International Journal of Research
Available at https://edupediapublications.org/journals
e-ISSN: 2348-6848
p-ISSN: 2348-795X
Volume 04 Issue 14
November 2017

Analysis and classification for Software as a Service Reviews Using Naive Bayes

Swathi & Smitha Karpe
[1]M. Tech Student, Department of CSE, MallaReddy Engineering College For Women's, Maisammaguda, Dhulapally, RangaReddy, Telangana, India.

[2]Associate Professor, Department of CSE, MallaReddy Engineering College For Women's, Maisammaguda, Dhulapally, RangaReddy, Telangana, India.

Abstract— *with the quick development of cloud administrations, there has been a noteworthy increment in the quantity of online customer audits and sentiments on these administrations on various web-based social networking stages. These audits are a wellspring of important data as to cloud showcase position and cloud shopper fulfilment. This investigation investigates cloud purchasers' audits that mirror the client's involvement with Software as a Service (SaaS) applications. The surveys were gathered from various web-based interfaces, and around 4000 online audits were broke down utilizing feeling examination to distinguish the extremity of each survey, that is, regardless of whether the assessment being communicated is sure, negative, or impartial. Likewise, this examination builds up a model for anticipating the feeling of Software as a Service customers' audits utilizing a managed learning machine called a help vector machine. The opinion comes about demonstrate that 62% of the surveys are sure which shows that purchasers are in all likelihood happy with SaaS administrations. The outcomes demonstrate that the expectation exactness of the SVM-based Binary Occurrence approach (3-crease cross approval testing) is 92.30%, showing it performs better in deciding supposition contrasted and different methodologies (Term Occurrences, TFIDF). This work additionally gives important understanding into online SaaS surveys and offers the exploration group the main SaaS extremity dataset.*

Index Terms*—SaaS reviews, sentiment analysis, sentiment classification, supervised machine learning, SaaS polarity datase*

I. INTRODUCTION

With of the wide acknowledgment of cloud administrations and the exponential development in the quantity of cloud purchasers, online cloud purchasers' audits have expanded as of late crosswise over various web stages. Online networking locales (survey destinations, web journals, gatherings) give countless audits and shoppers utilize these stages to express their perspectives and feelings on distinctive parts of cloud administrations utilizing normal dialect. These surveys assume a noteworthy part in controlling the buys of potential customers on cloud administrations [1].

Giving the expanded number of cloud buyer audits and cloud audit locales on the Internet, it is impossible that potential customers will experience each of these audits physically to recognize a helpful audit. Distinguishing the helpfulness of surveys is a critical undertaking, the same number of associations might want to know customers' conclusions on their administrations. Likewise, potential customers depend on these surveys while picking items or, then again benefits. A case of this is lodging surveys or travel goal surveys the same number of individuals depend on other explorers' encounter while picking their travel goals and lodgings[2]. This genuine case demonstrates unmistakably the basic idea of surveys in driving the commercial center in various spaces.

When all is said in done, an audit is a free arrangement content written in common dialect, in which individuals unreservedly express their feelings and feelings. A case of a purchaser audit is "Their client benefit is unpleasant. More than 24 hours to get a call. No one answers the telephone. I attempted to scratch off my membership and they continue charging my card. I've squandered more than 2 hours attempting to wipe out this administration. I won't prescribe this organization. Be careful! I presume they are not in the USA." [3]. These surveys can be subjective or objective; a subjective audit is one where the analyst communicates their feelings and gives criticism about a specific administration or item while a target audit is one where the analyst presents a reality or thought regarding the service[4]. Along these lines, investigating the content of the surveys to distinguish its

tone is not a simple undertaking. As of late, conclusion mining has turned out to be exceptionally notable as a sub-train of information mining to dissect and arrange individuals' sentiments. Supposition mining alluded to as computational semantics and is utilized to extricate client's sentiments from the World Wide Web (WWW). In the writing, the term conclusion mining is utilized conversely with notion examination, as notion investigation is utilized to decide the extremity and subjectivity of individuals' conclusions.

This exploration performs notion examination on Software as Service (SaaS) surveys that have separated from audit sites. It is remotely overseen and conveyed by the cloud suppliers what's more, individuals can subscribe to the application and utilize it whenever [5]. Feeling examination is a programmed investigation of regular dialect content utilizing normal dialect handling (NLP) furthermore, machine learning (ML) procedures to distinguish the content extremity, that is, regardless of whether it is certain, nonpartisan, or negative [6].A extensive measure of writing has been distributed on feeling mining and supposition investigation of purchasers' surveys. These examinations concentrate on the investigation of audits and the elements furthermore, designs removed from the audit message in a wide range of uses (i.e. business, vacationer and instructive), for example, in [7], [8]. Notwithstanding, in the distributed computing area, insufficient consideration has been paid to the examination of online feelings. A few ponders utilize cloud administrations to perform assessment investigation and arrangement as it facilitates the procedure work process. Sayeedunnissa et al. utilized Google's Platform as a Service (PaaS), in particular the Google App Engine (GAE) to perform sentiment examination on Twitter surveys and after that assembled a classifier utilizing the ML strategy [9]. Different examinations have exploited cloud administrations to build up an ongoing sentiment mining application [10].

To the best of our insight, to date, there have been no considers that emphasis on breaking down and compressing buyers' online sentiments concerning cloud administrations. This demonstrates the need to acquire more prominent knowledge into cloud buyers' surveys, contemplating the expanded measure of cloud surveys on the Internet.

The reason for this progressing study is to fill the hole in the writing by mining and outlining SaaS buyers' surveys utilizing notion examination and grouping procedures. All the more particularly, this investigation makes utilization of the Blue Pages Audit Dataset, which is an accumulation of Software as a Service audit information. The dataset has been accumulated from various web entrances (cloud audits, Get App) and is accessible online at www.Bluepagesdataset.com. Around 4000 audits have been dissected, and the slant of each survey has been resolved. Along these lines, this investigation makes a huge commitment to the exploration on the cloud conclusion mining area by

- giving a synopsis of customers feelings on SaaS administrations.

- giving a first extremity SaaS dataset.

- giving a Support Vector Machine (SVM) arrangement model to decide the estimation for SaaS surveys where the state of mind in obscure.

The paper is sorted out into six segments as takes after: Section 2 gives the foundation to the exploration, Section 3 portrays the dataset, Section 4 points of interest the system, Section 5 presents the outcomes and dialog, and Section 6 gives the conclusion and the headings for future work.

II. RELATED WORKS

The advancement of Web 2 and web-based social networking have changed the Internet into a wide talk room, where individuals trade their thoughts, encounters, and assessments on various viewpoints [11], what's more, it has likewise added to the extension of online business. The Internet has turned into an extensive online commercial center, particularly since the rise of cloud administrations. Showcasing for on the web administrations varies in nature from conventional promoting as it is difficult to monitor online customers. There are a vast what's more, assorted exhibit of online customers who have a tendency to be unequivocally affected by different people groups encounters and remarks while picking an administration/item [12], as appeared in the case of picking an inn [2].

In the writing, terms, for example, input, feelings, assessments furthermore, surveys are utilized to allude to individuals' composed remarks to impart what they think

about specific items and administrations. Distinctive philosophies, for example, slant investigation, conclusion mining or subjective/include investigation, have been utilized to empower PCs to break down the importance/feeling of a composed content [13].

A substantial and developing assemblage of writing has examined people groups audits, for example, in[14], and a great part of the current writing on mining audits gives careful consideration to on the web administrations (counting inns appointments, tourism, and day by day utilize items). The vast majority of these investigations have utilized distinctive social media online stages as their principle wellspring of surveys over diverse areas [15], [16], [17].

In late years, cloud administrations have turned out to be across the board also, have generally supplanted in-house registering assets by offering an on the web and on-area benefit. With the rise of these administrations, an expanding number of audit sites have showed up on which cloud shoppers can put surveys furthermore, input [18][19]. Like whatever other business chiefs, cloud chiefs and suppliers are quick to know what individuals think about their administration/item and to get to input (additionally alluded to as the buyers' voice) [20].

Assumption examination, sentiment mining and subjective/highlight investigation are likewise phrasings utilized as a part of the writing to allude to procedures which empower the PC to translate the meaning/feeling of the composed content. Basically, opinion examination, which is the concentration of this exploration, is the procedure of utilizing PC based systems to consequently dissect a bit of content (Twitter, audits, article) to distinguish whether the scholars sentiments/suppositions are sure, impartial, or negative [21]. Notion investigation utilizes two surely understood methodologies in particular: Characteristic Language Processing (NLP), and Machine Learning (ML).

As showed in the writing, these methodologies have been connected effectively in a wide scope of utilizations and different spaces including tourism and training [2], [22], [23]. Grammatical form (POS), the N-gram calculation, and the sack of-words are outstanding NLP systems that are generally used to distinguish the extremity of content.

In wide software engineering terms, ML can be characterized as a set of techniques/calculations which consequently recognize an example from the information [24], characterizing normal learning into administered furthermore, unsupervised learning. furthermore, it arranged in regards to normal of learning into Supervised and Unsupervised. The objective of regulated learning is to prepare the model to consequently outline input x to yield y, given an arrangement of information sources and the coveted

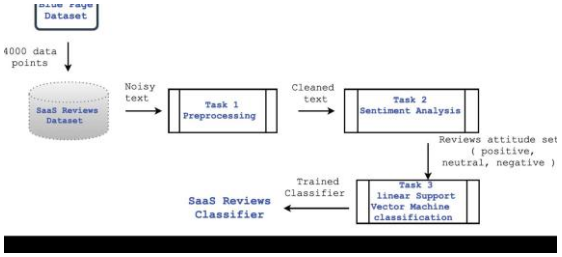

Figure 1. Framework of the Model

yield. Not at all like regulated learning, unsupervised learning points to find fascinating examples in the information just given the info without being recounted the coveted yield for each info. For case, Qiang et al. investigated explorers' surveys on seven diverse areas in Europe and America. In their investigation of travel web journals, Qiang et al. distinguished negative and positive surveys utilizing conclusion order systems. They additionally looked at the vary ent machine learning calculations (i.e. bolster vector machine, Naive Bayes) in light of the Ngram demonstrate [25]. So also, different creators [26] led distinctive tests utilizing SVM learning and survey datasets from various spaces. Utilizing supposition mining and notion investigation, they decided if the estimation of the SaaS audits were sure or negative and produced the preparation dataset for regulated learning. They connected Support Vector Machine learning (SVM) calculation and directed extraordinary tests utilizing a few weighting outline.

In this paper, we concentrate on breaking down and ordering SaaS purchasers' surveys on a particular cloud area. We apply assessment examination to decide the slant of SaaS audits also, we create the preparation dataset. Utilizing this dataset, we prepare the SVM calculation to construct an assumption characterization demonstrate.

III. PROPOSED SYSTEM

Our essential research objective is to condense buyers encounter of utilizing SaaS administrations. To do this, we apply supposition examination and grouping utilizing NLP and ML on the Blue Pages Reviews dataset (BPR) that we created by slithering SaaS customers audits from various web-based interfaces. This dataset is accessible online through www.bluepagesdataset.com. A point by point portrayal is given in the accompanying area.

A. Blue Pages Reviews Dataset We arranged the Blue Pages dataset to increase facilitate knowledge into the SaaS administrations offered over the WWW. Utilizing a web

scrubber, we gathered around 6000 SaaS offered subtle elements (i.e. benefit name, benefit cost, etc.)from diverse online interfaces for example, Cloud Reviews and Get App. In light of the gathered information, we created two datasets: a Service dataset and a Surveys dataset. The Service dataset contains data on SaaS, for example, benefit name, specialist organization, supplier URL and administration classification, while the Reviews dataset contains shopper audits alongside the commentators' data, such as the analyst's name, commentator's remarks, the rating of the analysts, and so on.

In our examinations, we utilized the information gathering containing 4000 audits. Conclusion investigation is utilized to decide the analyst's state of mind and whether the audits' conclusion grouping was certain, unbiased or negative. The slant of the audits is resolved utilizing a rating framework, for instance, audits with a rating of five or four can be considered to express a positive notion, while a survey with a rating of one and two can be considered to express a negative estimation, else it can be thought to be nonpartisan. In our examination, we apply estimation investigation based NLP procedures which gives more precise outcomes in connection to buyers' suppositions. Besides, we prepare a SVM arrangement calculation to fabricate a forecast model to arrange SaaS audits. The subtle elements of the look into technique are given in the following area.

IV. EXPERIMENTAL RESULTS

The point of this work is to look at SaaS customer's surveys to recognize whether they are negative, positive, or unbiased in estimation and afterward to prepare a ML classifier to construct a arrangement model to assess SaaS surveys. To do this, we make utilization of both NLP and ML by utilizing the Rapid Miner condition and a SaaS application called Semantria, created by Lexalytics. Quick Miner is an advancement stage for information mining, machine learning, and business examination apparatuses furthermore, is utilized as a part of both scholarly community and industry [27]. It involves diverse mining instruments and procedures including streamlining, 55 perception, and approval. Semantic is a SaaS application for content mining investigation that gives an assessment score to each survey to decide if the analyst's conclusions are negative, positive, or unbiased [24]. To execute this, we play out the three assignments appeared in Figure 1

A. Assignment 1: Pre-processing Reviews

The surveys accessible in the Blue Pages Reviews dataset were gathered from open online interfaces, which have been composed by open, so they contain some commotion. The clamour comes as unstructured content, for example, HTML labels, accentuation, erroneous spelling, non-standard words off base capitalization and so forth. This pre-processing stage is a pivotal stride in acquiring a slant of audits to accomplish most extreme exactness while applying the investigation. We make utilization of the Text Handling Plug-in Rapid Miner to play out this errand, and we utilize one of its apparatuses called "Process Document from Data". This instrument is a settled administrator that includes the accompanying errands:

1) Extract content administrator: to remove structure content from HTML code

2) Tokenization administrator: to transform the archives into a rundown of words and to separate dashes and spaces.

3) Transform Cases administrators: to swing everything to lower cases

4) Filter stop word: to expel words, for example, articles, digits, what's more, recommendations, and so on.

5) Stem(Porter): to apply the English Porter stemming calculation on the surveys to lessen each word to their root.

In the wake of pre-processing the audits and expelling the clamor, the following errand is to decide the extremity of each audit utilizing Semantria for Excel.

eviews	polarity_confidence	polarity
y company has b	0.24248223	positive
Sense has consist	0.457113117	positive
ındles huge DB c	0.441015244	positive
e spent about 6	0.413340598	positive
Sense helps us v	0.94254607	positive
ho Reports succe	0.267500013	positive
ıe integration wi	-0.227008581	negative
ho Reports succe	0.212381914	neutral
e compared a lot	-0.102500021	negative
ter setting up Da	0.080556497	neutral
ır use case is cer	0.167504683	neutral
ıu can have a gre	0	neutral
gicMonitor as a	0.319127381	positive
ıth LogicMonitor	-0.245000005	negative
gicMonitor is a g	0.582140028	positive
you really want	0.282369673	positive
vanted this app	0.452391326	positive
ention is a tool th	0.339855999	positive
ı are a niche adv	0.895999968	positive

Figure 2. Analysis of SaaS reviews

Table I
SUMMARY OF SENTIMENT ANALYSIS

Polarity of Reviews	Number of Reviewers	Polarity Threshold
Positive	2487	greater than 0.22
Neutral	1312	between 0.21 and 0.04
Negative	201	less that 0.05

B. Errand 2: Sentiment Analysis

The point of this errand is to quantify the enthusiastic tone of each SaaS post/survey in our dataset. We play out this undertaking utilizing Semantria conclusion examination that gives an assessment scoring for each survey. This examination can be connected at the record level or element level. In our situation, we apply the examination at the record level, as each survey is an archive. The feeling phrases are recognized in each record and are then scored to recognize positive, negative, or nonpartisan tones. At long last, the scores of the expressions are consolidated to acquire the general score for the record. The consequence of this undertaking is a SaaS extremity dataset that is exceptionally valuable for preparing regulated machine learning calculations.

C. Assignment 3:

Linear Support Vector Machines grouping (SVM) In this assignment, we utilized a directed machine student in view of preparing dataset In this assignment, we utilize a managed machine learning calculation in view of a preparation dataset that we made beforehand in assignment 2. Our preparation dataset contains audit messages alongside assessment scores (negative, positive, and impartial). For the arrangement procedure, we make utilize of audits with negative and positive polarities anyway we bar those with an unbiased extremity. To play out this assignment, we make utilization of the Rapid Miner apparatus called Classification and Relapse, and we apply one managed machine learning calculation called Linear help vector machines. Vapnik [28] to begin with built up the Linear help vector machines (SVMs) for twofold characterization. It is a straight order which implies it has two measurements that can be isolated by a line. SVMs are connected effectively for content grouping which more often than not gives adequate outcomes.

In this examination, we make utilization of SVMs with three diverse ways to deal with word vectors, in particular Binary Occurrence, Term Frequency and Term Frequency-opposite report recurrence (TF-IDF). The Binary Occurrence approach employments the paired an incentive to decide whether the term is in the report, where one means exhibit in the report, generally zero. The Term Frequency approach figures how frequently the term happens in the record. The TF-IDF approach is the term recurrence computed by word recurrence in the archive and the whole corpus. To prepare and test the classifier, we connected three-crease cross-approval, which included utilizing 1/3 of the preparation dataset for preparing purposes, so this procedure kept running with an alternate overlay as the preparation set.

V. RESULT

We led feeling examination utilizing the SaaS application Semantria as per the trial system in Figure 1.

Table II
TERM OCCURRENCES

	Actual negative	Actual negative	class precision
predict negative	6	30	16.67%
predict positive	195	2457	92.65%
class recall	2.99%	98.79%	

Table III
TFIDF

	Actual negative	Actual negative	class precision
predict negative	7	20	25.93%
predict positive	194	2467	92.71%
class recall	3.48%	99.2%	

Table IV
BINARY TERM OCCURRENCES

	Actual negative	Actual negative	class precision
predict negative	3	9	25.00%
predict positive	198	2478	92.60%
class recall	1.49%	99.64%	

Table V
COMPRESSION OF 3 WORD VECTOR APPROACHES WITH 3 FOLDS

Approach	Accuracy	classification error
TFIDF	91.63% +/- 0.24%	8.37% +/- 0.24%
Binary Term Occurrences	92.30% +/- 0.18%	7.70% +/- 0.18%
Term Occurrences	92.04% +/- 0.21%	7.96% +/- 0.21%

An outline of the assessment tests is given in Figure 2, incorporating content audits with extremity esteems and extremity marks (negative, positive, or impartial). The semantic devices score the audits and feature each gathering with a particular shading, for example, dim green means to a great degree positive and pink is negative. The outcomes demonstrate that there is a particular range edge for every extremity classification as takes after: negative extremity < 0.05, positive extremity > 0.22, also, impartial range between (0.21 - 0.04). The consequences of this examination exhibit that SaaS commentators will probably share a positive involvement in connection to a SaaS application (62%). Additionally, a critical number of analysts gave nonpartisan input about their experience utilizing SaaS (33%), while few commentators posted negative remarks (5%).

Furthermore, to direct managed machine learning, we built up a SaaS extremity dataset (produced in Task 1) utilizing the Rapid Miner stage. We directed a 3-overlap cross approval in this test, which implies the information is separated into three, one being the trying set and the other two being the preparing sets. We utilized three diverse ways to deal with create the word vectors. To assess the supposition grouping of our models, we utilized the regular record for content order counting precision and grouping blunder for each approach. Table II, Table III, and Table IV delineate the execution assessment. Class review and class exactness allude to the review proportion and exactness proportion for audits and the figuring for both negative and positive surveys. Table V looks at the SVM display based ways to deal with producing word vectors. The exactness alludes to the general correctnesses of various approaches, the outcomes showing that Binary Occurrence, Term Occurrence give practically a similar outcome, and TF-IDF is the most noticeably bad alternative of these. Parallel Occurrence gave the most elevated precision (92.30%).

VI. CONCLUSION

In this paper we have exhibited another approach called Learning Analytics and Predictive investigation to recognize scholastically at-hazard understudies and to anticipate understudies learning results in instructive organizations. The prescient models will help the educator to see how well or how ineffectively the understudies in his/her class will perform, and thus the teacher can pick appropriate educational and instructional intercessions to upgrade understudy learning results. It likewise encourages teachers to foresee about understudies achievement and disappointment in examination and furthermore they can give appropriate advices to avoid disappointment in the examinations. Data resembles Attendance, Class test, Seminar and Assignment marks were gathered from the understudy's past database, to forecast execution toward the finish of the semester. This investigation will help to the understudies and the instructors to enhance the division of the understudy. This investigation will likewise work to recognize those understudies which required exceptional thoughtfulness regarding diminish come up short apportion and making proper move for the following semester examination.

REFERENCES

[1] Q. Ye, R. Law, and B. Gu, "The impact of online user reviews on hotel room sales of product and consumer characteristics," Journal of marketing, vol. 74, no. 2, pp. 133–148, 2010.

[2] F. Zhu and X. Zhang, "Impact of online consumer reviews on sales: The moderating role," International Journal of Hospitality Management, vol. 28, no. 1, pp. 180–182, 2009.

[3] Q. Roller, "Quote roller reviews, pricing comparison of alternatives," GetApp, vol. 02-, 2015

[4] S. Model, "Software as a service (saas) from the gartner IT glossary," Gartner IT Glossary, vol. 04-, 2012. [Online].," International Journal of Computer Applications, vol. 56, no. 13, pp. 1–6, 2012.

[5] N. Mishra and C. Jha, "Classification of opinion mining techniques Available: http://www.gartner.com/it-glossary/software-as-a-service-saas/

[6] D. Song, H. Lin, and Z. Yang, "Opinion mining in e-learning system," in Network and Parallel Computing Foundations and trends in information retrieval, vol. 2, no. 1-2, pp. 1–135, 2008.

[7] B. Pang and L. Lee, "Opinion mining and sentiment analysis, Workshops, 2007. NPC Workshops. IFIP International Conference on. IEEE, 2007, pp. 788–792.

[8] S. F. Sayeedunnissa, A. R. Hussain, and M. A. Hameed, "Supervised opinion mining of social network data using a bag-of-words approach on the cloud," in Proceedings of Seventh International Conference on Bio-Inspired Computing an aspect-based opinion mining approach," Procedia Computer Science, vol. 22, pp. 182–191, 2013.

[9] E. Marrese-Taylor, J. D. Vel´asquez, F. Bravo-Marquez, and Y. Matsuo, "Identifying customer preferences about tourism products using: Theories and Applications (BIC-TA 2012). Springer, 2013, pp. 299–309.

[10] M. Cooke and N. Buckley, "Web 2.0, social networks and the future of market research," Journal of Applied Sciences, vol. 14, no. 14, p. 1653, 2014.

[11] N. Bharathi, G. Brindha, B. Santhi, and P. Neelamegam, "Integration of fpga, cloud and opinion mining," International Journal of Market Research, vol. 50, no. 2, p. 267, 2008.