Received: April 13, 2020

International Journal of
Intelligent Engineering & Systems

http://www.inass.org/

Self-Supervised Learning Based Anomaly Detection in Online Social Media

Sujatha Arun Kokatnoor[1]* **Balachandran Krishnan[1]**

[1]*Department of Computer Science and Engineering, School of Engineering and Technology, Christ (Deemed to be University), Bangalore, India*
* Corresponding author's Email: sujatha.ak@christuniversity.in

Abstract: Online Social Media (OSM) produce enormous data related to the human behaviours based on their interactions. One such data is the opinions expressed and posted for any specific issue addressed in the OSM. Majority of the opinions posted would be categorized as positive, negative and neutral. The lighter group's opinions are termed anomalous as it is not conforming the regular opinions posted by other users. Though, lot of conventional classification and clustering based learning algorithms works well under supervised and un-supervised environment, due to the inherent ambiguity in the tweeted data, anomaly detection poses a bigger challenge in text mining. Though the data is un-supervised, for the learning purpose it is treated as Supervised Learning by assigning class labels for the training data. This paper attempts to give an insight into various anomalies of OSM and identify behavioural anomalies for a Twitter Dataset on user's opinions on demonetization policy in India. Through Self-Supervised learning, it is observed that 86% of the user's opinions did agree to the demonetization policy and the remaining have posted negative opinions for the policy implemented.

Keywords: Online social media (OSM), Structural anomalies, Behavioural anomalies, Content anomalies, Classification algorithms, Clustering algorithms, Statistical approaches, Demonetization, Logistic regression, Random forest, Naïve bayes, Support vector machine, VADER, TextBlob, K-means, Self-supervised.

1. Introduction

Anomaly detection deals with detecting data elements from a data set which is different from all the other data elements in a set. Anomalies can arise due to different reasons such as trending topics, negative opinions or feedbacks, evolution of new communities, hate speeches, flash mobs and unexpected contents in an Online Social Media [1].

It is not easy to define exactly what is an anomaly or an outlier. Fig. 1 illustrates anomalies in a simple 2-dimensional dataset. The data has two normal regions, N1 and N2, since most observations lie in these two regions. Points that are sufficiently far away from the regions, e.g., point's O_1 and O_2, and points in region O_3, are anomalies.

People use Online Social Media to socialize with friends and acquaintances, and to share information, photos, and videos. Increased usage invites illegal or legal activities like sarcasm, and local community deviations.

The anomalies can be broadly categorized into: Content Anomalies, Behavioural Anomalies and the Structural Anomalies in OSM as shown in Fig. 2.

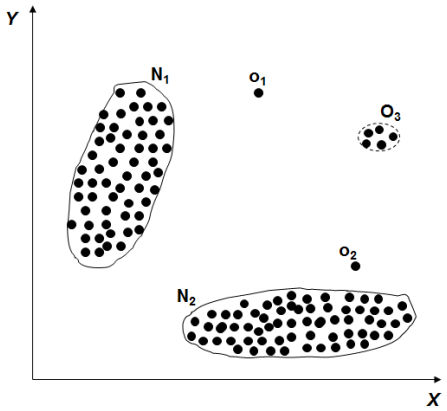

Figure. 1 A simple example of anomalies in a 2-dimensional data set

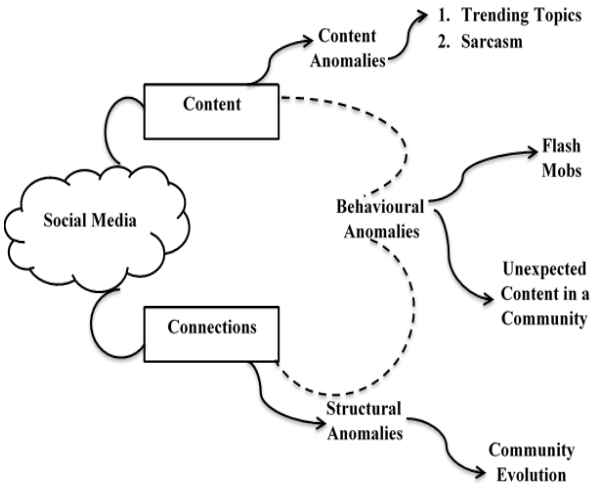

Figure. 2 Various Anomalies in OSM

Content Anomalies: Content or Point anomalies occur for data points that are considered abnormal when viewed against the whole dataset. It is the abnormal instances in data with respect to the implicit data alone. The identification includes Trending Topics, Topic Modelling and Sarcasms.

Behavioural Anomalies: This is related to the individual human behaviour. The identification includes Flash Mobs, and Unexpected Content in a Community. The behavioural attributes also define the non-contextual characteristics of an instance.

Structural Anomalies: When dealing with the graphical structures like in social networks, anomalies can be classified according to the graphical properties as well [2, 3, 4]. These are called Structural Anomalies. The identification includes the Evolution of Communities in Online Social Media.

Examining user behaviours to identify outliers in OSM is on demand. This emerging need is based on assumptions that: (1) the user behaviour and network patterns carry useful information for the social network analysers; and (2) the patterns can be linked to abnormal activities.

Twitter is one of the popular OSM where users interact with the Twitter system with messages known as Tweets. On an average more than 7, 00, 000 tweets are posted by various users per day. These tweets are the means for the users as an individual to express their opinions, thoughts, feedbacks and their feelings on various subjects and issues. The subject could be market analysis, issues related to national concern, politics, well known personalities and so on. In this paper, people's opinion on implementation of demonetization policy as posted in Twitter is considered for the study.

The opinions available are in the text format and involves text mining process for analyzing the same [5]. Text Mining is a process of extraction of patterns from a huge set of text documents. The data available is unstructured and therefore the identification of the trends and patterns for analyzing the textual data is challenging.

Once the text is processed, detection of anomalous user behaviours is generally done using supervised machine learning techniques provided, the dataset is labelled. As the dataset considered for the study is unlabelled, unsupervised machine learning approaches like K-Means and Hierarchical Clustering is applied to create clusters of similar data elements. These methods give good performance if the textual dataset considered has diversified topics for creating a dense feature vector. As the demonetization dataset considered for the study has limited topics, unsupervised approaches fail in creating appropriate clusters thereby misleading anomalous user behaviours to the normal ones. Therefore a hybrid model namely Self-Supervised learning approach is required for accurately classifying the given dataset into anomalous and non-anomalous user behaviours.

Self-Supervised Learning from the textual dataset allows the model to learn the sentiments exhibited by the users in their opinions posted in OSM with no need of the human annotated dataset. It is a two stepped process. In the 1st phase, the labelling is provided based on the semantic knowledge and the sentiments exhibited by the users. In the 2nd phase, a standard supervised classifier is used for binary classification of user behaviours into normal and anomalous behaviours. Since data can be obtained and labelled without human intervention, the advantages of using Self-Supervised Learning are that the models can be modified or trained entirely from scratch.

The following is the organization of the paper. The different types of anomalies, the various algorithms used for its identification are explored and surveyed in section 2. Section 3 is about considering one of the anomaly – behavioural anomalies for the study. Section 4 discusses the experimentation process and its results and conclusion in the section 5.

2. Related work

In this paper, different algorithms for various anomalies as mentioned in Fig. 2 are surveyed, identified different applications of OSM and its suitable algorithms. The survey incorporates recent developments in OSM. Anomalous applications of

online social networks are observed in three types of data mining approaches [6].

1. Supervised Learning Techniques
2. Semi-Supervised Learning Techniques
3. Unsupervised Learning Techniques.

2.1 Detection of structural anomalies

Structural Anomalies were detected using two phases. In the first phase simple conjugate Bayesian models were used for discreet time counting procedures and to monitor all the node's pairs in the graph. Whereas in the second phase standard network inference methods were used on a reduced subset of potentially anomalous nodes [7]. For Bayesian Models, NP-Complete structural preparation is computationally expensive and approximate.

Subdue approach [8] was used for locating blog flu groups as it used polynomial time for search. In Graph-based structure mining, for finding the similarity index, Pearson correlation coefficient was used.

Two graph metrics: ego and egonet were used and later Power-Law curve was determined to identify anomalous nodes in the given graph [9]. Anomaly scores were computed by fitting the power-law curve in the network: if a node was more distant from the curve, then the node was considered anomalous.

Detection for static and dynamic labelled and unlabelled dataset, it was generally done through statistical approaches namely for static unlabelled anomalies: detection of a star network structure, cliques, fitting a power-law curve in the anomalous network and signal processing techniques were used. For static labelled anomalies, detection of ego-nets, bipartite networks (Fraud Eagle Algorithm) and information theoretic approaches were used. For dynamic unlabelled anomalies: scan statistics, Bayesian inference, auto-regressive moving average (ARMA) model and link prediction methods were used and finally for dynamic labelled anomalies: signal processing approaches were used [10].

Network embedding methodology was used for structural anomaly detection. AScore was employed to seamlessly connect structural contradictions with the embedding and to differentiate anomalies from regular nodes using a single parameter [11]. An entropy based approach namely Shannon entropy method was used and later power law distribution was used [12] to detect structural malwares.

To detect the structural anomalies, the vertices of the network were encoded to its vector representation by a method called Clique Embedding". This method reduced the pairwise distance of vertex representation of the network. Later Reservoir sampling method was used to compute vector representation. Finally K-Means clustering algorithm was used to detect anomalous vertices/nodes [13, 14].

DBSCAN (Density Based Spatial Clustering of Applications with Noise) algorithm [15] was used to track those data points which doesn't fit in any of the clusters in a given network structure thereby detecting anomalous structures present in a network.

2.2 Detection of content anomalies

Latent Dirichlet Allocation (LDA)-based statistical topic models were used to evaluate the textual logs ' contents [16, 17] in order to identify the different topics / concepts within these logs at different theme levels.

Infinite LDA by considering the bias role of words were also used to divide the topics [18] for identifying the trending topics. TPMTM, a two-phase modeling method, which combined statistical modeling (LDA) with regular pattern mining were used which provided more detailed descriptions of rich subjects and semantics [19]. Mixture of Gaussian Mixture Models (GMM) and LDA Algorithm can also be used to detect anomalous topics [20].

For continuous datasets, statistical test charts were used to resolve the problems of anomaly detection. C4.5, using the entropy knowledge principle, is an improved version of ID3 decision trees based on training data was used for content anomaly detection [21].

Spam mail detection was proposed by the K-Nearest Neighbor method, integrating a correlation factor of Spearman with the distance measurement [22]. SVM classification method was used for labelled dataset for topic modeling [23].

Textual anomaly detection takes place through the design and enhancement, through an enhanced Expectation Maximization (EM) algorithm, of a multinomial Naïve Bayes classification. It used large quantities of unlisted data and improved the precision of the classification of Naive Bayes by using an EM algorithm. This method was used to identify irregularities in the text in a binary classification context [24].

Analytic Hierarchy Process (AHP) and Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) [25] were used for content anomaly detection. C –Means and an optimized K – Means approach were also used for the same [26].

Unsupervised machine learning approaches like Hierarchical Clustering and K-means algorithm were used in the first row [27] in which the original

collection from the middle was structured to efficiently group the microblog texts. So big clusters were subdivided in a text of noise and small clusters were separated [28, 29]. For topic modeling, the required textual dataset should comprise of diversified topics thereby leading to accurate clusters by K-Means and Hierarchical unsupervised approaches. If the dataset has minimum topics, the model poorly performs.

2.3 Detection of behavioural anomalies

In order to classify respondent responses, descriptive statistics were used. The data set was subsequently qualitative analysed and in order to assess variables associated with a change in relationship status on Facebook, the logistic regression model was used [30]. Logistic Regression approach linearly separates the dataset into normal and anomalous behaviours of the users for a labelled dataset. But most of the data extracted in the form of user's opinions or feedback from an online social media is unlabelled which is insufficient to train the Logistic Regression model.

Detection of Behavioural Anomalies was done using restricted Local Differential Privacy (LDP) for sanitizing user activity logs in social media network. Later Bayesian anomaly detection technique was implemented to the remodelled stream of users and with this, outliers were detected with respect to the duration of calls made between the individuals [31]. The errors and ε are significantly larger compared in the LDP Model. ε is called the privacy budget or privacy loss parameter which is used for controlling in detection ratio of anomalous output from a given dataset. LDP with Bayesian model results in poor accuracy as the performance of the model is sensitive to the skewed dataset.

Machine learning and statistical methods (Regression Model) were used for the discrimination among depressed communities by means of mood, psycholinguistic processes and content issues derived from the posts created by their members [32]. With regression model, the relationship between cause and effect between the variables is believed to remain unchanged. This assumption cannot always hold true and thus estimating the values of a variable based on the equation of regression can result in misleading and erroneous results.

A CVAR (Competitive Vector Auto Regression) model predicted presidential and conference elections using a combination of real-world poll information and online social multimedia data [33] and Proximity based (Nearest Neighbour) method, K-Means clustering based method, Density based (DBSCAN) [34, 35] method and Classification based (Neural Networks), SVM and Bayesian methods were explored [36, 37] for the detection of behavioural anomalies. Though CVAR combines visual information with textual dataset from an OSM and can give good prediction results, the limitation is the inference of the statistics associated with it which may give misleading results if the data is highly persistent.

The OSM textual data is unstructured and unlabelled. Due to the unavailability of the labels, finding anomalous data through unsupervised approach (like K-Means, Hierarchical Clustering, DBSCAN, LDA) results in incorrect detection of anomalies. This is due to the natural language used by the users. The proper knowledge of the lexicon is very important as it relies on the polarity of the lexical variable for deciding the polarity of a text. Nevertheless, certain lexical things appear positive in one domain document, but negative in another domain. Therefore the conventional unsupervised methods and standalone supervised classifiers (like SVM, Random Forest, Decision Trees, and Bayesian Model) will not be able to resolve this issue and hence the need for using a hybrid Self-Supervised model. The main idea is to use certain inputs or transformed inputs as labels in datasets.

3. Proposed work: Behavioural anomaly detection through self-supervised learning

Behavioural anomalies are the anomalies representing the unusual behaviours of the users in terms of the tweets posted in the OSM, which does not conform to the regular tweets posted by non-anomalous users. Detection of behavioural anomalies can be used to prevent any suspicious activities thereafter. In addition, the knowledge obtained from the patterns from the tweets posted by the users in OSM is in most cases essential and workable. It is necessary to identify the normal behaviour in order to detect an anomaly. For a given crowd sourced data collected from tweets for a specific concern, issue or a topic, if the major portion of the data shows similar patterns after analysis then it is termed "Normal".

In this paper, behavioural anomaly detection is done for demonetization dataset. The "demonetization" efforts of the Indian Government raised such an awareness about the need of cash less transaction. When Indian government announced on November 8, 2016 that 500 and 1000 rupees denomination notes were not valid after certain period. This measure was done to promote transparency. Following the announcement, people started expressing their views and opinions on the

effects of the implemented demonetization policy. The decision and introduction of demonetization is considered as a long-term historic step in India's political and economic future.

The tweeting analysis is an attempt to understand the general public's view of the government of India's demonetization policy. As opinions expressed in the tweets does not have class labels, attempt is being made to assign output class label as Positive, Negative or Neutral based on the content of message/tweet.

A Self-Supervised based learning approach is used for detection of anomalies in this research paper. The dataset obtained is unlabelled. In the first level, Rule Based approach is used for classifying the tweets into positive, negative and neutral. Once the tweets are classified, a training dataset is provided to build a feature based model. This feature based model upon giving the test data, classifies into anomalous tweets. The proposed architecture for the Self-Supervised Learning is as shown in Fig. 3.

3.1 Corpus creation

The dataset was obtained using a Twitter API from 22nd November 2016 till 21st April 2017. This dataset has 14940 user's opinions on demonetization and its effects on India. The dataset is a textual dataset and is unstructured and unlabelled. The following Table 1 shows sample tweets by few Twitter Users:

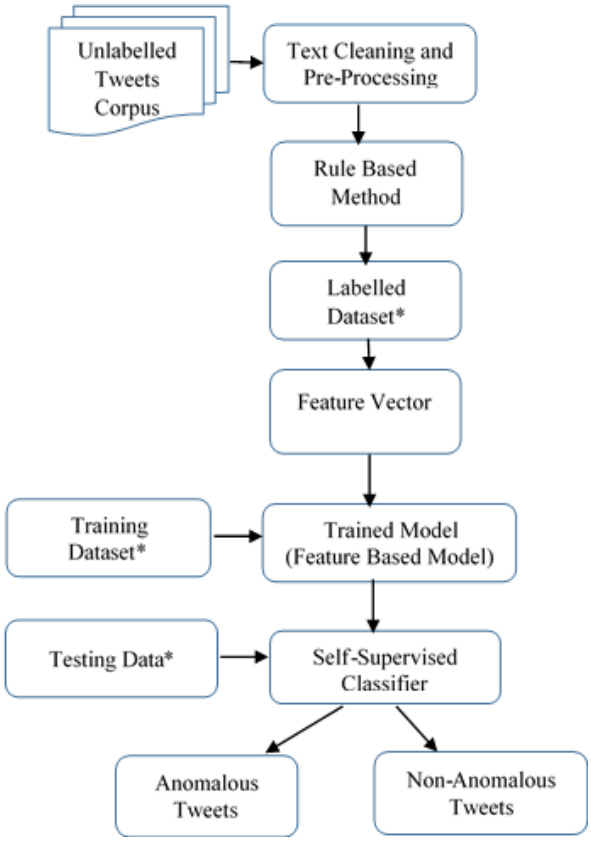

Figure. 3 Proposed architecture for the self-supervised learning

Table 1. Sample tweets

India is so rich that PM Narendra Modi had to implement demonetization to find out who is actually…
People call #Demonetization as poor policy with disproportionate negative impact.
Now country is divided into two religion #Modi & #AntiModi, whose riot potentially can way bigger than other riots.
@narendramodi I am a banker in UCO bank.sir we worked very hard in demonetization.so will hope the wage revision will be good for us.

3.2 Data pre-processing

The dataset collected contains URLs, Hashtags(#), explanations(@), emoticons and ampersand(&) symbols. The dataset needs to be pre-processed. The following steps were used to create a standard dataset for further processing:

a. The text was converted to lowercase.
b. Removal of punctuations and the following symbols: #, @, &.
c. Stopwords and Stemming words were removed.
d. For the feature selection, TF-IDF (Term frequency and Inverse Document Frequency) model was used.

3.3 Rule based methods

As a representation of information, a rule-based structure uses laws. Those laws are coded as if-then-else statements into the program. The key concept for a rule-based system is to capture an expert's expertise in a particular area and to integrate it into a computer program.

In this paper, the following rule based methods were explored for converting an unlabelled dataset into a labelled one.

- TextBlob
- VADER

3.3.1. TextBlob

TextBlob is a built-in library used in Python for processing text dataset and accepts a string as an

input. This simple API provides Natural Language Processing (NLP) functionalities such as tokenization, lemmatization, POS tagging (Part-of-Speech), extraction of nouns and its phrases, analysis of the sentiments and does classification.

sentiment() function is used for the sentiment analysis. The function returns a tuple with two parameters: 'polarity' and 'subjectivity'. The polarity value ranges from [-1, 1] and subjectivity value ranges from [0, 1] where 0 is considered most objective and 1 is considered most subjective. Sample Sentiment Calculation for a sample tweet as listed in Table 1.

text=TextBlob("Now country is divided into two religion #Modi & #AntiModi, whose riot potentially can way bigger than other riots.")
text.sentiment
Sentiment (polarity = - 0.04167, subjectivity = 0.625)

As the polarity value is negative, the sample tweet is classified into Negative. 0.625 Subjectivity value refers that it is a public feedback and not an accurate value.

text =TextBlob("@narendramodi I am a banker in UCO bank.sir we worked very hard in demonetization.so will hope the wage revision will be good for us.")
text.sentiment
Sentiment (polarity = 0.1604, subjectivity = 0.6521)
Here the polarity value is positive.

3.3.2. VADER

Valence **A**ware **D**ictionary and s**E**ntiment **R**easoner (VADER) is a rule based and unsupervised method for text classification. VADER is mainly used for the social media textual data. VADER tool provides the facility to find the sentiment score of a text using SentimentIntensityAnalyzer function.

SentimentIntensityAnalyzer() is used to find the polarity of a text sentence. It helps in calculating the positive or negative score and along with that a sentence's positivity or negativity. It returns a tuple with four parameters: 'compound', 'neg', 'neu' and 'pos'. The 'compound' value ranges from [-1, 1]. Percentage value of 'neg', 'neu' and 'pos' indicate the text's negative, neutral or positive polarity respectively. Their value ranges from [0, 1].

Compound Score is analysed as follows:
If its value is >= 0.05 then its polarity is considered as positive.
If its value is > - 0.05 and < 0.05 then its polarity is considered as neutral
If its value is <= - 0.05 then it is considered as negative polarity.

Sample polarity is calculated as follows:

text="Now country is divided into two religion #Modi & #AntiModi, whose riot potentially can way bigger than other riots."
a = SentimentIntensityAnalyzer()
Vscore = a.polarity_scores(text)
Vscore = {'compound': -0.7845, 'neg': 0.289, 'neu': 0.711, 'pos': 0.0} indicates negative polarity.

text="@narendramodi I am a banker in UCO bank.sir we worked very hard in demonetization.so will hope the wage revision will be good for us."
a = SentimentIntensityAnalyzer()
Vscore = a.polarity_scores(text)
{'compound': 0.6258, 'neg': 0.059, 'neu': 0.737, 'pos': 0.204} classified as positive polarity.

3.4 Supervised learning based model training

The unlabelled textual dataset is converted into a labelled dataset with three labels: positive, neutral and negative. 70% of this dataset is used for training the model and 30% is used for testing. In this paper, four supervised machine learning algorithms were used for comparative studies. Namely Logistic Regression (LR), Naïve Bayes (NB), Random Forest RF) and Support Vector Machine (SVM).

3.4.1. Logistic regression (LR)

A systematic method of predicting binary classes is the logistic regression. The outcome is dichotomous in nature (just 2 classes). In this paper, it is used to classify the demonetization dataset into anomalous and non-anomalous based on the user's opinions. Bernoulli distribution is followed by the dependent variable in logistic regression. Estimate is carried out with the greatest probability. Fitness model is determined by the statistics and concordance.

The equation to calculate Linear Regression is given in the Eq. (1).

$$Y=\alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots. + \alpha_n X_n \quad (1)$$

Whereas Y is a dependant variable and X_1, $X_2 \ldots X_n$ are explanatory variables. α_0, is a constant which moves the curve to the left in the Logistic Regression and $\alpha_1, \alpha_2, \ldots, \alpha_n$ also called the slope defines how steep the curve would be. Later Sigmoid Function on the Linear Regression using Eq. (2) is applied.

$$P = 1 / (1+e^{-Y}) \quad (2)$$

The sigmoid function, also referred to as logistic, provides a 'S' shaped curve that can be used to map any real-life number to a value from 0 to 1. If the value of the sigmoid is more than 0.5, the value can be classified as non-anomalous, and if it is less than 0.5, then it is categorized as anomalous.

3.4.2. Naïve bayes (NB)

A Naive Bayes classifying system is a probabilistic classification learning model. The cluster's foundation rests on the theorem of Bayes. Bayes ' theorem considers the probability of an occurrence having taken place in terms of the possibility of another occurrence. It is calculated by using Eq. (3).

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)} \quad (3)$$

Where Y is a class variable and X is the independent feature vector.
P (Y|X) is called the posterior probability of class (Y, target) given predictor (X, feature vector).
P (X|Y) is called the likelihood which gives the probability of the predictor X for the given class.
P (Y) is called the prior probability of the class.
P (X) is called the predictor prior probability.

3.4.3. Random forest (RF)

As the name suggests, the random forest contains a broad variety of decision-making trees that function as an ensemble. Every single tree sprays a class prediction in the random forest and the class with the most votes is the prediction of the model. RFs develop several individual training decision-making trees. Predictions from all trees are grouped together to make a final prediction. If they make a final choice using a series of data, they are called ensemble techniques. Gini Index is used to decide on how nodes on a decision tree branches and is given by Eq. (4).

$$G = 1 - \sum_{k=1}^{c} (pi)^2 \quad (4)$$

Whereas *pi* is the relative frequency of the class present in the dataset and *c* is the total number of classes.

3.4.4. Support vector machine (SVM)

SVM is the supervised research model for two-group classification problems. They are able to categorize a new text after they have been provided the model, a trained data set. In an N-dimensional space (N–the number of features) the SVM aims at finding a hyperplane that specifically classifies the data points. There are two types of SVM classifiers: Linear and Non-Linear. An appropriate kernel function is important for better classification. In this paper, an RBF was used for the classification as shown in Eq. (5) because it produces local and finite responses along the x-axis.

- Radial Basis Function (RBF):
  $K(W, X) = \exp(-\alpha \|W - X\|^2), \alpha > 0$ (7)

Here α is the kernel parameter. It can be considered as the region of decision. When α is small, the "curve" of the decision limit is very small and the decision area is very large. If the value of α is high, the 'curve' of the decision limit is high and the decision-making regions surrounding data points are formed. W is a training vector and is mapped into a higher dimensional space by the function ϕ and $K(W, X) \equiv \phi(X)$, is known as kernel function. $\|W\text{-}X\|^2$ is the squared Euclidean Distance between training and the testing feature vector.

3.5 Self - supervised learning

Given an unstructured text dataset along with labels, supervised machine learning approach can classify it. The dataset in the form of user's opinions collected through OSM doesn't come with labels such as a positive opinion or a negative one. Collecting the labels manually is expensive and time consuming. Also using an unsupervised machine learning approach results in less efficiency. Self-Supervised is a hybrid supervised learning where the labels are determined with the help of the input data. Once the labels are found, a standard supervised classifier is used for appropriate classification. In this research work, the labels are found using rule based method as explained in 3.3 followed by a feature based model as covered in 3.4 for binary classification of the demonetization dataset into anomalous and non-anomalous user's opinions.

4. Experimental results and discussions

This section provides performance evaluation of the proposed research methodology. The dataset considered for the study was the user's opinions on demonetization and its effects on India after the policy was implemented. The dataset has 14940 user's opinions in the form of text which is unstructured, unlabelled, has grammatical and

spelling mistakes leading to syntactic and semantic ambiguities. The corpus creation and the pre-processing of the text dataset is discussed in 3.1 and 3.2 section. Before the proposed Self-Supervised learning method, the conventional human annotation and K-Means unsupervised learning were applied on the processed text dataset for the comparative purpose.

Initially the unlabelled dataset was manually annotated with the help of crowdsourcing. It was annotated into three categories. Positive, Negative and Neutral. Later K-Means clustering algorithm with K=3 which is an unsupervised approach was run on the dataset resulting into three clusters. The clusters were named again as positive, negative and neutral. The following python library tool was used for the same.

KMeans (n_clusters=3, init='k-means++', n_init=10, max_iter=200)
n_clusters = 3: indicates three clusters to be formed and also three centroids to be generated each or a cluster.
init='k-means++': selects the initial centroid values in a smart way in order for the algorithm to speed up its convergence.
n_init=10: indicates the number of times the algorithm will run for different initial centroid values.
max_iter=100: indicates the total number of iterations for a single run.

The results captured are as shown in Fig. 4. In manual annotation process, it takes time to manually process and analyze a given text. It is a slow process in which a person must read and determine how each text is organized.

Using K-Means clustering algorithm, which is an unsupervised machine learning approach too failed in appropriately identifying the anomalous tweets. It's accuracy of correctly detecting positive user's

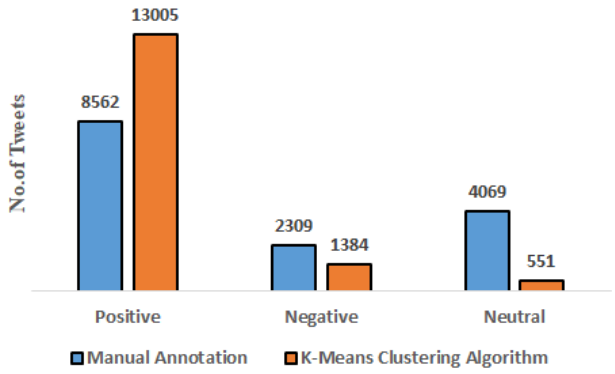

Figure. 4 Labelling through manual annotation and unsupervised kearning using k-means clustering algorithm

opinions is 66%. The reason behind this is when the text data was converted into a feature vector, majority of the data points were near to the 1st centroid which is termed as positive in our case. The Sum of Squared Error (SSE) for most of the tuples were zero for the 1st centroid and very few tuples had high values of SSE thereby putting those into other clusters. This is mainly because the text did not comprise of diversified topics.

For example for the text given below, classifying the text into two different clusters by K-Means is easy as the text contains two topics namely sports and place or cricket and traveling.

Document = ["This is the best lovely place on the earth.", "This player's skills in cricket is greater than tennis.", "Hi buddy! how was your Switzerland trip last month?", "This player in his entire career, once scored more than 200 runs in a single cricket innings."]

But for the text as shown in Table 1, the topics are unique and not much different apart from the emotions exhibited. Though K-Means was iterated for 200 times, there was no change in the centroids observed. Also K-Means is sensitive to initial seed value (probability = $3!/3^3$ = 22% as k=3) and the outliers present in the dataset. Considering the drawbacks of these two methods, a hybrid Self-Supervised approach was used in this paper.

In the first level, an unsupervised, rule-based approach was used for labelling the unstructured and unlabelled dataset on Demonetization Policy Implementation in India. In the second level, a supervised, feature-based machine learning approach was used for training the model thereby classifying the dataset into anomalous and non-anomalous tweets.

In the first phase, both TextBlob and VADER was applied for the pre-processed dataset. The following was imported from NLTK (Natural Language Processing Tool Kit) from python library tools for analyzing text using VADER and TextBlob respectively.

from nltk.sentiment.vader import SentimentIntensityAnalyzer
from textblob import TextBlob

It was observed that VADER tool had given more positive weightage to the tweets than TextBlob tool as shown in the Fig. 5. This was possibly due to a lack of coverage of emoticons, slangs, acronyms, as well

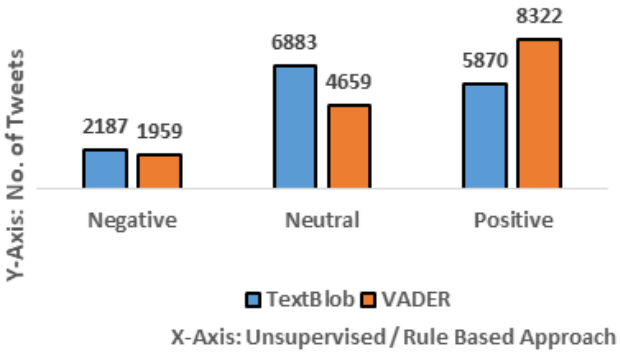

Figure. 5 Labelling of Tweets by VADER vs. TextBlob

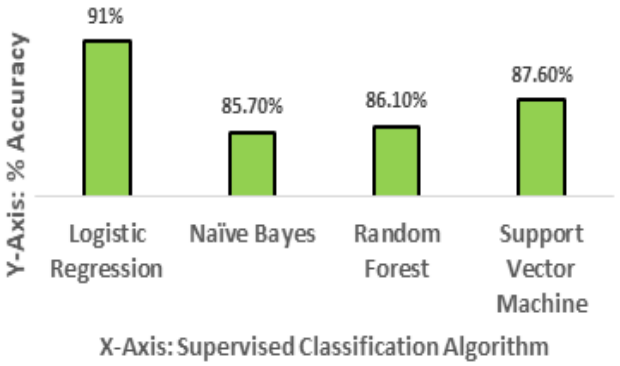

Figure. 6 Result of Accuracy

as sentiment-oriented language in social media text.

In the second phase, 70% of the labelled tweets obtained from VADER tool were fed into Feature-Based Supervised Machine Learning Classifiers. Four Supervised Learning algorithms were used for the comparative study: Logistic Regression (LR), Naïve Bayes (NB), Support Vector Machine (SVM) and Random Forest (RF). It was observed that Logistic Regression obtained 91% accuracy as compared to SVM – 87.60%, RF – 86.10%, and NB - 85.70% as shown in Fig. 6.

Logistic regression had performed well as the feature set was linearly separable and was less prone to overfitting. Its accuracy had increased due to its direction of its associativity towards binary classification of the input dataset into anomalous and non-anomalous.

5. Conclusion

Self-Supervised learning learns useful knowledge representations without having to use an annotated and labelled dataset. Rather, unlabelled data can be used and predefined pretext tasks are optimized. These features can then be used for new tasks that are scarce in data. In this analysis, output labels were generated by revealing a relationship between the object parts or different views of the object internally from the input textual dataset. In this paper, Self-Supervised Learning improved the robustness of the model by without applying any unsupervised machine learning clustering algorithm for labelling the text dataset. The experimental results showed that using Self-Supervised Learning approach resulted in 91% accuracy when compared to a conventional unsupervised approach - K-Means algorithm resulting in a 66% accuracy in classifying the results. So with the proposed model, it was found that 13.11% of the user's opinions did not conform the implementation of demonetization policy in India and have posted negative opinions for the policy implemented.

The future enhancement can include automated labelling for the textual dataset as it would reduce the labelling time that is calculated in the first level of Supervised-Learning Process. The most important, data is categorized, listed, grouped, sorted by a particular tag, expected user behavior, and established precautions based on automated data labelling.

Acknowledgements

Authors wishes to acknowledge the technical and infrastructural help rendered by the faculty members of CSE department of CHRIST (Deemed to be University), Bangalore, India.

References

[1] M. Elahi, K. Li, W. Nisar, X. Lv, and H. Wang, "Efficient Clustering-Based Outlier Detection Algorithm for Dynamic Data Stream", In: *Proc. of the 5th International Conf. on Fuzzy Systems and Knowledge Discovery*, Shandong, China, pp. 298–304, 2008.

[2] K. Henderson, B. Gallagher, L. Li, L. Akoglu, T. Eliassi-Rad, H. Tong, and C. Faloutsos, "It's who you know: Graph Mining using Recursive Structural Features", In: *Proc. of the 17th ACM SIGKDD International Conf. on Knowledge Discovery and Data mining*, San Diego, United States, pp. 663–71, 2011.

[3] L. Akoglu, M. McGlohon, and C. Faloutsos, "Oddball: Spotting Anomalies in Weighted Graphs", In: *Proc. of the Pacific-Asia Conf. on Knowledge Discovery and Data Mining*, Berlin, Heidelberg, pp. 410-421, 2010.

[4] A. Rezaei, Z. M. Kasirun, V. A. Rohani, and T. Khodadadi, "Anomaly Detection in Online Social Networks using Structure-Based Technique", In: *Proc. of the 8th International Conf. on Internet Technology and Secured Transactions*, London, United Kingdom, pp. 619–22, 2013.

[5] B. Viswanath, M. A. Bashir, M. Crovella, S. Guha, K. P. Gummadi, B. Krishnamurthy, and A. Mislove, "Towards Detecting Anomalous User Behavior in Online Social Networks", In: *p\Proc. of the 23rd USENIX Security Symposium*, San Diego, United States, pp. 223-238, 2014.

[6] V. J. Hodge and J. Austin, "A Survey of Outlier Detection Methodologies", *Artificial Intelligence Review*, Vol. 22, No. 2, pp. 85-126, 2004.

[7] N. A. Heard, D. J. Weston, K. Platanioti, and D. J. Hand, "Bayesian Anomaly Detection Methods for Social Networks", *The Annals of Applied Statistics*, Vol. 4, No. 2, pp. 645-662, 2010.

[8] C. D. Corley, D. J. Cook, A. R. Mikler, and K. P. Singh, "Text and Structural Data Mining of Influenza Mentions in Web and Social Media", *International Journal of Environmental Research and Public Health*, Vol. 7, No. 1, pp. 596-615, 2010.

[9] A. Rohani and T. Khodadadi, "Anomaly Detection in Online Social Networks Using Structure-Based Technique", In: *Proc. of the 8th International Conf. for Internet Technology and Secured Transactions*, London, United Kingdom, pp. 619-622, 2013.

[10] D. Savage, X. Zhang, X. Yu, P. Chou, and Q. Wang, "Anomaly Detection in Online Social Networks", *Social Networks*, Vol. 39, No. 1, pp. 62-70, 2014.

[11] R. Hu, C. C. Aggarwal, S. Ma, and J. Huai, "An Embedding Approach to Anomaly Detection", In: *Proc. of the 32nd International Conf. on Data Engineering*, Helsinki, Finland, pp. 385-396, 2016.

[12] M. S. L. Yellari, M. Manisha, J. Dhanesh, M. S. Rao, and Dr. S. Suhasini, "Identifying Malicious Data in Social Media", *International Research Journal of Engineering and Technology*, Vol. 4, No. 3, pp. 1732-1738, 2017.

[13] W. Yu, W. Cheng, C. C. Aggarwal, K. Zhang, H. Chen, and W. Wang, "NetWalk: A Flexible Deep Embedding Approach for Anomaly Detection in Dynamic Networks", In: *Proc. of the 24th ACM SIGKDD International Conf. on Knowledge Discovery and Data Mining*, London, United Kingdom, pp. 2672-2681, 2018.

[14] L. Tran, L. Fan, and C. Shahabi, "Distance-Based Outlier Detection in Data Streams", In: *Proc. of the VLDB Endowment*, Vol. 9, No. 12, pp. 1089-1100, 2016.

[15] A. S. Dokuz, "Anomalous Activity Detection from Daily Social Media User Mobility Data", *Omer Halisdemir University Journal of Engineering Sciences*, Vol. 8, No. 2, pp. 638-651, 2019.

[16] A. Mahapatra, N. Srivastava, and J. Srivastava, "Contextual Anomaly Detection in Text Data", *Algorithms*, Vol. 5, No. 4, pp. 469-489, 2012.

[17] G. Xu, Y. Meng, Z. Chen, X. Qiu, C. Wang, and H. Yao, "Research on Topic Detection and Tracking for Online News Texts", *IEEE Access*, Vol. 7, pp. 58407–58418, 2019.

[18] Y. Fang, H. Huang, P. Jian, X. Xin, and C. Feng, "Self-Adaptive Topic Model: A Solution to the Problem of 'Rich Topics get Richer", *China Communications*, Vol. 11, No. 12, pp. 35-43, 2014.

[19] T. T. Wai and S. S. Aung, "TPMTM: Topic Modeling over Papers' Abstract", *Advances in Science, Technology and Engineering Systems Journal*, Vol. 3, No. 2, pp. 69-73, 2018.

[20] B. L. Abraham and A. P. Nair, "Anomalous Topic Discovery Based on Topic Modeling from Document Cluster", *International Research Journal of Engineering and Technology*, Vol. 5, No. 2, pp. 966-972, 2018.

[21] Y. A. S. Prasad and G. Ramakrishna, "Statistical Anomaly Detection Technique for Real Time Datasets", *International Journal of Computer Trends and Technology*, Vol. 6, No. 2, pp. 89-94, 2013.

[22] A. Sharma and A. Suryawanshi, "A Novel Method for Detecting Spam Email using KNN Classification with Spearman Correlation as Distance Measure", *International Journal of Computer Applications*, Vol. 136, No. 6, pp. 28-35, 2016.

[23] Y. Liu and S. Xu, "Detecting Rumors Through Modeling Information Propagation Networks in a Social Media Environment", *IEEE Transactions on Computational Social Systems*, Vol. 3, No. 2, pp. 46-62, 2016.

[24] C. Steyn and A. de Waal, "Semi-Supervised Machine Learning for Textual Anomaly Detection", In: *Proc. of the International Conf. on Pattern Recognition Association of South Africa and Robotics and Mechatronics*, Stellenbosch, South Africa, pp. 1-5, 2016.

[25] R. Kaur, S. Singh, and H. Kumar, "AuthCom: Authorship Verification and Compromised Account Detection in Online Social Networks using AHP-TOPSIS Embedded Profiling based Technique", *Expert Systems with Applications*, Vol. 113, No. 4, pp. 397-414, 2018.

[26] M. Mei, X. Guo, B. C. Williams, S. Doboli, J. B. Kenworthy, P. B Paulus, and A. A. Minai, "Using Semantic Clustering and Auto Encoders for Detecting Novelty in Corpora of Short

Texts", In: *Proc of the International Joint Conf. on Neural Networks*, Rio de Janeiro, Brazil, pp. 1-8, 2018.

[27] D. Shah, M. Hurley, J. Liu, and M. Daggett, "Unsupervised Content-Based Characterization and Anomaly Detection of Online Community Dynamics", In: *Proc. of the 52nd Hawaii International Conf. on System Sciences*, Grand Wailea, Hawaii, pp. 2264-2273, 2019.

[28] C. Xiao, D. M. Freeman, and T. Hwa, "Detecting Clusters of Fake Accounts in Online Social Networks", In: *Proc. of the 8th ACM Workshop on Artificial Intelligence and Security*, Denver, United States, pp. 91–101, 2015.

[29] X. Geng, Y. Zhang, Y. Jiao, and Y. Mei, "A Novel Hybrid Clustering Algorithm for Topic Detection on Chinese Microblogging", *IEEE Transactions on Computational Social Systems*, Vol. 6, No. 2, pp. 289-300, 2019.

[30] O. L. Haimson, N. Andalibi, M. De Choudhury, and G. R. Hayes, "Relationship Breakup Disclosures and Media Ideologies on Facebook", *New Media & Society*, Vol. 20, No. 5, pp. 1931-1952, 2017.

[31] R. Aljably, Y. Tian, M. Al-Rodhaan, and A. Al-Dhelaan, "Anomaly Detection over Differential Preserved Privacy in Online Social Networks", *Plos One*, Vol. 14, No. 4, pp. 1-20, 2019.

[32] T. Nguyen, D. Phung, B. Dao, S. Venkatesh, and M. Berk, "Affective and Content Analysis of Online Depression Communities", *IEEE Transactions on Affective Computing*, Vol. 5, No. 3, pp. 217- 226, 2014.

[33] Q. You, L. Cao, Y. Cong, X. Zhang, and J. Luo, "A Multifaceted Approach to Social Multimedia-Based Prediction of Elections", *IEEE Transactions on Multimedia*, Vol. 17, No. 12, pp. 2271-2280, 2015.

[34] C. Nagamani and S. Chittineni, "Efficient Neighborhood Density Based Outlier Detection inside a Sub Network with High Dimensional Data", *Ingenierie des Systemes d'Information*, Vol. 24, No. 1, pp. 107-111, 2019.

[35] A. Coluccia, A. D'Alconzo, and F Ricciato, "Distribution-Based Anomaly Detection in Network Traffic", In: *Biersack E., Callegari C., Matijasevic M. (eds) Data Traffic Monitoring and Analysis, Springer Lecture Notes in Computer Science*, Vol. 7754. pp. 202-216, 2013.

[36] R. Kaur and S. Singh, "A Survey of Data Mining and Social Network Analysis based Anomaly Detection Techniques", *Egyptian Informatics Journal*, Vol. 17, No. 2, pp. 199-216, 2016.

[37] L. P. D. Bosque and S. E. Garza, "Prediction of Aggressive Comments in Social Media: an Exploratory Study", *IEEE Latin America Transactions*, Vol. 14, No. 7, pp. 3474-3480, 2016.