A Social Media Complaint Workflow Automation Tool using Sentiment Intelligence

Akshi Kumar, *Member, IAENG,* Vikrant Dabas

***Abstract*—Providing a great, consistent & responsive customer service is imperative to build any successful business. This is primarily because high service quality advocates customer satisfaction which is directly associated to business revenue generation. The business industry has adopted social media as a platform to generate buzz. But listening to the social media & catching the negative word of mouth is equally important. The volatile banking industry has also embraced the social media culture for implementing the customer service aptly. Customers are using Social Media as a complaint channel, but there is no automated process that allows the social media tools such as Facebook, Twitter and LinkedIn to capture, inspect, appropriately manage and process complaints, which is indispensable to a quick response strategy. This paper introduces an end-to-end complaint handling tool that automatically captures incoming posts and/or comments to the posts at the social media source, uncovers the post type as complaint based on sentiment and finally directs the complaints to various departments of the bank where they can be instantly assessed, prioritized and responded. The problem is envisaged as a classification task which is resolved by introducing a tool based on machine learning algorithms. It implemented to social media posts of HDFC Bank Limited, India & a comparative performance analysis for same has been performed.**

***Index Terms*—Customer, Complaints, SERVQUAL, Sentiment, Classifier**

I. INTRODUCTION

CUSTOMERS are the foremost partakers to gauge business performance, as success inevitably depends on customer satisfaction & achieving the right customer service quality. This implies that customer satisfaction is the key business performance indicator & better quality of service has a positive influence on satisfaction, directly contributing to the business revenue.At the same time, the pervasive & voluminous social media is evolving as a potential source of information for business intelligence. By analysing the large volumes of data available on social media, companies can gain deep insights into customer behaviour, sentiments and needs which may facilitate improvement of both product and service. The banking industry has also embraced the social media technology as an ‘enabler’ &‘driver’ of business. As customers increasingly use social media to share opinions on financial products and services, banks must listen, learn and respond, as well as incorporate their social activities into their overall corporate strategies. In most cases, this requires banks to rethink their core business strategies to make them more customer-centric [1].

The provision of customer service via social media channels has become axiomatic, especially due to its cheap, convenient and public nature. With approximately two billion people using social media around the world, a disappointed, unhappy or frustrated customer may choose it as a medium to voice his dissatisfaction in the form of complaint or negative word-of-mouth. According to recent reports [2], platforms such as Twitter [3] and Facebook [4] are being viewed as the perfect vehicles for customers to escalate complaints. The problem manifolds due to the customer’s anticipated response time. Nearly, 42% customers expect response within 60 minutes [5]. Thus, the need of the hour is to intelligently support the customer service department to enable quick & transparent response. Having a dedicated Social Media Manager, who keeps checking the media platforms every ten minutes may seem a solution but in a B2C industry like banking, manually keeping track &monitoring the velocity of complaints in a time-bound manner will be tedious &inefficient. Moreover, a documented escalation policy will be desired to map the complaint to the respective department, the employees & their contact details. Doing all this manually will compromise with the higher expected responsiveness, which has recently been hyped as the metric for customer satisfaction. Motivated by this, we in this paper propose a Social Media Complaint Workflow Automation tool. The idea was to draw correspondence to thewell-known SERVQUAL model (5-dimensional: Tangibles; Reliability; Responsiveness; Assurance; Empathy) [6, 7] which helps evaluating service quality in banking industry, as studies have indicated that out of its five dimensions, responsiveness & empathy can be improved with the help of social media analytics. These service quality factors (SERVQUAL model dimensions) are defined as follows:

Responsiveness: refers to the willingness of service providers to help customers and provide prompt service.

Empathy: refers to the provision of caring and individualized attention to customers.

The proposed Social Media Complaint Workflow Automation tool depicts the online complaint management as a two-level classification problem that uses sentiment analysis on the social media posts to quickly resolve complaints. Sentiment analysis or Opinion mining, as it is sometimes called, is one of many areas of computational studies that deal with opinion oriented natural language processing [8].It normally involves the classification of text into categories such as “positive”, “negative” and in some cases “neutral”.The tool expounded in this paper, first automatically captures incoming posts and/or comments to bank’s post at the social media source. Not all posts are

Manuscript received March01, 2016
Akshi Kumar is a faculty with the Dept. of Computer Science & Engineering, Delhi Technological University, Delhi, India (e-mail: akshikumar@dce.ac.in).
Vikrant Dabas is an under-graduate student with the Dept. of Computer Science & Engineering, Delhi Technological University, Delhi, India

complaints and so as the first classification task, we uncover the type of post into complaints (negative feedback), queries (neutral) or positive feedback/irrelevant posts by perceiving the sentiment and then finally direct the complaints to various departments of the bank using another classifier so that the workflow is complete for a prioritized & quick response. The tool has been implemented to social media posts of HDFC Bank Limited, India [9] (Facebook & not Twitter, because tweets about complaints were very less as compared to the posts on Facebook page) & tested for various classification models. A performance analysis interpreting the results is also presented which justifies the need to automate the process that reduces the service-quality gap & incorporates sentiment intelligence within the business intelligence model.

The paper is organized as follows: Section 2 gives a brief explanation of related background work in thisdirection. Section 3 describes the proposed workflow model of the proposed Social Media Complaint Workflow Automation Tool using Sentiment Intelligence. Section 4 gives a brief overview of the implementation and analysis using a case study of HDFC Bank. Section 5 concludes the research undertaken while discussing the potential applications & directions of future work.

II. Related Work

According to Capgemini [10] worldwide,the social network user base is expected to increase from 1.47 billion in 2012 to 2.55 billion in 2017. Alsosocial media yields such a wealth of customer-based, customer-voiced information which banks can harness to better understand &enhance the overall customer experience.Studies aligning customer satisfaction to the levels of service quality are available much in literature. Gautam and Singh [11] as well as Lohani and Bhatia [12] have used SERVQUAL model to identify the service quality gaps in banking sector. Moreover, according to them,Empathy & Responsiveness,are found to be the most vital and strategic determinants of service quality and customer satisfaction for banks.Santhiiyavalli [13] also concluded that among the five dimensions 'Responsiveness' and 'Empathy' are the major factors responsible for customer satisfaction. So efforts should be made to reduce the service quality gap by improving these factors. Another study conducted by Rathee, Deveshwar and Rajain[14], concluded that the highest gap was found in the dimension of reliability and empathy. And banks have to reduce this gap giving individual personal attention to understand customer specific needs. Moreover, results of a study by Brahmbhatt and Panelia[7] provide evidence that the SERVQUAL dimensions are a useful tool to predict overall satisfaction.

Social media should be embedded into a bank's entire ecosystem because it impacts numerous areas, such as customer relationship management, risk management, product and service design, customer education, etc. [1]. Sentiment analysis of micro blogging or social networking websites has been a wide area of research in the present decade [15]. But its usage hasn't been much extended in the banking sector to improve the service quality and customer satisfaction. Most of the process, if ever done, for improving the service quality and customer satisfaction with the help of social media has been done manually rather than opting for an automated process. So to counter this problem, this paper suggests an automated method for capture individual posts at the social media source, classify & identify the complaints and route them to the appropriate resolution owners, supervisory levels and management. The next section describes the proposed workflow tool.

III. The Proposed Social Media Complaint Workflow Automation Tool

The Social Media Complaint Workflow Automation tool proffered here depicts an online complaint management as a two-level classification problem that uses sentiment analysis on the social media posts and/or comments to quickly resolve complaints. We define Sentimental intelligence as a conjunction of intelligence, empathy and sentiments to enhance opinion and comprehending social interactive dynamics. It is the capability of voicing and monitoring public opinion enabling apt classification of sentiments which guides and impacts thinking attitude and aptitude. Instituted on this the proposed tool first automatically captures incoming posts & comments to bank's post at the social media source and classifies them as enquiries, positive feedback/irrelevant post or complaints by perceiving the sentiment and then finally directs the complaints to various departments of the bank using another classifier so that the workflow is complete for a prioritized & quick response. For both levels of classification, Naïve Bayes Classifier, its 3 variants, namely Multinomial Naïve Bayes, Bernoulli Naïve Bayes and Gaussian Naïve Bayes & Support Vector Machine are employed. The tool is implemented to the social media posts of HDFC Bank Limited, India & tested for various classification models.The following figure 1 illustrates the basic workflow architecture of the proposed tool.

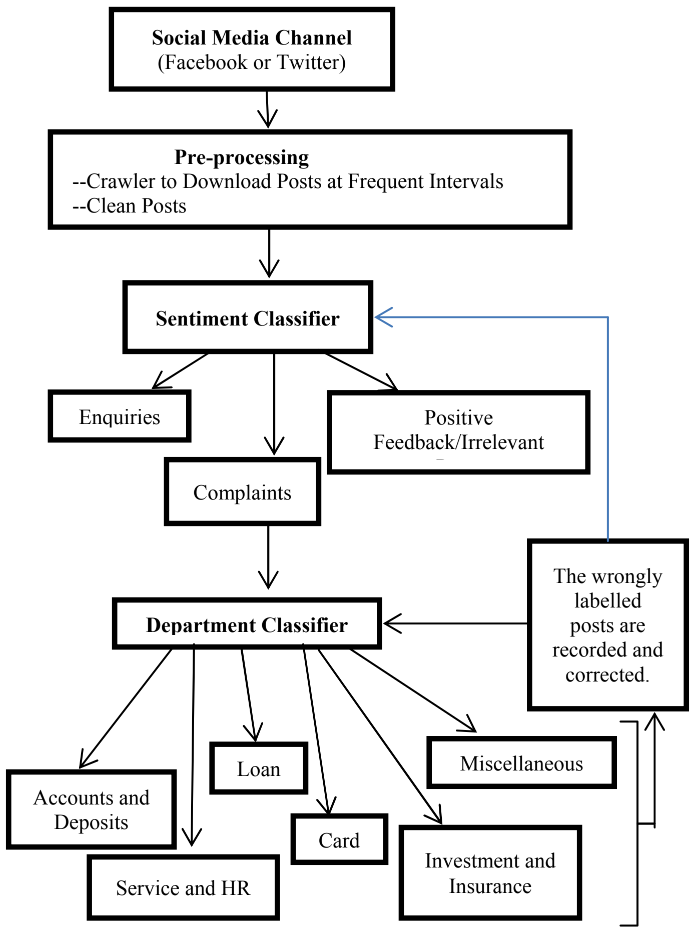


Fig 1. The proposed complaint workflow automation

ISBN: 978-988-19253-0-5
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

The following sub-sections expound the details pertaining to each component of the proposed workflow:

A. *Pre-processing*

Large publically available social media datasets are crawled to fetch data using explicit APIs. Data extracted in its original form is hard to use for analysis directly. It is often incomplete, noisy & inconsistent [16]. Thus, the pre-processing consists of the following two steps:

1) *Crawling social media channels to capture posts*

According to the Internet Society's "Global Internet Report 2015" [19], there are more than 3 billion people online globally who use Internet for communicating, sharing, finding information, publishing & commerce amongst others. Statistics also reveal that more than 2 billion people worldwide have social accounts (Facebook, Twitter etc.). The Social networking websites like Facebook and Twitter which are a substantial, unparalleled sourceof user-generated content are easy to crawl and extractingonline posts from them can be done in a straight forward manner.To get large publically available Facebook datasets, we use Facebook's graphAPI.We form a URL & hit the graph API to receive JSON object responsescontaining data about Likes, talking about etc. Required data is extracted from these JSON objects and saved in a separate file for further processing.To get large publically available Twitter datasets, we use Twitter API. In order to make authorized calls to twitter's APIs, the application must first obtain an OAuth access token on behalf of a Twitter user or can issue Application-only authenticated requests when user context is not required. For implementation & analysis Facebook was selected over Twitter, because tweets about complaints were very less as compared to the posts on Facebook page.

2) *Cleaning Module*

Data cleaning for quality is significantly desirable for higher accuracy & better results. The data extracted from Facebook containsmany undesirable symbols, words, URLs, E-mail ids, etc. that may slow down the applied algorithms or even make them perform incorrectly. Hence, it becomes necessary to incorporate a data cleaning module.The following strategy has been identified to clean the downloaded data for processing:

- Removal of # tags: Facebook has incorporated hashtags, which works on personal profile posts, fan page posts, group posts, event posts, and all comments.
- Removal of "@": @ symbol *is* used for tagging or to "mention" someone on Facebook.
- Removal of URLs & E-mail ids.
- Removal of numbers, digits, alphanumeric data, punctuations, symbols.
- Expanding short-forms (acronyms), if any.
- Removal of non-English posts by encoding them into ASCII and if encoding produces any exception, then reject them, otherwise, select the posts.
- Removal of stop-words.

The original posts& the cleaned posts are kept temporarily at each step. The cleaned post is used for processing & the original post are stored for further use.

B. *Classification of posts*

Cleaned posts and/or comments to the post are labeled and trained over the first classifier (Sentiment Classifier). These are classified into three major categories, namely:

- *Complaints*: an expression of dissatisfaction on a consumer's behalf to a responsible party.
- *Inquiries*: a request for information from the concerned department of the organization.
- *Positive feedback/ irrelevant posts*: posts that are basically feedback from customers that show a positive response or a set of posts that are either for advertisement purposes or are completely irrelevant.

The sentiment classifier extracts complaints from all the posts and the department classifier separates the complaints into different department files. Each file issent to its particular department. The department class files which are created are explained as follows:-

- *Accounts and Deposits:* This department handles everything from opening to closing of various types of accounts including Salary Accounts, Current Accounts, and Saving Accounts etc.
- *Loans:* This department handles all the work related to processing all kinds of loans, be it a personal loan or car loan.
- *Cards:* This department is responsible for issuing credit/debit or any other type of cards and maintains information about them.
- *Investment and Insurance:* Getting a life insurance to managing Mutual Funds, all the related activities are handled by this department.
- *Customer Service:* This is the department which supports the customers to approach all the other departments. It provides services such as Net Banking, Phone enquiry.
- *Miscellaneous:* All the other departments like Recovery, Lockers, etc. can be grouped into this category. Moreover, there are departments that are unique to a particular bank. So those departments are also grouped into this category.

These sub-categories don't exactly reflect one department (like the sub-category Miscellaneous) but tell us to which department these complaints should be forwarded too.

The sentiment and department classifiers aretrained over a labeled data set. Once the complaints have been labeled department-wise, the original posts that were mapped to the cleaned posts during pre-processing are sent to the respective departments for further action & proceedings and the cleaned posts are discarded. In case, they reach a wrong department, i.e., if it is found to be incorrectly classified, the posts are labeled/marked with the correct department and are used to update the classifiers.

For both classifiers, the learning algorithmsconsidered were Naïve Bayes (NB), its 3 variants, namely, Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB) and Gaussian Naïve Bayes (GNB) & Support Vector Machine (SVM) [17]. Naïve Bayes classifier uses a bag-of-words [18] approach for sentiment analysis on the content. The Bernoulli Naive Bayes can be used when in our problem the absence of a particular word matters. For example Bernoulli is commonly used in Spam or advertisement detection with very good results.

The Multinomial Naive Bayes is suitable for classification with discrete features &is typically used when occurrence of the word matters more than the frequency. For example it does not really matter how many times someone mentions the words “bad”, “awful”, “terrible”, but rather only the fact that he does.ABinarized Multinomial Naive Bayesis used for the sentiment classification (first classifier), i.e., it has binary weighting function where the value 1 means that the word occurs in the particular document, and 0 means that the word does not occur in this document. A tf-idf Multinomial Naive Bayes is used for the department classification (second classifier). The tf-idf approach assumes that the importance of a word is inversely proportional to how often it occurs across all documents.The Gaussian Naïve Bayes & support vector machine were also used for a comparative performance analysis.

IV. Implementation and Performance Analysis

To clearly illustrate the effectiveness of the proposed method, a case study of HDFC Bank India’s Facebook page is presented. The following figure 2 shows the HDFC Bank, India Facebook page with some sample posts.

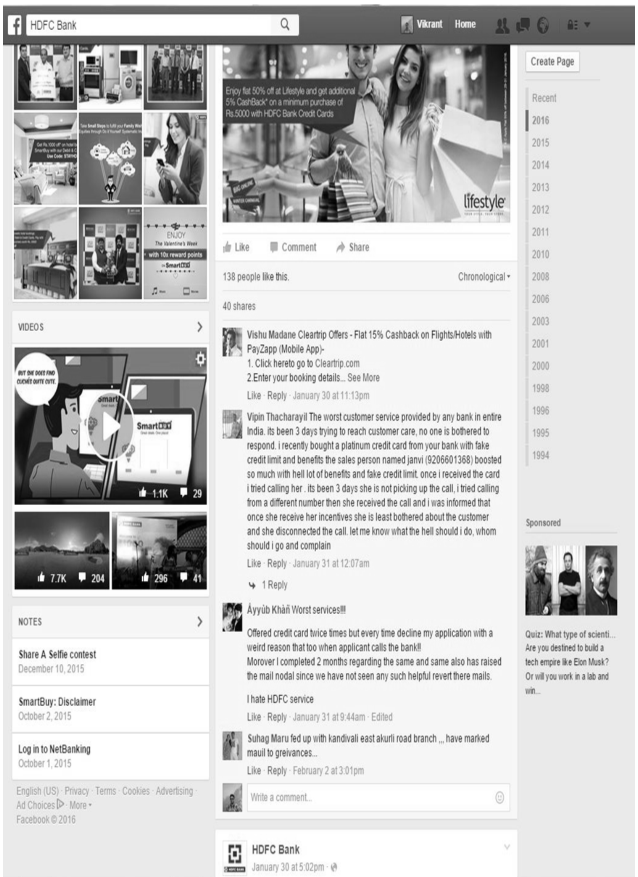


Fig 2. Snapshot of HDFC Bank, India Facebook page with some sample posts.

Each of the components of the workflow is exemplified to demonstrate how it actually works by firstly identifying the complaints from the Facebook post & then referring it to the appropriate department which can cater a quick response, thus automating the entire process in an efficient manner.

A. Pre-processing

1) Crawling social media channels to capture posts

To download the Facebook data, an access token was generated by registering as a developer on Facebook. The next step was to make an HTTP GET request to the Facebook graph API using the urllib.request module in python. As a response, a JSON dump of the data was received and useful information (user who posted, post text, time of posting) wasextracted & stored in a CSV file.

2) Cleaning

The cleaning process as discussed in the previous section, involved removing hashtags like #HDFCBank etc., tagging symbol @ (@HDFC_Bank), URL supporting additional information (links to bank’s website) and any e-mail ids that are included in the post, example: creditcards@hdfcbank.com. It also scrubs numbers (account numbers/mobile numbers/dates etc.) and very common stop-words like “a, an, the, this, that”. Short-forms like A/C, a/c (account) & CC (Credit-card) are expanded too.

B. Classification

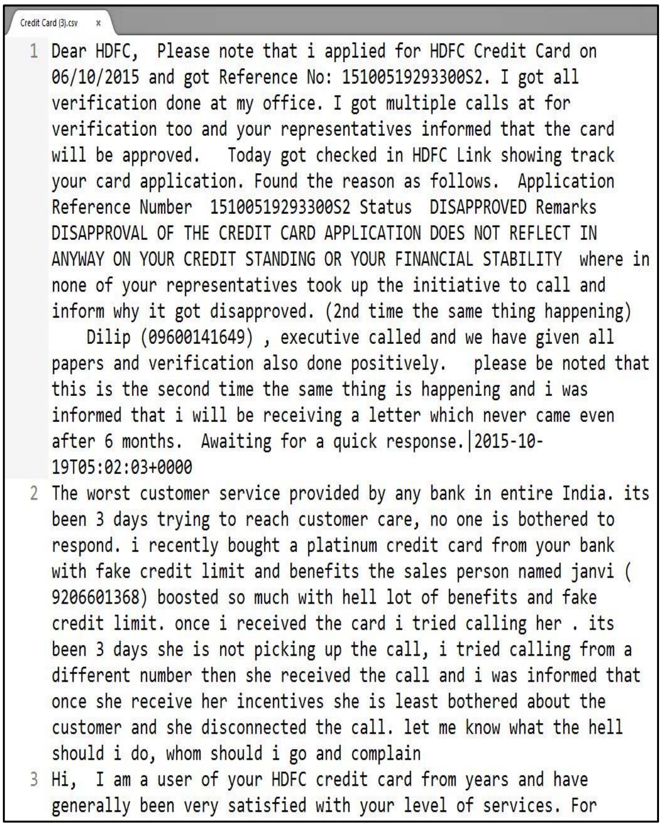
As discussed in the previous section, for both levels of classification, Naïve Bayes Classifier (NB), its 3 variants, namely Multinomial Naïve Bayes (MNB), Bernoulli Naïve Bayes (BNB) and Gaussian Naïve Bayes (GNB)& Support Vector Machine (SVM) have been employed. The results from the first classification task for finding sentiment from the post are shown below in table 1:

TABLE I
SAMPLE RESULTS OF SENTIMENT CLASSIFIER

POST	SENTIMENT	CLASS
This is a warning to #HDFC #HDFCBank which has the worst credit card division and i am going to file a case against them for torturing me in a credit card case. When you people have already credited an amount of 92k from my bank account six months back making a lean on my account. This was the final payment and as you people told me that i will be getting a NOC. My credit card was blocked from 2-5 years approx. Even now i am receiving credit card statements every weak. This is really ridiculous. I request you to issue me a letter for no dues. Ajay Katnoria +91-9892622230 katnoria.ajay@gmail.com	NEGATIVE	COMPLAINT
This is the worst bank i have ever seen. Today when i payment on amazon app.the amount was deducted from my account. But	NEGATIVE	COMPLAINT

amazon get any payment from my side. When i called at hdfc customer care they told me it will take 2-4 days. How could be possible...that is my money. They have to send in my account in an hour... And even thy didnt called me that the amount is deducted so thy will refund me.		
What documents required for converting the Savings account in to salary account..I have offer letter, and pay slip with me..pls comment	NEUTRAL	ENQUIRY
HI, Is payzapp available for windows phone?	NEUTRAL	ENQUIRY
Team engage4more partners HDFC Bank in their brand new performing arts talent based contest 'HUNAR' . A PAN India turnkey Mandate includes conceptualisation, format creation and production. Great to see this wave of corporate belief in engaging employees around their passion. #hr #employeeengagement #india #performingarts pics from regional finale west. India finale follows.	POSITIVE	POSITIVE FEEDBACK /IRRELEVANT
Best Bank in India. Finest banking service.Outstanding performance very very Happy customer.	POSITIVE	POSITIVE FEEDBACK /IRRELEVANT

The department classifier categorizes the complaints acquired into various departments of the bank for further necessary action on the complaints. A snapshot for the complaints fitting to the Credit Card department is given below:


Credit Card (3).csv

1 Dear HDFC, Please note that i applied for HDFC Credit Card on 06/10/2015 and got Reference No: 15100519293300S2. I got all verification done at my office. I got multiple calls at for verification too and your representatives informed that the card will be approved. Today got checked in HDFC Link showing track your card application. Found the reason as follows. Application Reference Number 15100519293300S2 Status DISAPPROVED Remarks DISAPPROVAL OF THE CREDIT CARD APPLICATION DOES NOT REFLECT IN ANYWAY ON YOUR CREDIT STANDING OR YOUR FINANCIAL STABILITY where in none of your representatives took up the initiative to call and inform why it got disapproved. (2nd time the same thing happening) Dilip (09600141649) , executive called and we have given all papers and verification also done positively. please be noted that this is the second time the same thing is happening and i was informed that i will be receiving a letter which never came even after 6 months. Awaiting for a quick response.|2015-10-19T05:02:03+0000

2 The worst customer service provided by any bank in entire India. its been 3 days trying to reach customer care, no one is bothered to respond. i recently bought a platinum credit card from your bank with fake credit limit and benefits the sales person named janvi (9206601368) boosted so much with hell lot of benefits and fake credit limit. once i received the card i tried calling her . its been 3 days she is not picking up the call, i tried calling from a different number then she received the call and i was informed that once she receive her incentives she is least bothered about the customer and she disconnected the call. let me know what the hell should i do, whom should i go and complain

3 Hi, I am a user of your HDFC credit card from years and have generally been very satisfied with your level of services. For

Fig 3. Snapshot of result from department classifier with complaints fitting to the Credit Card department

C. Results& Performance Analysis

This section discusses the result & does a performance analysis for both the classification tasks with all learning algorithms exercised.

The following table 2 shows the performance analysis of the sentiment classifier(firstclassifier) on the basis of accuracy and time. The Multinomial Naive Bayes Classifier gave the highest accuracy. Bar graphs are shown for the same in figures 4 and 5 respectively.

TABLE II
ANALYSIS OF SENTIMENT CLASSIFIER

Algorithm	Accuracy(%)	Time(sec)
Naive Bayes Classifier	78.667	0.1562
Multinomial Naive Bayes Classifier	83.667	0.0312
Bernoulli Naive Bayes Classifier	79.667	0.0468
Gaussian Naive Bayes Classifier	67.667	0.0156
Support Vector Machines	76.333	0.0312

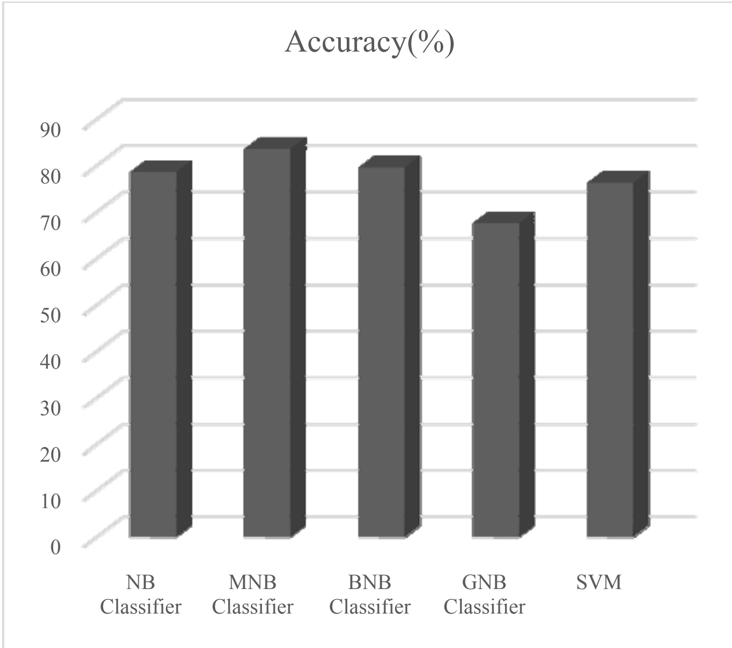


Fig 4. Accuracy Analysis of Sentiment Classifier

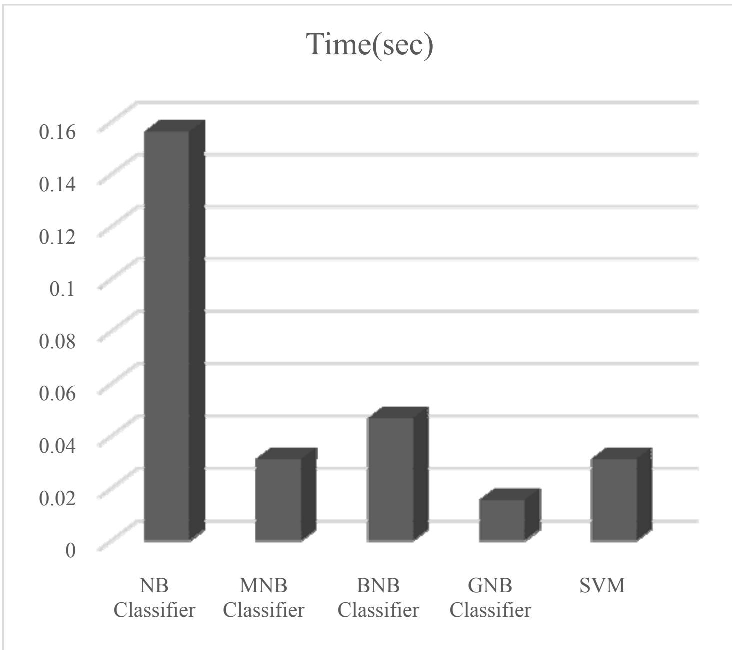


Fig 5. Time Analysis of Sentiment Classifier

For the department classifier (second classifier), the following table 3 shows its performance on the basis of accuracy and time. Here too, the Multinomial Naive Bayes Classifier gave the highest accuracy and Gaussian Naïve Bayes gave the lowest accuracy.Gaussian Naïve Bayes works well for continuous numerical data and owing to our discrete dataset, its performance was compromised. Bar graphs are shown for the same in figures 6 and 7 respectively.

TABLE III
ANALYSIS OF DEPARTMENT CLASSIFIER

Algorithm	Accuracy(%)	Time(sec)
Naive Bayes Classifier	70.667	0.2035
Multinomial Naive Bayes Classifier	74.667	0.1002
Bernoulli Naive Bayes Classifier	73.667	0.0989
Gaussian Naive Bayes Classifier	61.667	0.0311
Support Vector Machines	69.333	0.0899

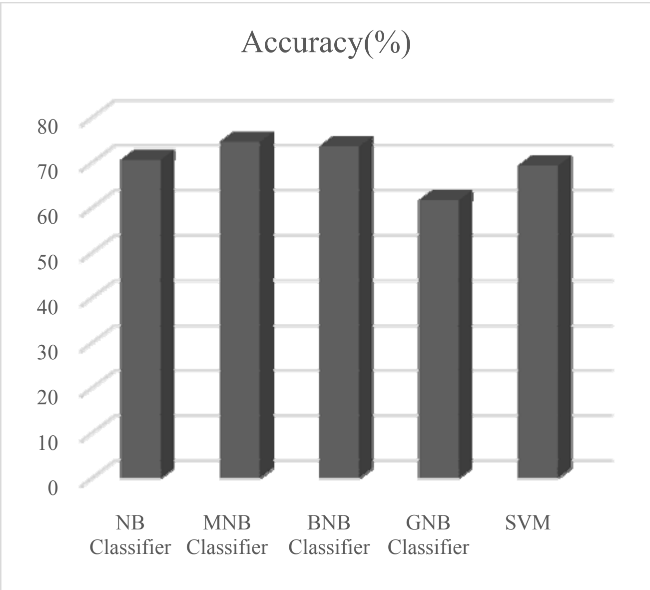


Fig 6. Accuracy Analysis of Department Classifier

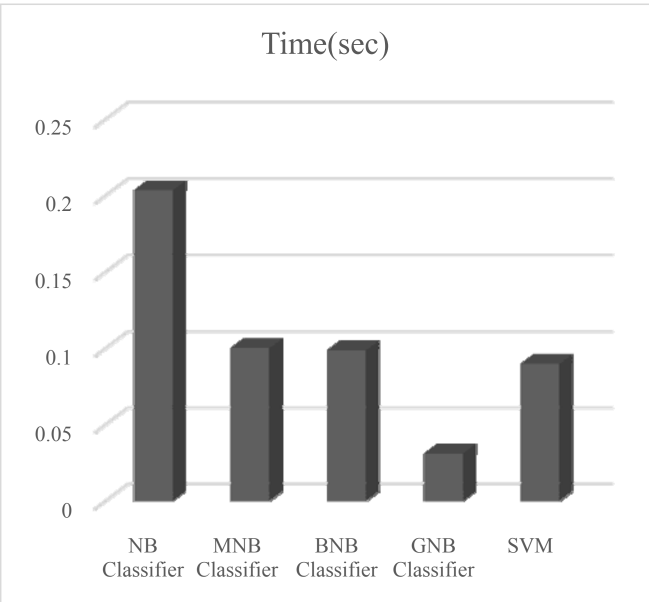


Fig 7. Time Analysis of Department Classifier

V. CONCLUSION

Over the past decade banks have been compelled to examine new business models & technological innovations to match the tech-savvy, socially mindful Gen-Y. Thus tracking social media for negative word-of-mouth (complaints) should be regarded as a pro-active response policy. The tool proposed in this paper is an automation of the process that manages complaints at a social media sourceby firstly capturing incoming posts and/or comments to the posts, extracting the complaints by uncovering the post-type based on sentiment and finally directing them to the concerned departments of the bank where they can be instantly assessed, prioritized and responded. A comparative performance analysis for implementationon HDFC Bank Facebook page posts gave encouraging results with a highest accuracy of approximately 83% for sentiment classification & approximately 75% for department classification.To further improve the accuracy results & optimize results in terms of response time, we intend to examine association mining and/or co-information metric for relation analysis amongst co-occurring terms. We also plan to extend this workflow for department-wise enquiry classification. A fine-grain sentiment analysis, probably emotion-based is envisioned too. The idea is to yardstick public mood toward services, at regular intervals, say quarterly, which may help in brand development & reputation management.

REFERENCES

[1] D.Hazarika, S. Nag , "How Banks Can Use Social Media Analytics to Drive Business Advantage", *Cognizant Insights Whitepaper,*2014.
[2] J. Causon, "Customer complaints made via social media on the rise", *The Guardian*, May 2015.
[3] Twitter Homepage: www.twitter.com
[4] Facebook Homepage: www.facebook.com
[5] J. Baer, "42 Percent of Consumers Complaining in Social Media Expect 60 Minute Response Time", *Convince&Convert*, Dec 2015.
[6] M. Nagabhushanam, "A study on Customer Service Quality of Banks in India", *Research Report,*www.iibf.org.
[7] M. Brahmbhatt, D. Panelia, "An Assessment of Service Quality in Banks", *Global Management Review*, 2(4), 2008, 37-41.
[8] A. Kumar, T.M. Sebastian, "Sentiment analysis: A perspective on its past, present and future",*International Journal of Intelligent Systems and Applications (IJISA)* 2012; 4(10): 1-14.
[9] Official Homepage of HDFC Bank: www.hdfcbank.com; Facebook page of HDFC Bank: www.facebook.com/HDFC.bank
[10] V.K. Suvarna, B. Banerjee, "Social Banking: Leveraging Social Media to Enhance Customer Engagement", *Capgemini*, 2014.
[11] S. Gautam, A.Singh, "To identify service Quality gaps in Banking Sector: A Comparative Study of Public Sector Banks And Foreign Banks", *Global Journal of Finance & Management*, Vol. 6, No. 5, 2014, 439-444.
[12] M. B. Lohani, P. Bhatia, "Assessment of Service Quality in Public and Private Sector Banks of India with Special Reference to Lucknow City", *International Journal of Scientific and Research Publications*, Vol. 2, Issue 10, 2012
[13] G. Santhiyavalli, " Customer's perception of service quality of State Bank of India-A Factor Analysis", *International Journal of Management & Business Studies*, Vol.1, Issue 3, 2011.
[14] R.Rathee, A.Deveshwar, P.Rajain "To Identify Service Quality Gaps in Banking Sector: A Study of Private Banks", *International Journal of Emerging Research in Management & Technology*, Vol.3, Issue 7, 2014, 101-106.
[15] A. Kumar, T.M. Sebastian, " Sentiment Analysis on Twitter", *International Journal of Computer Science Issues*, Vol. 9, Issue 4, No 3, 2012,372-378.
[16] G. J. Williams, S. J. Simoff,*Data Mining: Theory, Methodology, Techniques and Applications*, Springer, 2007.
[17] R.Schapire, "Machine Learning Algorithms forClassification", Princeton University
[18] Rennie J., Shih, L., Teevan, J., &Karger, D., "Tackling the Poor Assumptions of NaïveBayes Text Classifiers" *Proc. of ICLM-2003.*