JOURNAL OF CREATIVE WRITING
VOLUME 4, ISSUE 2
2020, PP 60 – 80
ISSN 2410-6259

Automatic Method to Rate Websites Based on Terms of Services

Mehraj U'Din Bhat[1], Dr. Satwinder Singh[2]

ABSTRACT

Terms of services agreement of a website, though neglected by most of the users, plays a major role in deciding whether the website policies are designed by considering the developer /owner and users rights and needs. This paper focuses on developing an automatic tool that ranks websites based on their terms of services agreements using concept of natural language processing. This is the first such attempt in the field. The developed tool uses bag of words text classification approach and 2-layered artificial neural network. The method works in two phases: first phase consists of training and machine learning. It classifies terms into good, bad and neutral classes. This cond phase defines rating scale and ranks websites into classes A, B, C, D and E. Around 65.5% accuracy observed during testing phase opens the doors for research in developing such tools.

Keywords: artificial neural network, bad terms, bag of words, machine learning, natural language processing, terms of services, website ranking.

INTRODUCTION

Today the Internet boasts of millions of websites luring users with their schemes and policies, which may be both real and fake. Often user send up caught in the spider web of these schemes and policies, the most probable reason for this is negligence of "Terms of Services (TOS)" provided by every website. TOS, also referred as "Terms of Use", "Terms and Conditions," are set of rules that user must agree too by to use a service provided by a web site or search engine. TOS is mostly used for legal purposes by web portals and Internet service providers which store user's personal data such as user ids and passwords for different social networking sites, email accounts, online transaction of money etc. It is necessary to establish proper arrangement between two parties involved in any deal. ATOS agreement acts as a legal bond between website owner and user. When a user clicks on "I Agree" button provided below the terms of services document of a web portal, it accepts all the terms and conditions of that web portal, just similar to

[1] M.Tech Central University of Punjab Bathinda, bhatmehraj71@gmail.com
[2] Head of Department, Department of computer science and technology , Central University of Punjab Bathinda, satwinder.singh@cup.edu.in

signing of a deal between two business parties. Usually users skip the terms of services document and randomly agree to the service, which is wrong. As a responsible user, every user should go through each detail of TOS document, so that it knows about all the rules and regulations, its rights and uties. Merely just clicking the "I Agree" button does not legally bind user and website owner into a legal contract.

In this paper, we have developed an automatic tool that ranks the websites based on terms of terms of services.

Objectives

I. To define labeling scheme for various keywords used in terms of services.
II. To create a Meta dictionary of keyword with labels.
III. To define rating scale for terms of services.
IV. To rate the websites on the basis of their terms of services

RELATEDWOR

We reviewed the literature in three parts. In first part, we evaluated existing text categorization techniques so we could choose the best labeling scheme to classify the terms. In second part, we evaluated the impact of TOS on users and in third part; we observed how TOS is important for a website.

Text Categorization Techniques

Many text categorization techniques introduced and discussed over past few decades. A different adaptive experimental design (MAED) for text categorization suggested (Deng Cai, Member, IEEE and Xiaofei He, Senior Member, IEEE, 2010).

Unlike previous active learning approaches that inspect either Euclidean or data-independent non-linear structure of the data space, MAED surveys fundamental manifold structure. An efficient approach FSKNN recommended (Jung-Yi Jiang, Shian-Chi Tsai and Shie- Jue Lee, 2011). This method uses fuzzy similarity measure (FSM) and kNN, for multi-label text categorization. A new feature selection algorithm, called CMFS proposed by (Jieming Yang, Yuanning Liu, XiaodongZhu, ZhenLiu and Xiaoxu Zhang, 2012).

This method examines the importance of a term for within the category and for between two or more categories also. Comparison of feature selection is based on term and document frequency (Nouman Azam and Jing Tao, 2011). Comparative frequencies of both frequencies considered. A two-level term selection and term extraction to escalate the performance of text categorization introduced by (Harun Uguz, 2011), in which less useful terms are ignored, and term selection and term extraction methods applied regarding high value.

A refined K-Nearest Neighbor (kNN) algorithm to indemnify the less capable traditional kNN algorithm suggested (Shengyi Jiang, Guansong Pang, Meiling Wu and Limin Kuang, 2012). In the proposed method, the classification models built by combining single pass clustering algorithm and kNN text categorization. Emphasis laid on reducing the notation cost of verified text classification (Mossaab Bagdouri, William Webber, David D. Lewis and Douglas W. Oard 2013). A framework presented for joint reduction of training and test notation that retains the analytical credibility of power estimates, and produces a real interpretation of a superlative ration of notations to education and experimental data. A forerunner based classifier introduced for text categorization, in which set of standards used to embody a document category (Jianfei, LifeiChen, GondeGuo, 2013). A unique informative technique that exploits staged learning-based resource allocation Network introduced by Wei Song, Peng Chan and Soon Cheol Park (2014).

SLRAN is made of two phases; preliminary learning phase and refined learning phase, and it has many advantages like it developed the compressed structure, due to which, the estimation involvement of network reduces and learning ability hikes. The most popular available method of text categorization that is "Bag-Of-Words" (BOW) discussed by (Roberto H.W. Pin heiro, George D.C. Cavalcanti, TsangIng Ren, 2014). BOW expresses each document as factor vector where each vector position accounts for a word. Two refining methods used for factor selection in text categorization are MFD (maximum F functions per document) and MFDR (maximum F functions per document-reduced). The aim of factor selection is to reduce the size of factor vector without damaging the conduct of categorization. A literal access for text categorization of medical documents introduced (Rajni Jindal and Shweta Taneja, 2015). Authors used lexical nearest neighbor K (LKNN) algorithm in which tokens used to personify the documents. Term class purpose to gauging the importance of a term in classifying the document into a class has been initiated (DS Guru and Mahamad Sahil, 2015). The recommended measure used to weigh the weight of importance of a given term, as a product of class term weight and class term density. A new feature selection method using particle swarm optimization in text categorization suggested by (Mehdi Hossein zadeh Aghdam and Set are h Heidari, 2015). A compound feature selection approach based on Genetic Algorithm (GA) suggested (Abdullah Saeed Ghareb, Azuraliza Abu Bakar, Razak Hamdan, 2015). This method based on compound search technique that utilizes the unification of advantages of filter selection methods and Enhanced GA (EGA) in a wrapper approach to handle the soaring magnitude of the feature space and concurrently enhances categorization performances. An experiment described to test that impetus-feedback association or a general rule robotized during considerable proceeding at intuitive categorization (Jessica L. Roeder, F. Gregory Ashby, 2016). In this experiment, 27 users participated, each of them performed 12,300 trials either of intuitive categorization, which were rule-based (RB) categories or in which procedural learning required. R F Boost;

AdaBoost.MH boosting algorithm's improved version suggested (Bassam Al Salemi, Shahrut Azman Mohd Noah, Mohd Juzaiddin, 2016). In R F Boost, the low information based on elevating a limited fixed number of rated in each boosting round, instead of using all features like Ada Boost. MH. Text normalization and semantic indexing to enhance instant messaging and SMS spam filtering suggested (Tiago A. Almeida, Tiago P. Silva, IgorSantos, Jose M. Gomez Hiddlgo, 2016. An ensemble multi-label text categorization based on rotation forest and semantic indexing suggested by (Haytham Elghazel, Alex Aussem, Ouadie Gharroudi, Wafa Saadaoui, 2016). A method based on cooperation on an integrated frame work suggested improving text summarization and classification (Hyoungil Jeong, Young joong Ko, Jungyun Seo, 2016). In this mode, factor-measuring method for text categorization uses a language model that associates factor divisions in each section and text, and one for text classification, which gives sentence importance scores, predicted from text categorization. The focus laid on a register-specific meaning categorization of linking adverbials in English (Zihan Yin, 2016). A new classification system based on both phonological and realistic approaches suggested. A new method for categorizing tweets by blending content and basic learning proposed (J.M Cotelo, F.L. Cruz, F. Enriquez, J.A. Troyano, 2016). In this approach, to integrate two key aspects of a tweet, proper textual content and its underlying structural information explored.

Impact of TOS on Users

This section discusses how the user gets affected by terms of service. It also addresses user's rights and responsibilities. Spotlight reposed on presence in transaction by consumers and suppliers- their contributions at various stages of the value creation process (Guilherme D. Pires, Alison Dean, and Muqqadas Rehman, 2014). User satisfaction of China's air line market investigated (Hongwei Jiang, and Yahua Zhang, 2016). A fuzzy multiple criteria decision-making (MCDM) model has been developed to assess the customer value for three specific service providers in Taiwan based on shipper's perspective (Ji-Feng Ding, Wen-Hwa Shya, Chun-Tsen Yeh, PI-Hui Ting, Chung-Te Ting, Chien-Pang Lin, Chien-Chang Chou, and Su-Sin Wu, 2016). An exploratory study on online customer service and emotional labored by (KumiIshii, and KrisM. Markman, 2016). Focus laid on utilizing customer satisfaction in ranking prediction for personalized cloud service selection (Shuei Ding, Zeyuan Wang, Desheng Wu, and David L. Olson, 2016). Research done to evaluate customer social participation in the social networking services and its impact on the client's equity of global fashion brands (Heeju Chae, and Eunjo Ko, 2016). A study done to examine the relationship between online victimization and user's activity views on personal information security on social networking sites (George Saridakis, Benson Vladlena, Ezingeard Jean-Noel, and Tennakoon Hemamali,2016).

Impact of TOS on Websites

This section lays focus on how terms of service agreement act as a catalyst in ensuring the safety of websites. Network dependent text analysis examined to evaluate trends in Microsoft's security innovations (Tabitha L. James, Lara Khansa, Deborah F. Cook, Olga Bruyaka, and Kellie B. Keeling, 2013). Research has been done to check how safety and privacy concerns affect the educational use of cloud service (Ibrahim Arpaci, KeremKiliur, and Salih Bardakci, 2015. The view of the foreign tour got examined regarding E-commerce services that bring satisfaction, belief and loyalty among customers (MutiaSobihah, Mahadzirah Mohamad, Nor Azman Mat Ali, Wan Zulqurnain Wan Ismail, 2015). An experiment has been done to check how users read privacy policies online (Nili Steinfeld, 2015). A report has been written to observe the impact of government websites (Flavio Perazzo Barbosa Mota, Carlo Gabriel Porto Bellini, Juliana Morais da Silva Souza, Terezinha De Jesus Nogueira Oliveira, 2016). Implications of Payment Services Directive II (PSD II) regarding payments in the digital market strategy of European Commission evaluated (Mary Donnelly, 2016).

METHODOLOGY

We executed the work in six phases. In first phase, we examined and compared popular text categorization methods. We focused on choosing the best text categorization method based on evaluation metrics. It involved intense research of various text categorization methods, going through literature surveys of these methods, and selecting the apt one that easily gels with the NLP toolkit and python. In second phase, we analyzed TOS of social networking sites and used in India. It included creating comparison frameworks of different social networking sites based on user account and user content clauses of terms of service agreement. Then, from comparison frameworks, we built filtration tables consisting of good, bad and neutral terms per user's perspective. In third phase, we designed Meta dictionaries of TOS to store different keywords and phrases along with their labels (good, bad and neutral). We prepared them using filtration tables created in second phase. We used them as training data set. In fourth phase, we defined a rating scale that would used to assign the class to website based on keywords found in terms of services document. In fifth phase, we developed a tool that rates websites according to their terms of services automatically. In sixth phase, we examined the precision, accuracy and error rate of developed tool.

Working of the Tool

The tool works in 2 major phases. First phase, built up of training phase and machine learning phase, emphasizes on intra TOS document categorization in to good, bad and neutral services.

Second phase counts good, bad and neutral terms obtained in first phase and categorizes website on the basis of percentage of bad terms in its TOS document**.**

First Phase

In this phase, intra classification of a TOS document is done. It consists of training phase and machine learning phase. In training phase, bag of words text classification approach has been used to create the training data set. Two libraries; NLTK (Natural Language Toolkit) is used for stemming and tokenization, and JSON (JavaScript Object Notation) is used to store synaptic weights calculated during the machine learning phase. Three classes of training data defined-good, bad and neutral. First data of training set is arranged into proper structures, each sentence is being looped, and words in every sentence is tokenized, followed by addition of words in word list, further if belongs to, in any of three classes of training set, and stemming and removal words is executed, then a pattern is considered, that also initially considered as zero. Then each word in pattern undergoes stemming process. Output weighed one for current tag and zero for other tags. If W exists in word, pattern is appended by one, else by zero. This leads to creation of training data, on which bag of words technique implemented. Initially, the bag taken; each training sentence is converted into array of zeros and ones against array of unique words stored in corpus. Finally, output obtained in the form of single class, multi class or none.

In machine learning phase, we considered two-layered artificial neural network (ANN). Numpy and sigmoid functions of ANN used. Synaptic weights generated respective to zeros and ones array obtained against array of distinctive words in corpus.10 neurons taken, value of alpha (gradient descent parameter) taken 0.1. Synaptic weights stored in j son file and tool constructed. Classification function defined and terms in TOS documents of websites categorized into good, bad and neutral.

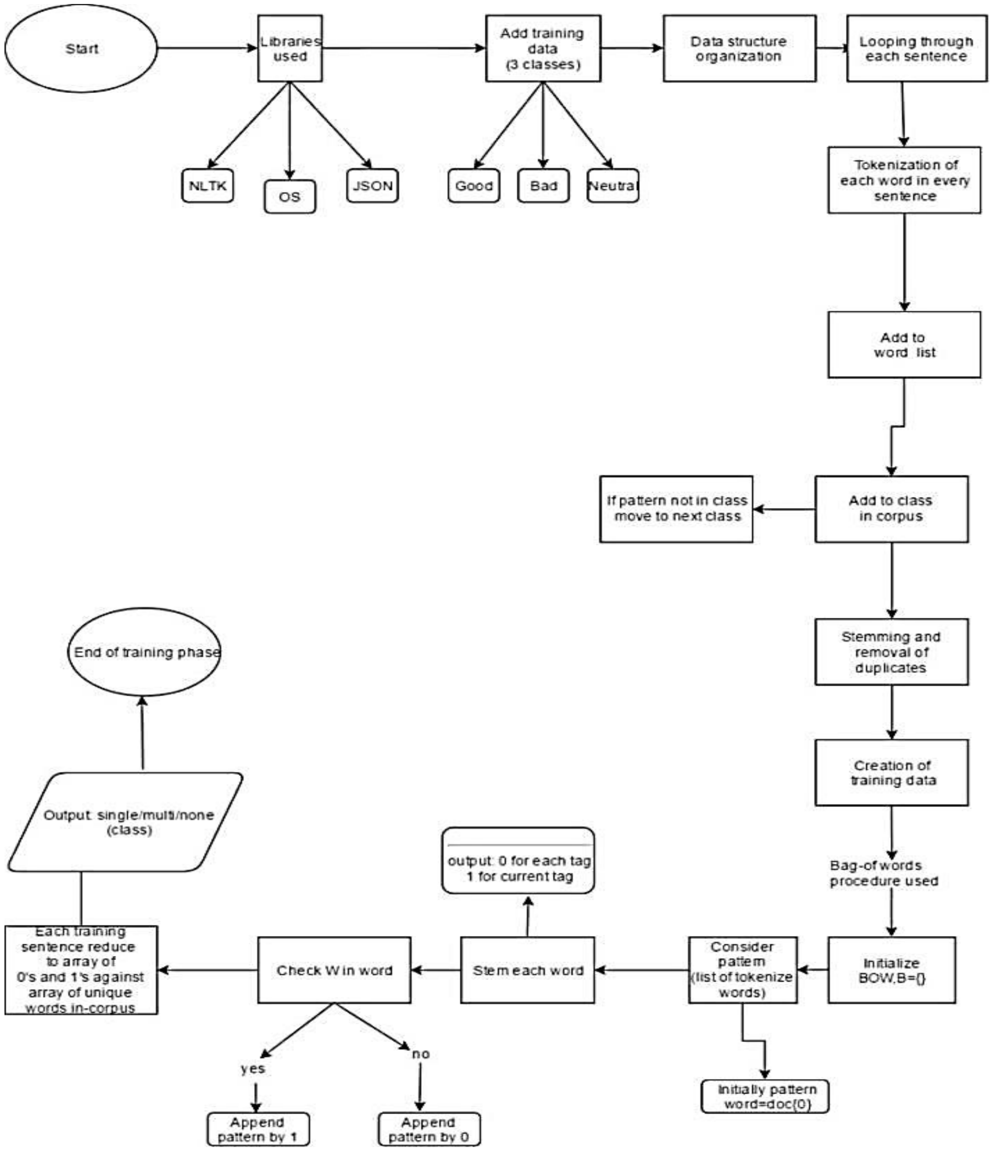

Fig. 1. Working of Training Phase

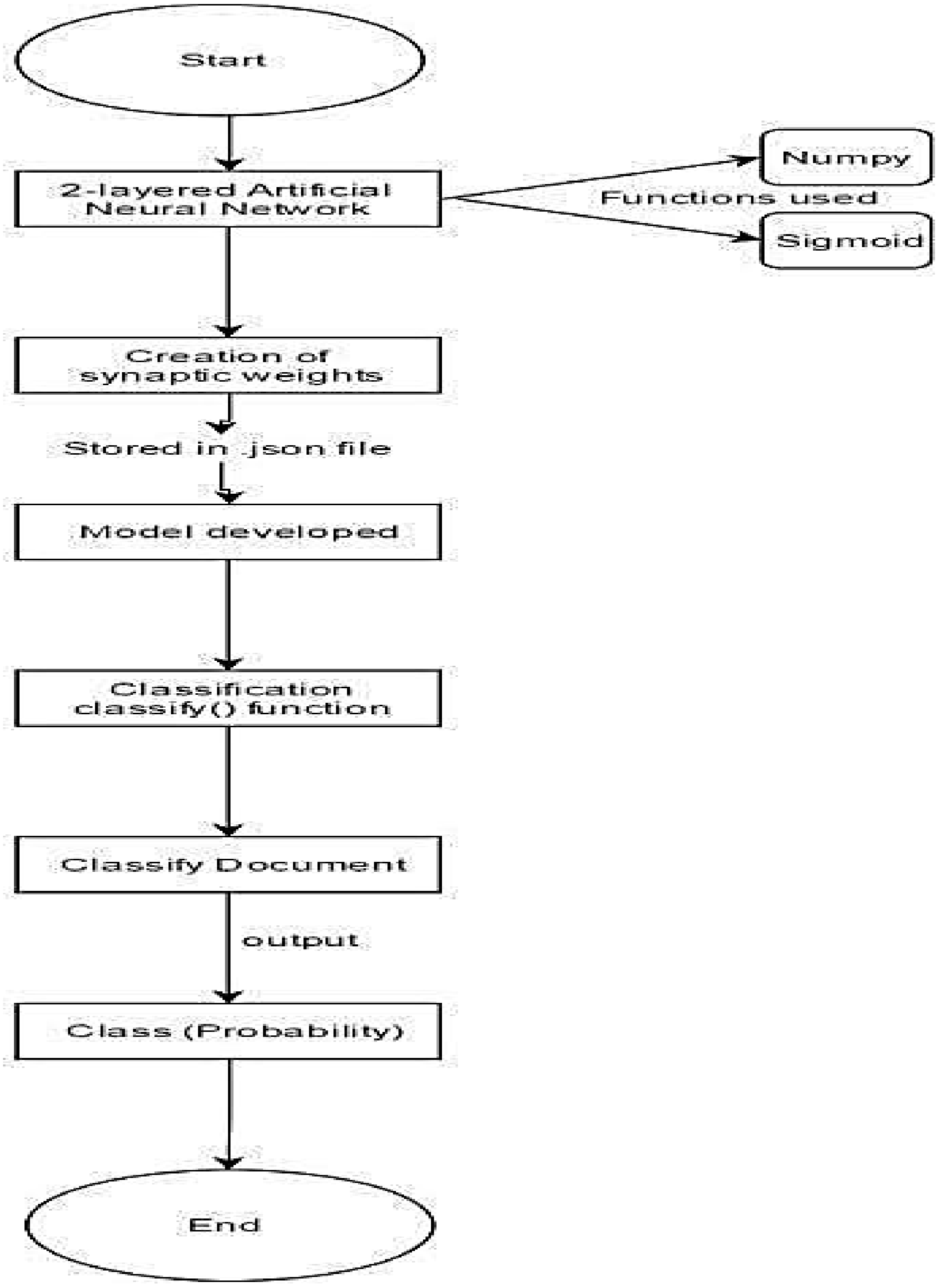

Fig. 2. Working of Machine Learning Phase

Secondary Phase

In second phase, inter TOS document categorization of websites has been done. The number of good, bad and neutral terms attained for each website during first phase is counted, and respective percentages of good, bad and neutral terms has been calculated.

I. Good term percentage = Number of good terms * 100 / Number of good terms +number of bad terms + number of neutral terms

II. Bad term percentage = Number of bad terms*100/

Number of good terms + number of bad terms + number of neutral terms

III. Neutral term percentage = Number of neutral terms*100

/ Number of good terms + number of bad terms + number of neutral terms

Based on bad term percentage, rating scale is defined. The website with lowest bad term percentage is assigned the best class A, whereas with highest bad term percentage is allotted the worst class E.

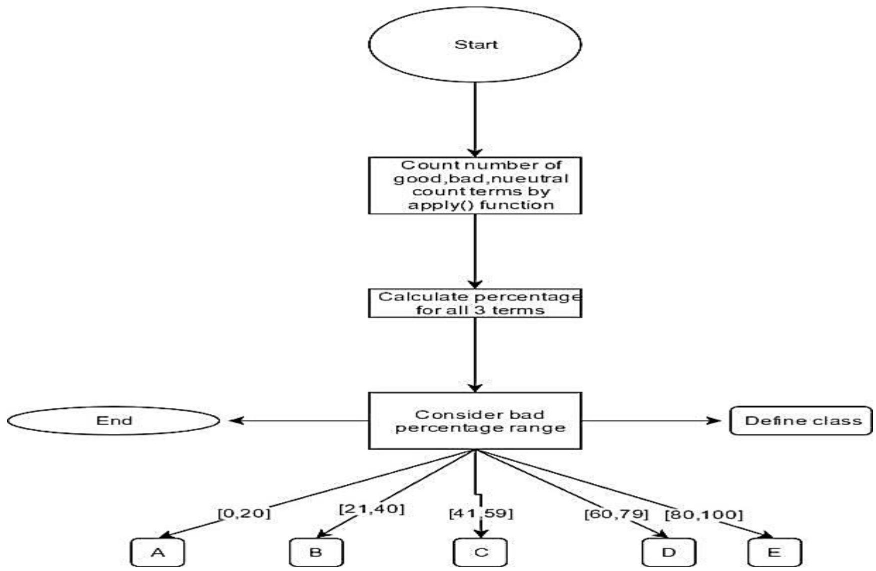

Fig. 3. Working of Second Phase

Table 1. Defined Rating Scale

BAD TERM PERCENTAGE RANGE	CLASS	REMARKS
0 -20	A	Very Good
21-40	B	Good
41-59	C	Average
60-79	D	Bad
80-100	E	Needs Attention!

RESULTS

Defined Labeling Scheme

We emphasized on both intra document classification; that is examining good, bad and neutral terms within a TOS document of website, and inters document classification; that is calculating the major class for each web portal based on their legal term agreements, and intra document categorization of web portal(s) into good, bad and neutral services make use of combination of bag-of- words text classification technique and artificial neural network. Bag-of-words is used as text categorization method and artificial neural network (ANN) is used in machine learning phase of the tool. The inter document classification is done based on percentage of bad term in the website; least percentage leads to highest class.

Meta Dictionaries

Total six Meta dictionaries, three each consisting of good, bad and neutral terms keywords and phrases from both user account and user content clauses from 10 social networking sites (SNS) namely Facebook, Twitter, Whatsapp, Instagram, LinkedIn, Tumblr, Google Plus, BigoLive, Tagged, Like and Snapchat.

User Account

Table 2. User Account Good Terms Meta Dictionary

User can deactivate account anytime
User control communication
Service announcements
Administrative messages
No content/data accessible after account deactivation
Recover account using email

User right to enter term of jurisdiction
Website not to be used for commercial purpose
Application seeks permission of user before using its account
User controls settings of its account
Personal information used for limited purposes
User manage emails it receives within account settings
Part time jobs for users
Feedback System

Table 3. User Account Bad Terms Meta Dictionary

User needs to create account
User name and password required to use services
User responsible for safeguarding account
No part of account should be transferred.
SNS has the right, without compensation to user or others, to serve ads near its content
Userresponsibleforallactivityonitsaccountunlessitclosesitor reports misuse
Associated email address must be updated
Age limit on website usage
Strong secure password
No sharing of password
Only one personal account
Website usage in compliance with law
No use of other person username
No violation of other people's rights
No use of services for unlawful purpose
No abusing, harassing, threatening, impersonating or intimidating other users
Other account creation prohibited on termination of existing account
License terminated on unauthorized use

Table 4. User Account Neutral Terms Meta Dictionary

User can deactivate account anytime
User control communication
Service announcements

User Content

Table 5. User Content Good Terms Meta Dictionary

User owns all content
User controls its content
Application asks for user's permission
User retains rights of intellectual property it posts
User can choose who can view its content and activities
SNSwillnotincludeuser'scontentinadvertisementswithoutuser'sseparate consent
User has legal right to enter terms of use in jurisdiction
SNS grants user a worldwide, revocable, non-exclusive, non-sub licensable, and non-transferable license to download, store, view, display, perform, redistribute, and create derivative works of Content
User can report inappropriate content or any scam it comes across
User can change its browser settings
Part time optional work for users
Brings together people from diverse backgrounds on same platform
User need not pay any money for part time work
SNS has no right to use content after user removes it
User can use invite friends feature
SNS uses industry standards to prevent misuse and improper access of user data
User can update/correct information
User can cancel its account
SNS contacts user through email
SNS doesn't claim ownership of user content
User can end license for specific content

Table 6. User Content Bad Terms Meta Dictionary

User owns all content
User controls its content
Application asks for user's permission
User retains rights of intellectual property it posts
User can choose who can view its content and activities
SNSwillnotincludeuser'scontentinadvertisementswithoutuser'sseparate consent
User has legal right to enter terms of use in jurisdiction
SNS grants user a worldwide, revocable, non-exclusive, non-sub licensable, and non-transferable license to download, store, view, display, perform, redistribute, and create derivative works of Content
User can report inappropriate content or any scam it comes across
User can change its browser settings
Part time optional work for users
Brings together people from diverse backgrounds on same platform
User need not pay any money for part time work
SNS has no right to use content after user removes it
User can use invite friends feature
SNS uses industry standards to prevent misuse and improper accessof user data
User can update/correct information
User can cancel its account
SNS contacts user through email
SNS doesn't claim ownership of user content
User can end license for specific content

Table 7. User Content Neutral Terms Meta Dictionary

Back up content regularly
No violent and pornographic content be posted on web page
No sending of spam messages and emails
Website has right on all other content except user content
Usage of user information by website to improve services
No use of robot spider scrap or unautomated access to web portal
Website reserves right but no obligation to monitor disputes

Website not to be used for commercial purpose
Promotion of websites by company
Part time work discontinued if work turns out to be fake.

Ratings of the Websites by the Developed Tool

We calculated the ratings for randomly selected 50 websites. Maximum number of websites belonged to class B, followed by class A and C. Only one website belonged to class D none to class E.

Table 8. Tool Ratings of 50 Websites

S. No	*Website*	*Class*	*No. Of Good Terms*	*No. Of Bad Terms*	*No. Of Neutral Terms*
1.	*500px*	*A*	*30*	*13*	*27*
2.	*Act Corp*	*A*	*2*	*1*	*5*
3.	*Adobe*	*B*	*21*	*17*	*14*
4.	*Amazon*	*C*	*4*	*9*	*7*
5.	*Archive*	*A*	*22*	*7*	*12*
6.	*Ask Fm*	*B*	*28*	*15*	*28*
7.	*Bill desk*	*B*	*8*	*7*	*17*
8.	*Bitly*	*B*	*8*	*12*	*13*
9.	*Bloomberg*	*B*	*7*	*11*	*22*
10.	*Couch Surfing*	*B*	*23*	*13*	*16*
11.	*Coursera*	*B*	*4*	*3*	*4*
12.	*Dailymotion*	*C*	*9*	*14*	*5*
13.	*Dainik Bhaskar*	*B*	*22*	*17*	*17*
14.	*Delicious*	*D*	*2*	*4*	*0*
15.	*Disney*	*A*	*8*	*4*	*8*
16.	*Divya Bhaskar*	*B*	*18*	*18*	*11*
17.	*Dropbox*	*B*	*7*	*6*	*10*
18.	*Ebuddy*	*A*	*16*	*7*	*16*
19.	*Entrance Corner*	*B*	*2*	*7*	*15*
20.	*Evernote*	*A*	*21*	*12*	*63*

21.	*Expedia*	*B*	*4*	*8*	*22*
22.	*Fare Compare*	*A*	*3*	*2*	*12*
23.	*Filmi Beat*	*B*	*2*	*3*	*8*
24.	*First Cry*	*A*	*6*	*4*	*42*
25.	*First Post*	*B*	*0*	*3*	*7*
26.	*Flipkart*	*B*	*13*	*14*	*25*
27.	*Gaana*	*B*	*27*	*22*	*28*
28.	*Google*	*B*	*9*	*9*	*6*
29.	*Gravatar*	*A*	*1*	*1*	*4*
30.	*Hotstar*	*B*	*13*	*8*	*16*
31.	*Imdb*	*C*	*4*	*6*	*3*
32.	*Indiabix*	*C*	*8*	*10*	*4*
33.	*Indigo*	*C*	*7*	*9*	*5*
34.	*Jabong*	*B*	*8*	*7*	*9*
35.	*Lastpass*	*A*	*23*	*10*	*17*
36.	*Learn Python*	*B*	*7*	*6*	*9*
37.	*Medium*	*C*	*1*	*7*	*5*
38.	*Microsoft*	*B*	*34*	*27*	*38*
39.	*Mouthshut*	*C*	*9*	*13*	*7*
40.	*Nabble*	*C*	*2*	*4*	*2*
41.	*Nytimes*	*B*	*8*	*9*	*15*
42.	*Rajnikant Vs CID jokes*	*B*	*5*	*3*	*3*
43.	*Saavn*	*A*	*9*	*2*	*8*
44.	*Tvf Play*	*B*	*19*	*17*	*24*
45.	*Whatsapp*	*A*	*9*	*5*	*15*
46.	*Wikipedia*	*B*	*27*	*20*	*23*
47.	*Wordpress*	*C*	*13*	*30*	*19*

48.	*Yahoo*	*B*	*15*	*13*	*17*
49.	*Youtube*	*B*	*6*	*9*	*10*
50.	*Zomato*	*A*	*19*	*9*	*27*

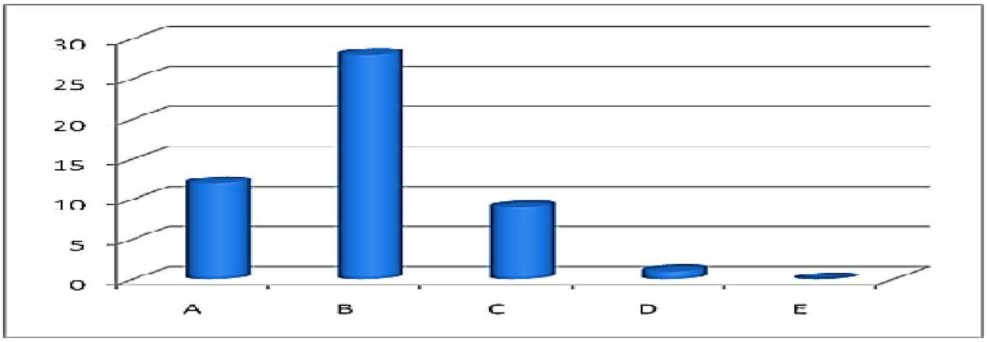

Fig. 4. Graph showing number of websites belonging to each class

Evaluation of Developed Tool
To check the accuracy, precision and error rate of the developed tool, we manually analyzed and assigned classes to the 50 websites, and compared it with the tool ratings.

Table 9. Tool Defined Ratings Vs Manually Defined Ratings

S.No	*Website*	*Class Defined By Tool*	*Manually Defined Class*
1	*500px*	*A*	*A*
2	*ActCorp*	*A*	*A*
3	*Adobe*	*B*	*C*
4	*Amazon*	*C*	*C*
5	*Archive*	*A*	*A*
6	*AskFm*	*B*	*B*
7	*Billdesk*	*B*	*B*
8	*Bitly*	*B*	*B*
9	*Bloomberg*	*B*	*B*
10	*Couch Surfing*	*B*	*B*
11	*Coursera*	*B*	*C*
12	*Dailymotion*	*C*	*B*
13	*Dainik Bhaskar*	*B*	*B*

14	***Delicious***	***D***	***D***
15	***Disney***	***A***	***A***
16	***Divya Bhaskar***	***B***	***C***
17	***Dropbox***	***B***	***B***
18	***Ebuddy***	***A***	***A***
19	***Entrance Corner***	***B***	***B***
20	***Evernote***	***A***	***A***
21	***Expedia***	***B***	***B***
22	***Fare Compare***	***A***	***B***
23	***Filmi Beat***	***B***	***C***
24	***First Cry***	***A***	***A***
25	***First Post***	***B***	***B***
26.	***Flipkart***	***B***	***B***
27.	***Gaana***	***B***	***B***
28.	***Google***	***B***	***C***
29.	***Gravatar***	***A***	***B***
30.	***Hotstar***	***B***	***B***
31.	***Imdb***	***C***	***C***
32.	***Indiabix***	***C***	***C***
33.	***Indigo***	***C***	***C***
34.	***Jabong***	***B***	***A***
35.	***Lastpass***	***A***	***A***
36.	***Learn Python***	***B***	***B***
37.	***Medium***	***C***	***C***
38.	***Microsoft***	***B***	***B***
39.	***Mouthshut***	***C***	***B***
40.	***Nabble***	***C***	***C***
41.	***Nytimes***	***B***	***B***
42.	***Rajnikant Vs CID jokes***	***B***	***B***
43.	***Saavn***	***A***	***A***
44.	***Tvf Play***	***B***	***C***
45.	***Whatsapp***	***A***	***B***
46.	***Wikipedia***	***B***	***B***
47.	***Wordpress***	***C***	***B***
48.	***Yahoo***	***B***	***B***
49.	***Youtube***	***B***	***C***
50.	***Zomato***	***A***	***B***

Table 10. Performance Metrics of Developed Tool

Accuracy %	Precision %	Error Rate%
65.5	**84.3**	**34.6**

On comparing the tool-defined ratings with manually defined ones, we achieved 65.5 % accuracy, 84.3 % precision and 34.6 % error rate.

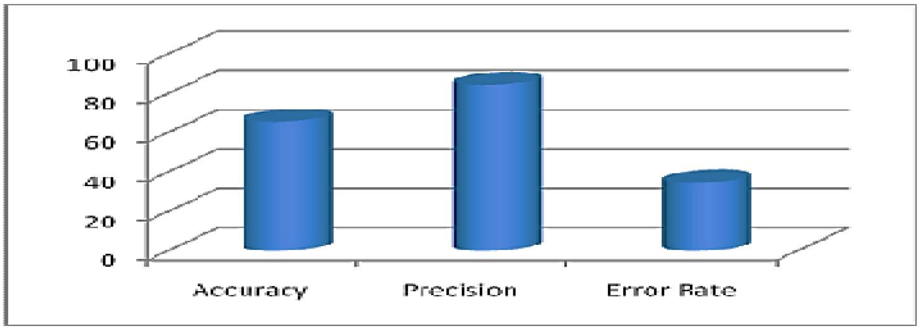

Fig. 5. Graph Showing Performance Metrics of Developed Tool

CONCLUSION AND FUTUREWORK

A method developed to rank the websites based on their terms of services belonging to user account and user content clauses. The method designed and implemented in two phases. Intra document classification of the terms of services agreement done in first phase, which consist of training phase and machine learning phase. Bag of words text categorization technique used in the training phase to prepare training data, which is implemented in the machine- learning phase. Two-layered artificial neural network used in machine learning phase calculated synaptic weights of the output achieved in the training phase. The first phase classified each term in the terms of service document into good, bad and neutral. In second phase, inter terms of service document categorization of websites using outcome obtained in the first phase has been done and finally websites are allotted classes A, B, C, D and E, based on the percentage of bad terms in the document as classified by the tool. Fair percentage of performance metrics achieved.

Future Work:

The training data set can be expanded by incorporating new terms of services belonging to not only user account and user content clauses but also from other clauses in the terms of service agreement of web page that directly or indirectly affects the user. This will lead to building of more accurate, efficient tool to rank websites based on their terms of services.

ACKNOWLEDGMENT
We would like to thanks Central University of Punjab for providing us the required software and hardware tools for successful implementation of the research work.

REFERENCES

A. S. Ghareb, A. A. Bakar, and A. R. Hamdan, "Hybrid feature selection based on enhanced genetic algorithm for text categorization," *Expert Systems with Applications*, vol. 49, pp. 31–47, 2016.

B. Al-Salemi, S. A. M. Noah, and M. J. A. Aziz, "RFBoost: An improved multi-label boosting algorithm and its application to text categorization," *Knowledge-Based Systems*, vol. 103, pp. 104–117, 2016.

D.CaiandX.He,"ManifoldAdaptiveExperimentalDesignfor Text Categorization," *IEEE Transactions on Knowledge and Data Engineering*, vol. 24, no. 4, pp. 707–719,2012.

D. S. Guru and M. Suhil, "Term-class-max-support (TCMS): A simple text document categorization approach using term-class relevance measure," *2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2016.

F.C.A.P.B.Mota,C.G.P.Bellini,J.M.D.S.Souza,andT. D. J. N. Oliveira, "The influence of civic mindedness, trustworthiness, usefulness, and ease of use on the use of governmentwebsites,"*RevistadeAdministração*,vol.51,no.4, pp. 344–354, 2016.

G. D. Pires, A. Dean, and M. Rehman, "Using service logic to redefine exchange in terms of customer and supplier participation," *Journal of Business Research*, vol. 68, no. 5, pp. 925–932, 2015.

H. Chae and E. Ko, "Customer social participation in the social networking services and its impact upon the customer equity of global fashion brands," *Journal of Business Research*, vol. 69, no. 9, pp. 3804–3812, 2016.

H. Elghazel, A. Aussem, O. Gharroudi, and W. Saadaoui, "Ensemble multi-label text categorization based onrotation forest and latent semantic indexing," *Expert Systems with Applications*, vol. 57, pp. 1–11, 2016.

H. Jeong, Y. Ko, and J. Seo, "How to Improve Text Summarization and Classification by Mutual Cooperation on an Integrated Framework," *Expert Systems with Applications*, vol. 60, pp. 222–233, 2016.

H. Jiang and Y. Zhang, "An investigation of service quality, customer satisfaction and loyalty in Chinas airline market,"JournalofAirTransportManagement,vol.57,pp.80– 88, 2016.

I. Arpaci, K. Kilicer, and S. Bardakci, "Effects of security and privacy concerns on educational use of cloud services," *Computers in Human Behavior*, vol. 45, pp. 93–98, 2015.

J. Cotelo, F. Cruz, F. Enríquez, and J. Troyano, "Tweet categorization by combining content and structural knowledge," *Information Fusion*, vol. 31, pp. 54–64,2016.

J. L. Roeder and F. G. Ashby, "What is automatized during perceptual categorization?" *Cognition*, vol. 154, pp. 22–33, 2016.

J.-F. Ding, W.-H. Shyu, C.-T. Yeh, P.-H. Ting, C.-T. Ting,C.-

J.-Y. Jiang, S.-C. Tsai, and S.-J. Lee, "FSKNN: Multi-label text categorization based on fuzzy similarity and k nearest neighbors,"*ExpertSystemswithApplications*,vol.39,no.3,pp. 2813–2821, 2012.

K. Ishii and K. M. Markman, "Online customer service and emotional labor: An exploratory study," *Computers in Human Behavior*, vol. 62, pp. 658–665,2016.

M. Donnelly, "Payments in the digital market: Evaluating the contribution of Payment Services Directive II," *Computer Law & Security Review*, vol. 32, no. 6, pp. 827–839,2016.

M.H.AghdamandS.Heidari,"FeatureSelectionUsingParticle Swarm Optimization in Text Categorization," *Journal of Artificial Intelligence and Soft Computing Research*, vol. 5, no. 4, Jan. 2015.

N. Azam and J. Yao, "Comparison of term frequency and document frequency-based feature selection metrics in text categorization," *Expert Systems with Applications*, vol. 39, no. 5, pp. 4760–4768, 2012.

P. Lin, C.-C. Chou, and S.-S. Wu, "Assessing customer value forexpressserviceproviders:Anempiricalstudyfromshippers' perspective in Taiwan," *Journal of Air Transport Management*, vol. 55, pp. 203–212, 2016.

R. H. Pinheiro, G. D. Cavalcanti, and T. I. Ren, "Data-driven global-ranking local feature selection methods for text categorization," *Expert Systems with Applications*, vol. 42, no. 4, pp. 1941–1949, 2015.

R. Jindal and S. Taneja, "A Lexical Approach for Text Categorization of Medical Documents," *Procedia Computer Science*, vol. 46, pp. 314–320,2015.

S. Ding, Z. Wang, D. Wu, and D. L. Olson, "Utilizingcustomer satisfaction in ranking prediction for personalized cloud service selection," *Decision Support Systems*, vol. 93, pp. 1–10,2017.

S. Jiang, G. Pang, M. Wu, and L. Kuang, "An improved K- nearest-neighbor algorithm for text categorization," *Expert Systems with Applications*, vol. 39, no. 1, pp. 1503–1509,2012.

T.A.Almeida,T.P.Silva,I.Santos,andJ.M.G.Hidalgo,"Text normalization and semantic indexing to enhance Instant Messaging and SMS spam filtering," *Knowledge-Based Systems*, vol. 108, pp. 25–32,2016.

T. L. James, L. Khansa, D. F. Cook, O. Bruyaka, and K. B. Keeling, "Using network-based text analysis to analyze trends in Microsoft's security innovations," *Computers &Security*, vol. 36, pp. 49–67, 2013.