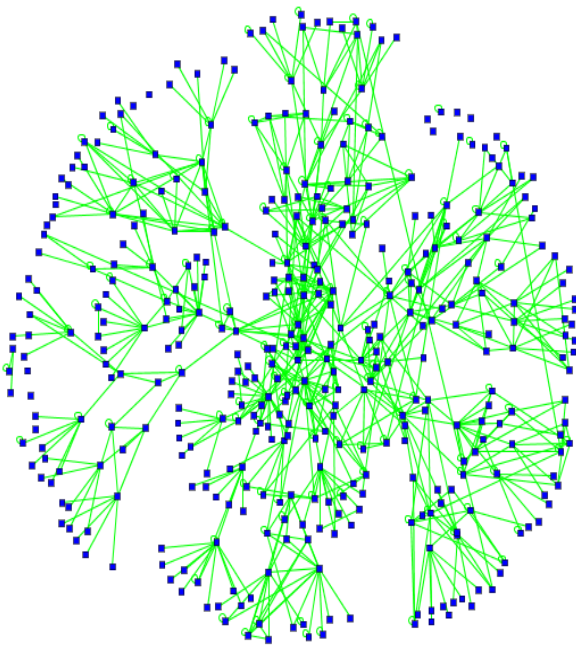


Fig. 1. Social Network Example from GUESS

A little social media might be demonstrated by general charts, for example, that of a little world system [5]. For a colossal very much associated arrange, most hubs could achieve each other hub through few connections. The possibility of six level of detachment proposes that, by and large, every two people are connected by six bounces [6]. The circumstance in Social Networking Service (SNS) isn't vastly different. The normal separation on Facebook in 2008 was 5.28 jumps, while in November 2011 it's 4.74. In the MSN delegate organizes, which contains 180 million customers, the middle and the 90th percent level of partition are 6 and 7.8 separately. On Twitter, the middle, normal, and 90th percent remove between any two customers are 4, 4.12

and 4.8, separately [7]. To some things up, the level of partition shifts on various SNS stages as well as on various time yet it's very little. An informal community is a scale free system [8] for which the degree dissemination asymptotically takes after a power law. On Twitter, up to 105 of the quantity of followings/adherents fit the power-law dissemination with the type of 2.276. The quantity of being re-tweeted and specified by customers on Twitter additionally takes after a power law [9].

## 3. SOCIAL MEDIA

Social networking contains stages to make and trade customer created matter [10]. Some of the time social networking is called shopper delivered media (CGM). Social networkings are not the same as customary media, for example, daily paper, books, and TV, in that nearly anybody could distribute and get to data cheaply utilizing social networking. Conversely, conventional media (which is additionally eluded as old media or inheritance media) requires huge assets to distribute matters. In any case, social networking and conventional media are not totally particular. For instance, significant news channels have official records on Twitter and Facebook.

There are numerous types of social networking that incorporate web journals, person to person communication locales, virtual social universes, communitarian ventures, matter groups and virtual diversion universes [11]. A few types of social networking do not have an informal organization. In this manner in blogspot.com, which is a celebrated blog stage, there are no social connections among bloggers.

Social networking has a few or these seven capacity squares: character, discussions, sharing, nearness, connections, notoriety, and gatherings [12]. Distinctive types of social networking have diverse purposes of core interest. For instance, cooperative ventures, for example, Wikipedia generally think about sharing and notoriety, And in virtual amusement universes, character, nearness, notoriety, and gatherings are of the best concern.

As of late, social networking assumed critical part in unfurling newsworthy occasions. For instance, in the outcome of the Tohoku Earthquake in Japan individuals utilized social networking to contact companions, trade emergency data, and discover vital assets.

### 3.1 Social media research and applications

Social networking information is obviously the hugest, wealthiest and most unique confirmation base of human conduct, bringing new chances to comprehend people, gatherings and society. Creative researchers and industry experts are progressively discovering novel methods for consequently gathering, consolidating and breaking down this abundance of information. Normally, doing equity to these spearheading social networking applications in a couple of passages is testing. Three illustrative regions are: business, bioscience and sociology.

The early business adopters of social networking examination were regularly organizations in retail and back. Retail organizations utilize social networking to outfit their image mindfulness, item/customer benefit change, publicizing/promoting systems, arrange structure investigation, news spread and even misrepresentation location. In back, social networking is utilized for estimating market slant and news information is utilized for exchanging. As an outline, [31] estimated conclusion of arbitrary example of Twitter information, finding that Dow Jones Industrial

Average (DJIA) costs are related with the Twitter opinion 2–3 days sooner with 87.6 percent precision. They utilized Twitter information to prepare a Support Vector Regression (SVR) model to anticipate costs of individual NASDAQ stocks, finding 'noteworthy preferred standpoint' at determining costs min in the potential.

In the biosciences, social networking is being utilized to gather information on enormous accomplices for behavioral change activities and effect checking, for example, handling smoking and corpulence or observing maladies. An illustration is Penn State University scholars [32] who have created imaginative frameworks and methods to track the spread of irresistible illnesses, with the assistance of news Web locales, web journals and social networking.

Computational sociology applications include: observing open reactions to declarations, addresses and occasions particularly political remarks and activities; experiences into group conduct; social networking surveying of (difficult to contact) gatherings; early location of developing occasions, as with Twitter. For instance, utilize computational etymology to naturally anticipate the effect of news on people in general impression of political competitors. utilize film audit remarks to contemplate the impact of different methodologies in removing content highlights on the precision of four machine learning techniques—Naive Bayes, Decision Trees, Maximum Entropy and K-Means bunching. Finally found that Facebook's Gross National Happiness (GNH) displays pinnacles and troughs in accordance with significant open occasions in the USA.

### 3.2 Social media overview

For this research paper, we amass online networking apparatuses into Social media information social networking information writes (e.g. informal organization media, wikis, websites, RSS channels and news, and so on.) and groups (e.g., XML and JSON). This incorporates informational indexes and progressively imperative ongoing information sustains, for example, monetary information, customer exchange information, telecoms and spatial information.

- Social media programmatic access: information administrations and devices for sourcing and scratching (literary) information from interpersonal interaction media, wikis, RSS channels, news, and so on. These could be helpfully subdivided into:
- Data sources, services and tools: where information is gotten to by instruments which secure the crude information or give basic examination. Cases include: Google Trends, Social Mention, Social Pointer and Social-Seek, which give a surge of data that totals different online networking bolsters.
- Data feeds via APIs: where informational indexes and nourishes are open through programmable HTTP-based APIs and return labeled information utilizing XML or JSON, and so on. Cases incorporate Wikipedia, Twitter and Facebook.
- Text cleaning and storage tools: devices for cleaning and putting away printed information. Google Refine and Data Wrangler are cases for information cleaning.
- Text analysis tools: individual or libraries of apparatuses for dissecting online networking information once it has been scratched and cleaned. These are for the most part characteristic dialect handling, examination and characterization devices, which are clarified beneath.

- Transformation tools: straightforward apparatuses that could change literary information into tables, maps, graphs (edge, pie, disperse, bar, and so on.), course of events or even movement (activity over timetable, for example, Google Fusion Tables, Zoho Reports, Tableau Public or IBM's Many Eyes.
- Analysis tools: further developed investigation instruments for dissecting social information, recognizing associations and building systems, for example, Gephi (open source) or the Excel module NodeXL.
- Social media platforms: conditions that give far reaching online networking information and libraries of apparatuses for investigation. Illustrations include: Thomson Reuters Machine Readable News, Radian 6 and Lexalytics.
- Social network media platforms: stage that give information mining and investigation on Twitter, Facebook and an extensive variety of other social media sources.
- News platforms: stages, for example, Thomson Reuters giving business news files/sustains and related investigation

Social media technique to scrutinize the two noteworthy obstacles to utilizing online networking for scholarly research are right off the bat access to thorough informational collections and besides apparatuses that permit 'profound' information examination

## 4. SOCIAL NETWORKING SERVICE

Long range informal communication benefit is an arrangement of social destinations and applications, which at any rate comprise of three sections: customers, social connections, and intuitive correspondences. Truth be told, SNS is a subset of social networking, which incorporate the social media.

On SNS, correspondence is intuitive. For example, for unadulterated web journals, a non-SNS social networking, for example, blogspot.com, the customers' significant inspirations could be recording one's everyday life, giving critique and suppositions, communicating feeling, exhibiting thoughts through content, and keeping group. The initial four inspirations are all data sharing. For small scale blogging, a regular SNS, the customer aim could be generally characterized into three classes: data sharing, data chasing, and companionship upkeep [13].

All SNS suppliers have two center concentrations: social relations and customer created matters. Regarding social relations, they may mirror the social media of people, in actuality, manufacture new social associations in light of interests and exercises, or both. For customer delivered matters, they give a simple method to make, offer, and rank and trade data.

## 5. SOCIAL MEDIA ANALYTICS

Social networking examination could be depicted as the way toward gathering information from the social networking sites and breaking down that information to settle on business choices. Social networking investigation is for the most part used to mine customer opinion so as to help promoting and customer benefit exercises. Information examination could be ongoing or disconnected investigation, including elements, for example, impact, reach, and importance of reasonable estimations. Time contemplations are imperative to understanding the setting of information

being examined. The significance of social networking examination could be viewed as the scientists at AT&T built up a logical programming to listen in customers organize issue grumblings on Twitter. The groups will be sent to settle the issue by removing time, area, and sort, from the tweet [14]. Association's devotion to serving the mass with this level of need makes it all the more intriguing and makes rivalry among the association. Associations have been concentrating on research and advancement in examination in view of the assets they as of now have.

### 5.1 Offline Social Media Analytics

Disconnected information investigation alludes to the uninvolved examination of information, the most part of which is to be utilized for advanced showcasing channels. The disconnected information is the particular information caught, which is created by the customer or from disconnected sources, for example, CRM information records. The caught disconnected information of a specific customer from social networking has been extremely valuable and the results of the information examination shed light to the revealed factors. The significance of disconnected examination could be viewed as the greatest presidential race happens in the USA, where hopefuls generally crusade through social networking. Analysts display a solid determining framework for US presidential races and US house race, named Competitive Vector Auto Regression (CVAR) [15]. CVAR analyzes the fame of different contending applicants by consolidating visual data with printed data from the Flickr social networking. This sort of framework could give crusade bits of knowledge to the applicant with the goal that hopeful could chip away at their shortcomings and could additionally enhance their self.

Aside from the races, investigation has been utilized to anticipate securities exchange costs. The stock exchange decides the financial estimation of the nation, numerous individuals every day share good and bad times of securities exchange costs via social networking media. Analysts recommended that securities exchange value developments could be anticipated through social networking examination by proposing an Energy Cascade Model (ECM) [16]. ECM could viably anticipate center term directional securities exchange value developments, accomplishing a normal exactness of 67.7% for upward stock value developments. A comparative approach [17] was utilized to investigate the two noteworthy occasions of securities exchange value change and exchange volume. Trust data separated from the Twitter bunch was contrasted and Dow Jones Industrial Average (DJIA). By keeping trust data into account, the outcomes demonstrate that value change and exchange volume are more related than simply checking the quantity of tweets and exchange volume is more groundedly connected than value change.

However another investigation has been helped out through people in general miniaturized scale blogging long range informal communication website Twitter. The execution and mental prosperity of sprinters [18] had been followed, by checking Twitter tweets of sprinters gathering. The 925,825 messages of sprinters who utilized Nike + wellness GPS beacon were investigated. Analysts found that fitness devices were most popular in North America

less than 2% runners consistently ran for at least 150 min a week, which is recommended by Centers for Disease Control and Prevention

Physical activity lowered on Friday as the clients may need to be relaxed.

The sprinters have been recorded for 3 months in length; this by one means or another demonstrates that the old records could be utilized for the examination purposes. Be that as it may, it might be a major test to change over the records into some helpful frame before investigation.

From the viewpoint of dialect and history, digitization of million chronicled books and an investigation of the prior period 200 years had been finished by the tech goliath Google, demonstrating a difference in dialect utilization, elements of notoriety, oversight, and time pressure of aggregate memory. The puzzles of social networking before the Internet age could be tackled through more endeavors in this field. The enormous disconnected information investigation is some of the time simpler to perform on the grounds that information and commotion introduce in the information are predictable. The colossal volume and high speed of information could be a genuine test, and numerous specialists have done sublime work in the continuous social networking investigation.

### 5.2 Real-Time Social Media Analytics

Constant examination means the ability to utilize every single accessible datum and assets when required. The examination of information is done progressively and reports are created with no postponement. For the most part continuous examination is utilized for geographic area and following purposes. These days, individuals right away offer via social networking media about circumstances like catastrophic events, henceforth the constant investigation of social networking may give life-sparing data. Ongoing social networking examination of streams and charts called as Milano Design Week (MSW13) [19], it prescribes settings to guests of geo-and transiently limited city-scale occasions in Milan included 681 scenes for facilitating 1,127 occasions went to by 500,000 guests in a single week. By consolidating deductive and inductive stream thinking procedures, this framework dissects Twitter's tweets registered with its delivered excellent connection predicts. As specified before, individuals invest more energy in social networking and offer whatever is going on in the encompassing, regardless of whether it's a seismic tremor, auto crash, tidal wave, or avalanche.

A multilevel issue investigation of Twitter tweets gathered about landfall recommends that noteworthy data was less demanding to discover while seeking along hash labels. The utilization of Twitter in precrisis phases of a climate occasion could be useful for the crisis administration office's [20]. Another case of continuous social networking examination is the observing of flare-ups through the intermediary of customers look. Google's Flu Trends and Dengue Trends give evaluations of influenza and dengue in light of pursuit designs. Additionally, Google Trends could precisely anticipate the case disconnected achievement in view of the rating of social notices of individual films and the check of the inquiry made on YouTube. From the above investigations, couples of things are concerned, for example, the responses of the general population to the circumstance. The conduct of people groups differs as indicated by the circumstance, which could choose that on whom they will trust indiscriminately or may go out on a limb to put stock in others.

A thorough and quantitative meta-examination was directed to explore the observational proof of the most compelling elements, trust, and hazard which influence the individual conduct towards social networking stages. The discoveries proposed that both hazard and trust had noteworthy impacts, yet trust had a more grounded impact.

The impacts of hazard and trust have been plainly noticeable on the social networking. Trust is firmly identified with the satisfaction of the human conduct and generally more joyful people are more reliable.

At a development rate of  $\square 8\%$ , Internet customers are presently over 40% of whole total populace. Social networking ceaselessly assuming an awesome part to achieve that checks and has touched the numerous parts of human life. With this, the social networking is in charge of a radical new pattern that is important for associations, discovering and developing interesting patterns in human conduct.

## 6. PREDICTION SUBJECTS

In this segment, we depict regions where predict with social networking might be made. For the most part, a subject, that could be well unsurprising with social networking, must meet the accompanying prerequisites.

Right off the bat, the expectation subject must be human related occasion. Via social networking media, customers distribute their suppositions and convictions. Expectation techniques investigate, remove and coordinate the data, and after that as per the impact of people to the anticipated subject, make the prediction. In any case, if the subject is non-human-related occasion, for example, overshadow, despite the fact that there might be huge amounts of customers talk about this theme via social networking media, the customers' considerations have nothing to do with the improvement of that occasion. Thus, the information via social networking media couldn't be utilized to foresee normal occasions whose improvement is autonomous of human activities.

Furthermore, if masses of individuals are included, the appropriation of organization of included people via social networking media ought to be the same as or like that in certifiable. Since not every person in true will utilize social networking, the customers via social networking media could be dealt with as tests of the included masses much of the time. Be that as it may, the inspecting procedure is wild, which may prompt examples with worked in predisposition. Despite the fact that we can't avoid one-sided tests totally, we should ensure the extent of the one-sided tests is in the adequate and sensible range.

In conclusion, the included occasions ought to be anything but difficult to be talked in broad daylight. Something else, the issues via social networking media would be one-sided. For instance, three is social accord that giving fitting tips is great and too much low tipping is discourteous and inadmissible. Under such social weight, nearly no one will concede that he/she paid tips that were too low. The mysterious mode could be utilized as a part of finding a solution to this issue yet such an unknown mode will have no data about pertinent informal organization structures.

### 6.1 Marketing

There is some confirmation that there is solid relationship between spikes in closeout rank and the quantity of related blog entries. Be that as it may, in the meantime, in light of blog notices, foreseeing whether tomorrow's bartering rank for a specific thing will be higher or lower than the present sell-offs rank seems, by all accounts, to be hard [21]. There are two conceivable purposes behind these apparently opposing conclusions. On one hand, there might be a postponement between the expansion of blog notices and the increment of closeout. Then again, the quantity of blog notices may foresee the difference in sell off. Yet, the

difference in closeout around one item does not really change the sale rank of different items.

Despite the fact that there is connection, couple of analysts takes a shot at the prediction of sales with socianetworking. Since the everyday barterers information may include business secrecy, and there are such a significant number of approaches to buy items, it's almost difficult to get exact day by day barterers information. Thus, analysts want to take a shot at a small scale level that is on the item selection by customers.

Utilizing social networking, customers with great degree of positive or negative encounters will probably express their emotions and assessment, and contrasted these with direct encounters. On Twitter, association or item marks are eluded in around 19% of the considerable number of tweets, over 80% of which don't demonstrate any huge opinion [22]. The mass electronic Word of Mouth (eWOM) allows us to research how eWOM and person to person communication takes a shot at item reception and anticipate its potential selection. The verbal (WOM) is enormously persuasive on the principal buying of item or administration, particularly when the WOM is from companions or associates. For the most part, the negative eWOM is more intense than the positive ones. As far as social proposals, we find clashing exploration comes about. Some exploration brings up that, the more duplicates of a similar message are gotten and the higher likelihood that one will receive that development. In any case, in another exploration, extreme proposal apparently had a negative impact. At first the likelihood of obtaining increments with more suggestion however after some edge, the likelihood drops and remains on a moderately low level.

Long range informal communication likewise influences the creation reception extraordinarily. The likelihood of a person's buys increments if the item has been firmly received by companions [23]. Furthermore, customers with fewer companions are all the more effortlessly affected into reception [23]. Contrasted and eWOM, which is one sort of unequivocal suggestion, long range informal communication impacts reception as sort of certain proposal.

The impact of eWOM and person to person communication on item selection could be incompletely clarified by Heider's adjust hypothesis. In adjust hypothesis, companions have a tendency to accomplish and keep up consistency in enjoying and loathing of articles. The consistency will then prompt comparable or same item appropriation.

The eWOM and person to person communication do have some effect on item selection. Be that as it may, request starts things out in appropriation [23]. What's more, people could impact only a couple of companions, as opposed to everyone they know. So to foresee item selection, we should utilize social networking as an assistant predict instrument instead of the unequivocal one.

### 6.2 Movie box-office

To foresee motion picture film industry with social networking is a standout amongst the most contemplated region. Notwithstanding the conventional predict factors, for example, MPAA rating and number of screens [24], social networking matters could likewise be viable to anticipate film industry. There are numerous reasons that foreseeing film industry a decent subject to inquire about.

Right off the bat, there is volumes of information about motion pictures and related social networking. As per IMDB.com, in excess of 200 element films, which start in the U.S.A and have U.S.A film industry record, were discharged each year. Moreover, films are generally chatted

via social networking media. For instance, there are in excess of 100,000 tweets for each observed motion picture [25]. Subsequently, there is sufficient information to be broke down.

Furthermore, the movies are anything but difficult to be scrutinized and assessed. On one hand, the gross pay and opening end of the week pay can be effectively got from Internet Movie Database (IMDB). Then again, the pay on opening end of the week ordinarily represents around 25% of aggregate sales. So we could get the estimated film industry records soon after the opening end of the week. Now and again, the prediction about the high-netting motion pictures is much exact than that about low-earning motion pictures. Despite the fact that most analysts regard the movies as consistent variable, in some cases discretization is connected to separate the movies into classes as indicated by their sum.

Ultimately, there is a reasonable sensible relationship between social networking matters and film industry. The customers who post something before the film discharge are without a doubt inspired by the motion picture and subsequently they are probably going to watch the motion picture. The 1-week pre-discharge information has the most grounded connection with net than the information in some other pre-discharge eras. After the motion picture discharge, customer posts, particularly 6, swing to be somewhat eWOM, which would impact other potential customers. The ones with the assumption [26] swing to be somewhat eWOM, which would impact other potential customers.

All things considered, there are some remarkable impediments in look into on film industry. Normally, the names of motion pictures are likewise utilized as different implications in interchanges. For instance, the exceptionally well known motion picture "The God father" by Francis Ford Coppola has a title which would be utilized as a part of bunches of different cases. In reality, it's difficult to discover the related tweets however seek with "The God father" as catchphrases. Additionally, a few motion pictures hold a similar title. For example, there are motion pictures with a similar title of "Adoration". In such cases; it's very difficult to recognize the particular film related social networking information and different issues.

### 6.3 Elections

Decision expectation utilizes the review of general feeling on political gathering or government official from a specific example to anticipate the race result. Customarily, the race surveys should be possible by means of phone reviews. In any case, a large number of calls effortlessly prompt cost as high as a huge number of dollars. As a recently developing strategy, web review with social networking furnishes a chance to do that with minimal effort.

Essentially, the quantity of related online networking matters might be a substantial indicator for fruitful race. In 2008 U.S.A Presidential Primaries, simply the quantity of Facebook supporters could foresee the outcome effectively. In 2009 German government decision, despite the fact that 4% of all customers are in charge of over 40% of the issues, the quantity of messages on Twitter still could have anticipated the race result, and it even approached the convention race survey's precision. The conclusion could likewise be useful to do expectation, however not considerably [27].

In the meantime, there is a progressing banter on whether right now social networking is successful in gathering general sentiments in a fair way and foresees the decision result. For instance, in British Columbia's 2001

commonplace race, the quantity of notices on web message sheets did not demonstrate the relative quality of gatherings.

In such research, when contrasting the web overview and conventional surveys, specialists did not give the date when the customary surveys were made. Predict with social networking was generally made near the decision. On the off chance that the customary surveying was made a long way from that occasion, the examination is uncalled for and aimless. This exploration additionally avoided little gatherings. Including the little gatherings could have changed the predictions.

The specialists did not call attention to why a particular day and age to gather social networking matters was picked. The expectation result changes vigorously are relying upon the time period. Counting just extra couple of days prompts an extensive increment in the mean supreme mistake. No exploration gives any rule to pick a sensible and exact time window.

The benchmark to legitimize the online networking overview strategy ought to not generally be the arbitrary pick [28]. In the U.S.A congressional races, officeholders won 91.6% of the races in 2008 and 84.5% of every 2010. In the event that simply showing every one of the occupants would win, the precision would be higher than 80%. Utilizing comparative techniques on various informational collections delivers more regrettable outcomes than the ones in the previous papers with the mean normal blunder of 17.1% for utilizing negligible Twitter volume, 7.6% for the estimation investigation, and just 2-3% for conventional expert surveying administration.

It ought to be noticed that online networking does not mirror the socioeconomics of the general public. Regarding age, in 2000, 36% of U.S.A subjects in the vicinity of 18 and 24, half of residents in the vicinity of 25 and 34, and 68% of those more than 35 voted [28], yet on Twitter, over 60% of customers are under 24. In this way irregular inspecting via social networking media is one-sided examining. Additionally, it's difficult to know the time of social networking customers, in light of the fact that the customer's profile is classified. Appropriately, it's close to totally unthinkable for factually fair inspecting via social networking media, as far as age, and likewise different characteristics, for example, district and ethnicity. Then again, when connected to political issue via social networking media as a rule and Twitter specifically, the exactness of assessment investigation techniques, utilized as a part of some predict models, is superior to an irregular classifier to demonstrate the political introduction of the customers [28]. One conceivable clarification is that the vocabulary in most belief assessment framework is intended for elegantly composing Standard English, as opposed to the short posts via social networking media [27].

### 6.4 Macroeconomic

The macroeconomic incorporates the provincial, national or worldwide economies. A few specialists are endeavoring to utilize social networking to manage its patterns, for example, financial lists and securities exchanges. For the most part the social networking couldn't be utilized to precisely decide these patterns alone, yet could help the analysts to catch or foresee patterns.

Regarding financial records, scientists have utilized social networking to specifically anticipate or aid expectation. For Gallup Organization's "Monetary Confidence" file and Index of Consumer Sentiment (ICS) from Reuters/University of Michigan Surveys of Consumers, a few proportions in view of opinion data of social

networking can catch the wide patterns in customary financial surveys [27]. In any case, the coefficient between supposition information and shopper certainty fluctuates significantly at various time.

In securities exchanges, examine in light of arbitrary walk hypothesis and Efficient Market Hypothesis (EMH) proposes that stock costs are flighty. Be that as it may, late research, from the viewpoint of Socioeconomic Theory of Finance and behavioral financial aspects, propose that stock costs could be anticipated to some degree.

The huge number of postings on fund message sheets, for example, Yahoo! Fund, predicts negative ensuing stock returns. The relationship between them is measurably noteworthy. Yet, the impact is monetarily little yet the volume of postings is useful to anticipate the stock unpredictability. For posts that incorporate particular emotive words, for example, expectation, stress and dread, the aggregate number of them is more prescient to stock files than the number and extent of their sending times and unique writers' supporters [29]. Essentially, the aggregate number of such words has a solid relationship with other money related market patterns, for example, gold value, oil cost and cash conversion standard. Despite the fact that it needs strong proof and legitimization, it gives the idea that a non-direct relationship is probably going to exist between social networking and securities exchange. Non straight representations, for illustration, Support Vector Machine (SVM), could superior utilize social networking for prediction.

## 7. THE PREDICTORS

In this part, we will catalog the main metrics regarding social media utilized in prediction. Typically these metrics only don't have adequate prediction supremacy but their permutations work enhanced. These analysts might be separated into two classes: message characteristics and social network characteristics.

### 7.1 Message characteristics

Message characteristics concentrate on the post themselves, for instance the opinion and time series metrics. If the study concentrates on common entities, every the accessible messages are carried with time seals. Or else, the investigate outcome with defined keywords is chosen.

#### (a) opinion metrics

The opinion metrics are the fixed attributes of messages. As well as the common opinions conversed in the subsequent, there are a few precise opinion categories, such as gladness and nervousness, on a case-by-case base. Since they not have generalization, we don't inspect them thoroughly. However the idea, mining and utilization of them are identical as these of common ones. With qualitative sentiment study scheme, the messages might be tagged as +ve, -ve, or impartial sentiments. Thus obviously the amounts of +ve, -ve, impartial, non-neutral, and total messages are 5 basic matter analysts. These metrics might have diverse prediction supremacy at diverse phases. For a predicted occasion in common and movie box-office in scrupulous, the amount of positive references compares with the occasion outcome enhanced than total calculation in the pre-event phase. However in the message occasion phase, the total calculation is improved.

moreover, we might estimate the ratios amongst them, generally including the ratio among the amounts of positive and entire messages, the ratio among the amounts of negative

and entirety messages [30], the ratio among the amounts of impartial and entirety messages, the ratio among the amounts of non-neutral and entirety messages, the ratio among the amounts of neutral and non-neutral messages [25], and the ratio among the amounts of positive and negative messages [25] [27]. These ratios imitate the comparative potency of these opinions. More compound, we might merge these basic components to analyze the opinions variation [30] and sentiments index  $I_{sent}$ .

$$Sent\_deff = \frac{N_{positive} - N_{negative}}{N_{total}} \quad \dots (1)$$

$$I_{sent} = 100 * \left( \frac{\frac{N_{positive} - N_{negative}}{N_{total}}}{2} + 0.5 \right) \quad \dots (2)$$

The  $N_{positive}$ ,  $N_{negative}$  and  $N_{total}$  signify the amount of positive message, negative messages and total messages correspondingly. The sentiments index is verified to have robust connection with rating of IMDB and be valuable in guess prizes of Oscar when utilized for cinema.

#### (b) Time series metrics

Time series metrics attempt to inspect the messages vigorously, together with the velocity and method of the message production. The posts producing rate signifies how promptly the messages are formed. It's effortlessly figured as the subsequent:

$$Post\_rate = \frac{N_{total}}{Time\ window\ size} \quad \dots (3)$$

According the time windows range, the producing rate might be predictable diversely, for instance hourly, day by day or weekly. Through higher posts producing rate, more individuals are regarding it, and the subject is more striking. A few testing demonstrated that, the day by day producing rate previous to release is a excellent analyst for movie box-office [25].

The cause configuration of messages in a time window is an additional time series metrics. At this point the time might be actual or virtual. For instance, in election on digg.com, we might delight every vote as a Digg subsequent. So in the primary 10 Digg subsequent, the votes are collected of supporter votes and non-supporter votes. If the fan votes engage awesome percentage of the total, these messages will at last gather less votes than others. Since for every gap among two supporter votes, the more non-supporter votes are in the gap, which shows the message's appeal to community, the less supporter votes' quantity is, and the extra the concluding votes will be.

### 7.2 Social network characteristics

Social network characteristics compute configuration characteristics. We as well entitle these features as metrics/computes in social network analysis. Being extended studied, there are numerous features that it's unfeasible to record and inspect all of them here. So we presently itemize and provisionally converse the mainly utilized ones in predictions.

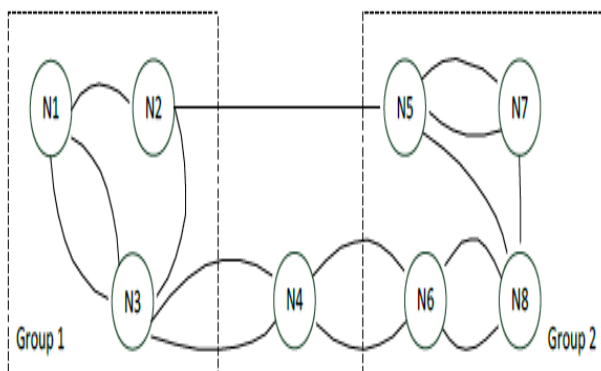


Fig. 2: Directed social network

(a) Terminology

Now we commence some unique terms utilized in social media. As diverse social networking websites have diverse utilities and IDs for the social connections, we join the namespace to make simpler the conversations. In the subsequent, we will concentrate on the directed network. As well as the undirected social network is comparable. Except precised, all the conversations are with directed networks.

**Node:** each node symbolizes one exclusive unit in the social networking. For instance, on Twitter, the users might be conveyed as nodes. Although on a venture trading network, the companies are symbolized as nodes.

**Follow:** if node A specifies to have an affiliation with node B, A follows B, which is signified as a directed edge from A to B in graph. For instance, in the Figure 2, N1 follows N2. The affiliation might signify another way in diverse websites. On YouTube, it is subscription, whereas on Twitter it is a supporter. Moreover, the follow might be unidirectional or bidirectional. In unidirectional follow, for instance Google+, node A might follow node B, with no being followed by B. conflictingly in bidirectional manner, for instance Facebook, the follow from node A to node B will convoy with the follow from B to A.

**Follower:** if A follows B, A is B's follower. For illustration, in the Fig. 2, N5 has N2 being its follower.

**Followee:** followees are the following units. In Fig. 2, N5 and N3 are N2's followees. And N1 is the follower and followee of N3 at the same time.

(b) Degree

Degree is the amount of attachments to/from other nodes in the network, together with in-degree, out-degree and total-degree. The in-degree and out-degree are obtainable merely when the social network is directed. Occasionally, the degree of a node is as well known as its degree centrality.

**In-degree:** in-degree of a node is the amount of directed edges to this node. That is, the in-degree is the calculation of followers. Typically, every user has 85 followers on Twitter. Treating the calculation of messages as the action index, with the in-degree growing, the action initially boosts, and then after about 300 followers as threshold, turn into steady.

**Out-degree:** out-degree of a node is the amount of its finish endpoints. Or we can say, the out-degree is equivalent to the amount of followees. According to Standard, every Twitter user follows 80 other users. As the study only illustrations a division of Twitter users moderately than all of them, the middling values of followers and followees aren't the similar. Both the in-degree and out-degree alone specify disorder about the user's influence.

**Total-degree:** total-degree is the total of the in-degree and out-degree. In the undirected network, this is the merely one degree metric.

(c) Density

**Density:** density is the quantity of existing edges calculation relative to greatest potential edges calculation. In a directed social networking with n nodes and e edges, the density is calculated as:

$$\text{Density} = \frac{e}{n(n-1)} \dots\dots (4)$$

And in an undirected social networking with similar constraints, the density is approximated as:

$$\text{Density} = \frac{e}{n(n-1)/2} \dots\dots (5)$$

The network of followers/followees is extremely impenetrable. But in 2009, 77.9% user duos have one-way association. That is, only 22.1% user duos have shared associations. Thus the network of authentic friends, who converse jointly and directly, is greatly lighter and easier.

(d) Centrality

Centrality determines the comparative significance of a node inside a network. Betweenness centrality and closeness centrality are two centrality metrics, which are extensively utilized. Also, as well to the degree centrality on the node level, which is conversed in section (b), the group degree centrality is initiated.

**Betweenness centrality:** the betweenness centrality quantitatively determines the control of a node on the messages among other nodes in the social networking. In a network with n nodes set V, the betweenness centrality of a node v is:

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\delta_{st}(v)}{\delta_{st}} \dots\dots (6)$$

Where t is the sum of shortest paths among node s and node t and  $\delta_{st}(v)$  is the amount of these paths passing via v. The betweenness centrality might be normalized by (V-1) for directed network or (V-1)/2 for undirected network.

**Closeness centrality:** in a network, the space of two nodes is the span of the shortest path among them. The farness of a node is the amount of its spaces to all other nodes. And its closeness centrality is the opposite of the farness:

$$C_C(v) = \frac{1}{\sum_{s \neq v \in V} \text{length}(SP_{sv})} \dots\dots (7)$$

Where  $SP_{sv}$  signifies the shortest path among node s and v. For node v, the closeness centrality approximates how extensive it will acquire to broadcast information from node v to each and every node in the network.

**Group degree centrality:** coverage the degree from the node level to group level,

Additionally to the metrics stated about, there is still countless other social media feature, for instance network diameter. Other than they aren't broadly utilized or established to be potent in calculation. So we don't record them in comprehensive here.



## 8. KEY METHODS

We start with definitions of some of the key techniques related to analyzing unstructured textual data:

- Natural language processing: (NLP) is a field of software engineering, computerized reasoning and phonetics worried about the collaborations amongst PCs and human (characteristic) dialects. In particular, it's the procedure of a PC extricating significant data from normal dialect input as well as creating characteristic dialect yield.
- News analytics: the estimation of the different subjective and quantitative characteristics of literary (unstructured information) news stories. Some of these characteristics are: assessment, pertinence and curiosity.
- Opinion mining: opinion mining (sentiment mining, opinion/sentiment extraction is the region of research that endeavors to make programmed frameworks to decide human sentiment from content written in normal dialect.
- Scraping: gathering social information from online networking and other Web destinations as unstructured content and furthermore known as website scraping, web collecting and web information extraction.
- Sentiment analysis: assumption examination alludes to the utilization of regular dialect preparing, computational etymology and content investigation to distinguish and remove subjective data in source materials
- Text analytics: includes information recovery (IR), lexical examination to think about word recurrence appropriations, design acknowledgment, labeling/comment, data extraction, information mining systems including connection and affiliation investigation, perception and prescient investigation.

## 9. RESEARCH CHALLENGES

Social media scraping and analytics provides a rich source of academic research challenges for social scientists, computer scientists and funding bodies. Challenges include:

- Scraping—albeit online networking information is available through APIs, because of the business estimation of the information, the greater part of the significant sources, for example, Facebook and Google are making it progressively troublesome for scholastics to acquire extensive access to their 'crude' information; not many social information sources give reasonable information offerings to the scholarly community and analysts. News administrations, for example, Thomson Reuters and Bloomberg regularly charge a premium for access to their information. Interestingly, Twitter has as of late reported the Twitter Data Grants program, where analysts could apply to gain admittance to Twitter's open tweets and chronicled information to get bits of knowledge from its monstrous arrangement of information (Twitter has in excess of 500 million tweets every day).
- Data cleansing: cleaning unstructured printed information (e.g., normalizing content), particularly high-recurrence spilled continuous information, still exhibits various issues and research challenges.

- Holistic data sources: analysts are progressively uniting and joining novel information sources: online networking information, constant market and customer information and geospatial information for investigation.
- Data protection: once you have made a 'major information' asset, the information should be secured, proprietorship and IP issues settled (i.e., putting away scratched information is against the vast majority of the distributors' terms of administration), and customers furnished with various levels of access; generally, customers may endeavor to 'suck' all the important information from the database.
- Data analytics: modern examination of social networking information for supposition mining (e.g., assumption investigation) still raises a bunch of difficulties because of outside dialects, remote words, slang, spelling blunders and the normal developing of dialect.
- Analytics dashboards: numerous online networking stages expect customers to compose APIs to get to bolsters or program examination models in a programming dialect, for example, Java. While sensible for PC researchers, these aptitudes are normally past most (sociology) specialists. Non-programming interfaces are required for giving what may be alluded to as 'profound' access to 'crude' information, for instance, designing APIs, blending online networking nourishes, consolidating all encompassing sources and creating expository models
- Data visualization: visual portrayal of information whereby data that has been disconnected in some schematic frame with the objective of imparting data unmistakably and successfully through graphical means. Given the extent of the information included, perception is winding up progressively critical.

## 10. RESEARCH SCOPE

As a developing examination point, predict with social networking faces numerous difficulties. Here we call attention to some earnest and imperative potential works.

At present, analysts pick indicators utilizing the experimentation technique. We know neither why these indicators are superior to others, nor how these indicators could foresee the outcome. Not knowing the foundation rationale between these measurements and the last expectation result, we simply utilize an accumulation of measurements to be prepared on test information, discover which ones have the most elevated coefficients, and utilize them to make the predict display. Thusly, deficient with regards to a strong supporting hypothesis, we can't make sure that one model, which functions admirably in one case, could be connected to different circumstances with a similar exactness. That is the reason a few models demonstrate great execution in one race predict, however totally explode in another. To ensure our model has great execution in all cases, we have to know the rationale and hypothesis behind the model.

Most analysts utilize straightforward techniques, for example, direct relapse investigation. These techniques are known to function admirably under a few conditions. Social networking is created on a mind boggling framework and in

this manner probably the indicators and predicts results have non-direct relationship. Moreover, mix of strategies may prompt leap forward. In such mix, a surface learning specialist, for example, immediately prepared neural systems, rapidly adjusts to new modes and developing patterns via social networking media. What's more, a profound realizing specialist centers on long haul designs. More or less, we should attempt some non-direct techniques and discover the appropriate strategies or potentially mixes for every predict domains.

## 11. CONCLUSIONS

Social networking is characterized as electronic and portable based Internet applications that permit the creation, access and trade of customer delivered matter pervasively. Social networking is particularly imperative to inquire about into computational sociology that researches questions utilizing quantitative strategies (e.g., computational insights, machine learning and multifaceted nature) thus called enormous information for information mining and recreation displaying. This has prompted various information administrations, instruments and examination stages. Be that as it may, this simple accessibility of social networking information for scholastic research may change essentially because of business weights open. In this paper, we introduced a review of expectation utilizing social networking. We additionally gave a diagram of predict elements and strategies and recorded testing issues and regions for additionally inquire about. In spite of the fact that expectation utilizing social networking is just a rising exploration subject and its outcomes have moderately low exactness, it has made another route for us to gather, extricate and use the shrewdness of group in a target way with ease and high proficiency.

## 12. REFERENCES

- [1.] Alexa Internet Inc, "Alexa Top 500 Global Sites". <http://www.alexa.com/topsites>. [Accessed Jan 24, 2018].
- [2.] Facebook, "Statistics". <http://www.facebook.com/press/info.php?statistics>. [Accessed Jan 14, 2018]
- [3.] Twitter, "#numbers". <http://blog.twitter.com/2011/03/numbers.html>. [Accessed Jan 24, 2018].
- [4.] Eytan Adar, "GUESS, the Graph Exploration System". <http://graphexploration.cond.org/>. [Accessed Jan 24x, 2018].
- [5.] Kleinberg, Jon. "The small-world phenomenon: An algorithmic perspective." In Proceedings of the thirty-second annual ACM symposium on Theory of computing, pp. 163-170. ACM, 2000.
- [6.] Watts, Duncan J. Six degrees: The science of a connected age. WW Norton & Company, 2004.
- [7.] Kwak, Haewoon, Changhyun Lee, Hosung Park, and Sue Moon. "What is Twitter, a social network or a news media?." In Proceedings of the 19th international conference on World wide web, pp. 591-600. ACM, 2010.
- [8.] Watts, Duncan J. "The "new" science of networks." *Annu. Rev. Sociol.* 30 (2004): 243-270.
- [9.] Cha, Meeyoung, Hamed Haddadi, Fabricio Benevenuto, and P. Krishna Gummadi. "Measuring client influence in twitter: The million follower fallacy." *Icwsn* 10, no. 10-17 (2010): 30.
- [10.] Romero, Daniel M., Wojciech Galuba, Sitaram Asur, and Bernardo A. Huberman. "Influence and passivity in social media." In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 18-33. Springer, Berlin, Heidelberg, 2011.
- [11.] Kaplan, Andreas M., and Michael Haenlein. "Clients of the world, unite! The challenges and opportunities of Social Media." *Business horizons* 53, no. 1 (2010): 59-68.
- [12.] Kietzmann, Jan H., Kristopher Hermkens, Ian P. McCarthy, and Bruno S. Silvestre. "Social media? Get serious! Understanding the functional building blocks of social media." *Business horizons* 54, no. 3 (2011): 241-251.
- [13.] Java, Akshay, Xiaodan Song, Tim Finin, and Belle Tseng. "Why we twitter: understanding microblogging usage and communities." In Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis, pp. 56-65. ACM, 2007.
- [14.] Ikeda, Kazushi, Gen Hattori, Chihiro Ono, Hideki Asoh, and Teruo Higashino. "Early detection method of service quality reduction based on linguistic and time series analysis of twitter." In Advanced Information Networking and Applications Workshops (WAINA), 2013 27th International Conference on, pp. 825-830. IEEE, 2013.
- [15.] You, Quanzeng, Liangliang Cao, Yang Cong, Xianchao Zhang, and Jiebo Luo. "A multifaceted approach to social multimedia-based prediction of elections." *IEEE Transactions on Multimedia* 17, no. 12 (2015): 2271-2280.
- [16.] Zhang, Wenping, Chunping Li, Yunming Ye, Wenjie Li, and Eric WT Ngai. "Dynamic business network analysis for correlated stock price movement prediction." *IEEE Intelligent Systems* 30, no. 2 (2015): 26-33.
- [17.] Ruan, Yefeng, Lina Alfantoukh, and Arjan Duresi. "Exploring stock market using twitter trust network." In Advanced Information Networking and Applications (AINA), 2015 IEEE 29th International Conference on, pp. 428-433. IEEE, 2015.
- [18.] He, Qian, Emmanuel Agu, Diane Strong, Bengisu Tulu, and Peder Pedersen. "Characterizing the performance and behaviors of runners using twitter." In Healthcare Informatics (ICHI), 2013 IEEE International Conference on, pp. 406-414. IEEE, 2013.
- [19.] Balduini, Marco, Alessandro Bozzon, Emanuele Della Valle, Yi Huang, and Geert-Jan Houben. "Recommending venues using continuous predictive social media analytics." *IEEE Internet Computing* 18, no. 5 (2014): 28-35.
- [20.] Lachlan, Kenneth A., Patric R. Spence, Xialing Lin, Kristy Najarian, and Maria Del Greco. "Social media and crisis management: CERC, search strategies, and Twitter matter." *Computers in Human Behavior* 54 (2016): 647-652.
- [21.] Gruhl, Daniel, Ramanathan Guha, Ravi Kumar, Jasmine Novak, and Andrew Tomkins. "The predictive power of social chatter." In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining, pp. 78-87. ACM, 2005.
- [22.] Jansen, Bernard J., Mimi Zhang, Kate Sobel, and Abdur Chowdury. "Twitter power: Tweets as electronic word of mouth." *Journal of the Association for Information Science and Technology* 60, no. 11 (2009): 2169-2188.
- [23.] Bhatt, Rushi, Vineet Chaoji, and Rajesh Parekh. "Predicting product adoption in huge-scale social networks." In Proceedings of the 19th ACM international conference on Information and knowledge management, pp. 1039-1048. ACM, 2010.
- [24.] Zhang, Wenbin, and Steven Skiena. "Improving movie gross prediction through news analysis." In Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on, vol. 1, pp. 301-304. IEEE, 2009.

- [25.] Asur, Sitaram, and Bernardo A. Huberman. "Predicting the potential with social media." In Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology-Volume 01, pp. 492-499. IEEE Computer Society, 2010.
- [26.] Lica, Liviu, and Mihaela Tuta. "Predicting product performance with social media." *Informatica Economica* 15, no. 2 (2011): 46.
- [27.] O'Connor, Brendan, Ramnath Balasubramanyan, Bryan R. Routledge, and Noah A. Smith. "From tweets to polls: Linking text sentiment to public opinion time series." *Icwsn* 11, no. 122-129 (2010): 1-2.
- [28.] Metaxas, Panagiotis T., Eni Mustafaraj, and Dani Gayo-Avello. "How (not) to predict elections." In *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on, pp. 165-171. IEEE, 2011.
- [29.] Zhang, Xue, Hauke Fuehres, and Peter A. Gloor. "Predicting stock market indicators through twitter "I hope it's not as bad as I fear"." *Procedia-Social and Behavioral Sciences* 26 (2011): 55-62.
- [30.] Zhang, Wenbin, and Steven Skiena. "Improving movie gross prediction through news analysis." In *Web Intelligence and Intelligent Agent Technologies, 2009. WI-IAT'09. IEEE/WIC/ACM International Joint Conferences on*, vol. 1, pp. 301-304. IEEE, 2009.
- [31.] Bollen, Johan, Huina Mao, and Xiaojun Zeng. "Twitter mood predicts the stock market." *Journal of computational science* 2, no. 1 (2011): 1-8.
- [32.] Salathe, Marcel, Linus Bengtsson, Todd J. Bodnar, Devon D. Brewer, John S. Brownstein, Caroline Buckee, Ellsworth M. Campbell et al. "Digital epidemiology." *PLoS computational biology* 8, no. 7 (2012): e1002616.