Hybrid Classifier for Analyzing Customer Satisfaction

MSc Research Project

Data Analytics

Yash Sanjay Palwe

Student ID: x20143869

School of Computing

National College of Ireland

Supervisor: Dr. Pramod Pathak & Dr. Paul Stynes

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Yash Sanjay Palwe		
Student ID:	x20143869		
Programme:	Data Analytics	**Year:** 2021	
Module:	MSc Research Project		
Supervisor:	Dr. Pramod Pathak & Dr. Paul Stynes		
Submission Due Date:	16/12/2021		
Project Title:	Hybrid Classifier for Analysing Customer Satisfaction		
Word Count:	**4700**	**Page Count: 14**	

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: Yash Sanjay Palwe

Date: 15th December 2021

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Hybrid Classifier for Analyzing Customer Satisfaction

Yash Sanjay Palwe
x20143869

Abstract

Sentiment analysis can be used for identifying customer satisfaction from online reviews. Sentiment analysis determines the polarity of the sentence written in natural language whether its positive, negative, and neutral. The challenge is to identify actual opinions about a service or product instead of just determining the polarity of reviews. This research proposes a hybrid classifier for identifying the scale of sentiment at granular level. The hybrid classifier uses techniques such as linguistic feature-based sentence filtering and hybrid feature selection. The model was trained on Amazon Product Reviews dataset. The dataset has 34k reviews. The model was compared based on precision, accuracy and recall and F1-score for analyzing and evaluating customer satisfaction. Results demonstrated that Random Forest with F4 feature outperformed all the models by showing accuracy of 99.5% with 99.54% F1score which is 20% greater than the state of the art. The proposed classifier will be useful for e-commerce organizations for understanding the needs of customer base also product's shortcomings and best features can be analyzed easily and processing load of business intelligence system can be decreased.

1. Introduction

Text analysis techniques are used for categorization and analysis of sentiments expressed in textual data which is referred to as sentiment analysis. Sentiment analysis helps to understand the customer's opinion, emotions, feelings as well as polarity of opinions expressed online whether it is positive, negative, or neutral. Determining the sentiments of opinions loses all information and actual reason of customer satisfaction. Hence, this research proposes a hybrid classifier which can identify the customer satisfaction at more granular level and comprehend them for actual opinion about a product or service.

The study of language in a scientific manner is called linguistics. It includes study related to its phonetics, semantics and grammar. Atul Khedkar et. al (2018) defines Linguistic feature based sentence filtering as a technique to filter out irrelevant information by analyzing the sentences linguistically for identification of praises and complaints by matching linguistic properties of positive and negative reviews. Hybrid feature selection technique as explained by Atul Khedkar et. al (2018) uses features such as meta, synthetic, content and semantic features for identifying the best features of the sentences.

The aim of this research is to investigate to what extent a hybrid classifier model can identify more granular levels of customer satisfaction other than just positive, neutral and negative. The major contribution of this research is a novel hybrid classifier that combines Linguistic feature-based sentence filtering and hybrid feature selection techniques for measuring a more granular level of customer satisfaction from the online opinions.

Section 2 of the research discusses about the related work carried out. Research methodology is discussed in section 3. Design Specification is discussed in section 4. Section 5 talks about implementation. Evaluation of the research carried out is presented in section 6. Research conclusion and future work are discussed in section 7.

2. Related Work

Customer Satisfaction specifies how a company's product or services meet the expectations of the customers. Online reviews help in discovering the customer satisfaction by providing insights. However, the reviews are present in unstructured text which requires processing techniques such as natural language processing, sentiment analysis, data mining, etc to discover the opinions expressed. Sentiment analysis can be performed with two kind of techniques such as lexicon based and learning based. Lexicon based techniques involves calculating the sentiment from the semantic orientation of words or phrases that occur in a text. Learning based technique involves application of different machine learning algorithms on annotated data.

Saad et. al (2021) implemented learning-based and lexicon-based methods for sentiment analysis. They have checked fficiency of drugs reviews posted on healthcare web forums was determined using sentiment analysis. TF,TF-IDF and union of TF and TF-IDF were used for feature engineering. Observations showed that AFFIN extracted negative sentiments more as compared to VADER and TextBlob dictionary. VADER extracted the least number of neutral sentiments. Models were evaluated using linear regression, extra tree classifier, AdaBoost Classifier, and multilayer perceptron were implemented. 96% accuracy was achieving from their approach using TextBlob, MLP and TF-IDF. Future work of involves combining sentiment lexicons for general purpose such as, AFFIN TextBlob and VADER which will expand corpus of lexicons and the same time provides a way to uniquely analyze the online reviews.

Atul Khedkar et. al (2018) proposed an ensemble classifier which classifies the customer reviews into complaint or praise. Implementation of the ensemble classifier was done on the Airport review dataset. The first step involved data processing in which unwanted and redundant information was removed from the dataset using techniques such as tokenization, removal of stop words, POS tagging. Moving forward linguistic feature-based sentence filtering was performed based on the sentiment score of AFFIN dictionary for filtering out the neutral sentences and for identification of praise/complaint sentences by matching with the linguistic properties. Observations showed praise sentences has a greater number of nouns, adjectives, conjunctions and were longer in length. Similarly, complaint sentences had a greater number of past tense, few nouns, few adjectives and negative words. A new formula was used for calculating sentiment score of praise and complaint sentences. Four different hybrid features namely FS1, FS2, FS3, FS4 which considered different hybrid features like meta, synthetic, etc were considered. The Linguistic analysis of positive and negative sentences showcased that 70.3% were praise and 31.9% were complaint sentences. The reviews were mainly classified into praise or complaint classes by the ensemble classifier. The ensemble classifier was able to achieve 79.83% accuracy for automatically classifying sentence as praise or complaint. The future research involves extending the base classifier as Multiclass single label-based praise/complaint classifier based on the aspect present in the sentence. Improving performance of the classifier by using pre-trained GLOVE/Word2Vec embeddings.

Kavita Ganesan et. al (2016) studied the characteristics of sentences involving complaints and praise . It was observed that praise is derived from the pool of positive reviews and Complaint are derived from negative reviews. More occurrences of adjectives, nouns, intensifiers were present in the praise sentences and were longer in length as compared to the average length of sentence. Conversely, fewer occurrences of nouns and adjectives were present, and a greater number of conjunctions and past tense were present in complaint. (Mowlaei et. al. 2018) proposed aspect-based lexicon generation approach which was quicker in execution and was proposed for identifying sentiments of online e-commerce customer reviews. In FBSA algorithm, frequency of positive and negative words is counted at fast pace which are available in training dataset of negative and positive words. The proposed algorithm searches every lexicon in every sentence where the word appears and finds the closest aspect available later increases the count of polarity either positive or negative. Results showed restaurant reviews for size of window as one had highest score as compared AFFIN, NRC Hashtag, Sentiment 140, Lui Lexicon and SentiWordNet dictionaries. Accuracy was enhanced by 3% when the dictionaries were combined. Future work involves improvement in lexicon generation and building of portal which can categorize and calculate the polarity of reviews based on aspects and decide window size automatically.

Al-Qudah et al.(2020) proposed a method for combining neutrality detector model with genetic algorithm for doing analysis of sentiments and XGBoost algorithm. Along with it predicted reviews an e-payment service for the users. Arabic social networking site was used for gathering the data. The proposed model eliminates the neutral reviews form the dataset which used the pre-defined dictionary. Genetic algorithm was mainly used deciding the optimal parameters and saving time by making execution faster. Results proved that combining genetic algorithm and XGBoost enhanced the performance as compared to the model with just XGBoost. The combined model outperformed the KNN and XGBoost by reflecting a 10% increase in the accuracy.

Xu et al. (2019) addressed the problem of evaluating polarity of the sentences accurately by identifying that lack of volume of sentiment words in sentiment dictionary. Hence, proposed a modified sentiment dictionary containing basic , polysemic sentiment words which are resulted in increasing the precision for sentiment analysis. Implementation of Naïve Bayesian classier distinguished the polarity of the sentiments for text classification. However, the research excluded passive and active words weights which needs to be more polished. It is limited to only particular fields and cannot be generalized.

Yang et al. (2020) research implemented a new model SLCABG based on CNN and sentiment lexicon and bidirectional gated recurrent unit. This model combined deep learning and sentiment lexicon. Opinion sentiments were enhanced using the lexicon. The main attributes and weights are extracted using CNN and FRU algorithm. Final stage involved weighted sentiment categorization. Their model consists convolutional layer, embedded pooled, and attention layer with 10-fold cross validation has resulted higher accuracy as compared to the existing model. However, the research classified the sentiments into only positive and negative sentiments which does not provide any in-depth vital insights about the opinions expressed by the customers. Thus, the future work of research requires study in fine tuning the classification of text based on the sentiments.

Phan et al. (2020) identified that majority ensemble models on the tweets have implemented syntactic knowledge of words without using sentiments. It was observed that fuzzy sentiment has been ignored majority of the study. The method proposed improves the performance by using three staged approaches. First stage identifies the fuzzy sentiments from the tweets. The second stage uses ensemble features and creates embeddings of tweets by combining feature vectors found out in the previous stage and last stage classifies the sentiments with the help of

CNN model. The tweets are classified into of 5 sets such as strong positive, positive, negative, neutral, and strong negative tweet sets. Results of the experiments projected 18% accuracy increase for analysing the tweet sentiments with fuzzy sentiment.

Kastrati, Imran and Kurti, (2020) researched on the aspect-based sentiment analysis for providing feedback to students on large scale . Hence, a framework was proposed which uses automatically investigate the of students conveyed in the reviews. The framework is based on aspect-level sentiments and aims to identify the sentiment expressed towards MOOC. Resulted in less manual effort annotating the data. Implementation of model was done on the dataset of Coursera more than 100k student reviews. The framework performs finely for both the aspect and aspect-sentiment categorization. Furthermore, led to increase in accurate results. However, the proposed research cannot be generalised for any domain other than education domain. Application of it on to other domains requires modifying for the parameters given as input.

Bouazizi and Ohtsuki (2018) proposed sentiment analysis model which deals with multiclass classification by addressing sentiments truly expressed by user rather than considering overall text sentiment. As the research progressed, quantification which is a twitter dataset as used. The term quantification refers to the sentiment identification in one post rather than just assigning a single label of sentiment. A unique method which assigns different scores to sentiment present in a tweet automatically was proposed and selects the sentiments having highest score in the conveyed text in the latter stage. In the model first stage consists of labelling the data which was done manually, and the results compared against the human annotation. As a result, the model was suitable and projected an F1 score of 45.9%.

Liu, Cao and Yin (2019) developed model for calculating sentiment score for short texts. The proposed model uses bi-level attention model for sentiment identification in short text. They have used latent topic knowledge for semantic representation of text but informativeness of sentiments was not considered

In conclusion, from the literature review it can be inferred there exists various types of sentiment analysis approaches which mainly focuses on polarity classification of reviews into three main classes namely neutral, positive, and negative and ignores the informativeness of reviews. Opinions which deal with strong and extreme opinions like extremely negative or extremely positive reviews are very challenging to identify. Furthermore, the customer reviews contain ratings which are not taken into consideration which performing sentiment analysis on reviews. There is a need of classifier which can classify the reviews into a more continuous and granular scale of sentiment for gaining vital insights from the reviews. Also, there is a need for incorporating user ratings into sentiment analysis for making it more meaningful and informative. This research proposes a hybrid classifier for identifying the scale of sentiment at granular level with taking into consideration user ratings. The hybrid classifier using techniques such as linguistic feature-based sentence filtering and hybrid feature selection.

3. Methodology

The proposed research methodology consists of 6 stages (shown in fig.1) namely data understanding, data pre-processing, linguistic feature-based sentence filtering, hybrid feature selection, hybrid classifier model, evaluation, and visualization.

The first step, *Data gathering*, Amazon reviews product dataset[1] contains 34k reviews posted online on amazon about different products. The dataset is obtained from Kaggle.

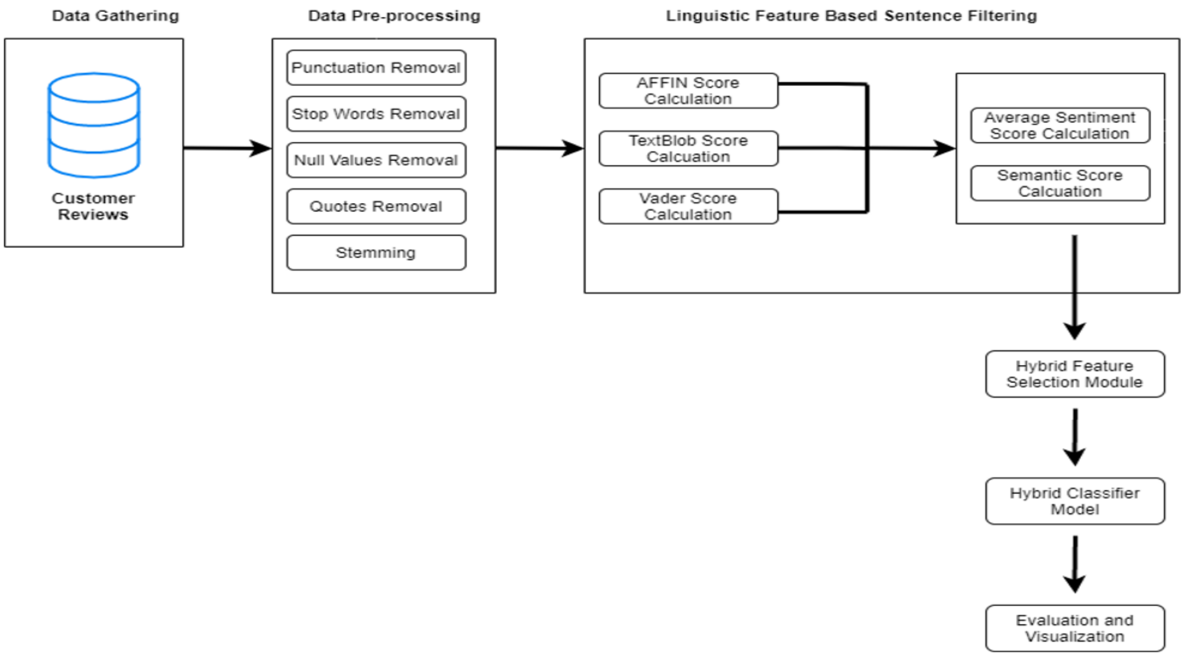

Figure 1 : Research Methodology

The second step, *Data Pre-processing* involves loading the reviews into data frame and elimination of redundant and unrelated information such as punctuation, stop words, null values and quotes from the dataset. Then the reviews are split into sentence level. The dataset initially contained 23 features such as id, date, brand, etc after cleaning only 3 features review, rating and label based on the ratings were left.

The third step, *Linguistic feature-based sentence filtering module* calculates the overall average sentiment score of the sentence by combining scores from three different libraries namely AFFIN, TextBlob and VADER (Valence Aware Dictionary for Sentiment Reasoning) for linguistically identifying whether the sentence is positive, negative or neutral. Furthermore, semantic score is generated which combines the overall average sentiment score and review ratings which helps to eliminate the neutral sentences and classify the positive and negative sentences into two new more granular level classes such as extremely positive and extremely negative. The formula devised for generating semantic score is given by following equation 1:

$$\text{Semantic Score} = (\text{Overall Average Sentiment Score} * \text{Review Ratings})/10 \qquad (1)$$

For further analysis only positive, extremely positive, negative, and extremely negative sentences are considered.

The fourth step, *Hybrid feature selection module* Atul Khedkar et. al (2018) *selects* relevant subset of features. It uses hybrid features such as meta features which is the length of sentence, number of stop words, and semantic score which was generated from the previous module.

[1] https://www.kaggle.com/datafiniti/consumer-reviews-of-amazon-products?select=Datafiniti_Amazon_Consumer_Reviews_of_Amazon_Products_May19.csv

The fifth step, *Hybrid classifier model* involves encoding labels and classes creation, training of different models such as KNN, Logistic Regression, SVC, Random Forest, Ensemble, etc with training and testing dataset split into 80:20 ratio.

The sixth step, *Evaluation and Visualization* involves performance evaluation of different machine learning models used for classification of reviews using precision, recall and accuracy. The different models were compared amongst each other for finding the optimal model.

4. Design Specification

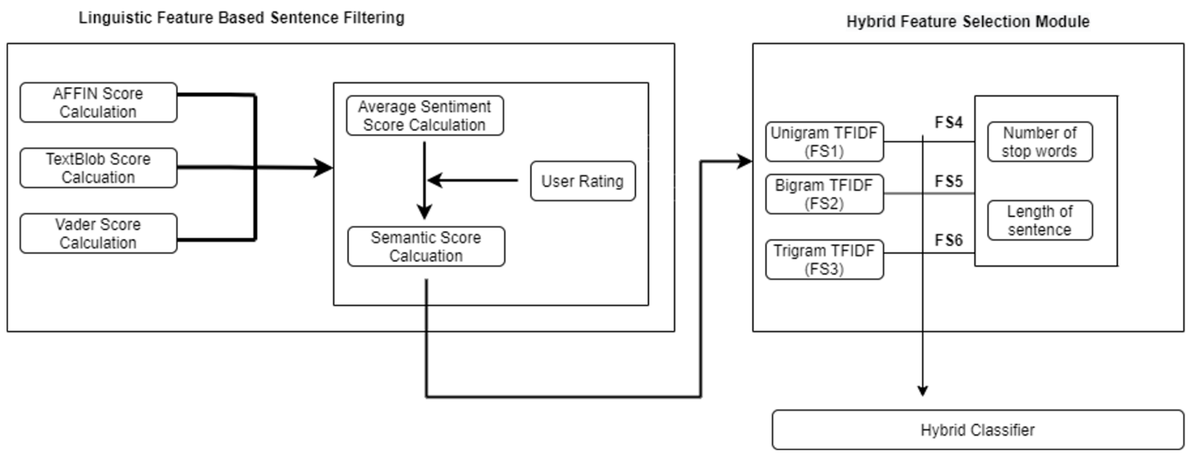

Fig. 2: Linguistic Feature and Hybrid Feature Selection Module

The proposed classifier combines linguistic feature-based sentence filtering and hybrid feature selection techniques shown in fig.2. The linguistic feature consists of libraries such AFFIN, TextBlob and VADER Saad et. al (2021) which helps in calculating average sentiment score. Incorporating all the three libraries and averaging them out helps to eliminate biasness in sentiment score calculation. As Saad et. al (2021) experiment showcased negative sentiments were extracted more by AFFIN than TextBlob and VADER because of the negative words in the dictionary. Similarly, TextBlob extracted a greater number of positive sentiments amongst all. The linguistic feature-based sentence filtering starts with calculation of average sentiment score generated with the help of all the three libraries. Later semantic score is calculated given by equation 1 incorporating ratings provided by users. Review sentences having semantic score 0 are eliminated as sentimentally neutral sentences doesn't contribute much towards analysis. Next, threshold value is set for the classification of sentiment score as extremely positive(score>1), positive (0> score <1), negative(-0.10< score <0) and extremely negative(score <0).

The hybrid feature selection module selects subset of relevant features. It contains N-gram features which is any sequence of occurrence of "n tokens" Atul Khedkar et. al (2018). Based on the occurrence of unigrams, bigrams and trigrams are considered as base features which can be used to identify opinionated text from a set of review sentences. More informative context is obtained from bag of n-grams than using bag of words as only words are given and not the context. To represent text as numerical data for machine learning algorithm TF-IDF is used feature representation techniques for unstructured text data. It assigns term weight based on its frequency of occurrence in the document. Meta features such as length of sentence and number of stop words along with semantic score are combined to generate hybrid features. As shown in Table 1, in this study the following set of Hybrid Features based on Linguistic properties of positive and negative sentences are considered.

Table 1 : Description of Proposed Linguistic Base and Hybrid Features

Type	Model Features	Descriptions
Base Features	**FS1**	Unigram TF-IDF (1-gram)
	FS2	Bigram TF-IDF(2-gram)
	FS3	Trigram TF-IDF(3-gram)
Hybrid Features	**FS4**	FS1 (Unigram TF-IDF) + (Hybrid Features = Meta + Semantic features)
	FS5	FS2 (Bigram TF-IDF) + (Hybrid Features = Meta + Semantic features)
	FS6	FS3 (Trigram TF-IDF) + (Hybrid Features = Meta + Semantic features)

The output from the hybrid feature selection module is then passed on to the hybrid classifier. In this work review sentences are classified as extremely positive, positive, negative and extremely negative classes.

5. Implementation

The hybrid classifier for analyzing customer satisfaction was implemented in python programming language using Jupyter notebooks on the Amazon product reviews dataset. The dataset had 34k customer reviews about various product hosted on amazon. Necessary libraries such as pandas, matplotlib, sklearn, nltk, etc were imported first in jupyter notebook. Next, AFFIN, TextBlob, VADER natural language processing libraries were imported for calculating the average sentiment score. Later the process mentioned in research methodology were followed. Finally, the proposed hybrid classifier was trained on the Saad et. al (2021) dataset by splitting the data in 80:20 ratio of train and test respectively and various machine learning algorithms such as KNN, SVC, Random Forest, etc were implemented for selecting the best performing machine learning algorithm. Results showcased that Random Forest with unigram-TF-IDF, and hybrid feature outperformed the state of the art by having a very high accuracy of 99.57% and F1 score of 99.5%.

6. Results

The aim of this experiment is to study the properties of positive and negative reviews posted online by using supervised machine learning algorithms Random Forest, SVC, KNN ,Naïve Bayes and implement the state of the art model. The model was trained to recognize and classify review sentences as praise or complaint considering only AFFIN score along with base and hybrid features. Result showed that dataset had 11330 positive and 5750 negative sentences after processing the data and applying linguistic analysis. Out of all the positive sentences present, praises are 7860 and plain positive sentences were 3470. From all the negative sentences, complaint sentences were 3250 and 2330 plain negative sentences were found. Highest accuracy of 79.83% was achieved by bigram with hybrid features using ensemble algorithm. It was observed that praise were longer in length and used more nouns and adjectives. Complaints were longer in length as well and contained very less adjectives and more past tense.

The aim of the second experiment is to generate an average sentiment score by using AFFIN, VADER and TextBlob libraries and include user ratings in calculating semantic score for classification of reviews using base features only. The results of experiment are showcased in table 2. It was observed that SVC with bigram tfdif outperformed all the other algorithms by showing an accuracy of 83.192% whereas Naïve Bayes showed the least accuracy with unigram tfidf features.

Table 2: Results of Experiment second considering only Base features

	Machine Learning Algorithm	Accuracy	F1 Score	Precision	Recall
Unigram TFIDF (FS1)	SVC	83.192	79.592	83.192	83.192
	Random Forest	83.192	79.592	83.192	83.192
	Naïve Bayes	76.666	66.786	76.666	76.666
	Logistic Regression	80.422	75.226	80.422	80.422

	Machine Learning Algorithm	Accuracy	F1 Score	Precision	Recall
Bigram TFIDF (FS2)	SVC	82.159	78.111	82.159	82.159
	Random Forest	82.159	78.111	82.159	82.159
	Naïve Bayes	76.713	66.877	76.713	76.713
	Logistic Regression	80.563	75.283	80.563	80.563

	Machine Learning Algorithm	Accuracy	F1 Score	Precision	Recall
Trigram TFIDF (FS3)	SVC	82.159	78.054	82.159	82.159
	Random Forest	82.159	78.054	82.159	82.159
	Naïve Bayes	76.901	67.313	76.901	76.901
	Logistic Regression	80.798	75.652	80.798	80.798

The aim of third experiment is to include hybrid features in the classifier for classification of reviews. The models SVC, Random Forest, MLP, AdaBoost, KNN, Ensemble were trained on dataset. Results from Table 3 show that Random Forest had the highest accuracy of 99.577% with FS4 (Unigram + Hybrid features).

Table 3: Results of Third Experiment

	Machine Learning Algorithm	Accuracy	F1 Score	Precision	Recall
Unigram TFIDF + Hybrid Features (FS4)	SVC	98.779	98.777	98.777	98.777
	Random Forest	99.577	99.574	99.577	99.577
	MLP	95.868	95.853	95.868	95.868
	Ensemble	98.779	98.747	98.779	98.779
	KNN	93.56	93.301	93.568	93.568
	Logistic Regression	97.323	96.99	97.323	97.323

	Machine Learning Algorithm	Accuracy	F1 Score	Precision	Recall
Bigram TFIDF + Hybrid Features (FS5)	SVC	98.967	98.97	98.967	98.967
	Random Forest	99.154	99.147	99.154	99.154
	MLP	95.352	95.272	95.352	95.352
	Ensemble	98.591	98.541	98.591	98.591
	KNN	93.943	93.65	93.943	93.943
	Logistic Regression	97.323	96.949	97.323	97.323

	Machine Learning Algorithm	Accuracy	F1 Score	Precision	Recall
Trigram TFIDF + Hybrid Features (FS6)	SVC	99.107	99.11	99.107	99.107
	Random Forest	99.436	99.431	99.431	99.436
	MLP	95.305	95.23	95.305	95.305
	Ensemble	98.685	98.635	98.685	98.685
	KNN	94.084	93.709	94.084	94.084
	Logistic Regression	97.323	96.973	97.323	97.323

On analysing word cloud of positive sentences (fig.3) it was observed most of the customers were happy with their purchase and most of the positive sentences gave an overview about the purchase by mentioning words like love, appease, nice etc. Customers like the easy to use nature of Alexa. The device had good screen and was perfect item to use.

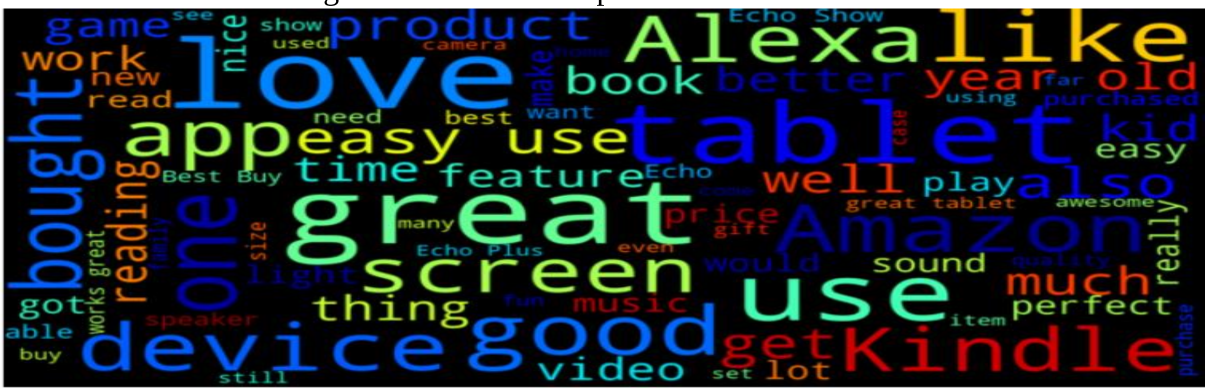

fig.3 : Word cloud of positive reviews

The word cloud of extremely positive sentences (fig. 4)gave more detailed analysis like the picture quality of screen was amazing and had better price. Speaker was the best feature of Alexa. It had great sound. It was available at better price. It was fast and easy to use.

Fig.4 : Word Cloud of extremely positive reviews

The word cloud of negative(fig. 5)and extremely reviews (fig. 6) showcase that the customers were not satisfied with the build quality of tablet. The customers complaint about the battery life and were not able to read much on Amazon kindle. The tablet needs frequent charging.

fig. 5: Word cloud for negative reviews

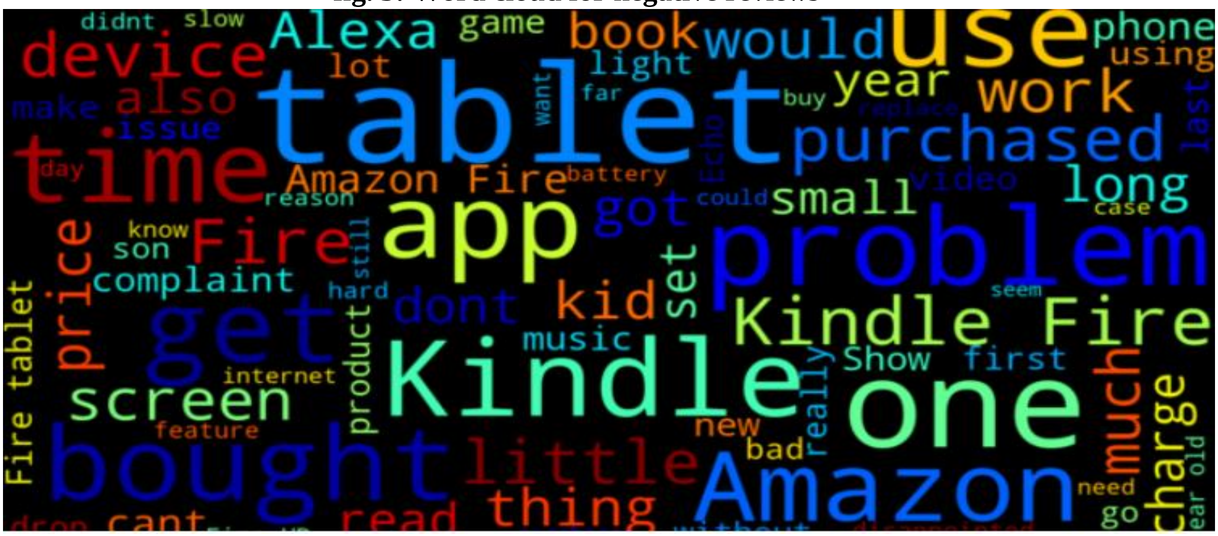

Fig. 6 : Word cloud for extremely negative reviews

7. Conclusion and Future Work

The aim of this research was to propose a hybrid classifier for analysing customer satisfaction from online reviews at more granular level of sentiments. The research proposes a hybrid classifier that combines linguistic feature-based sentence filtering and hybrid feature selection techniques. Results demonstrate that with the inclusion of hybrid features the accuracy of the classifier has been increased and is shown by Random Forest by achieving 99.295% with Unigram TFIDF+ Hybrid features (FS4) outperforms the state of the art classifier having accuracy of 78.9%. This research can potentially enhance the customer satisfaction from online reviews.

For future work, this research can be extended using deep learning algorithms for achieving better accuracy. Aspect based opinion mining can be implemented also a framework can be developed on the proposed classifier which can automatically select the top 10 best and worst reviews based on aspect-based opinion mining. Lastly, emoticons can be considered which determining the polarity of the reviews and stylometric analysis can be incorporated for restricting users to post only negative and positive reviews.

References

Saad, E., Din, S., Jamil, R., Rustam, F., Mehmood, A., Ashraf, I. and Choi, G.S. (2021). Determining the Efficiency of Drugs Under Special Conditions From Users' Reviews on Healthcare Web Forums. *IEEE Access*, 9, pp.85721–85737.

Atul Khedkar, S. and Shinde, S.K. (2018). Customer Review Analytics for Business Intelligence. *2018 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC).*

Mowlaei, M.E., Abadeh, M.S. and Keshavarz, H. (2018). A Lexicon Generation Method for Aspect-Based Opinion Mining. *2018 IEEE 22nd International Conference on Intelligent Engineering Systems (INES).*

Yang, J. and Yang, J. (2020). Aspect Based Sentiment Analysis with Self-Attention and Gated Convolutional Networks. *2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS).*

Li, M., Ma, Y. and Cao, P. (2020). Revealing Customer Satisfaction With Hotels Through Multi-Site Online Reviews: A Method Based on the Evidence Theory. *IEEE Access*, 8, pp.225226–225239.

Hutto, C.J. and Gilbert, E. (2014). *VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text*. [online] Available at: http://comp.social.gatech.edu/papers/icwsm14.vader.hutto.pdf.

Al-Qudah, D.A., Al-Zoubi, A.M., Castillo-Valdivieso, P.A. and Faris, H. (2020). Sentiment Analysis for e-Payment Service Providers Using Evolutionary eXtreme Gradient Boosting. IEEE Access, 8, pp.189930–189944.

Xu, G., Yu, Z., Yao, H., Li, F., Meng, Y. and Wu, X. (2019). Chinese Text Sentiment Analysis Based on Extended Sentiment Dictionary. *IEEE Access*, 7, pp.43749–43762.

Yang, L., Li, Y., Wang, J. and Sherratt, R.S. (2020). Sentiment Analysis for E-Commerce Product Reviews in Chinese Based on Sentiment Lexicon and Deep Learning. IEEE Access, 8, pp.23522–23530.

Phan, H.T., Tran, V.C., Nguyen, N.T. and Hwang, D. (2020). Improving the Performance of Sentiment Analysis of Tweets Containing Fuzzy Sentiment Using the Feature Ensemble Model. *IEEE Access*, [online] 8, pp.14630–14641. Available at: https://ieeexplore.ieee.org/document/8949358 [Accessed 18 May 2020].

Kastrati, Z., Imran, A.S. and Kurti, A. (2020). Weakly Supervised Framework for AspectBased Sentiment Analysis on Students' Reviews of MOOCs. *IEEE Access*, 8, pp.106799– 106810.

Lin, Y., Li, J., Yang, L., Xu, K. and Lin, H. (2020). Sentiment Analysis With Comparison Enhanced Deep Neural Network. IEEE Access, 8, pp.78378–78384.

Bouazizi, M. and Ohtsuki, T. (2018). Multi-Class Sentiment Analysis in Twitter: What if Classification is Not the Answer. IEEE Access, 6, pp.64486–64502.

Liu, W., Cao, G. and Yin, J. (2019). Bi-Level Attention Model for Sentiment Analysis of Short Texts. IEEE Access, 7, pp.119813–119822.

Adil, B., Abdelhadi, F., Mohamed, B. and Haytam, H. (2019). A Spark Based Big Data Analytics Framework for Competitive Intelligence. *2019 1st International Conference on Smart Systems and Data Science (ICSSD).*

Bouazizi, M. and Ohtsuki, T. (2019). Multi-class sentiment analysis on twitter: Classification performance and challenges. *Big Data Mining and Analytics*, 2(3), pp.181–194.

Almeida, A.M.G. de, Barbon, S. and Paraiso, E.C. (2016). Multi-class Emotions Classification by Sentic Levels as Features in Sentiment Analysis. *2016 5th Brazilian Conference on Intelligent Systems (BRACIS)*.

Shi Yuan, Junjie Wu, Lihong Wang and Qing Wang (2016). A hybrid method for multi-class sentiment analysis of micro-blogs. *2016 13th International Conference on Service Systems and Service Management (ICSSSM)*.

Rane, A. and Kumar, A. (2018). Sentiment Classification System of Twitter Data for US Airline Service Analysis. *2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC)*.

Yilmaz, S.F., Kaynak, E.B., Koc, A., Dibeklioglu, H. and Kozat, S.S. (2021). Multi-Label Sentiment Analysis on 100 Languages with Dynamic Weighting for Label Imbalance. *IEEE Transactions on Neural Networks and Learning Systems*, pp.1–13.

Ganesan, Kavita, and Guangyu Zhou. "Linguistic Understanding of Complaints and Praises in User Reviews." Proceedings of NAACL-HLT. 2016