Journal of Information Systems Engineering and Management
2025, 10(18s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Effective Machine Learning Approach to Track Polarity Change in Social Media Data

Amit K Jadiya[1], Ramesh Thakur[2], Archana Thakur[3]
[1]*Institute of Engineering and Tecchnology, Devi Ahilya University, Indore, India*
[2]*International Institute of Professional Studies, Devi Ahilya University, Indore, India*
[3]*School of Computer Science & IT, Devi Ahilya University, Indore, India*

ARTICLE INFO

Received: 15 Dec 2024

Revised: 26 Jan 2025

Accepted: 15 Feb 2025

ABSTRACT

An enormous amount of textual data is generated daily on various social media platforms, which may relate to election campaigns, political discussions, reviewing government policies for the public, launching new utility products, and respective review comments, etc. With these textual data, sentiment analysis helps to understand the behavior and attitudes of users and the public, whether they say good or bad things about the brand, products, and services. However, because we are such emotional creatures, our emotions influence everything we do and serve as the foundation for our future behavior. So, over time, prior positive polarity may change to negative due to some incidents or phenomena. In this paper, we aim to identify the change in polarity over time using a machine learning approach so that respective organizations can analyze possible incidents and circumstances to investigate the reasons behind the change. When the polarity moves from positive to negative, it is essential to identify the reason behind this and rectify or take necessary actions as soon as possible to overcome the negative impact. In other words, respective organizations must work to discover pain points for the polarity change, especially if the change is positive or negative. In this paper, we suggest a framework to identify the change in polarity, and in case a polarity change from positive to negative is detected, the proposed questionnaire helps to get insights into the responsible factors for this change. Results show that the proposed framework, studying daily statistics of the sentiment classification outcomes and the ML model's interpretable capabilities, gives better accuracy for the change in polarity evaluation.

Keywords: Sentiment analysis, Sentiment change, Social media data, Natural Language Processing (NLP), Naive Bayes.

INTRODUCTION

One of the most important sources of data for sentiment analysis is Social media data, which is a powerful way to reach the public or customers. Good reviews and posts on social media encourage the public or customers to utilize the service or buy from the company. But the reverse is also true. Negative social media posts or reviews can be very costly to the respective organization and the business, which can also cause big financial losses.

As mentioned above, negative comments and reviews have a significant negative impact. However, over time, due to any incident or phenomenon, it is essential to know the insights behind this negative impact as soon as possible if the positive polarity shifts into negative polarity over time. This change in the polarity of sentiments from positive to negative or negative to positive over time can be termed as 'polarity shifting over time' or 'polarity change over time'. Also, the term 'polarity shift' refers to a linguistic phenomenon in which the polarity of sentiment can be reversed (i.e., positive to negative or vice versa) by some special linguistic structures called polarity shifters, e.g., negation ("I don't like this movie") and contrast ("Fairly good, but not my style") [1-2].

There is a difference in the term polarity shift used in a few other research works and for this study. Many researchers used the term polarity shift as a problem because polarity shift produces sentiment opposite text as it changes the sentiment orientation, so in the BOW model representation, two sentiment-opposite texts are to be considered very similar because the BOW model represents text in the vector of independent words format [2]. For example, two sentiment-opposite texts, “I don’t like this picture” and “I like this picture,” are regarded to be very similar in the BOW representation [1]. However, in our study, we calculate the change in the polarity over time, which may be due to some incident or phenomenon, and this change is termed ‘polarity shift over time’ or ‘polarity change over time’.

For the identification of polarity change, the first step is to identify the polarity of the statements that are extracted from multiple social media platforms. The sentiment analysis process is utilized to ascertain whether a particular text contains neutral, positive, or negative emotions. Sentiment analysis typically leverages textual data from the mentioned areas and is considered an important source for identifying the polarity of sentiments by using various machine learning techniques, natural language processing (NLP), and text analytics approaches. Other names for sentiment analysis include "opinion mining" and "emotion artificial intelligence". This is a kind of text analytics that uses numerous types of AI methods. There are many benefits of sentiment analysis with machine learning approaches, such as removing bias through consistent analysis, processing data at speed, and saving time with automation. Also, one major task is to deal with the Big Data challenges arising in social network platforms streaming applications concerning Volume, Variety, Veracity, and Velocity [3]. This can be overcome by big data processing techniques like distributed database HDFS (Hadoop Distributed File System), etc.

There are a few challenges in identifying changes in the polarity of social media data. A vast amount of structured, semi-structured, and unstructured data that is diverse, anomalous, and complicated in nature is produced due to the enormous expansion in data and the availability of divergent data sources [4-7]. If the dataset is unbalanced, this also need to be taken care and apply appropriate techniques for balancing of data [8]. Also, the change in the polarity is reflected in multiple social media platforms. However, data gathering is difficult because it requires extracting and compiling data from multiple data sources with different schemas. Other than these, due to constraints like jargon, allusions to preexisting information or ideas, and abbreviations, sentiment analysis in social media is more complex than in other text forms. Additionally, there are challenges because of human interaction behavior on social media, such as using specialized language, short attention span, and desire for immediacy [9]. As a result, the raw data sources must be converted into relevant and valuable features that embody a certain attribute of the observations. During preprocessing also there are challenges like noise and irrelevant data, HTML characters, different data formats, URLs, missing values etc. [10].

As it is very important to know which incidents or phenomena are responsible for polarity changes, this paper proposes an effective framework using a machine learning approach to identify polarity changes in Social media data. The proposed approach uses all the tools and techniques that are required for building a classification model, such as data extraction, preprocessing, feature engineering, and the ML model’s capability of interpretability. Base machine learning models like k-NN, Naive Bayes, Linear regression, Logistic regression, SVM, Decision tree, and Random forest are studied with the proposed framework, and the best suitable ML model is selected according to evaluation metric log-loss. Also, these models have different benefits in terms of their interpretable capability [11].

The contribution of this study can be summarized as follows:

1. To present an efficient ML model for the classification of sentiments that are extracted from social big data.
2. To propose an efficient framework for identifying the change in the polarity in the sentiments over time for offline and real-time streaming data.
3. To propose a set of questionnaires and respective solutions for evaluating change in the polarity. This helps to understand the responsible factors behind the polarity change.

We have studied the interpretable capabilities of various baseline models to get more insights and answers to the proposed questionnaire. The k-NN is highly interpretable because it uses simple distance metrics and makes predictions based on similarity to neighboring data points. Some optimization methods, such as deep neural networks, have complicated, nonlinear transitions that make them difficult to understand. Since k-NN’s performance gets worse in high-dimensional spaces due to computational constraints for very high-dimensional data, it is not helpful in our scenario. In this study, the dataset consists of text data from multiple social media, and the corresponding machine learning features are high dimension, so k-NN is not appropriate.

The interpretation of decision trees is straightforward: begin at the root node, go to the following nodes, and the edges give information about the subsets we are checking out. The leaf node will give the expected result once we get there. It indicates that decision trees are highly interpretable, but it’s true that they are interpretable as long as they are short. As depth grows, the number of terminal nodes rapidly rises and it gets harder to understand a tree's decision criteria.

Naive Bayes has a low computational cost and is based on simple mathematical computations. As a result, it can operate rapidly even with big datasets. This study is best suited for Naive Bayes since it performs well with high-dimensional data and is chosen for text-based problems like text classification. The assumption of independence makes the Naive Bayes model interpretable. Since we can interpret the conditional probability, how much each attribute contributes to a specific class prediction is very clear.

Additional interpretable models are Linear regression, Logistic regression, and Linear SVMs. These models are interpretable because each input feature has a weight that directly influences the model output. The absolute value of the weight can be used to determine the importance of features. These important features help to understand the model's interpretability. However, interpretability is difficult in the case of Kernal SVM.

LITERATURE SURVEY

In literature, we have found a number of methodologies and techniques proposed by different authors for sentiment analysis and polarity change methods.

Alattar et al. [12] introduced a method to identify the most representative tweets and respective candidate reasons for the change in sentiments. First, an effective sentiment analysis classifier for short texts is prepared, and then, for interpreting sentiment variations on Twitter data, a filtered LDA framework is proposed. For better understanding and relationships between Reason Candidates, Hierarchical Pachinko Allocation (HPA) and Multi Grain Latent Dirichlet Allocation (MG-LDA) are used and have shown outputs for one month of subject-specific tweets. The limitation is that their work mainly considered data from Twitter only, but there are many social media platforms where sentiment data may be in different schemas. Our study used an SBD preprocessor to select data from multiple social media platforms.

Tasoulis et al. [3] presented a method that does not require any offline phase or training to identify sentiment shifts in Twitter data streams. They have proposed using lexicon to describe tweets and detection algorithms to identify shifts in sentiment scores over time. The two main components of their suggested methodology are gathering and classifying tweets based on their sentiment and identifying notable shifts in the sentiment score time series. They have employed the cumulative sum (CUSUM) technique for both online and offline change detection. Additionally, they have displayed word clouds that plot the frequency of words that appear in the relevant set of Twitter posts. They used the Term Frequency Inverse Document Frequency (TFIDF) technology to preprocess the text.

Giachanou et al. [13] worked on Twitter data that span over nine months and applied conventional time series analysis methods for sentiment trend analysis. They leveraged time series analysis on public opinions to understand the data and identified patterns, trends, and seasonality. Also, they have worked to detect outliers and investigated their usability in detecting important events or reasons for the polarity change.

Using data from the web, Yu Jiang et al. [14] conducted research on sentiment change analysis. From a collection of documents addressing a particular topic, their suggested method can analyze sentiment and detect changes in sentiment. Additionally, their approach can spot trending occurrences that could be the reason for a shift in opinion. There are two modules in their suggested solution. The first module is the Topic Sentiment Analysis, which has two steps. The first step is to extract topics from the corpus and classify document contents into these topics. The second step is to analyze the sentiment of each extracted topic. Based on the first module's findings, the second module does the topic sentiment change analysis.

With the study of polarity change of sentiments, we also studied recent work on capability and approaches for ML model interpretability. Liu et al. [15] proposed a framework, INAFEN (Interpretable Automated Feature Engineering), for logistic regression, which automatically transforms the nonlinear relationships between numerical features and labels into linear relationships and extracts information from black-box models. The Black Box Problem refers to the difficulty in understanding and interpreting the internal functioning of ML models. It makes it challenging to depend on ML models for crucial applications and choices by raising concerns about trust, accountability, bias, fairness, and regulatory compliance. They have also talked about two approaches to solving black-box problems: interpretable and explainable.

Tabassoum et al. [16] worked to classify the news data into respective news categories like Business, Health, World, Crime, Sports, Media, etc., using k-NN, Random Forest, Decision Tree, and Logistic regression ML models. With the classification, they also used the Explainable Artificial Intelligence (XAI) technique of Local Interpretable Model-agnostic Explanations (LIME) to support the interpretability of their best-performing ML model, which is Random Forest.

We also reviewed work related to polarity shift problems where polarity shifting was a challenge for sentiment classification and is considered one of the main reasons why the standard machine learning algorithms make inaccurate predictions [1-2], [17]. Our proposed method for identification of sentiment change and finding possible reasons behind the polarity change is using base machine learning models and their capability to interpret the outcomes for further analysis.

PROPOSED WORK

In this paper, we have presented a framework for identifying changes in polarity and analyzing the reasons behind these changes in the context of social media data. Figure 1 shows three sections in the proposed framework: ML modeling, identification of polarity change, and evaluation of polarity change factors.

ML Modeling

First, the polarity of sentiments from offline social media datasets must be calculated to identify sentiment changes over time. The proposed framework's first step is to design and build an effective Machine-learning model for classifying sentiments into positive, negative, and neutral categories. Based on this offline social media dataset, an efficient ML model will be developed, which will be further used to classify real-time comments/reviews. Any organization that wants to evaluate public sentiments can post their product/service related post to multiple social media platforms. This textual data, including comments and reviews for specific subjects, needs to be extracted from multiple social media platforms. After extracting, data needs to be preprocessed and integrated due to the different attributes of different platforms. The polymorphic SBD preprocessor can do this integration [4], which is an efficient method for preprocessing and integrating data from multiple sources. Once the integrated dataset is ready, text data needs to be converted into a numerical form to develop a machine-learning model. By using various feature engineering tools and approaches, new features can also be derived to improve the classification accuracy of the machine learning model. The text data to numerical or mathematical form transformation can be done through Response coding, one-hot encoding, TFIDF weighted vector and word2vec, etc. The preprocessed and transformed dataset needs to be divided into train, test, and cross-validation. After hyper-parameter tuning and testing, the designed final classification model (M) should be deployed in production. According to the nature of the data, the final classification model (M) may be Naive Bayes, Linear Regression, Logistic regression or Linear SVM, Random forest, or any other stacking model.

The production deployed model (M) needs to be used for further sentiment evaluation, which is coming in real-time from multiple platforms. With the designed ML model (M) and the social media offline dataset, percentage values of each day's polarities are calculated, which needs to be used to evaluate change in polarity for the nth day. Here,

P_o = percentage value of positive polarity

Q_o = percentage value of negative polarity

R_o = percentage value of neutral polarity

Identification of polarity change

As shown in Figure 1, the second stage of the proposed framework is identifying the polarity change. The final ML classifier model, which was deployed for social media data sentiment analysis in the previous stage, is used to evaluate real-time text reviews/comments daily for a number of days. For the evaluation, the real-time generated text data needs to be stored in a database, from where these are preprocessed and further fed into the production deployed classification model to evaluate the polarity. Since a large amount of data is generated daily, older data must be moved to the archive database to improve efficiency.

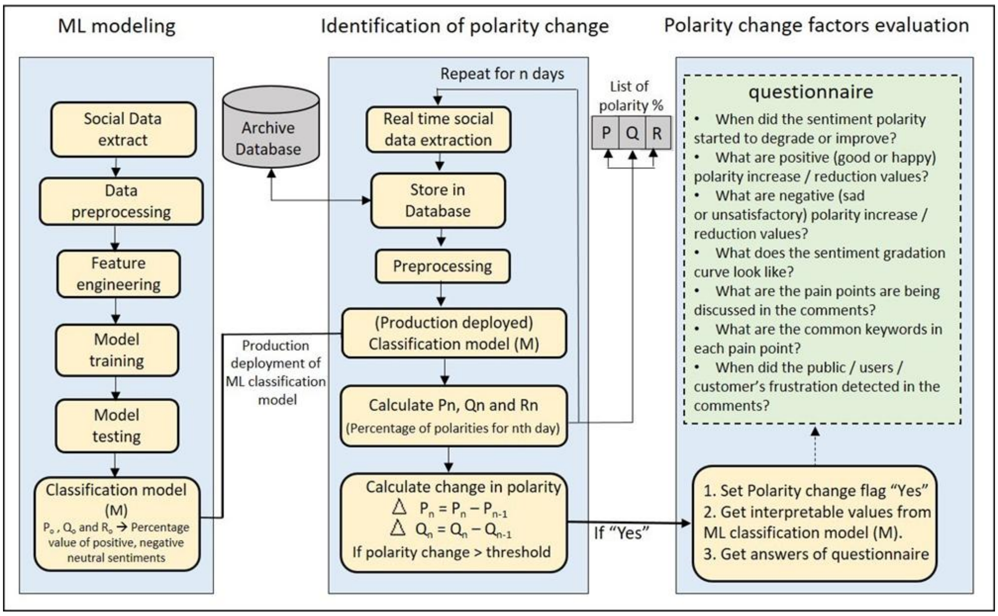

Figure 1: Framework for polarity change identification and evaluation

For the specific subject, real-time text reviews and comments are generated daily, which may be positive, negative, or neutral in nature. If the percentage of daily positive/negative polarity is approximately the same as in older days, it is acceptable and considered stable. However, if it drops or increases suddenly or continuously, this is the point that should be noted, and what happened recently should be tracked. We need to get answers to a few specific sets of questions, which are discussed in the next section, 'polarity change evaluation'. We have divided polarity change categories into five categories: positive, negative, continuous-positive, continuous-negative, and stable. The threshold value (T) is the percentage value that decides whether there is a change in polarity or not. If there is a change in polarity, what is the category of this change? The threshold value may differ in different scenarios, depending on many parameters like the nature of review comments, political or organizational expectations, etc. For example, the polarity distribution on n^{th} day is as percentage positive polarity is P_n, percentage negative polarity is Q_n, and percentage neutral polarity is R_n then on $(n+1)^{th}$ day, polarity change category will be positive if positive polarity reviews increase by more than threshold (T) value ($P_{n+1} > P_n + T$) or negative polarity reviews decrease by threshold value ($Q_{n+1} < Q_n - T$). We have considered the percentage change threshold value for our experiments to be five. It means if the percentage of positive polarity decreases by more than 5% the next day, the polarity change category will be negative; if the percentage of negative polarity increases by more than 5%, then the polarity change category will also be negative. Similarly, other categories are also calculated, as discussed in Algorithm 1. If the percentage change value is less than five percent, it will be considered stable for that day. If there is no significant increase or decrease in polarity in a single day, but if there is a continuous decrease or increase in the polarity value that spans over a few days, this also needs to be identified. Equations- 1, 2, and 3 represent the polarity change value for streaming reviews.

$$\Delta P_n = PS_n - PS_{(n-1)} \quad (1)$$

$$\Delta Q_n = QS_n - QS_{(n-1)} \quad (2)$$

$$\Delta R_n = RS_n - RS_{(n-1)} \quad (3)$$

Where Δ P, Δ Q, and Δ R are percentage values of change in polarity between the nth and its previous day's value for positive, Negative, and neutral polarity, respectively. PS, QS, and RS are the percentage value polarity of the streaming data set.

If the polarity change class is positive or negative or continuous positive or continuous negative, set the polarity change flag as "Yes" for further analysis. Also, if there is a continuous drop in polarity by the value of T (threshold), then also set up the polarity change flag as "Yes". For each day, percentage change is stored in the list P (positive polarity list), Q (negative polarity list), and R (Neutral polarity list). For our study, we did not find the significance of the value of changes in the neutral polarity. However, according to the requirements, this can also be included whenever needed.

Polarity change evaluation

Once it is identified that there is a polarity change value greater than the threshold, the reason behind the change in polarity from its previous day's polarity and offline data to real-time data (online streaming data) needs to be identified for further actions. To get these insights, we have prepared a questionnaire that helps to identify the reasons and take appropriate actions to improve the positive polarity score. As mentioned in this study, we considered the percentage change threshold value = 5. Pseudo code for polarity change evaluation is mentioned in Algorithm 1.

If a polarity change is identified, further insights need to be achieved by using interpretable values from the Machine learning model used for classification. The below questionnaire is helpful for determining the reason behind the polarity change.

Questionnaire:

Q 1. When did the sentiment polarity started to degrade or improve?
Q 2. What are Positive (good or happy) polarity increase/reduction values?
Q 3. What are negative (sad or unsatisfactory) polarity increase/reduction values?
Q 4. What does the sentiment gradation curve look like?
Q 5. What are the pain points are being discussed in the comments?
Q 6. What are the common keywords in each pain point?
Q 7. When did the public/users/customer's frustration detected in the comments?

Out of the above seven questions, getting insights for questions no 5, 6, and 7 depends on which sentiment classifier model is being used as there are pros and cons of each baseline model like k-NN, Naive Bayes, Logistic regression, Linear regression, Linear SVM, Decision tree and Random Forest, etc. Questions 1, 2, 3, and 4 do not depend on the classification model.

As per the proposed framework shown in Figure 1, a history list of positive polarity percentages (P), a list of negative polarity percentages (Q), and a list of neutral polarity percentages (R) is being prepared. With available data in lists P, Q, and R from 1^{st} day to the n^{th} day, details about questions 1 to 4 can be gathered using utility functions like graphs, etc., and can be studied further, including manual analysis.

Algorithm 1. Pseudo code for polarity change evaluation

Input: D_o = Offline social media text data set, C_o = Class of sentiments, positive, negative, neutral), n = number of days for polarity change evaluation, DS (Dataset of streaming texts from social media platforms) = [DS_1, DS_2,------, DS_n], T = Threshold.

Output: Sentiments classification using ML model (M), Polarity percentage for offline dataset: P_o (% Positive polarity for offline data), Q_o(% Negative polarity for offline data), R_o(% Neutral polarity for offline data), Polarity percentage for streaming data: PS (% Positive polarity for streaming data) = [PS_1, PS_2, ------, PS_n], QS (% Negative polarity for streaming data) = [QS_1, QS_2, ------- , QS_n], RS (% Neutral polarity for streaming data) = [RS_1, RS_2, ------ , RS_n]

Method:

step 1. Train ML model M

1.1 $D_o \rightarrow$ D_o_train_X, D_o_train_Y, D_o_test_X, D_o_test_Y

1.2 Preprocessing (D_o_train)

01.3 Get features: Feature generation (D_o_train)

01.4 Hyperparameter tuning : parameters + features $\rightarrow$ M

1.5 Test ML model: (D_o_test_X, D_o_test_Y) $\rightarrow$ M

1.6 Deploy ML model M

step 2. Evaluate percentage value P_o, Q_o, and R_o

P_o = count of positive polarity * 100 / total number of reviews

Q_o = count of -ve polarity * 100 / total number of reviews

R_o = count of neutral polarity * 100 / total number of reviews

step 3. Identification of polarity change

3.1 For each i = 1 to n

3.1.1 Streaming data of nth day to ML model M for getting polarity
$DS_i \rightarrow M$

3.1.2 Calculate % value PS_i, QS_i, and RSi of positive, negative, and neutral polarity reviews, respectively, from the streaming dataset for the i^{th} day.
PS_i = count of i^{th} day positive polarity * 100 / total reviews on i^{th} day

QS_i = count of i^{th} day negative polarity * 100 / total reviews on i^{th} day

RS_i = count of i^{th} day neutral polarity * 100 / total reviews on i^{th} day

3.1.3 Evaluate change in polarity.
If (i=1)
$\Delta P_i = PS_i - P_o$, $\Delta Q_i = QS_i - Q_o$ and $\Delta R_i = RS_i - R_o$
Else
$\Delta P_i = PS_i - PS_{(i-1)}$, $\Delta Q_i = QS_i - QS_{(i-1)}$, $\Delta R_i = RS_i - RS_{(i-1)}$
End-If

3.1.4 Identification of polarity change
If ((ΔP_i > Threshold) OR (($\Delta Q_i < 0$) AND ($|\Delta Q_i|$ > Threshold)))
Polarity change class = positive
Else if ((($\Delta P_i < 0$) AND ($|\Delta P_i|$ > Threshold)) OR (ΔQ_i > Threshold))
Polarity change class = negative
Else if (Δ Pi > 0 for continuous Threshold days)
Polarity change class = continuous positive
Else if (Δ Pi < 0 for continuous Threshold days)
Polarity change class = continuous negative
Else
Polarity change class = stable
End-If

3.1.5 Polarity Change Factors Evaluation
If (Polarity change class = positive OR negative OR continuous positive OR continuous negative)
Polarity change flag = "Yes"
Evaluate polarity change using the questionnaire
End-If

End.

EXPERIMENTS AND EVALUATION

Dataset

For our experiment, we have prepared a dataset from Facebook posts of CNN, BBC, and the York Times shared posts, which are downloaded by the site data.world.com (18). This contains news articles and text data from over a hundred thousand records. The dataset has multiple subject data, so we have extracted only subject-specific data. In this study, we also considered offline and online streaming data, so the extracted data was insufficient for training and evaluation. This is solved by manually adding additional data from other social media sources, such as comments from Twitter, YouTube, blogs, etc. From the datasets, a few points that are included for sentiment change identification are like the discussion of foreign policy and immigration policy plans of political parties.

Experimental setup

Python with the nltk and sklearn libraries are used to implement the proposed framework on a machine with an Intel Core i7 processor and 16 gigabytes of RAM. The three stages mentioned in the proposed framework, as shown in Figure 1, are implemented to identify polarity change. For the first part of ML modeling, data preprocessing is done using Polymorphic SBD preprocessor [4]. After getting the final dataset, text features and categorical features are converted into numerical format using one-hot encoding. For this study, 25% of data is used for initial ML model building, and a further 75% of data is used for simulating streaming data for thirty days. Baseline models k-NN, Naive Bayes, Logistic Regression, Linear SVM, Decision tree, and Random Forest were studied and developed ML models using 25% of data, including for train, test, and cross-validation. Each ML model has pros and cons, but Naive Bayes gave the best accuracy for the extracted social media dataset with an evaluation metric log-loss value of 0.6847. The remaining 75% of data was used to simulate real-time data (streaming data) for the next thirty days of study and fed into ML model M. For these thirty days' data, the difference in the percentage polarity change is calculated using the proposed approach.

All the review comments with a polarity change flag = "Yes" were identified, and all answers to seven questions from the proposed questionnaire were gathered. For questions 1, 2, 3, and 4, answers were prepared from the statistics and graph as shown in Figure 2. If the designed ML model is interpretable, then details can be gathered from the deployed classification model M itself for questions 5, 6, and 7. The classification model's interpretability will be helpful for further questions no. 5, 6, and 7. The user's pain points, keywords, etc., are extracted for further analysis by getting interpretable values. Among the four machine learning models we studied, the final classifier model will be selected based on the evaluation metric log loss value, which works quickly even on large datasets. For the social media dataset, the solutions of questions 5, 6, and 7, Naive Bayes found most suitable as its results are interpretable. As conditional probability can be interpretable, it is very clear for each feature how much it contributes towards a certain class prediction. Similarly, Linear Regression, Logistic Regression, and Decision Trees can also be used as interpretable classifiers to get further details about the data points. In this way, all details were gathered for all seven questions from the questionnaire.

Evaluation metrics

In the proposed model, multiple processes are involved in the identification of polarity change, and one of the most important processes is the classification of sentiments as positive, negative, or neutral. If classification accuracy is high, then polarity changes also will be identified correctly. As shown in equations 1, 2, and 3, the change in polarity ΔP_n, ΔQ_n, and ΔR_n are evaluated using the polarity generated by the classification model. In other words, we can say that the accuracy of polarity change identification is proportional to the accuracy of the classification model. The performance of the classification model is evaluated using the evaluation metric log-loss because this is the best evaluation metric for the binary and multiclass classification models.

RESULTS AND DISCUSSION

The extracted data includes comments and reviews about one particular election for the country. The dataset is divided into two parts to simulate offline and real-time data. First is a set of 25%, which is prepared by selecting the first 25% of data according to date and time, which is termed as D_o (Offline dataset), and second is a set of the remaining 75% termed as DS (Streaming dataset) which is considered for real-time data and this is divided across 30 days (DS_1, DS_2, ---- , DS_n) where n = 30. As per the proposed framework, the first stage is to identify an efficient Machine learning model (M) and, after training, deploy the model M into a production environment. The dataset D_o is divided into train, test, and cross-validation for this stage. The dataset is preprocessed, and textual data is converted into numerical formats using one hot encoding technique. For the experiment with the offline dataset, the machine learning model (M) is designed using Naive Bayes, Logistic regression, Linear SVM, and Random forest model and calculated values of P_o (% value of positive polarity), Q_o (% value of negative polarity) and R_o (% value of neutral polarity) as shown in Table 1.

As per Table 1, Naive Bayes gives the best accuracy in terms of log-loss value 0.6847. So, the finalized trained Naive Bayes model (M) is deployed in production, and this deployed model (M) is used for further evaluation of the sentiments for the streaming dataset DS.

With the trained Naive Bayes model (M) for the dataset, the calculated percentage values of positive, negative, and neutral polarity sentiments are P_o = 50.1%, Q_o = 26.45%, and R_o = 23.45%.

Table 1: Log loss and polarity percentage values with respective ML models.

Machine Learning Model	Test log loss	Positive polarity P_o%	Negative polarity Q_o%	Neutral polarity R_o%
Logistic regression	0.7343	49.52	23.18	27.3
Naive Bayes	0.6847	50.1	26.45	23.45
Linear SVM	0.8384	43.26	28.39	28.35
Random Forest	0.7422	46.75	32.84	20.41

Later, as per the next 30 days dataset DS_i (i=1 to 30), the percentage value of positive, negative, and neutral polarity sentiments PS_i and QS_i are calculated for streaming data of 24^{th} day.

Δ Pi and Δ Qi are calculated using P_o, Q_o, PS_i, and QS_i to identify the polarity change value, as shown in Figure 2.

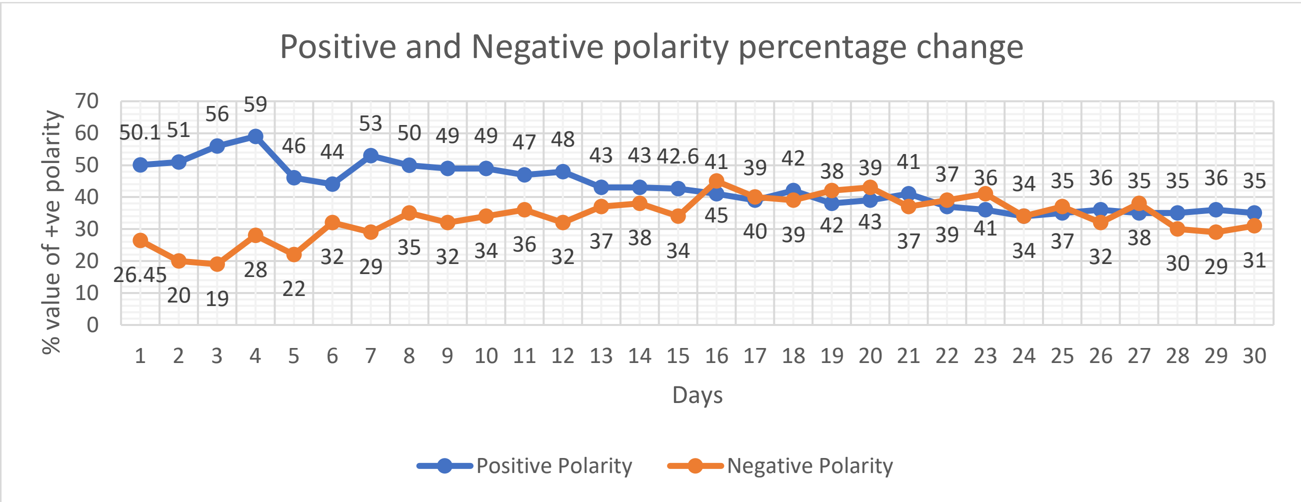

Figure 2: Graphical representation of polarity change over time

Below are answers and details for questions 1, 2, 3, and 4 are gathered using the graph shown in Figure 2.

Q 1. When did the sentiment polarity started to degrade?

- Positive polarity is falling down from the fourth day onwards. After the fourth day, positive polarity did not reach more than or equal to 59%
- Negative polarity is increasing from the third day onwards. After the third day, negative polarity did not reach less than or equal to 19%.

Q 2. What are Positive (good or happy) polarity increase/reduction values?

- Highest positive polarity (59%) on fourth day
- Lowest positive polarity (34%) on 24^{th} day
- From day 1 to day 30^{th}, decrease in positive polarity (50.1 to 35) = 30.1%. From day 1 to day 30^{th}, increase in positive polarity = Not Applicable
- One-day positive polarity highest increase (44% to 53% from sixth to seventh day) = 20.45%
- One-day positive polarity highest decrease (59% to 46% from fourth to fifth day) = 22.03%

Q 3. What are negative (sad or unsatisfactory) polarity increase/reduction values?

- Highest negative polarity (45%) on 16th day

- Lowest negative polarity (19%) on 3rd day
- From day 1 to day 30th, decrease in negative polarity = Not applicable. From day 1 to day 30th, increase in negative polarity (26.45 to 31) = 17.2
- One-day negative polarity highest increase (34% to 45% from 15th to 16th day) = 32.34%
- One-day negative polarity highest decrease (38% to 30% from 27th to 28th day) = 21.05%

Q 4. What does the sentiment gradation curve look like?

- The sentiment gradation curve is shown in Figure 2, which shows all risings and downfalls in the polarity.

For questions 5, 6, and 7, details were gathered from the machine learning model (M) interpretable capability. As discussed above in our study, the Naive Bayes model was found to be the most efficient, so below are the outcomes gathered from its interpretability for questions 5, 6, and 7.

Q 5. What are the pain points are being discussed in the comments?

- First, identify the list of review comments for the day where the polarity change flag = 'Yes'. For those review comments, retrieve the most important features and get interpretable values. These will be compared against the previous day's data, and collectively, these important features give insights into what is being discussed and what pain points are responsible for changing the polarity.

  From the dataset, below are a few points identified for polarity change and were discussed mostly:

  a. "She doesn't call me a racist' amid foreign workers row," and "'Don't call me a racist' for talking about immigration." She defends her plans to make firms do more to employ British people.
  b. His anti-immigration message resonated with conservative voters, while his opposition continued her success by courting the minority.
  c. Is immigration out of control? Migration added 330,000 people to the country's population last year. But how many came from the EU?

Q 6. What are the common keywords in each pain point?

- Identified common words from all features retrieved in the previous step. The dataset provides examples of words such as immigration, jobs, employment, population, fear, migration, migrants, economy, security, workers, border, home, worried, etc.

Q 7. When did the public/users/customer's frustration detected in the comments?

- Collected all words that show the customer's frustration and searched the retrieved review comments for further insights about the behaviors.

CONCLUSION AND FUTURE WORK

For the identification of polarity change in social media data, this study includes data extraction from multiple social media platforms, preprocessing using polymorphic SBD preprocessor, feature engineering using one-hot encoding, and four baseline ML classification models Logistic regression, Naive Bayes, Linear SVM, and Random Forest were trained. For the evaluation of daily real-time review comments, the Naive Bayes model was selected due to its efficiency in terms of log loss value 0.6847. The proposed algorithm calculates the change in polarity by comparing the percentage value of offline data polarity and day-wise real-time data polarity. If a change in polarity is identified, a questionnaire is suggested to evaluate and analyze the factors behind the change in polarity. Also, using the interpretable capabilities of classification ML models, answers to all seven questions are prepared, which helps to analyze the reason for the polarity change so that the organization can take necessary actions. This work can be extended with more feature engineering techniques, and text data can be transformed into vectors using word2vec or a similar approach where semantic relationships can be considered. Also, instead of baseline ML models, deep learning can be used to improve efficiency and interpretability.

REFRENCES

[1] Xia, R., Xu, F., Yu, J., Qi, Y., & Cambria, E. (2015). Polarity shift detection, elimination and ensemble: A three-stage model for document-level sentiment analysis. Information Processing & Management, 52(1), 36–45. https://doi.org/10.1016/j.ipm.2015.04.003

[2] Kolekar, N., Gauri Rao, P., Dey, S., Mane, M., Jadhav, V., & Patil, S. (2016). SENTIMENT ANALYSIS AND CLASSIFICATION USING LEXICON-BASED APPROACH AND ADDRESSING POLARITY SHIFT PROBLEM. Journal of Theoretical and Applied Information Technology, 90(1), 118–125. Retrieved from https://www.jatit.org/volumes/Vol90No1/13Vol90No1.pdf

[3] Tasoulis, S. K., Vrahatis, A. G., Georgakopoulos, S. V., & Plagianakos, V. P. (2018). Real Time Sentiment Change Detection of Twitter Data Streams. 2018 Innovations in Intelligent Systems and Applications (INISTA) (pp. 1–6). https://doi.org/10.1109/inista.2018.8466326

[4] Jadiya, A. K., Chaudhary, A., & Thakur, R. (2020). POLYMORPHIC SBD PREPROCESSOR: A PREPROCESSING APPROACH FOR SOCIAL BIG DATA. Indian Journal of Computer Science and Engineering, 11(6), 953–961. https://doi.org/10.21817/indjcse/2020/v11i6/201106169

[5] Cui, J., Wang, Z., Ho, S., & Cambria, E. (2023). Survey on sentiment analysis: evolution of research methods and topics. Artificial Intelligence Review, 56(8), 8469–8510. https://doi.org/10.1007/s10462-022-10386-z

[6] Bhairavi Rewatkar, Aditya Barhate, & Verma, P. (2024). Application of Machine Learning Algorithms for Analyzing Sentiments Using Twitter Dataset. 2024 2nd International Conference on Sustainable Computing and Smart Systems (ICSCSS), 1392–1397. https://doi.org/10.1109/icscss60660.2024.10624990

[7] Kukkar, A., Mohana, R., Sharma, A., Nayyar, A., & Shah, M. A. (2023). Improving sentiment analysis in social media by handling lengthened words. IEEE Access, 11, 9775–9788. https://doi.org/10.1109/access.2023.3238366

[8] ElMassry, A. M., Abdulla Alshamsi, Abdulhameed, A. F., Zaki, N., & Abdelkader Nasreddine Belkacem. (2024). Machine Learning Approaches for Sentiment Analysis on Balanced and Unbalanced Datasets. 2024 IEEE 14th International Conference on Control System, Computing and Engineering (ICCSCE), 18–23. https://doi.org/10.1109/iccsce61582.2024.10695972

[9] Sánchez-Rada, J. F., & Iglesias, C. A. (2019). Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison. Information Fusion, 52, 344–356. https://doi.org/10.1016/j.inffus.2019.05.003

[10] Jadiya, A. K., & Thakur, R. (2019). Efficient workflow for social big data processing. Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA). https://doi.org/10.2139/ssrn.3365531

[11] Molnar, C. (2024, July 31). 5.7 Other interpretable models | Interpretable Machine Learning. Retrieved from https://christophm.github.io/interpretable-ml-book/other-interpretable.html

[12] Alattar, F., & Shaalan, K. (2021). Using artificial intelligence to understand what causes sentiment changes on social media. IEEE Access, 9, 61756–61767. https://doi.org/10.1109/access.2021.3073657

[13] Giachanou, A., & Crestani, F. (2016). Tracking sentiment by time series analysis. Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval. https://doi.org/10.1145/2911451.2914702

[14] Jiang, Y., Meng, W., & Yu, C. (2011). Topic Sentiment Change Analysis. In Lecture notes in computer science (pp. 443–457). https://doi.org/10.1007/978-3-642-23199-5_33

[15] Liu, M., Guo, C., & Xu, L. (2024). An interpretable automated feature engineering framework for improving logistic regression. Applied Soft Computing, 153, 111269. https://doi.org/10.1016/j.asoc.2024.111269

[16] Tabassoum, N., & Akber, Md. A. (2024). Interpretability of Machine Learning Algorithms for News Category Classification Using XAI. 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), 770–775. https://doi.org/10.1109/iceeict62016.2024.10534385

[17] Mathew, D., & Krishnan, H. (2017). Polarity Shift Handling Techniques: A Survey. International Journal of Computer Applications, 158(6). Retrieved from https://www.ijcaonline.org/archives/volume158/number6/mathew-2017-ijca-912802.pdf

[18] 2012-2016 Facebook Posts - dataset by martinchek. (n.d.). Retrieved from data.world website: https://data.world/martinchek/2012-2016-facebook-posts