TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

Stock Trading Recommendations using Opinion Mining

[1]S Shabana Begum, Research Scholar, Rayalaseema University, Kurnool, Andhra Pradesh.
[2]Dr.N. Kasiviswanath, Professor&Head, Dept. of Computer Science and Engineering,
G.Pulla Reddy Engineering College, Kurnool, Andhra Pradesh.
Email: shabanasphd@gmail.com & hodcse@gprec.ac.in.

Article Info
Volume 82
Page Number: 16641 - 16655
Publication Issue:
January-February 2020

Article History
***Article Received**: 18 May 2019*
***Revised**: 14 July 2019*
***Accepted**: 22 December 2019*
***Publication**: 28 February 2020*

I. Abstract:

The prediction of stock trading factorsis critical in making decisions for investors. Most of them depend on disclosures of news for making decisions in selling or buying stocks. Nevertheless, precise stock market modelling trends through disclosures of news is a thought-provoking task, deliberating the ambiguity & intricacy of natural languages utilized. Unlike former work besides with the research that implements bag of words typically for extracting several features for building prediction method, we project a method based on opinions posted in twitter trends, which are highly influential towards a certain sector of the stock trading. A novel method that defines the meta-heuristics for trading factors listed as a quantity of stocks, opening cost, closing cost, aggregate cost and status. These meta-heuristic values of the corresponding factors recommend by their correlation with opinions exhibited in corresponding twitter trends. The experimental study was carried on the data extracted from the yahoo ticker, and the tweets posted during the contemporary period in twitter trends that are correlated to the yahoo ticks collected from the sector of the stock trades. The experimental study outcomes denoting the significance of the proposed stock trading recommendations using opinion mining (STROM).

Keywords: UGC (user-generated content), yahoo ticker, Sensitivity, Accuracy, UGC sentiment analysis, opening cost, closing cost

I. INTRODUCTION

An individual can broadcast a brief statement to some or entire social network members by twitter. The work [1] presents that twitter is one of the social networks that havethe huge potentiality of an audience, which attracts average users of 271 million per month. The work [2], [3] presents that current research explored to whether the effect of twitter is substantial economically or not. The impact of twitter is exhibited mainly associated with practical media products such as electronic games, music and many more; and usually, these were the products over which immediate success is needed. For example, the Bru no, failure, it is a movie of multimillion-dollar that attributed to negative-sentiment regarding movie stated on twitter. On the other dimension, the positive sentiments are observed the cause of unanticipated opening karate kid remake success [3].

The methods based on twitter could be built then for accumulating the combined population opinions. They could be utilized for estimating further trends when attaining resourceful inputs into individual conduct [4]. The services of micro-blogging offered by social networks enable the fast spread of UGC (user-generated content) from one to many people all over the world as videos, images or short-text. The platforms of micro-blogging have exponentially grown, which observed to be indispensable information sources [5] & are attaining the popularity rapidly among researchers, organizations & users in several disciplines.

The micro-blogging popularity could be explained through its divergent features like accessibility and convenience that enables users for responding & disseminating the information instantly with confined or no limitations on the content [6]. Here, it could be stated that prominently social-media influenced our routine lives & has modified the business performance &individual’s way, awareness is created and advice is pursued [6]. Hence, it is not astonishing that the continuous & rich data mass available with these platforms could

TEST Engineering & Management

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

be harnessed by researching group or individual behaviour and also especially global patterns in respect to sentiment over products, current news, political, brands, events issues. The work [7] presents that currently twitter is 10^{th} popular website that is globally over 300 million monthly active users.

Twitter could be updated several times in data through content changing from the individual every-day updates towards universal events & news. The twitter enables users for creating personal-profiles, which others might subscribe to it or follow and status updates are published called 'tweets' confined to minimum characters of 140 and to interact with others by replies. Here,a common practice could be 're-tweeting', where the user might select to send the tweet that they identified exciting for the followers while believing on original author, enabling well-known posts for travelling beyond original creator network. Hence, it is noticed that maximum re-tweeted posts impact the opinions of the global community of twitter [8]. Also, twitter encourages the hashtags utilization that enables tweets for organizing in the form of thread, which could be utilized for succeeding particular topics & events. Mostly stated hashtags & phrases tracked by twitter are posted under trending topics list that is regularly updated and enables users to keep track of what could be well-known at any specified time.

Initially, Twitter was set as a communication platform type designed for enabling friends to place tabs on 1 other. Acknowledges the API availability that stores tweets, where researchers could be accessed, & their appropriate features like filtering through variables such as keywords & location. The work [6] presents that twitter encouraged researches to explore and have an interest beyond social-network. The work [9], [10] assesses the twitter structural properties as studies of social network concentrates on influentially of the user.

The research of content analysis concentrates on examining vitality, tweets motivations & content. The work [11] classified tweets motivation into succeeding categories: conversations, news reporting, information sharing, & daily-chatter. Researchers of SA concentrate on utilization of chatter sentiment of twitter to estimate the behaviour. The work [12] states that even though every tweet depicts a unique view, an accumulative sample needs to offer a precise depiction of people mood.

II. Related work

The work [12] presents that in extracting the information from socialmedia, there is a prominent impact in literature. These researchers derived 6 dimensional depiction of moods such as happy, calm, kind, vital, sure & alert by mining a huge amount of tweets. The outcome of mood is utilized for estimating DoW Jones Index movement direction. Nevertheless, their claim related to the estimation of every-day down/up movement of the market at 86.7% accuracy with an extensive scepticism in a computational community of finance. Primarily, the accuracy estimation is stated was the optimal of 8 methods tried in the contribution and is reviewed to be discriminating report. Next, the period of testing deliberated by researchers is short and only trading days of 15 with 13/15 days estimated exactly. Finally, the researchers test the causality of granger among financial & sentiment time-series. Nevertheless, the causality of granger estimates covariance of stationary over time that is not in the instance of market prices, evincing the characteristics of non-stationary. The researchers could not explain whether they are flexible for this estimation or not, for example by devising data-windows.

Regardless of this, the thought of associating extracted information from the news feeds towards a movement of the market through atransitional stage of sentiment characterizing is a prominent one. Entire emotions both at wider society & trader's groups' level were confined to impact the financial markets conduct. When the individual investor's behaviour may be modulated heavily by the framework [12], exactly assume that entire combined signal impacts the behaviour of market trends that is present and this might be recoverable through an exchange of mining messages on the socialmedia. Some concerns are addressed in contribution [12] through testing on heterogeneous, voluminous sets of data.

III. PREDICTING STOCK TRADING FACTORS BY OPINION MINING

Size of UGC sentiment analysis could be an augment research area with a possibility of several implementations. The UGC is been utilized for estimating success & outcomes of movies [13], political elections [14] & music albums, books [15]. The more dynamic research area is the utilization of UGC for estimating the movements of the financial market. Various contributions introduced association among unusual returns of market [16], [17] & maximum levels of activity yet none of the predictive values is found. Identical to [12], [8], [18-23], have identified chatter of twitter is more successful in introducing predictive association related to the performance of the financial market.

These researches are classified into 2 important categories: sentiment or volume-driven. The work [21] presents that utilizing volume analysis exhibit that every data count of tweets stating 500 S&P are prominently associated with 500 S&P indicators of stock and could be utilized for predicting market movements & sector level by the accuracy of 68%. The work [19] recognized no prominent casual association among predictions of stock market & volume and work [20] presents that by comparing the methods it recommends that sentiment could be robust stock market movement predictor than volume

The work [19] concentrates on general emotions levels & exposed that there could be a predictive association among movements of stock & mood of twitter. The work [18] employed an identical model and identified that positive association present among the volume of close prices & positive tweets, but in 48h time, and negative tweets were the optimal indicators. The work [12] employed refined sentiment collection method and as an outcome are capable of estimating daily variations in the closing price of DIJA with an accuracy of 87.6%.

The contemporary contribution titled "Predicting Stock Market Volatility using Financial News" [24] is using the opinions reflecting on financial news to predict the stock market trading factors. This method deriving the correlation between market opinions related to finance to predict stock trading factors. Though, the method proven to be significant under the context of data considered for experimental study, the method has not considered the diversified sectors of the stock trading and trends related to current affairs.

In order to improve the prediction of stock trading factors, a novel recommendation system titled as Stock Trading recommendations using Opinion Mining (STROM) has proposed, which adopts the correlation between stock trading factors and the opinions presented on trends of current affairs, which have posted as tweets on twitter.

IV. METHODS AND MATERIALS

This section details the model of the proposal and the materials and methods involved in the proposed model. The description of the objective includes the context and scope of the proposed recommendation system in stock trading, which involves the sentiments related to diversified events trending in the newscast and social media domain. The overall process of the proposed model aligned the recommendations in stock trading towards a brand, product, and business sectors. The other context of the recommendations considering the tenures (day, week, month, and custom) of the stock trading state.

The Data Structure

The given data of the stock trading has to partition into diversified business sectors, and for each sector, partitions the records into the tenures (day, week, month. quarterly, half-yearly, and another custom number of days) given as input, and the trends of the corresponding tenures. Further, prepares training records in the following format.

Record ID	Quantity of stocks	Opening cost	Closing cost	aggregate cost	status	positive trend n-grams	negative trend n-grams

In order to prepare the training corpus, the following tasks evolve in sequence

The n time-intervals (like days, weeks, fortnights, months or custom range) are the primary

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

input of the model. Further, for each time interval $\{t_i \exists i = 1,2,3,...n\}$, extract the stocks trading information $st(t_i)$ of the given sector, which have observed during the time interval t_i. Similarly, the tweets $tw(t_i)$ from the trends that are highly correlated to the sector of the stock trading, which published with in the time interval t_i.

Further, partition the tweets into two groups, such that, the tweets having positive sentiment lexicons fall into one group $pG(t_i)$, and the tweets having negative sentiment lexicons as the second group $nG(t_i)$.Further find the positive trend n-grams, and negative trend n-grams, which shall be done by using NLP and distribution diversity assessment method. These outcomes further use to frame-up the records, which uses for the learning phase. The n number of records shall be framed from the given corpus, where the number n represents the number of time intervals possible from the given corpus. The notations used to represent the columns of the records framed to perform learning phase are:

- Record id (rid), which begins at 1 and ends with n
- Quantity of stocks (qS): the mean of a number of stocks related to the sector have traded each day during the time interval given,
- Open cost $(ocst)$: Indicates the minimal value of the opening costs $ocst(t_i)$ noticed for the stocks traded each day during the given time interval t_i
- Closing Cost $(ccst)$: Indicates the minimal value of the closing costs $ccst(t_i)$ noticed for the stocks traded each day during the given time interval t_i
- Aggregate cost $(acst)$ indicates the mean of the aggregate value of the stocks traded each day during the given time interval t_i
- Trading status (ts): Denotes the mean of the closing indices (up (1), stable (0), or down (-1)) of the stocks traded each day of the given time interval
- N-grams of the positive trends (nPt): denotes the terms (word tokens), which are highly correlated to the positive sentiment lexicons
- N-grams of a negative trend (nNt): denotes the terms (word tokens), which are highly correlated to the negative sentiment lexicons.

These resultant records have further used to derive the meta-heuristic scales from the features "Quantity of stocks (qS), Open cost $(ocst)$, closing cost $(ccst)$, aggregate cost $(acst)$, and positive n-grams (nPt), negative n-grams (nNt) to predict the status of the stock trading of the selected sector.

Data processing

This section details the data processing in relate to the stock trading of the given sector and the twitter trends that highly considered as influential factors of the stock trading in the corresponding sector

Data processing of the recommended twitter trends

- $\overset{n}{\underset{i=1}{\forall}} \{t_i \exists 1 \le i \le n\}$ Begin //For each time interval t_i
    - Let the notation $tw(t_i)$ denotes the tweets related to the trends that are highly correlated to the sector of the stock trading
    - Let the notation $st(t_i)$ denotes the trading ticks related to the sector of the stock trading
    - Let the notations pL, nL denote the set of positive lexicons and negative lexicons in respective order, which are often used to represent the opinion
    - Let the notation tkn_i is an empty set, contains unique elements
    -
    - $\overset{|tw(t_i)|}{\underset{j=1}{\forall}} \{tw_j \exists tw_i \in tw(t_i)\}$ Begin //For each tweet
        - $wv_j =$ process((tw_i) // performs text process on the tweet tw_j that results in word vector wv_j

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

- $tkn_i \leftarrow tkn_i \cup \{wv_j \setminus (pL \cup nL)\}$ // adding all the words that are not the sentiment lexicons to the hash set
- $wv_j^+ = 1 - \frac{1}{|(pL \cap vw_j)|}$ // normalizing the number of positive lexicons appeared in the word vector wv_j
- $wv_j^- = 1 - \frac{1}{|(nL \cap vw_j)|}$ //normalizing the number of negative lexicons appeared in the word vector wv_j
- $wv_j = \{wv_j \setminus (pL \cup nL)\}$ // discard all the sentiment lexicons from the word vector wv_j
- $if\,(wv_j^+ > wv_j^-)$
  - $P \leftarrow wv_j$ move the resultant words of the vector wv_j to the set P of positive tweets
- $else$
  - $N \leftarrow wv_j$ // move the resultant word vector wv_j to the set N of negative tweets

- End
- $\forall_{k=1}^{|tkn_i|} \{e_k \exists e_k \in tkn_i\}$ // for each entry e_k of the tokens tkn_i
  - $pos(e_k) = 1 - \frac{1}{\sum_{j=1}^{|P|}\{1 \exists wv_j \in P \wedge e_k \in wv_j \wedge e_k \in tkn_i\}}$ //positive occurrence score of the word e_k
  - $nos(e_k) = 1 - \frac{1}{\sum_{j=1}^{|N|}\{1 \exists wv_j \in N \wedge e_k \in wv_j \wedge e_k \in tkn_i\}}$ //negative occurrence score of the word e_k
  - $if\,((pos(e_k) - nos(e_k)) > 0)$ $nPt(t_i) \leftarrow e_k$ // add a token e_k to the positive n-grams nPt of the time-interval t_i
  - $elseif\,((pos(e_k) - nos(e_k)) < 0)$ $nNt(t_i) \leftarrow e_k$ // add a token e_k to the negative n-grams nNt of the time-interval t_i
  - $else$ discard the token e_k
- End

- End

Processing stock trading inputs

Collect count of stocks $qs(t_i)$ traded during the given time interval (usually equal to the $(i+1)^{th}$ time interval of the opinions extracted.

Further, find the opening cost of the stocks of the corresponding sector as follows:

- $ocst_A = \frac{\sum_{j=1}^{|t_i|}\{ocst(d_j)\}}{|t_i|}$ // average of the opening cost of the stocks observed in each trading day of the time interval t_i
- $ocst_D = \frac{\sum_{j=1}^{|t_i|}\left\{\sqrt{(ocst_A - ocst(d_j))^2}\right\}}{|t_i|}$ // deviation of the opening cost observed during the traded days of the given time interval
- $ocst(t_i) = ocst_A - ocst_D$ // the absolute distance between average opening cost and corresponding deviation observed in a given time interval t_i
- Similarly, the closing cost of the stocks of the given sector has to estimate, which is as follows,

- $ccst_A = \frac{\sum_{j=1}^{|t_i|}\{ccst(d_j)\}}{|t_i|}$ // average of the closing cost of the stocks observed in each trading day of the time interval t_i
- $ccst_D = \frac{\sum_{j=1}^{|t_i|}\left\{\sqrt{\left(ccst_A - ccst(d_j)\right)^2}\right\}}{|t_i|}$ // deviation of the closing cost observed during the traded days of the given time interval
- $ccst(t_i) = ccst_A - ccst_D$ // the absolute distance between average closing cost and corresponding deviation observed in a given time interval t_i

The trading Status of the given time interval denotes by the aggregate of the status referred as 1 (up), 0(stable), and -1 (down). If the resultant value is greater than zero, then the status referred as up, if less than zero, then the status referredas negative, else if the resultant value is equal to zero, the status denotes as stable.

The tweets posted in twitter trends fall into the categories (Government activities, political factors, international transactions, speculation and expectations, supply-demand, and health care) those influence the stock trading. Further partitions the data corpus stock trading and tweets from corresponding twitter trends.

Meta-heuristic Scales

Upon completion of the data collection and preprocessing, the resultant records will be used to derive meta-heuristic scales, which is as follows:

The following description details the process of identifying the correlation between the stock trading status and positive n-grams and negative n-grams.

The given processed records as corpus shall use to define the proposed recommendation system. In this regard, n-gram patterns of positive opinion and negative opinion has to derive initially, which is as follows:

1-gram patterns

The overall positive tokens of all records have to considered as a set pT without duplicate entries, that further referred to as 1-gram positive tokens. Similarly, the overall negative tokens shall consider as set nT without duplicate entries, which have done as follows

- $\forall_{i=1}^{|C|}\{pNt_i, nNt_i\}$ Begin
  - $\left.\begin{matrix} pT \leftarrow pT \cup pNt_i \\ nT \leftarrow nT \cup nNt \end{matrix}\right\}$ // finding one-grams as sets pT, nT of positive and negative tokens in respective order
- End

Further, for each one-gram of both positive and negative lists, list the records having the corresponding one-gram as sets, which is as follows:

- $\forall_{i=1}^{|pT|}\{t_i \exists t_i \in pT\}$ // Begin
  - $\forall_{j=1}^{|C|}\{nPt_j \exists nPt_j \in r_j \wedge r_j \in C\}$ // Begin
    - $if\,(t_i \in nPt_i)\ pRs(t_i) \leftarrow r_i$ // adding the record r_i to the set $pRs(t_i)$ that contains all the records having the one-gram t_i in positive tokens
  - End
- End
- $\forall_{i=1}^{|nT|}\{t_i \exists t_i \in nT\}$ // Begin
  - $\forall_{j=1}^{|C|}\{nNt_j \exists nNt_j \in r_j \wedge r_j \in C\}$ // Begin
    - $if\,(t_i \in nNt_i)\ nRs(t_i) \leftarrow r_i$ // adding the record r_i to the set $nRs(t_i)$ that contains all the records having the one-gram t_i in negative tokens
  - End
- End

Discovering n-gram positive and negative patterns

Do{

Let the notation tpT is an empty set

Let the notation $sz = |pT|$ denotes the size of the set pT

$\forall_{i=1}^{|pT|}\{t_i \exists t_i \in pT\}$ begin // for each gram exists

$\forall_{j=1}^{|pT|}\{t_j \exists t_j \in pT \wedge t_j \neq t_i\}$ // Begin

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

$t = t_i \cup t_j$
$if\left(t \notin pT\right)$ // Begin
$pRs(t) = pRs(t_i) \cap pRs(t_j)$
$tpT \leftarrow t$
End
End
End
$if\left(\left|tpT\right| > 0\right) \wedge \left(|pT| < |\left(pT \cup tpT\right)|\right)\ pT = pT \cup tpT$

} While $(sz < |pT|)$

After completion of the required iterations of the aforementioned loop, the notation pT contains all possible n-gram patterns of the positive tokens. similarly, the all possible n-gram tokens nT for negative label shall define, which is as follows

Do {
Let the notation tnT is an empty set
Let the notation $sz = |nT|$ denotes the size of the set nT
$\overset{|nT|}{\underset{i=1}{\forall}}\{t_i \exists t_i \in nT\}$ begin // for each gram exists
$\overset{|nT|}{\underset{j=1}{\forall}}\{t_j \exists t_j \in nT \wedge t_j \neq t_i\}$ // Begin
$t = t_i \cup t_j$
$if\left(t \notin nT\right)$ // Begin
$nRs(t) = nRs(t_i) \cap nRs(t_j)$
$tnT \leftarrow t$
End
End
End
$if\left(\left|tnT\right| > 0\right) \wedge \left(|nT| < |\left(nT \cup tnT\right)|\right)\ nT = nT \cup tnT$

} While $(sz < |nT|)$

Defining Heuristic-scales for Positive n-grams

Further discovers the heuristic scales for each of the other feature listed in the corpus, which is traced from the given input of stock trading transactions. The heuristic scales have to define as follows:

$\overset{|pT|}{\underset{i=1}{\forall}}\{t_i \exists t_i \in pT\}$ Begin // for each n-gram of positive token

//Finding metaheuristics for each observation of the stock trades of the respective sector

//finding heuristics of the metric "quantity of stocks"

$$qS_{ep} = \frac{\sum_{j=1}^{|pRs(t_i)|}\left\{qS_j \exists qS_j \in r_j \wedge r_j \in pRs(t_i)\right\}}{|pRs(t_i)|} //$$

finding the empirical probability of the quantity of stocks listed in the records $pRs(t_i)$ those contains n-gram positive token t_i

$$qS_{md} = \frac{\sum_{j=1}^{|pRs(t_i)|}\left\{\sqrt{\left(qS_{ep} - qS_j\right)^2} \exists qS_j \in r_j \wedge r_j \in pRs(t_i)\right\}}{|pRs(t_i)|}$$

// finding mean deviation of the quantity of stocks of the records listed in set $pRs(t_i)$

the heuristics of the quantity of the stocks are,

$$\left.\begin{array}{l} qS^+_{\min} = qS_{ep} - qS_{md} \\ qS^+_{\max} = qS_{ep} + qS_{md} \end{array}\right\}$$ //the quantity of stocks

towards n-gram positive pattern ranges betweenthe minimum quantity of stocks $qS^+_{\min}$ and the maximum quantity of stocks $qS^+_{\max}$

// finding heuristics of the opening cost is as follows:

$$ocst_{ep} = \frac{\sum_{j=1}^{|pRs(t_i)|}\left\{ocst_j \exists ocst_j \in r_j \wedge r_j \in pRs(t_i)\right\}}{|pRs(t_i)|} //$$

finding the empirical probability of the opening cost listed in the records $pRs(t_i)$ those contains n-gram positive token t_i

$$ocst_{md} = \frac{\sum_{j=1}^{|pRs(t_i)|}\left\{\sqrt{\left(ocst_{ep} - ocst_j\right)^2} \exists ocst_j \in r_j \wedge r_j \in pRs(t_i)\right\}}{|pRs(t_i)|} //$$

finding the mean deviation of the open cost of the records listed in set $pRs(t_i)$

$$\left.\begin{array}{l} ocst_{\min} = ocst_{ep} - ocst_{md} \\ ocst_{\max} = ocst_{ep} + ocst_{md} \end{array}\right\}$$ the opening cost

towards n-gram positive pattern ranges between the minimum opening cost of the stocks $ocst_{\min}$ and maximum opening cost $ocst_{\max}$

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

// finding heuristics of the closing cost is as follows:

$$ccst_{ep} = \frac{\sum_{j=1}^{|pRs(t_i)|}\{ccst_j \exists ccst_j \in r_j \wedge r_j \in pRs(t_i)\}}{|pRs(t_i)|} //$$

finding the empirical probability of the closing cost listed in the records $pRs(t_i)$ those contains n-gram positive token t_i

$$ccst_{md} = \frac{\sum_{j=1}^{|pRs(t_i)|}\left\{\sqrt{\left(ccst_{ep} - ccst_j\right)^2} \exists ccst_j \in r_j \wedge r_j \in pRs(t_i)\right\}}{|pRs(t_i)|}$$

// finding the mean deviation of the closing cost of the records listed in set $pRs(t_i)$

$$\left.\begin{aligned} ccst^{+}_{\min} &= ccst_{ep} - ccst_{md} \\ ccst^{+}_{\max} &= ccst_{ep} + ccst_{md} \end{aligned}\right\}$$ the closing cost towards n-gram positive pattern ranges between the minimum closing cost of the stocks $ccst^{+}_{\min}$ and maximum closing cost of the stocks $ccst^{+}_{\max}$

//The heuristics of the aggregate cost have estimated as follows:

$$acst_{ep} = \frac{\sum_{j=1}^{|pRs(t_i)|}\{acst_j \exists acst_j \in r_j \wedge r_j \in pRs(t_i)\}}{|pRs(t_i)|} //$$

finding the empirical probability of the aggregate cost listed in the records $pRs(t_i)$ those contains n-gram positive token t_i

$$acst_{md} = \frac{\sum_{j=1}^{|pRs(t_i)|}\left\{\sqrt{\left(acst_{ep} - acst_j\right)^2} \exists acst_j \in r_j \wedge r_j \in pRs(t_i)\right\}}{|pRs(t_i)|}$$

// finding the mean deviation of the aggregate cost of the records listed in set $pRs(t_i)$

$$\left.\begin{aligned} acst^{+}_{\min} &= acst_{ep} - acst_{md} \\ acst^{+}_{\max} &= acst_{ep} + acst_{md} \end{aligned}\right\}$$ the aggregate cost towards n-gram positive pattern ranges between minimum aggregate stock $acst^{+}_{\min}$ and maximum aggregate cost of stocks $acst^{+}_{\max}$

The index status under each positive n-gram has to estimate as follows:

$$sts = \sum_{j=1}^{|pRs(t_i)|}\{sts_j \exists sts_j \in r_j \wedge r_j \in pRs(t_i)\}$$

$if\,(sts > 0)\;\; sts^{+} = 1$

$elseif\,(sts < 0)\;\; sts^{+} = -1$

$else\; sts^{+} = 0$

- End

Finding Heuristic scales for Negative n-grams

Similarly, the meta-heuristics of the stock trading elements under the negative opinion of the twitter trends shall discover as follows:

$\forall_{i=1}^{|nT|}\{t_i \exists t_i \in nT\}$ Begin // for each n-gram of negative token

//Finding metaheuristics for each observation of the stock trades of the respective sector

//finding heuristics of the metric "quantity of stocks"

$$qS_{ep} = \frac{\sum_{j=1}^{|nRs(t_i)|}\{qS_j \exists qS_j \in r_j \wedge r_j \in nRs(t_i)\}}{|nRs(t_i)|} //$$

finding the empirical probability of the quantity of stocks listed in the records $nRs(t_i)$ those contains n-gram negative token t_i

$$qS_{md} = \frac{\sum_{j=1}^{|nRs(t_i)|}\left\{\sqrt{\left(qS_{ep} - qS_j\right)^2} \exists qS_j \in r_j \wedge r_j \in nRs(t_i)\right\}}{|nRs(t_i)|}$$

//finding mean deviation of the quantity of stocks of the records listed in set $nRs(t_i)$

the heuristics of the quantity of the stocks are,

$$\left.\begin{aligned} qS^{-}_{\min} &= qS_{ep} - qS_{md} \\ qS^{-}_{\max} &= qS_{ep} + qS_{md} \end{aligned}\right\}$$ //the quantity of stocks towards n-gram negative pattern ranges between the minimum quantity of stocks $qS^{-}_{\min}$ and the maximum quantity of stocks $qS^{-}_{\max}$

// finding heuristics of the opening cost is as follows:

$$ocst_{ep} = \frac{\sum_{j=1}^{|nRs(t_i)|}\{ocst_j \exists ocst_j \in r_j \wedge r_j \in nRs(t_i)\}}{|nRs(t_i)|} //$$

finding the empirical probability of the

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

opening cost listed in the records $nRs(t_i)$ those contains n-gram negative token t_i

$$ocst_{md} = \frac{\sum_{j=1}^{|nRs(t_i)|} \left\{ \sqrt{\left(ocst_{ep} - ocst_j\right)^2} \exists ocst_j \in r_j \wedge r_j \in nRs(t_i) \right\}}{|nRs(t_i)|}$$

// finding the mean deviation of the open cost of the records listed in set $nRs(t_i)$

$$\left.\begin{array}{l} ocst_{\min} = ocst_{ep} - ocst_{md} \\ ocst_{\max} = ocst_{ep} + ocst_{md} \end{array}\right\}$$ the opening cost towards n-gram negative pattern ranges between the minimum opening cost of the stocks $ocst_{\min}$ and maximum opening cost $ocst_{\max}$

// finding heuristics of the closing cost is as follows:

$$ccst_{ep} = \frac{\sum_{j=1}^{|nRs(t_i)|} \left\{ ccst_j \exists ccst_j \in r_j \wedge r_j \in nRs(t_i) \right\}}{|nRs(t_i)|}$$ //

finding the empirical probability of the closing cost listed in the records $nRs(t_i)$ those contains n-gram negative token t_i

$$ccst_{md} = \frac{\sum_{j=1}^{|nRs(t_i)|} \left\{ \sqrt{\left(ccst_{ep} - ccst_j\right)^2} \exists ccst_j \in r_j \wedge r_j \in nRs(t_i) \right\}}{|nRs(t_i)|}$$

// finding the mean deviation of the closing cost of the records listed in set $nRs(t_i)$

$$\left.\begin{array}{l} ccst^{-}_{\min} = ccst_{ep} - ccst_{md} \\ ccst^{-}_{\max} = ccst_{ep} + ccst_{md} \end{array}\right\}$$ the closing cost towards n-gram negative pattern ranges between the minimum closing cost of the stocks $ccst^{-}_{\min}$ and maximum closing cost of the stocks $ccst^{-}_{\max}$

//The heuristics of the aggregate cost have estimated as follows

// finding heuristics of the closing cost is as follows:

$$acst_{ep} = \frac{\sum_{j=1}^{|nRs(t_i)|} \left\{ acst_j \exists acst_j \in r_j \wedge r_j \in nRs(t_i) \right\}}{|nRs(t_i)|}$$ //

finding the empirical probability of the aggregate cost listed in the records $nRs(t_i)$ those contains n-gram negative token t_i

$$acst_{md} = \frac{\sum_{j=1}^{|nRs(t_i)|} \left\{ \sqrt{\left(acst_{ep} - acst_j\right)^2} \exists acst_j \in r_j \wedge r_j \in nRs(t_i) \right\}}{|nRs(t_i)|}$$

// finding the mean deviation of the aggregate cost of the records listed in set $nRs(t_i)$

$$\left.\begin{array}{l} acst^{-}_{\min} = acst_{ep} - acst_{md} \\ acst^{-}_{\max} = acst_{ep} + acst_{md} \end{array}\right\}$$ the aggregate cost towards n-gram negative pattern ranges between minimum aggregate stock $acst^{-}_{\min}$ and maximum aggregate cost of stocks $acst^{-}_{\max}$

//The index status under each negative n-gram has to estimate as follows:

$$sts = \sum_{j=1}^{|nRs(t_i)|} \left\{ sts_j \exists sts_j \in r_j \wedge r_j \in nRs(t_i) \right\}$$

$if\ (sts > 0)\ sts^{-} = 1$

$elseif\ (sts < 0)\ sts^{-} = -1$

$else\ sts^{-} = 0$

- End

Stock Trading recommendations using influential Twitter trends

The given tweets of the twitter trends linked to the corresponding stock trading sector have to use as input to predict the range of values for the trading factors such as opening cost, closing cost, aggregate cost, and index status. In this regard, the tweets posted under the trends acclaimed as influential factors of the stock trading have to consider as the input corpus of the proposed recommendation system. Further, perform the process detailed in the following description.

$\forall_{i=1}^{|T|} \{t_i \exists t_i \in T\}$ Begin //For each tweet $\{t \exists t \in T\}$ of the given corpus T,

Tokenize the tweet t as a word vector wv,

Remove noise,stop-words and sentiment lexicons. Further,

Perform stemming (removing ed and ing forms) on the leftover tokens of the word vector wv.

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

$vT = vT \cup wv$ // move all the tokens of the word vector wv to the set vT, which contains unique word tokens (no duplicates).

End

Later, build all possible n-gram patterns from the tokens listed in the set vT (see sections 3.3.1 and 3.3.2) and list them as a set nG

$\forall_{i=1}^{|nG|}\{ng_i \exists ng_i \in nG \wedge ng_i \in pT\}$ begin //For each n-gram pattern $\{ng_i \exists ng_i \in nG\}$ that exists in the set pT

Further, fetch the heuristics derived for the other trading factors (quantity of stocks, opening price, closing price, aggregate price, and status)

$\left.\begin{matrix} OC^{+}_{\min} \leftarrow ocst^{+}_{\min}(ng_i) \\ OC^{+}_{\max} \leftarrow ocst^{+}_{\max}(ng_i) \end{matrix}\right\}$ //move the open cost heuristics $ocst^{+}_{\min}, ocst^{+}_{\max}$ observed for positive n-gram $\{ng_i \exists ng_i \in pT\}$ to respective sets $OC^{+}_{\min}, OC^{+}_{\max}$, which discovered during the training phase of the proposed model

$\left.\begin{matrix} CC^{+}_{\min} \leftarrow ccst^{+}_{\min}(ng_i) \\ CC^{+}_{\max} \leftarrow ccst^{+}_{\max}(ng_i) \end{matrix}\right\}$ //move the closing cost heuristics $ccst^{+}_{\min}, ccst^{+}_{\max}$ observed for positive n-gram $\{ng_i \exists ng_i \in pT\}$ to respective sets $CC^{+}_{\min}, CC^{+}_{\max}$, which discovered during the training phase of the proposed model

$\left.\begin{matrix} AC^{+}_{\min} \leftarrow acst^{+}_{\min}(ng_i) \\ AC^{+}_{\max} \leftarrow acst^{+}_{\max}(ng_i) \end{matrix}\right\}$ //move the aggregate cost heuristics $acst^{+}_{\min}, acst^{+}_{\max}$ observed for positive n-gram $\{ng_i \exists ng_i \in pT\}$ to respective sets $AC^{+}_{\min}, AC^{+}_{\max}$, which discovered during the training phase of the proposed model

$STS^{+} \leftarrow sts^{+}(ng_i)$ //move the status sts^{+} observed for positive n-gram $\{ng_i \exists ng_i \in pT\}$ to respective set STS^{+}, which discovered during the training phase of the proposed model

End

// similarly list the all heuristics observed for negative n-gram patterns as follow,

$\forall_{i=1}^{|nG|}\{ng_i \exists ng_i \in nG \wedge ng_i \in nT\}$ begin //For each n-gram pattern $\{ng_i \exists ng_i \in nG\}$ that exists in the set nT

$\left.\begin{matrix} OC^{-}_{\min} \leftarrow ocst^{-}_{\min}(ng_i) \\ OC^{-}_{\max} \leftarrow ocst^{-}_{\max}(ng_i) \end{matrix}\right\}$ //move the open cost heuristics $ocst^{-}_{\min}, ocst^{-}_{\max}$ observed for negative n-gram $\{ng_i \exists ng_i \in nT\}$ to respective sets $OC^{-}_{\min}, OC^{-}_{\max}$, which discovered during the training phase of the proposed model

$\left.\begin{matrix} CC^{-}_{\min} \leftarrow ccst^{-}_{\min}(ng_i) \\ CC^{-}_{\max} \leftarrow ccst^{-}_{\max}(ng_i) \end{matrix}\right\}$ //move the closing cost heuristics $ccst^{-}_{\min}, ccst^{-}_{\max}$ observed for negative n-gram $\{ng_i \exists ng_i \in nT\}$ to respective sets $CC^{-}_{\min}, CC^{-}_{\max}$, which discovered during the training phase of the proposed model

$\left.\begin{matrix} AC^{-}_{\min} \leftarrow acst^{-}_{\min}(ng_i) \\ AC^{-}_{\max} \leftarrow acst^{-}_{\max}(ng_i) \end{matrix}\right\}$ //move the aggregate cost heuristics $acst^{-}_{\min}, acst^{-}_{\max}$ observed for negative n-gram $\{ng_i \exists ng_i \in nT\}$ to respective sets $AC^{-}_{\min}, AC^{-}_{\max}$, which discovered during the training phase of the proposed model

$STS^{-} \leftarrow sts^{-}(ng_i)$ //move the status sts^{+} observed for negative n-gram $\{ng_i \exists ng_i \in nT\}$ to respective set STS^{-}, which discovered during the training phase of the proposed model

End

Further, estimates the recommended values of the respective trading factors as follows

Find theaverages and deviations of the minimum quantity of stocks $\{rqs^{+}_{\min} \pm rqs^{+}_{dmi}\}$, open costs $\{roc^{+}_{\min} \pm roc^{+}_{dmi}\}$, closing costs $\{rcc^{+}_{\min} \pm rcc^{+}_{dmi}\}$, and aggregate costs $\{rac^{+}_{dmi} \pm rac^{+}_{dmi}\}$ observed from positive n-grams (see in Table 1).

Similarly, discover averages and deviations of the maximum quantity of stocks $\{rqs^{+}_{\max} \pm rqs^{+}_{dma}\}$,

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

open costs $\{roc^{+}_{\max} \pm roc^{+}_{dma}\}$, closing costs $\{rcc^{+}_{\max} \pm rcc^{+}_{dma}\}$, and aggregate costs $\{rac^{+}_{\max} \pm rac^{+}_{dma}\}$ observed from positive n-grams.

Table 1: Mean and deviation values of the trading factors under the positive opinion

	MIN	MAX
Quantity of stocks	$\{rqs^{+}_{\min} \pm rqs^{+}_{dmi}\}$	$\{rqs^{+}_{\max} \pm rqs^{+}_{dma}\}$
Open cost	$\{roc^{+}_{\min} \pm roc^{+}_{dmi}\}$	$\{roc^{+}_{\max} \pm roc^{+}_{dma}\}$
Closing cost	$\{rcc^{+}_{\min} \pm rcc^{+}_{dmi}\}$	$\{rcc^{+}_{\max} \pm rcc^{+}_{dma}\}$
Aggregate cost	$\{rac^{+}_{dmi} \pm rac^{+}_{dmi}\}$	$\{rac^{+}_{\max} \pm rac^{+}_{dma}\}$

Further discover these metrics for negative n-grams, which have listed in the Table 2,

Table 2: Mean and deviation values of the trading factors under the negative opinion

	MIN	MAX
Quantity of stocks	$\{rqs^{-}_{\min} \pm rqs^{-}_{dmi}\}$	$\{rqs^{-}_{\max} \pm rqs^{-}_{dma}\}$
Open cost	$\{roc^{-}_{\min} \pm roc^{-}_{dmi}\}$	$\{roc^{-}_{\max} \pm roc^{-}_{dma}\}$
Closing cost	$\{rcc^{-}_{\min} \pm rcc^{-}_{dmi}\}$	$\{rcc^{-}_{\max} \pm rcc^{-}_{dma}\}$
Aggregate cost	$\{rac^{-}_{dmi} \pm rac^{-}_{dmi}\}$	$\{rac^{-}_{\max} \pm rac^{-}_{dma}\}$

Further, derive the cumulative trading factors scope under both positive and negative n-grams as follows,

For positive n-grams:

- $\min_{+ve} = \{1-(rqs^{+}_{\min} \otimes roc^{+}_{\min} \otimes rcc^{+}_{\min} \otimes rac^{+}_{\min})\}$ //discovering the minimum heuristic value of all trading factors under positive opinion.
- $\max_{+ve} = \{1-(rqs^{+}_{\max} \otimes roc^{+}_{\max} \otimes rcc^{+}_{\max} \otimes rac^{+}_{\max})\}$ // discovering the maximum heuristic value of all trading factors under positive opinion.

For Negative n-grams:

- $\min_{-ve} = \{1-(rqs^{-}_{\min} \otimes roc^{-}_{\min} \otimes rcc^{-}_{\min} \otimes rac^{-}_{\min})\}$ // discovering the minimum heuristic value of all trading factors under negative opinion.
- $\max_{-ve} = \{1-(rqs^{-}_{\max} \otimes roc^{-}_{\max} \otimes rcc^{-}_{\max} \otimes rac^{-}_{\max})\}$ // discovering the minimum heuristic value of all trading factors under negative opinion.

Similarly, discover the deviation values of the minimum and maximum heuristics of all trading factors under both positive and negative opinions, which is as follows:

For positive n-grams:

- $\min^{d}_{+ve} = \{1-(rqs^{+}_{dmi} \otimes roc^{+}_{dmi} \otimes rcc^{+}_{dmi} \otimes rac^{+}_{dmi})\}$ //discovering the deviation of minimum heuristic value of all trading factors under positive opinion.
- $\max^{d}_{+ve} = \{1-(rqs^{+}_{dma} \otimes roc^{+}_{dma} \otimes rcc^{+}_{dma} \otimes rac^{+}_{dma})\}$ // discovering the deviation of maximum heuristic value of all trading factors under positive opinion.

For negative n-grams:

- $\min^{d}_{-ve} = \{1-(rqs^{-}_{dmi} \otimes roc^{-}_{dmi} \otimes rcc^{-}_{dmi} \otimes rac^{-}_{dmi})\}$ //discovering the deviation of minimum heuristic value of all trading factors under negative opinion.
- $\max^{d}_{-ve} = \{1-(rqs^{-}_{dma} \otimes roc^{-}_{dma} \otimes rcc^{-}_{dma} \otimes rac^{-}_{dma})\}$ // discovering the deviation of maximum heuristic value of all trading factors under positive opinion.

The recommendations of the trading factors under the influence of positive and negative opinions shall define follows:

Relation of heuristics discovered➔	$\{((\min_{+ve} > \min_{-ve}) \wedge (\max_{+ve} > \max_{-ve}))\}\ or$ $\left\{\begin{matrix}((\min_{+ve} > \min_{-ve}) \vee (\max_{+ve} > \max_{-ve})) \wedge \\ ((\min^{d}_{+ve} < \min^{d}_{-ve}) \wedge (\max^{d}_{+ve} < \max^{d}_{-ve}))\end{matrix}\right\}$		
		Min	max
Recommendations	Quantity of stocks	$\{rqs^{+}_{\min} \pm rqs^{+}_{dmi}\}$	$\{rqs^{+}_{\max} \pm rqs^{+}_{dma}\}$

to the expected values of the trading factors	Open cost	$\left\{roc^{+}_{\min} \pm roc^{+}_{dmi}\right\}$	$\left\{roc^{+}_{\max} \pm roc^{+}_{dma}\right\}$
	Closing cost	$\left\{rcc^{+}_{\min} \pm rcc^{+}_{dmi}\right\}$	$\left\{rcc^{+}_{\max} \pm rcc^{+}_{dma}\right\}$
	Aggregate cost	$\left\{rac^{+}_{dmi} \pm rac^{+}_{dmi}\right\}$	$\left\{rac^{+}_{\max} \pm rac^{+}_{dma}\right\}$
	status	$\left\{\begin{matrix} 1(up) \because if \left(\sum_{i=1}^{\|STS^{+}\|}\left\{sts_i \exists sts_i \in STS^{+}\right\}\right) > 0 \\ -1(down) \because if \left(\sum_{i=1}^{\|STS^{+}\|}\left\{sts_i \exists sts_i \in STS^{+}\right\}\right) < 0 \\ 0(no\,change) \because if \left(\sum_{i=1}^{\|STS^{+}\|}\left\{sts_i \exists sts_i \in STS^{+}\right\}\right) = 0 \end{matrix}\right\}$	
Relation of heuristics discovered➔	$\left\{\left(\left(\min_{+ve} < \min_{-ve}\right) \wedge \left(\max_{+ve} < \max_{-ve}\right)\right)\right\} or \left\{\begin{matrix}\left(\left(\min_{+ve} < \min_{-ve}\right) \vee \left(\max_{+ve} < \max_{-ve}\right)\right) \wedge \\ \left(\left(\min^{d}_{+ve} > \min^{d}_{-ve}\right) \wedge \left(\max^{d}_{+ve} > \max^{d}_{-ve}\right)\right)\end{matrix}\right\}$		
		Min	max
Recommendations to the expected values of the trading factors	Quantity of stocks	$\left\{rqs^{-}_{\min} \pm rqs^{-}_{dmi}\right\}$	$\left\{rqs^{-}_{\max} \pm rqs^{-}_{dma}\right\}$
	Open cost	$\left\{roc^{-}_{\min} \pm roc^{-}_{dmi}\right\}$	$\left\{roc^{-}_{\max} \pm roc^{-}_{dma}\right\}$
	Closing cost	$\left\{rcc^{-}_{\min} \pm rcc^{-}_{dmi}\right\}$	$\left\{rcc^{-}_{\max} \pm rcc^{-}_{dma}\right\}$
	Aggregate cost	$\left\{rac^{-}_{dmi} \pm rac^{-}_{dmi}\right\}$	$\left\{rac^{-}_{\max} \pm rac^{-}_{dma}\right\}$
	status	$\left\{\begin{matrix} 1(up) \because if \left(\sum_{i=1}^{\|STS^{-}\|}\left\{sts_i \exists sts_i \in STS^{+}\right\}\right) > 0 \\ -1(down) \because if \left(\sum_{i=1}^{\|STS^{-}\|}\left\{sts_i \exists sts_i \in STS^{+}\right\}\right) < 0 \\ 0(no\,change) \because if \left(\sum_{i=1}^{\|STS^{-}\|}\left\{sts_i \exists sts_i \in STS^{+}\right\}\right) = 0 \end{matrix}\right\}$	

V. EXPERIMENTAL STUDY

This section explains the significance of the proposed model STROM that compared to another benchmark method PSMV [24], and our earlier contribution "Distribution Diversity based Feature Optimization for Stock Trading Predictions using Supervised Learning (STPSL)". The data was collected from twitter trends [25] and stock trading information streams [26], which are highly correlated stock trading sector and twitter trends of current affairs.

Accuracy

The metric accuracy utilized for describing the approximations of measurement towards true-value, which is the ratio of the count of true values against the total count of records given for classification. The statistics observed for this metric from proposed model STROM, contemporary model PSMV, and earlier model STPSL are briefed in the following.

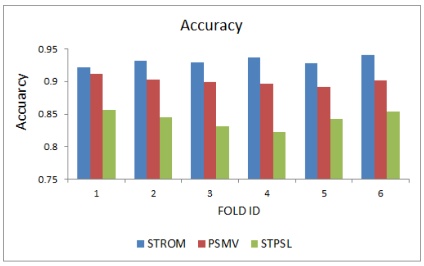

Figure 1: Depiction of the average standard deviation of accuracy for the STROM, PSMV and STPSLmethods over six-folds

January - February 2020
ISSN: 0193 - 4120 Page No. 16641 - 16655

In Figure 1, statistics are plotted between accuracy and six folds for the STROM, PSMV and STPSL methods. From the figure, we can evince that the average standard deviation for the STROM, PSMV, and STPSL methods are 0.931167 ± 0.00644, 0.900667 ± 0.006263 and 0.841833 ± 0.012047 in respective order. Therefore, it is envisioned from the statistics that, the STROM is more significant when compared to PSMV and STPSLmethods.

Precision

The metric precision denotes the positive predictive value, which is the ratio of the true positives against the aggregate of true positives and false negatives. The statistics observed for this metric from proposed model STROM and the other contemporary model PSMV, and earlier model STPSL has briefed in the following:

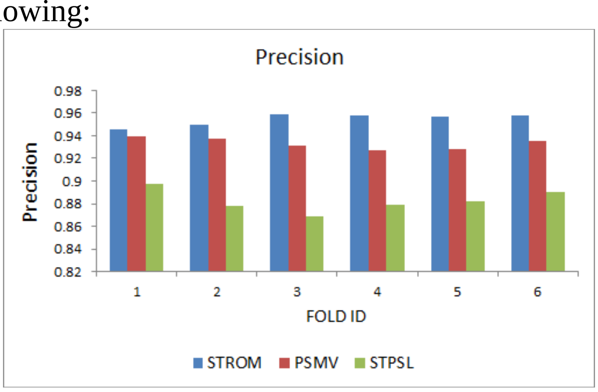

Figure 2: Depiction of the average standard deviation of precision for the STROM, PSMV and STPSL methods over six-folds

In Figure 2, the graph is drawn between precision and six folds for the STROM, PSMV and STPSL methods. The average standard deviation for the STROM is 0.954117 ± 0.005009, for the PSMV is 0.933367 ± 0.0046 and STPSL is 0.882583 ± 0.009149. Hence, it is exhibited that the STROMperforms better than PSMV and STPSL methods.

Sensitivity

The true-positive-rate of the classification process is denoted by the metric called Sensitivity, which is the ratio of true positives against the aggregate of the true positives and false negatives. The statistics of the sensitivity from the four folds classification carried by STROM, PSMV, and STPSL has briefed in the following figure.

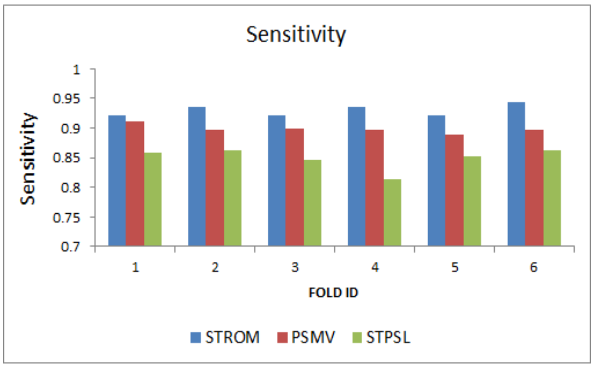

Figure 3: Depiction of the average standard deviation of sensitivity for the STROM, PSMV and STPSLmethods over six-folds

The graph is plotted between sensitivity and 6 folds for the STROM, comparison, and earlier methods. From Figure 3, we can evince that the average standard deviation for the STROM, PSMV, and STPSL methods are 0.929933±0.008756, 0.8987±0.006735 and 0.849433±0.016668 in respective order. Therefore, it is envisioned from the statistics that, the STROM method is more significant when compared to PSMV and STPSLmethods.

Specificity

The metric specificity is resourceful to identify the true-negative-rate of the classification process, which is the ratio of true negatives against the aggregate of true negatives and false positives. The specificity indicates the optimality of the classifier to identify the negative label. The statistics observed for the metric specificity from STROM, PSMV, and STPSL has briefed in the following figure:

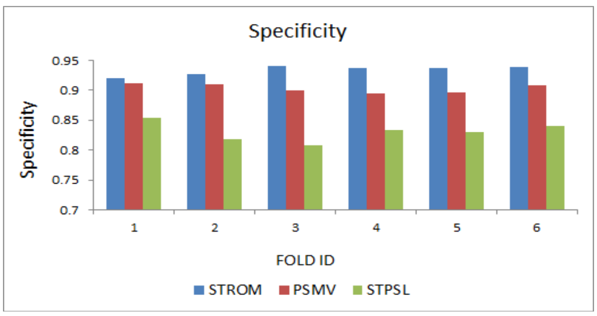

Figure 4: Depiction of the average standard deviation of specificity for the STROM, PSMV and STPSL methodsover six-folds

In Figure 4, the graph is drawn between specificity and six folds for the STROM, PSMV and STPSLmethods. The average standard deviation for the STROM is 0.933± 0.007191, for the PSMV is 0.9035± 0.006818 and STPSL is 0.830533 ± 0.014631. Hence, it is exhibited that the STROM method performs better than PSMV and STPSLmethods.

F-measure

The metric F-measure denotes the consistency of values portrayed for precision and sensitivity of binary classification process, which is double of the ratio of "product of precision and recall" against the sum of the corresponding precision & recall.The statistics of the F-measure from the four folds classification carried by the STROM, PSMV, and STPSL has briefed in the following figure.

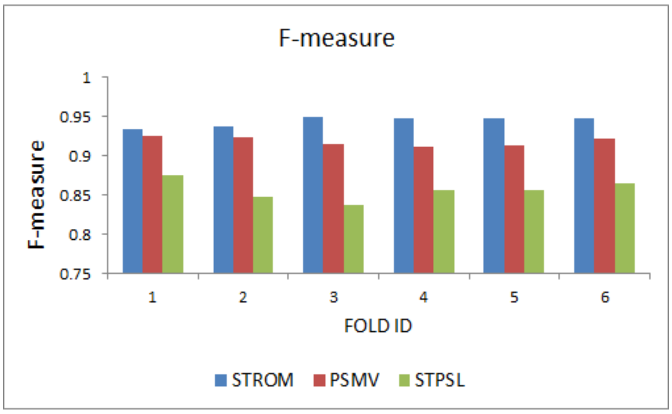

Figure 5: Depiction of the average standard deviation of F-measure for the STROM, PSMV and STPSLmethods over six-folds

The graph is plotted between F-measure and 6 folds for the STROM, PSMV and STPSLmethods. From Figure5, we can evince that the average standard deviation for the STROM, PSMV, and STPSL methods are 0.94345± 0.006115, 0.918183 ± 0.005752 and 0.85575 ± 0.011978 in respective order. Therefore, it is envisioned from the statistics that, the STROM is more significant when compared to PSMV and STPSLmethods.

Matthews's correlation coefficient (MCC)

The metric "Matthews correlation coefficient (MCC)" is utilized to denote the consistency of the binary classification against the diversified sizes of both classes. The stated metric derives the correlation between projected and predicted classes of the given test data. The MCC estimation covers as follows.

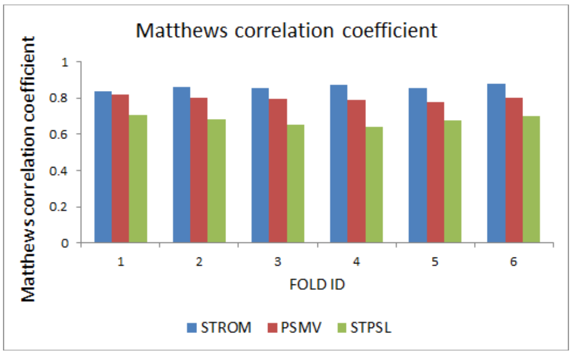

Figure 6: Depiction of the average standard deviation of MCC for the STROM, PSMV and STPSL methods over six-folds

The graph is drawn between MCC and six folds for the STROM, PSMV and STPSLmethods as portrayed in Figure 6. The average standard deviation for the STROM method is 0.857983 ± 0.013063, for the PSMV is 0.795833 ± 0.012709 and STPSL is 0.67445 ± 0.023693. Hence, it is exhibited that the STROM method performs better than PSMV and STPSLmethods.

VI. CONCLUSION

This manuscript presented a machine learning approach for "Stock Trading Recommendations by Opinion Mining (STROM)", which aimed to predict the stock trading factors through opinion mining on trends highly correlate to a business sector of stock trading. The outcomes exhibited that the opinion mining on twitter trends is optimal to perform recommendations on stock trading factors that compared to PSMV [24], and STPSL methods. The prediction accuracy observed for the proposed method 93%, which is significant that PSMC, and STPSL, which exhibited the accuracy as 90%, 84% in respective order. The future research shall endeavour to define a novel method to predict the stock trading factors through evolutionary techniques to handle the crux of the high dimensionality of opinions and stock trading factors

. As for return on equity, the relationship between this financial ratio and the share price is unidentified due to the contradicting results that were obtained from this study. The findings of this study support the results from past literature conducted by Sharma (2011), Almumani's (2014), Menike & Prabath (2014), Arshad et al (2015) and Sharif et al. (2015).

REFERENCES

1. Chan, K., T., Yap, V., C., & Nur Qasrina, B, A., (2012), Stock Performance of the Property Sector in Malaysia, Journal of Modern Accounting and Auditing, 8(2), pp. 241 – 246.
2. Donald, S., G., & Lang, K., (2007), Inference with Differences in Differences and Other
3. Panel Data, The Review of Economics and Statistics, 89(2), pp. 221-223.
4. Khan, M., N., & Amanullah, (2012), Determinants of Share Prices at Karachi Stock Exchange, International Journal of Business and Management Studies, 4(1), pp. 111 – 120.
5. Malhotra, N., & Tandon, K., (2013), Determinants of Stock Prices: Empirical Evidence from NSE 100 Companies, International Journal of Research in Management & Technology, Vol. 3, No. 3, pp. 86 – 95.
6. Wang, J., Fu, G., & Luo, C., (2013), Accounting Information and Stock Price Reaction of Listed Companies – Empirical Evidence from 60 Listed Companies in Shanghai Stock Exchange, Journal of Business & Management, Vol. 2, Issue 2, pp. 11 – 21.
7. Menike, M., G., P., D., & Prabath, U., S., (2014), The Impact of Accounting Variables on Stock Price: Evidence from the Colombo Stock Exchange, Sri Lanka, International Journal of Business and Management, Vol. 9, No. 5, pp. 125 – 137.
8. Sharma, S., (2011), Determinants of Equity Prices in India, Journal of Arts, Science & Commerce, 2(4), pp. 51 – 60.
9. Arshad, Z., Arshaad, A., R., Yousaf, S., & Jamil, S., (2015), Determinants of Share Prices of Listed Commercial Banks in Pakistan, Journal of Economics and Finance, Volume 6, Issue 2, Ver. III, pp. 56 – 64.
10. Turk, A., F., & Chapman, M., (2006), The Effect of Financial Ratios and Market Hype, Illinois Wesleyan University Honors Project.
11. Paul, R., K., (2006), Multicollinearity: Causes, Effects and Remedies, IASRI, New Delhi.
12. Ghasemi, A., & Zahediasl, S., (2012), Normality Tests for Statistical Analysis: A Guide for Non-Statisticians, International Journal of Endocrinology and Metabolism, 10(2), pp. 486 – 489.