



Advanced Machine Learning Techniques used for Deodorant Based Sentimental Anatomy

CMAK Zeelan Basha, Vaishnavi Yaramanchu, Y Spoorthy, Yaramanchu Divya

Abstract: Now a days there is lots of data floating in the life of world access i.e Internet which is unstructured data. To manage this unstructured data we are introduced some classification algorithms in machine learning to classify the data. Sentiment Analysis[5] is contextual mining of text from documents, reviews of customers which distinguishes and concentrates emotional data in source material. Assessment API works in fourteen unique dialects. We consider the issue of grouping records not by subject, however by generally speaking slant, e.g., deciding if an audit is certain or negative. Utilizing antiperspirants surveys as information, we locate that standard AI systems absolutely beat human-delivered baselines. The AI strategies we connected with for arrangement are Naive Bayes, maximum entropy[2] classification, and support vector machines classification algorithms for sentiment classification as on traditional topic-based categorization.[1].

Keywords : AI, Naive Bayes, API, Entropy.

I. INTRODUCTION

Now-a-days, a very a lot of data are accessible in on-line archives. Being a piece of the exertion for better arrange this data for clients, specialists have been effectively examining the issue of programmed content classification. [3]The greater part of such work has concentrated on topical order, endeavoring to sort archives as indicated by their subject related issue (e.g., sports versus legislative issues). Nonetheless, late years have seen a very fast development in on-line talk gatherings and audit locales (e.g., the New York Times' Books page) where a pivotal attributes of the posted articles is their assumption, or generally speaking sentiment towards the subject — for instance, regardless of whether an item survey is certain or negative can be resolved. Marking

these articles with their feeling would give exact and clear synopses to perusers; surely, these names are an a piece of the intrigue, which the two names antiperspirant audits that don't contain express appraising pointers and standardizes the non-comparative rating plans that individual analysts lock

in.[4]Feeling arrangement would likewise be generally useful in business insight applications (for example MindfulEye's Lexant system¹ which is a piece of business application) and recommender frameworks (e.g., some of recomanded frameworks are Terveen et al., Tatemura), where client info and criticism could be immediately characterized and abridged; without a doubt, when all is said in done, freestyle overview reactions given in common language arrangement could be handled utilizing the assessment order. Additionally, there are likewise productive standered calculations to message sifting in sentiment classification or categorization ; for instance, one may have the option to utilize assessment data to perceive and dispose of "flames"(Spertus,).[9] In this paper, we look at the viability of applying AI procedures and calculations to the opinion order issue. A gather part of this Classification issue that appears to recognize it from customary subject based characterization utilizing any conventional calculation is that while points are regularly recognizable by watchwords no one but, opinion can be communicated in an increasingly unpretentious way.[7] For example, the announcement like "what should anybody endure this coastline zone?" contains no single unit of word that is certainly not positive i.e negative.[8]Along these lines, assessment needs to require more understanding and compactness than the typical theme based grouping. In this way, our outcomes acquired by means of AI methods are delivered.

II. EXISTING METHODOLOGY

Paper[1]: Bo pang has written this paper using navie based machine learning algorithm and thr result for the classification accuracies resulting from using only unigrams.

Paper[2] : Kondru Ravi is the author of this paper. He used the SVM(support vector machine) algorithm. He classified the data using SVM with an accuracy of 95.95% on the data taken from twitter.

Paper[3] : Olena Kummer is the author of this paper. He used the algorithms like SVM, Neural Network, K-nearest Neighbours, etc. he suggested a method for overcoming the binary classification.

Paper[4]: Bruno Ohana written this paper .he used lexicon based approach to develop a classification model rather than machine learning technologies.

Manuscript published on November 30, 2019.

* Correspondence Author

CMAK Zeelan Basha*, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. Email: cmak.zeelan@gmail.com

Viashnavi Yaramanchu, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. Email: vaishnaviyaramanchu@gmail.com

Y Spoorthy, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. Email:

yarramreddyspoorthy@gmail.com

Yaramanchu Divya, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, AP, India. Email: divyayaramachu01@gmail.com

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

Paper[5] proposed lexicon based approaches for sentiment analysis.

Paper[6]:Dipak Kawade written the paper. he used machine learning approaches like naïve bayes,SVM e.t.c and he analyze twitter data using this strategies.

Paper[7]: Jovelyn C. Cuizon Written this paper.he used some scoring algorithm and he given results on restaurent dataset.

Paper[9]: suchita V written this paper .He used Naïve bayes,SVM,randomforest ad their hyperparameters for classiiation model.the results of this paper include which algorithm giving best accuracy.

III. PROPOSED METHODOLOGY

A. Constraints Of Deodorants:

For our experiment,we decided to work with Deodorant surveys. This space is tentatively advantageous on the grounds that there are enormous on-line accumulations of such surveys, and on the grounds that analysts frequently outline their general slant with a machine-extractable rating marker, for example, various stars; henceforth, we didn't have to hand-name the information for regulated learning or assessment purposesDevelopinng a model that will predict the Instant Liking of the following products:

- Deodorant B
- Deodorant G
- Deodorant J
- Deodorant F
- Deodorant H

B. Constraints:

Divide the data into test data and train data for classification

- Using train data develop a model
- Predict the model using test data
- Predict accuracy for developed model

IV. MACHINE LEARNING ALGORITHMS

This paper main goal in this with some machine learning strategies for sentimental classification just as an exceptional instance of point based order (with the "themes" being certain assessment and negative), or whether potential assumption classification techniques should be utilized. We explored different avenues regarding three standard calculations: Naive Bayes classification, most extreme entropy characterization, and bolster vector machines. The ways of thinking behind these three calculations are very unique, yet every ha been demonstrated to be viable in past content arrangement ponders

A. Naive Bayes

In general for any text classification or any other type of classifications we use Naïve bayes classification algorithm for working on datasets to develop a model.Admittedly the bayes classification algorithm uses probability concepts like posterior probability, joint probability e.t.c for classification of data.In naïve bayes classification it treats every variable or attribute are independent on one another.This algorithm[3] gives least accuracy than other classification algorithms.It

gives results by combining the test data prediction of all possible hypothesis,and their posterior probabilities.the optimal classification of instance can be maximum by using [2].

$$= \text{Argmax}_{\sum_{hi \in H} P\left(\frac{vi}{hi}\right)P\left(\frac{hi}{D}\right)}_{vj \in V}$$

To build up a credulous bayes model there are some constarints or instincts to consider a speculation space containing some theory like h1,h2,h3... ... And for given speculation we ought to figure back likelihood and appoint it.

B. Support Vector Machines:

Support vector machines (SVM) [6]can be used for efficient classification of documments .The objective behind support vector machine classification is to dissect the document vector in one class from the other by using other classification stragy i.e., maximum margin it is used to find the hyper-plane with maximum margin.Because of all this this classifier is called large margin classifier. This algorithms are not similar to probabilistic classifiers like Naïve bayes,Random forest e.t.c.in this classification the search leads to a constrained optimization problem.For instance consider the class $c_j \{1,-1\}$ be the correct class of topic level documents denoted by d_j ,the result can be represented by W in following equation.

$$W = \sum_j \alpha_j c_j d_j, \alpha_j \geq 0$$

Here the α_j 's are driven by solving a dual optimization problem. Those d_j such that α_j is greater than zero are named as support vectors, since they are the only document vectors contributing to w . Classification of test examples made up of simply determining which side of w 's hyperplane they fall on.

C. Svm Hyperparamteres:

Support vector machine deals with distint number of hyper parameters refered as[6]

- C cost
- sigma

The two parameters c and σ are used and engaged such a way that it gives a optimal results.Every parameter is effect on prediction .Firstly hyperparameter is c cost parameter of support vector machine ,from many past expriments we conclude that if c increases the accuracy of the model automatically increases.The next parameter σ which is corresponds to a kernal is used to map input data into a feature data space.As like first parameter if second parameter σ increases then accuracy of model drops from previous experiments.

D. Maximum Entropy

Most extreme entropy[2] grouping (MaxEnt, or ME, for short) is an elective strategy which has demonstrated successful in various characteristic language handling applications (Berger et al., 1996). Nigam et al. (1999) demonstrating that it now and again, however not generally, outflanks than Naive Bayes at standard content classification.or content arrangement,

$$P = \text{argmax} \left(- \sum_{k=0}^n (p(x) p\left(\frac{y}{x}\right) \log p\left(\frac{y}{x}\right)) \right)$$

V. METHODOLOGY

- Download the dataset from UCI Repository
- Dissect the dataset into training data and testing data
- Training data is used to develop an model
And testing data is used for validation of model
- To create a model we used classification algorithms such as Naïve Bayes,SVM e.t.c
- Using different classification algorithm we predict accuracy and finds which algorithm giving best accuracy for model

A. Data collection:

This paper uses the Deodorant liking dataset from UCI repository.It make standered use of dataset available in UCI repository.Dataset consists of 1300 rows with 3 coulmns regarding Deodorants.

B. Preferred classification algorithms:

We use three different claasification algorithms named naivebayes,svm,maximum entropy.

VI. EXPERIMENTAL SET-UP:

We used Deodorant instant liking dataset from kaggle(Repository) for classification.

Attributes of Deodorant Dataset:

- Respondant.ID
 - Product
 - Instant.Liking
1. Training set
 2. Testing set for validation
 3. Classification algorithms

[5]All the three algorithms are applies on the available sentiment analysis standard datasets. Standard Deodorants review datasets are used. All preprocessing is done on the dataset Deodorant review dataset. After preprocessing all algorithms are performed for classification of dataset. Algorithm makes the use of cross validation function. Cross validation is a good method to evaluate the performance of a developed model as it divides the data into two parts.[6] One part is used to train the model and build the classifier and the second part is used to test the accuracy of the predictions of the model. This process is done iteratively and different subsets of the data are used for training and testing at each iteration

VII. RESULTS:

A. SVM(Support vector machine)

Below table gives the brief idea of confusion matrix of support vector machine algorithm obtained after classification.the SvM classifier obtained 76% accuracy for used dataset.

Table :1 Confusion matrix of SVM

		Actual	
		Positive	Negative
Predicted	Positive	959	372
	Negative	273	853

Accuracy:0.767

B. Naive Bayes:

Below Table2 is the confusion matrix for the naive bayes classifier obtained after classification. The classifier has obtained accuracy of 62.29 % for dataset deodrants.

Table:2 confusion matrix of Naïve bayes

		Actual	
		Positive	Negative
Predicted	Positive	966	431
	Negative	352	874

Accuracy : 0.623

Comparison of accuracy with SVM And Naïve Naves Classifier:

Method	Accuracy
SVM	76.7%
Naïve	62.3%

VIII. CONCLUSION

Support vector machine model as high accuracy as compare to naïve bayes . The accuracy of svm model can further increased by use of hyperparameter. This paper makes the of machine learning algorithm such as naïve bayes, svm to do sentiments analysis of Deodorant review dataset. Different machine learning stratagies obtain different accuracy.

REFERENCES

1. Bo Pang and Lillian Lee and Shivakumar Vaithyanathan “Thumbs up? Sentiment Classification using Machine Learning Strategies”, Philadelphia, July 2002, pp. 79-86
2. P.Kalaivani,, Dr.K.L.Shunmuganathan, “Sentiment classification by machine learning strategies”, Indian Journal of Computer Science and Engineering , Aug-Sep 2013
3. CMAK. Zeelan Basha, K. M. Sricharan, Ch. Krishna Dheeraj, R. Ramya Sri,“ A Study on Wavelet Transform Using Image Analysis”,International Journal of Engineering & Technology ,vol.7,pp.94-96,2018.
4. CMAK Zeelan Basha, G Swetha , A Likhitha , J Bhargavi,“ Service Now Based Course Registration System”, International Journal of Innovative Technology and Exploring Engineering, Vol.9,Issue.12,pp.1643-1646.
5. CMAK Zeelan Basha, G Kalyan Jaya Simha, Y Vamsi Krishna, “An Efficient and Robust Fracture Detection in Femur Bones”, International Journal of Innovative Technology and Exploring Engineering, Vol.9,Issue.12,pp.1954-1957.
6. Gautami Tripathi , “Feature Selection and classification aproches for Sentiment Analysis”, Machine Learning and Applications: An International Journal (MLAIJ) ,Vol.2, No.2, June 2015
7. “Sentiment Classification of Reviews Using SentiWordNet” October 2009 , Conference: 9th. IT & T Conference, At Dublin

8. "A comparison of Lexicon-based approaches for Sentiment Analysis of microblog posts"
9. C. M. A. K. Z. Basha, M. Padmaja, G. N. Balaji, "Computer Aided Fracture Detection System", Journal of Medical Imaging & Health Informatics, vol. 8, no. 3, pp. 526-531, Mar. 2018

AUTHORS PROFILE



Mr. Cmak Zeelan Basha is working as Assistant Professor in department of CSE in Koneru Lakshmaiah University. His research area is Image processing. He has published several papers in area of image processing. He is having around 10 years of experience in teaching Area of interest in subjects are Image processing, Data mining and Data Warehousing, Data Structures, python programming, OOPS through Java etc.



Vaishnavi Yaramachu is working as PAT (Program Analyst Trainee) in Cognizant. She is passionate about research and her area of interest is image processing. She has Published several papers in many reputed journals. She is very much interested in coding.



Y Spoorthy is pursuing her education B.Tech in KL university. She is passionate about research and her area of interest are image processing, Networking and cloud computing. She has Published several papers in many reputed journals. She is very much interested in coding.



Y Divya is pursuing her education B.Tech in KL university. She is passionate about research and her area of interest are image processing, Networking and cloud computing. She has Published several papers in many reputed journals in the area of cloud computing, Image processing. She is very much interested in coding.