Predicting ratings of social media feeds: Combining latent-factors and emotional aspects for improving performance of different classifiers

Arghya Ray*
FORE School of Management
Area of Information Technology
Adhitam Kendra, B 18, NRPC Colony,
Qutab Institutional Area, New Delhi, Delhi-110016, India
Email: mailarghyaray@gmail.com
Phone: +91-9199166554

Pradip Kumar Bala
Indian Institute of Management Ranchi
Area of Information Technology and Business Analytics
Suchana Bhawan, 5th Floor, Audrey House Campus,
Meur's Road, Ranchi, Jharkhand, Pin-834008, India
Email: pkbala@iimranchi.ac.in

Nripendra P. Rana
College of Business and Economics,
Qatar University,
Doha, P.O. Box-2713, Qatar.
Email: nrananp@gmail.com

Yogesh K Dwivedi[a,b]
[a]Emerging Markets Research Centre (EMaRC),
School of Management, Room #323
Swansea University, Bay Campus,
Fabian Bay, Swansea, SA1 8EN, Wales, UK.
Email: y.k.dwivedi@swansea.ac.uk
[b]Department of Management,
Symbiosis Institute of Business Management, Pune & Symbiosis International (Deemed University),
Pune, Maharashtra, India

***Corresponding Author**

Abstract

Purpose – The widespread acceptance of various social platforms has increased the number of posts by users about various services based on their experiences about the service. Finding out the intended ratings of social media posts is important for both organizations and prospective users since these posts can help in capturing the user's perspectives. However, unlike merchant websites, the social media posts related to the service-experience cannot be rated unless explicitly mentioned in the comments. Additionally, predicting ratings can also help to build a database using recent comments for testing recommender algorithms in various scenarios.

Design/methodology/approach – In this study, we have predicted the ratings of social media posts using linear (Naïve Bayes, max-entropy) and non-linear (k-Nearest neighbour) classifiers, different features and sentiment and emotion scores.

Findings – Overall results of this study reveal that the non-linear classifier (kNN classifier) performed better than the linear classifiers (Naïve Bayes, Max-entropy classifier). Results also show an improvement of performance where the classifier was combined with sentiment and

emotion scores. Introduction of the feature ‘factors of importance’ or ‘the latent factors’ also shows an improvement of the classifier performance.
Originality/value – This study provides a new avenue of predicting ratings of social-media feeds by the use of machine learning algorithms along with a combination of different features like, emotional aspects, and latent factors.

Keywords: Emotional Aspects; k-Nearest Neighbours; Max-entropy; Naïve Bayes; Rating prediction; Social media feeds;

Paper type: Research Paper

1. Introduction

The availability of Internet and the widespread acceptance of social-media (SM) platforms (e.g. Facebook, Twitter, etc.) has led to the growth of online service providers and an increase in overall competition in almost every sector (Ray, Bala, and Dasgupta, 2019a). Knowing what a customer feels about a product or service based on the reviews or ratings provided by customers is of great strategic importance for organizations (Black and Kelley, 2009). While online reviews can impact sales (Li, Chen, & Zhang, 2020), online ratings reflect quality and consumer satisfaction (Engler et al., 2015) associated with a product. Both online reviews and ratings influence the decisions of prospective customers (Cui et al., 2012; Kostyra et al., 2016; Chatterjee, 2019; Ray et al., 2020b). Apart from posting reviews on merchandise or organizational pages, customers also post their views about a product or service in SM due to the increasing popularity of SM platforms and the easy means of sharing information (Simon et al., 2015; Chaffey, 2019). Customer behaviour can however be affected by several factors. Latent factor refers to the unobserved factor/s which although not observable can affect customer behaviour. These latent factors are however important for deciding customer behaviour, and we have termed it as ‘factors-of-importance’ for proposing a new feature for classifiers. Comments posted on SM about various products/services are rich in information (Chatterjee, 2019). A proper analysis of SM posts about products/services can reveal factors (latent factors) affecting customer behaviour (Ray et al., 2019a, 2019b) and can also throw light on the underlying sentiments and emotions involved (Chatterjee, 2019). Although researchers (e.g. Mukherjee and Bala, 2017) have worked on features such as content words, and function words, it increases the time and space complexity because of the use of a large bag-of-words for constant comparison and testing. However, no studies have utilized the 'factors-of-importance' as a feature. Although researchers (e.g. Dooms et al., 2014; Prasetyo and Winarko, 2016) have worked on rating prediction, the studies lack good prediction accuracy (Ray, Bala, and Jain, 2020a). The need for improving the accuracy of rating-prediction and to devise a way to reduce space and time complexity of classifiers motivated us to work on these gaps.

The business problems that drive this research are as follows: First, there is a lack of proper research on accurate rating prediction from SM posts about various products/services. For effective SM marketing, organizations need to understand consumer perspectives better and engage with them on SM platforms (Mosley, 2018). Due to the large number of SM posts about products/services on SM platforms (Simon et al., 2015), it is not easy for service-providers to segment the negative SM posts from the good ones (Oheix, 2018). But the SM posts have both strategic and marketing importance. When a new prospective customer reads a comment or post about a certain product/service, their decisions can be affected. Additionally, it has also been found that people do not usually spend much time going through a post (Patel, 2016; Read, 2016) but rather take a glance through the catchy words. However, when the ratings are available, it becomes easier for users to take a decision. Unlike merchandise websites, ratings are not available for SM posts. Thus, SM posts regarding a product/service will not be of much value to users who do not love reading lengthy posts. However, for the providers, each and every post is important for understanding customers' views and for solving the problems that users face. Inability to address the issues instantly may lead to loss of customers. Detecting ratings thus helps to provide a view of the users’ feelings and hence becomes an important part of customer-service. Second, deriving ratings from SM posts will help in generating datasets from most recent posts for training recommender systems (RSs) (Chambua and Niu, 2021). This will help to reduce the dependence on older datasets like MovieLens (Dooms, De Pessemier, and Martens, 2013). Third, since people post their views about a product/service in different platforms, detecting ratings and combing them from different platforms, will help users gain a better understanding about the product/service. This will also help providers to understand how they are performing in the market. Fourth, existing research has used features, like, content words, function words, etc. from the whole dataset for the classification algorithms (Mukherjee and Bala, 2017) which can be really time-taking. Existing research has not attempted at using the factors identified in empirical studies as features for classification algorithms which, we believe, can reduce both time and space complexity. Therefore, the research questions that drive this study are as follows:

> RQ1: Can the social media comments about a product/service be accurately rated?
>
> RQ2: Can we create a dataset containing user details, the product details, the customer comments and ratings from social media comments?
>
> RQ3: Can a combination of ”factors-of-importance” and textual emotional content improve the accuracy and performance of the classifiers for predicting ratings from social media posts?

This work has used various natural language processing (NLP) based approaches to classify the ratings of the SM posts. Initially, a web-mining process was followed to extract the posts. From the cleansed data, the emotion and sentiment scores from each comment were calculated. While the sentiment scores provide an overall score related to the sentiments expressed, the emotional score provides a deeper understanding of the feelings expressed in the user-reviews such as fear, anger, trust, etc. (Ray et al., 2020a; Chatterjee, 2019). Then various classifiers were used by combining the sentiment and emotion values to compare the performance. For training the classifiers, data from different provider-platforms were used. Once the model is trained properly it was able to classify the SM posts related to that provider-platform. This study has utilised three different contexts (i.e. online learning (eLearning), online food delivery services (OFDs) and online travel agency services (OTAs)) for comparing the results. Findings show that the non-linear classifier (kNN classifier) had a better performance. Results show an improvement of classifier performance when the sentiment and emotional values are combined. Introduction of 'factors-of-importance' also shows an improvement of classifier performance.

The paper is divided into the following sections. Section 2 following the introduction reviews literature related to rating prediction. Section 3, Section 4 and Section 5 contain the methodology, findings and the discussion. Section 6 concludes this study.

2. Literature Review

Online user feedback plays a very important role in this modern era. For prospective customers, the reviews and ratings help in taking decisions on whether to choose a certain product/service or not. For organizations, the reviews provide an idea of what the customers are looking for and the issues they face. Though reviews can provide an in-depth idea of what the customers feel, it may not be possible for service providers to go through each and every comment posted by users about the services on social platforms. Unlike merchandise or service-provider's websites, there is no provision of providing ratings to SM posts, unless explicitly mentioned by the user. Thus detecting ratings becomes an important task because it: (a) provides an easy view of the user's feelings; (b) helps to create recent databases on various contexts. This study proposes a method to determine ratings from SM textual posts (in this case, tweets) by using two different classifiers.

The relevant themes and associated papers are discussed in Table 1. Subsection 2.1 provides a brief overview of how the online reviews and ratings are important for various businesses. This sub-section also prepares the scenario as to how rating SM comments can be useful for businesses. The Subsection 2.2 discusses the seminal works on predicting ratings from SM feeds (tweets) by earlier researchers and the problems faced. To classify and predict the ratings from SM feeds, it is necessary to figure out the underlying sentiments and emotional aspects (refer Subsection 2.3). Subsection 3.5 provides an overview of the different classifiers used to classify and predict the ratings from SM feeds.

<<Insert Table 1 here>>

2.1 Online reviews and online ratings

Online reviews refer to the views shared by the customers on a certain aspect based on their experience or observation, which are written mainly in textual format. Duan et al. (2008) noted that user reviews play a similar role as word-of-mouth and helps in driving revenues accordingly. Online ratings on the other hand capture the customer's views in numerical form. For example in a rating of [1-5], '1' refers to high dissatisfaction whereas '5' refers to high satisfaction. The fact that "90% of the consumers read online reviews before visiting a business" (Saleh, 2015), "84% of the customers trust online reviews and ratings as much as they trust personal recommendations" (Rentpath, 2017) and "86% customers refrain from going to a business having negative reviews" (Saleh, 2015) shows the importance of online reviews for not only customers but also service providers (Willas, 2018). But online reviews may not be enough. Gavilan et al. (2018) found that online ratings in combination with online customer reviews have a positive impact on consumer's decision making in context of hotel booking. Additionally, in case of good ratings, customer's trust also gets affected by the number of reviews (Gavilan et al., 2018). However, in case of negative ratings, the number of reviews doesn't have any effect on the trust worthiness of the rating (Gavilan et al., 2018). Hence, there is a need to find out the online ratings of the user's views on SM.

Researchers have noted the importance of reviews and ratings in various contexts. Duan, Gu, and Whinston (2008) found that there is a negligible impact of movie ratings on revenues, and movie reviews on customers' intentions. However, in case of hotel reservations, Sparks and Browning (2011) found that customers are influenced by both negative and positive reviews and ratings. Customer's decision gets highly influenced when there are more negative reviews. However, when positive reviews and ratings are available, customer's intentions to book hotels increase (Sparks and Browning, 2011). Researchers have also found that various factors such as cost performance (Chen and Xu, 2016), customer's sentiment polarity (Geetha et al., 2017), social influence bias and ratings bubbles (Aral, 2014) affect the overall ratings of products or services. Researchers have also used different techniques like, tensor factorization method (Chambua et al., 2018), latent factor models (Gu et al., 2020), convolution neural networks (Khan and Niu, 2021), etc., to predict ratings from textual reviews. Since online reviews and ratings serve as great information sources (Engler et al., 2015) for users as well as providers (Chatterjee, 2019), determining ratings from SM posts can be really useful from the business perspective.

2.2. Predicting ratings from customer comments in SM

SM platforms (e.g. Facebook, Twitter) enable users to establish connections and express or share their or other's viewpoints in real-time (Lerman and Ghosh, 2010). Twitter, a micro-blogging site, is widely used by people all around the globe. More than eighty percent of Twitter users modify their accounts almost on a daily basis (Thelwall et al., 2011). Thus researchers usually use Twitter feeds for their research work. Table 2 summarizes the research works on prediction of ratings from SM data. Apart from using Google Scholar, we have used other academic literature databases like, Springer, Emerald, Blackwell-Wiley, and Web of Science to search the published articles. We have mainly focused on rating prediction from social-media posts using keywords {allintitle:

("rating" OR "ratings") "twitter" OR "Facebook" OR "Flickr" OR “LinkedIn” OR “Instagram”}. Chambua and Niu (2021) have highlighted the importance of rating prediction from textual data for RSs and have found the most common methods used are matrix factorization and deep learning techniques. Based on reviews on websites, Subroto and Christianis (2021) have used ensemble machine learning techniques and deep learning techniques for rating prediction.

Not much work has been done on predicting ratings from customer comments on social media. Twitter data has been used mostly by various researchers in various contexts like, promoting health literacy (Zhou et al., 2018), understanding customer’s food choices (Dondokova et al., 2018; Mostafa, 2018), understanding customer purchase intentions (Haque et al., 2019), etc. We have only selected those papers where tweets were used for predicting ratings. Scholars have utilized different techniques for rating prediction in different contexts like, TV program rating (Wakamiya et al., 2011; D’Souza et al., 2013; Kayahan et al., 2013, Mhaisgawali and Giri, 2014; Akarsu and Diri, 2016; Prasetyo and Winarko, 2016), movie rating (Dooms et al., 2013; Kesharwani and Bharti, 2017), books, video and music (Dooms et al., 2014). Wakamiya et al. (2011) have evaluated the overall relevance value based on different relevance measures (textual, temporal and spatial) and found an improved performance results. Prasetyo and Winarko (2016) also found an improved accuracy in similarity calculations (using cosine similarity and TF-IDF). Schmit and Wubben (2015) focused on utilising public opinion and tweets to predict ratings of newly released movies. Dooms et al. (2013, 2014) focused on building database from twitter extracts by assigning ratings to the posts in contexts like books rating, movie rating, music rating and video ratings. Razis and Anagnostopoulos (2014) examined ratings of twitter influencers from twitter data. Kesharwani and Bharti (2017) used an algorithm to assign weights to the positive and negative words for predicting movie ratings. However, findings showed much lesser rating as compared to rating on other websites. Researchers generally used topic modeling techniques for rating predictions and Mahadevan, and Arock (2020) have performed a thorough review on those studies. Ray et al. (2020a) noted an improvement in prediction accuracy by combining sentiment and emotion values to classify the user-generated content using different datasets. This study is an extension of the study by Ray et al. (2020a). Thus, there are very few studies on determining ratings from tweets in emerging contexts like education, food delivery and traveling services.

<<Insert Table 2 here>>

2.3. Sentiment and emotion analysis

Sentiment and emotion analysis have been used by research for understanding customer behaviour (Turney, 2002; Ray et al., 2020a) in different contexts like, article analysis (Alsmearat et al., 2015), etc. While sentiment analysis provides only valence scores (positive, negative or neutral) (Turney, 2002), emotion analysis provides scores capturing different emotions (anger, fear, disgust, sadness, joy, surprise, anticipation, and trust) from the textual data (Ray et al., 2020a). For example, the sentiment analysis of the sentence "The sun rises from the east" will just return the score 0.2449, whereas the emotional analysis will provide scores for different human emotions, like, anger=disgust=fear=sadness=0, and anticipation=joy=surprise=trust=0.125. Sentiment analysis and emotional analysis is used in different domains like, finance, marketing (Momtazi, and Layeghi, 2021), sports (Cohen-Kalaf et al., 2021), healthcare (Abualigah et al., 2020), etc. to understand person's opinion (Naqvi, Majid, and Abbas, 2021; Kumari, Agarwal, and Mittal, 2021), analyze product reviews (Doo, Shin, and Park, 2021), etc. Although researchers have utilized sentiment analysis in many studies, few studies have utilized emotional aspects (Siering et al., 2018). Ray et al. (2020a) have noted that combining the sentiment and emotion values can help to get more accurate predictions.

3. Research Methodology

An NLP-based approach is used for classifying the user reviews after training the classifier, because it is the most common method used by several researchers to understand customer perspectives from online textual data (Kunimoto and Saga, 2014). The argument for using NLP is in its ability to perform better linguistic analysis to help the machine understand the human-language better (Basic Semantics, 2016). However, earlier researchers have neither used classifiers nor the emotional content of textual data for predicting ratings. We attempt to utilize a combination of classifiers and textual emotional content since we feel that the emotional aspects, a key criterion to understand human-computer interaction (Cowie et al., 2001), if captured properly, will help the machine to learn and predict more accurately. The classifier is then utilized to predict the ratings from SM comments posted by users about a service. Determining the ratings from SM posts involves various steps, namely, data extraction, data pre-processing, data validation, feature extraction, and data analysis (prediction). These are discussed in the subsequent subsections.

3.1 Data extraction

For training the classifiers, three different types of datasets are used: Coursera 100k dataset (Ref.: Adona, 2019) (e-Learning services), Trip-Advisor dataset containing hotel reviews (Source: Datafiniti, 2019) (OTAs), and Zomato dataset containing ready-made food delivery reviews (Ref.: Mehta, 2019)(OFDs).

19,734 tweets were downloaded (Coursera and Udemy: 10,152 (e-learning); Zomato and Swiggy: 6,996 (OFDs); MakeMyTrip and GoIbibo: 2,586 (OTAs)) using hashtags and the provider's official twitter handles like, '#coursera', '@coursera' etc. R-Studio v1.0.136 was used for deriving the tweets.

3.2 Preprocessing raw data

The various steps involved in pre-processing are as follows: (a) re-tweets were deleted; (b) tweets written in English were only considered; (c) removing punctuations and unnecessary symbols; (d) performing stemming; and (e) tweets with minimum five words were used. Stop words were not deleted because some stop words act as function words (refer Table 3).This cleaning process resulted in 3,509 tweets (e-Learning:1421; OFDs:1,506; and OTAs:582). For reviews, this pre-processing, resulted in 417,136 reviews (e-Learning:86,760; OFDs:2,340; and OTAs:328,036). Out of these reviews, randomly 29,551 reviews were selected for training the classifiers in different contexts (Coursera:10,000; TripAdvisor:10,000; Zomato:9,551). The analysis was performed separately for each context. Each dataset was split in the ratio 3:1 (training:testing)(Schürer and Muskal, 2013), and a five-fold cross-validation was performed on each of the training set for ensuring that the results are more generalizable (Domingos, 2012).

3.3 K-fold cross-validation

It is usually conducted to ensure that there is no over-fitting in the training set. Over-fitting leads to distortion of the test-set results. We have used k=5, where the dataset is divided into five equal random folds. In every step, the four-folds are used for training and one fold for validation. This is performed again and again five times. However, every sub-fold is utilized for validation exactly once (Pennacchiotti and Popescu, 2011). This ensures the model's generalizability (Domingos, 2012; Mukherjee and Bala, 2017).

3.4 Feature extraction

Different features (refer Table 3) were extracted from the reviews and were used as the training dataset. Once the algorithm is trained, the same features are used in the test dataset to find out the performance of the classifier through metrics like accuracy and F-measure. Though feature extraction has not been used extensively in extant literature, authors Mukherjee and Bala (2017) have noted an improvement of performance in the classifier. Feature extraction helps to reduce the number of items/terms needed for describing a dataset (Guyon and Elisseeff, 2003). The major problem while working with a large dataset is that when the number of variables increase, there is an increase in the computation power and a reduction of classifier performance due to issues like over-fitting leading to poor generalizability. Feature extraction can help to get rid of the issues discussed above. Table 3 describes the list of features used in this study for the aid of creating a better classifier. Utilising multiple features or a combination of different features in classifiers the performance measures are compared.

<<Insert Table 3 here>>

3.5 Classification methods used

In natural language processing, classification methods are divided broadly into two categories, namely, generative and discriminative classifiers. A generative classifier learns the joint probability of the inputs (the features) and the outputs (the ratings) and makes the prediction utilising Bayes rule for selecting the closest rating (Yan and Yan, 2006). On the other hand, the discriminative classifiers utilise posterior probability (Ng and Jordan, 2002). Researchers have used both these types in extant literature. In this study we have used linear classifiers (Naïve Bayes classifier (generative) and Maximum Entropy (discriminative)) and a non-linear classifier (k-nearest neighbours). The intention of using classifiers of different types is for comparing between generative and discriminative classifiers and between linear and non-linear classifiers. Although there are many classifiers available, researchers have found Naïve Bayes, kNN and Maximum Entropy algorithms working better than other algorithms like ensemble classifiers (Kumar, and Goyal, 2018; Jitpakdee, and Uyyanonvara, 2017; Prasanna, 2019; Alamri, 2020).

3.5.1 Naïve Bayes classifier

It is mainly used for classifying documents (Argamon et al., 2007). For example, Tweet=(Feature$_1$, Feature$_2$,..., Feature$_n$). Here, 'Feature' can be any of the features mentioned in Table 3, which may be utilized in tweet classification. Thus, for a tweet, we have to determine which category (or class) it is best suited for, 1/2/3/4/5 for [1-5] scale ratings. Using Bayes' theorem, it can be written as:

$$p(Class|Feature_1, Feature_2, \dots, Feature_n) = \frac{p(Class)p((Feature_1, Feature_2, \dots, Feature_n)|\ Class)}{p(Feature_1, Feature_2, \dots, Feature_n)} \quad (1)$$

Naïve Bayes assumes that for a classification task, based on the assumption of independence, the equation can be written as:

$$p(Feature_1, Feature_2, \dots, Feature_n|\ Class) = p(Feature_1|Class) \dots p(Feature_n|\ Class) \quad (2)$$

Earlier researchers have found that this assumption works well (Yan and Yan, 2006). Combining equations (1) and (2), the new equation is:

$$p(Class|Tweet) = \frac{p(Class)p(Feature_1|Class) \dots p(Feature_n|Class)}{p\ (Tweet)} \quad (3)$$

Then the ratios of the posterior probabilities [*p(Class=1|Tweet)*, *p(Class=2|Tweet)*,etc.] are calculated for the various classes in a given document. The tweet is categorised in the class that has greater probability.

3.5.2 Maximum entropy classifier

This classifier assumes that features are not always independent of one another, and hence, is less restrictive unlike Naïve Bayes. Based on the maximum entropy principal, out of all the models used in training, it selects one with the highest entropy. However, the computation time of the maximum entropy classifier is more than that of Naïve Bayes due to optimization issues. A stochastic model is constructed, which represents the process behaviour. The contextual information (say, "c") of a document is taken as input (function words, etc.) and it produces the output value (say, “o”). The first step is to train the model that consists of data in the format (c_i, o_i), where c_i refers to the contextual information and o_i represents the class. In the next step, the training sample is summarised based on its probability distribution. The equation looks like:

$$p(c,o) = \frac{1}{N} \times number\ of\ times\ (c,o) appers\ in\ the\ training\ set \qquad (4)$$

where, N=size of training set.

Equation (4) is utilised to build the statistical model for random process that allocates the texts to the relevant class considering their contextual information. The function used is given below:

$$F_j(c,o) = \begin{cases} 1 & if\ o = L_i\ and\ c\ contains\ W_i \\ 0 & otherwise \end{cases} \qquad (5)$$

where, F_j=feature function. The output is ‘1’ when the function class (L_i) and document (c) contains the word W_i. The expected value of F_j for the distribution $p(c, o)$ is given below:

$$p(F_j) = \sum_{c,o} p(c) p\left(\frac{o}{c}\right) F_j(c,o) \qquad (6)$$

Using the principle of max-entropy, the model is given below:

$$p_{max} = \arg\max_{p \in L}\left(-\sum_{c,o} p(c) p\left(\frac{o}{c}\right) \log p\left(\frac{o}{c}\right)\right) \qquad (7)$$

Subject to constraints:

$$\sum_{c,o} p(c,o) F_j(c,o) = \sum_{c,o} p(c) p\left(\frac{c}{o}\right) F_j(c,o) \qquad (8)$$

. 3.5.3. k-nearest neighbours (k-NN) classifier

k-NN is a supervised machine learning algorithm for classification based on similar neighbor’s behaviour (Cunningham and Delany, 2007). The optimal value of “k” is data-specific. In general, researchers feel that a large value of ‘k’ is better since it reduces the overall noise, but in certain cases it can be the other way around. A good value of k ranges between ‘3-10’ (Cunningham and Delany, 2007). For classifying a case, a majority of votes of the neighbours is needed and the case gets classified to the class, which is the most common one among the k-nearest neighbours. The nearest neighbours are calculated using the distance function.

3.6. Calculating sentiment and emotional scores

While the sentiment score depicts the overall polarity of the sentence, the scores for the different emotions, like, anger, fear, surprise, trust, etc. provide an in-depth view of the user's feelings. The R-libraries "sentimentr" and "syuzhet" are used for finding the sentiment and the emotion values. "sentimentr" and "syuzhet" package is used for sentiment and emotion analysis because these packages takes into account not only the positive and negative words but also checks for valence shifters (negators, amplifiers, de-amplifiers, adversative conjunctions, inversions) (Sentimentr Document, 2019). While for calculating the sentiment score, Jocker's (2017a) dictionary is used, for finding the emotion involved, emotion lexicons are used (NRC Emotion library, n.d.; Mohammad & Turney, 2013; Jocker, 2017b). We have followed the steps as mentioned by Ray et al. (2020a) for calculating the sentiment and emotion values for each review/post.

4. Results

In this section, we work on training and classification of the SM tweets into the various classes ([1 to 5] or [1 to 10] based on the training set context). The classifier performance is examined using different metrics like, precision, recall, accuracy, and F-Measure (refer Ray et al., 2020a). Let us now look at the performance of the different classifiers using different combination of features and opinion scores (sentiment and emotion scores) context wise.

We look at the performance of the linear (generative and non-linear classifiers in the contexts under study (e-Learning services, OFDs, and OTAs), using different features and in different combinations (without use of sentiment or emotion values, using only sentiment scores, and a combination of sentiment and emotion values). In context of OFDs, even when we don't use sentiment or emotional scores to classify the textual data, the classifier has an accuracy of 100% while classifying the texts (refer Figure 1). This is because in context of OFDs, the reviews ratings are precise. For example, users who have given a review “Excellent” have rated 4.8, “very good” as 4, “average” as 3. This has helped the classifier to learn better and predict more accurately.

<<Insert Figure 1 here>>

But usually the cases are not that simple. In case of e-Learning and OTAs, users have given lengthy reviews and ratings as well. But the main issue is that sometimes when the sentiment and emotion scores suggest that the rating must be 3, the users have rated as 2 or 3 or 4. Let us check the classifier performance in case of e-Learning and OTAs separately. Figure 2 and Figure 3 show the performance of the various classifiers using different features without using sentiment or emotions in context of e-Learning and OTAs respectively. For e-Learning context (see Figure 2), kNN algorithm has achieved higher performance (average accuracy score=0.6690) results as compared to the linear classifiers. The highest performance is achieved (accuracy=0.691) when we use a combination of content and function words in case of kNNs. In case of Naïve Bayes and Max-entropy classifiers, the highest performance is achieved when utilising content words alone ($\text{accuracy}_{\text{Naive-Bayes}}$=0.5983, $\text{accuracy}_{\text{Max-entropy}}$=0.5984) closely followed by the use of content and function words ($\text{accuracy}_{\text{Naive-Bayes}}$=0.5823, $\text{accuracy}_{\text{Max-entropy}}$=0.5823).

In context of OTAs, the highest performance is achieved by Naïve Bayes and Max-entropy classifiers using parts-of-speech tags (accuracy in each case=0.7867). But when the average performance of the classifiers using different features is considered, the Naïve Bayes classifier (accuracy=0.7284) has performed better than the non-linear classifier kNN (accuracy=0.7143) and the Max-entropy classifiers (accuracy=0.6275). The best performance is achieved when the combination of POS tags, content words and function words is used.

<<Insert Figure 2 here>>

<<Insert Figure 3 here>>

Figure 4 and Figure 5 show the performance of the various classifiers using different features in combination with sentiment scores in context of e-Learning services and OTAs respectively. In this case, a slight increase in classifier performance is noticed. The Naïve Bayes classifier performs best when used with the feature n-grams (accuracy=0.796) in context of e-Learning services (see Figure 4). The second best performance is observed in kNN classifier when we use the features content and function words together (accuracy = 0.692). However, overall the Naïve Bayes (average accuracy=0.6788) and kNN (average accuracy=0.6712) classifiers outperform the max-entropy classifier (average accuracy=0.5749). The results are a bit different in context of OTAs (see Figure 5). The kNN classifier has clearly performed better than the other classifiers. The average of accuracy for kNN is 0.7013 while for Naïve Bayes and Max-entropy its 0.5646 and 0.5589 respectively.

<<Insert Figure 4 here>>

<<Insert Figure 5 here>>

When both sentiment and emotion scores are used for determining the ratings using different classifiers (see Figure 6 and Figure 7), in case of e-Learning, the performance of the classifiers is almost similar (average $\text{accuracy}_{\text{Naive-Bayes}}$=0.6841; average $\text{accuracy}_{\text{Max-entropy}}$=0.6344; average $\text{accuracy}_{\text{kNN}}$=0.6629) (see Figure 6). However, in context of OTAs, the kNN classifier outperformed the other two classifiers (average $\text{accuracy}_{\text{Naive-Bayes}}$=0.8374; average $\text{accuracy}_{\text{Max-entropy}}$=0.7203; average $\text{accuracy}_{\text{kNN}}$=0.6918) (see Figure 7).

The average performance scores are calculated by taking the average of the performance measures using different features in diverse contexts. For example, to calculate the average of the accuracy for kNN classifier in Figure 6, we used the below mentioned formula:

$$\text{Average Accuracy}_{kNN} = (\text{Accuracy}_{kNN\_using\_Content+Function\_words}+\text{Accuracy}_{kNN\_using\_ngrams}+\text{Accuracy}_{kNN\_using\_POS}+\text{Accuracy}_{kNN\_using\_Function\_words}+\text{Accuracy}_{kNN\_using\_Content\_words}+\text{Accuracy}_{kNN\_using\_all\_words})$$

<<Insert Figure 6 here>>

<<Insert Figure 7 here>>

Table 4 and Table 5 shows the classifier performance (accuracy measures) using the various combinations of sentiment, emotion and classifiers in context of e-Learning and OTAs. Results of the analysis in both cases show that there is a clear improvement of classifier performance when we use a combination of different features along with sentiment and emotion values while using Naïve Bayes and Max-entropy. However, for kNN classifier, we see a slight decrease in the performance. Another observation is that in both the cases, the overall performance (in terms of accuracy) is by using Naïve Bayes in combination with sentiment and emotion scores followed by Max-entropy and kNN classifiers.

<<Insert Table 4 here>>

<<Insert Table 5 here>>

We now introduce the use of factors identified by consumers as a feature in case of e-Learning and OTAs. We argue that utilising the factors previously identified by the consumers will help in reducing the computation time as well as improve classifier performance since using the "factors-of-importance" will help to compute the results using a focused set of words that the customers value in different contexts. In this work, we have used the most commonly used factors by earlier researchers in context of different e-services. Most commonly used factors in case of e-learning are 'course', 'great', 'like', 'content', 'concept', 'enjoy', 'good', 'love', and 'interest'. In context of OTAs, the factors are ‘make’, ‘great’, ‘love’, ‘price’, ‘nice’, ‘good’, ‘location’, ‘like’, ‘perfect’, and ‘cleanliness’. It is to be noted that we have used some of the most frequently used words, since these words can help to find the relevant ratings for the posts. Findings reveal that overall performance of the Naïve Bayes classifier (average accuracy=0.7038) is better than the kNN (average accuracy=0.6711) and Max-entropy (average accuracy=0.6787) classifiers (see Figure 8). In context of OTAs, the average performance (accuracy) of the max-entropy algorithm (average accuracy=0.7016) is better than Naïve Bayes (average accuracy=0.6659) and kNN classifiers (average accuracy=0.6792) (refer Figure 9). We now check the precision, recall and f-measures of the linear (Naïve Bayes) and non-linear (kNN) classifiers using the ‘factors-of-importance’ (see Table 6). Results show that classifier using the “factors-of-importance” in context of e-Learning and OTAs shows good performance in terms of precision, recall, and F-measures.

<<Insert Figure 8 here>>

<<Insert Figure 9 here>>

<<Insert Table 6 here>>

Results indicate that by using different features, different classifiers and using the sentiment and emotional scores in different combinations, we can classify and predict a SM feed into the relevant ratings. The best results are obtained using the features parts-of-speech (POS) tags, ‘factors-of-importance’, ngrams, in combination of content and function words, and content words. We have also observed that combining sentiment and emotion values improved the classifier performance and the non-linear classifier (kNN classifier) performed better than the linear classifiers (Naïve Bayes and Max-entropy).

6. Discussion

The primary objective of this study is to improve the prediction accuracy of the user’s intended rating from their SM textual comments. Although researchers have used different techniques for rating prediction like, using relevance score (Wakamiya et al., 2011), combination of sentiment and emotional scores (Ray et al., 2020a), etc., no studies have utilised a combination of classifiers (linear or non-linear), sentiment and emotional scores to predict ratings. Although Ray et al. (2020a) have noted an improvement in the accuracy of rating prediction when the sentiment and emotion values were used; this study extends the work by examining classifier performance when combined with sentiment score, emotion score and different features. For accomplishing the research objectives, we have utilised three classifiers namely Naïve Bayes, Maximum Entropy and k-nearest neighbour (kNN) classifier. Additionally, we have combined these classifiers with the sentiment and emotion scores and different features (content words, function words, n-grams, POS tags) to compare their performance. Earlier studies have usually used the entire bag-of-words and have not attempted at utilising "factors-of-importance" generated from earlier studies as a feature to reduce space and time complexity of different classifiers. The proposed feature "factors-of-importance" have improved the classifier performance in terms of performance and the time needed for training and classifying the reviews. Although utilising the feature 'factors-of-importance' may not reveal improvement in overall accuracy or F1 score, it will help to reduce the space and time complexities by reducing the number of words/terms being searched. The best results are obtained using the features POS tags, 'factors-of-importance', ngrams, the features content and function words, and content words. Additionally, findings also showed that the non-linear classifier (kNN classifier) performed better than the linear classifiers (Naïve Bayes and Max-entropy).

Findings reveal that SM comments about a product/service can be accurately rated based on our proposed approach. The combination of linear classifiers with sentiment and emotional scores has seen an increase in accuracy. Additionally, the combination of "factors-of-importance" and textual emotional content has seen a slight improvement in the accuracy and performance (space and time complexity) of the classifiers for determining ratings from the SM posts. This helps us to create a dataset containing user details, the product details, the customer comments and ratings from SM comments which will be helpful in creating recent vast datasets for training RSs (Dooms et al., 2014). This current research work is able to address the research gaps.

6.1 Contributions to theory

This work has four theoretical contributions. First, this study provides a new technique to understand the underlying customer perspectives from SM feeds better in context of e-Learning, OFDs and OTAs. This work will help future scholars trying to understand customer perspectives from online textual data, by providing them an easy approach to find the overall intended ratings. The use of sentiment and emotion scores in rating calculation will make it easier for researchers working on utilising valence scores for exploring factors affecting consumer behaviour in these contexts. The use of sentiment and emotion values in combination with classifiers for rating calculation enhances the classifier performance and also provides researchers a new approach for analysing textual data.

Second, determining ratings will help academicians in different research scenarios, like, recommending product/service, determining factors affecting usage behaviour, etc. Earlier researchers have noted that limited number of datasets used for RS contains old data, which may not be so useful while training models for handling recent marketing problems (Dooms et al., 2013, 2014). Determining ratings from SM comments will help to build datasets for training RSs using recent data. The available ratings from SM feeds will also help in understanding the user views about a product/service easily.

Third, usage of ‘factors-of-importance’ as features for classifying the textual data for predicting ratings is a novel approach. It will help future researchers working on textual data to perform an easy, faster and effective classification of the textual data for solving various related research problems. This approach also contributes to the classification algorithm literature by bringing a new dimension to improve the process of text classification.

Finally, this study helps to gain an understanding of the importance of customer sentiments and emotions in effective classification of textual data for determining the ratings a customer intends to give based on their SM post, thus adding a new dimension to textual classification literature.

6.2 Implications for practice

The practical implications are presented here. First, determining ratings from SM posts is important for providers because it can help them to form better strategies for building better customer relationships. This work will be helpful for classifying the SM posts containing poor ratings about the product/service from posts that contain good feedback from the customers, i.e., the highly rated feeds. In the present scenario, understanding customer perspectives by reading the customer posts takes a lot for time for customer service executives. Since, recruiting additional executives for chalking out bad-comments from good ones will increase the costs involved service providers. However, by utilising our proposed technique for determining ratings from SM feeds will help to save time, effort and money. Second, providers that rely on RSs will be benefitted because they can now utilise more recent huge datasets by determining the ratings from SM posts which can be used to train the RS algorithms. Since utilising recent data can help to develop a better understanding of the recent trends in user needs, RS algorithms will be able to provide better recommendations.

Third, this work will be beneficial for service providers in that if the overall ratings from all platforms are available regarding a particular product/service, it will help service providers to easily understand how their product/service is performing. This will help them to take corrective measures if they find the overall rating to be on the lower side. This study will also indirectly help prospective customers to understand how previous customers have expressed their views about a given service without having to take the pain to go through lengthy SM posts.

6.3 Limitations and future research directions

There are a few limitations in this study. First, we have used one type of classifier from the different types of classifiers, namely, linear classifiers (Naïve Bayes classifier (generative) and Maximum Entropy (discriminative)) and a non-linear classifier (k-nearest neighbours). This study has not explored the performance of advanced machine learning algorithms like support vector machines (SVMs), neural networks, or AdaBoosting. We intend to explore the performance of the feature 'factors-of-importance' and performance of different advanced machine learning algorithms in classification of SM posts in future. Second, we have not considered the personality traits of users while determining the ratings from their SM feeds. Understanding customer personality traits along with emotion and sentiment scores will help to provide a different view of how the user reacted in that particular scenario (Srivastava et al., 2021). Suppose a person who is usually silent reacts aggressively in one post, it must be that he/she is really disappointed with the service/product.

Third, this study has not captured the demographic data of the users. Capturing the demographic data will help to understand the location from where users have been happy or dejected regarding a particular service/product. Fourth, this study has not captured the time of SM posts. Capturing the time of the SM posts can help managers to understand the different scenarios regarding the various posts. In future, researchers can work on extracting the timestamp, location etc. from SM posts and improve the prediction accuracy by taking into consideration several factors such as political situation, predominant traditions in the locality, and urban/rural area. This work can also be extended to find out how people of different regions rate a product or service and which factors influence their decision.

7. Conclusions

Understanding user perspectives from a social-media post is an important task for organizations. The reviews and ratings are also important for new prospects because these help them to decide whether to use the product/service or not. However from social-media posts it is difficult to understand the ratings. Thus, accurately determining the ratings from SM feeds becomes an important research area. An NLP-based approach using both generative (Naïve Bayes) and discriminative (Maximum Entropy) classifiers are used to determine the ratings of SM posts based on various features, sentiment, and emotion scores of users. Results reveal that the non-linear classifier (kNN classifier) performs better than the linear classifiers (Naïve Bayes, Max-entropy classifier). Results show an improvement of classifier performance when the sentiment and emotion values are combined. Introduction of the ‘factors-of-importance’ feature also shows an improvement of the classifier performance. The uniqueness of this research lies in the fact that this research has utilised emotion and sentiment scores in combination with classifiers to predict the ratings of online textual data posted in SM platform (in our case Twitter). The use of such combination enhances the performance of the model. The major contributions of this research are: First, this study helps to create large datasets for building recommendation systems from the most recent data; Second, predicting ratings from social-media comments helps to separate the bad reviews from good ones easily. Third, the use of ‘factors-of-importance’ as features adds a new avenue for research on classifiers.

References

Abualigah, L., Alfar, H.E., Shehab, M., and Hussein, A.M.A. (2020). *"Sentiment analysis in healthcare: a brief review."* Recent Advances in NLP: The Case of Arabic Language, 129-141. Springer.

Adona, J.C.M (2019). *“100K Coursera's Course Reviews Dataset: 100K+ Scraped Course Reviews from the Coursera Website (As of May 2017)”*. Kaggle. [online]https://www.kaggle.com/septa97/100k-courseras-course-reviews-dataset (accessed June 2019)

Akarsu, C., and Diri, B. (2016). *"Turkish TV rating prediction with Twitter."* 2016 24th Signal Processing and Communication Application Conference (SIU). doi:10.1109/siu.2016.7495748

Alamri, H.L., Almuslim, R.S., Alotibi, M.S., Alkadi, D.K., Khan, U.I., and Aslam, N. (2020). *“Predicting Student Academic Performance using Support Vector Machine and Random Forest.”* 2020 3rd International Conference on Education Technology Management. doi:10.1145/3446590.3446607

Alsmearat, K., Shehab, M., Al-Ayyoub, M., Al-Shalabi, R., and Kanaan, G. (2015). *"Emotion analysis of Arabic articles and its impact on identifying the author’s gender."* 2015 IEEE/ACS 12th International Conference of Computer Systems and Applications (AICCSA). doi:10.1109/aiccsa.2015.7507196

Aral, S. (2014). "The Problem with Online Ratings." *MITSloan Management Review*. Vol. 55, No. 2, pp.46-52. [online]https://ruralengagement.org/wp-content/uploads/2015/08/reviews.pdf

Argamon, S., Koppel, M., Fine, J., and Shimoni, A.R. (2003). “Gender, genre, and writing style in formal written texts”, *Text – Interdisciplinary Journal for the Study of Discourse,* Vol. 23, No. 3, pp.321-346. doi:https://doi.org/10.1515/text.2003.014

Basic Semantics (2016). *“Natural Language Processing and Text Mining.”* Expert Systems. Accessed on 25 January 2020. [online] https://expertsystem.com/natural-language-processing-and-text-mining.

Black, H.G., and Kelley, S.W. (2009). “A storytelling perspective on online customer reviews reporting service failure and recovery.” *Journal of Travel & Tourism Marketing*, Vol. 26, No. 2, pp.169–179. doi:10.1080/10548400902864768

Chaffey, D. (2019). *Global social media research summary 2019*. [online] https://www.smartinsights.com/social media-marketing/social media-strategy/new-global-social media-research (accessed 20 December 2020)

Chambua, J., and Niu, Z. (2021). "Review text based rating prediction approaches: preference knowledge learning, representation and utilization." *Artificial Intelligence Review*. Vol. 54, pp.1171–1200 . https://doi.org/10.1007/s10462-020-09873-y

Chambua, J., Niu, Z., Yousif, A., & Mbelwa, J. (2018). Tensor Factorization Method based on Review Text Semantic Similarity for Rating Prediction. *Expert Systems with Applications*. Vol. 114, pp.629-638. doi:10.1016/j.eswa.2018.07.059

Chatterjee, S. (2019). Explaining customer ratings and recommendations by combining qualitative and quantitative user generated contents. Decision Support Systems. Vol. 119, pp.14-22. doi:10.1016/j.dss.2019.02.008.

Chen, R., and Xu, W. (2016). "The determinants of online customer ratings: a combined domain ontology and topic text analytics approach." *Electronic Commerce Research*, Vol. 17, No. 1, pp.31–50. doi:10.1007/s10660-016-9243-6

Cheng, M-H., Wu, Y-C., and Chen, M-C. (2016). “Television Meets Facebook: The Correlation between TV Ratings and Social Media.” *American Journal of Industrial and Business Management*, Vol. 6, No. 3, pp.282-290. doi:10.4236/ajibm.2016.63026

Church, K.W. (1989). *“A stochastic parts program and noun phrase parser for unrestricted text”*, International Conference on Acoustics, Speech, and Signal Processing, pp.136-143. doi: 10.1109/ICASSP.1989.266522

Cohen-Kalaf, M., Lanir, J., Bak, P., and Mokryn, O. (2021). *"Movie emotion map: an interactive tool for exploring movies according to their emotional signature."* Multimedia Tools and Applications, pp.1-22. Springer.

Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., and Kollias, S. (2001). “Emotion recognition in human-computer interaction,” *IEEE Signal Processing Magazine*, Vol. 18, No. 1, pp.32-80. doi:10.1109/79.911197

Cui, G., Lui, H. K., and Guo, X. (2012). “The effect of online consumer reviews on new product sales.” *International Journal of Electronic Commerce*, Vol. 17, No. 1, pp.39–57. doi:10.2753/JEC1086-4415170102

Cunningham, P., and Delany, S.J. (2007). *"k-Nearest Neighbour Classifiers."* [online] https://www.researchgate.net/profile/Sarah_Delany/publication/228686398_k-Nearest_neighbour_classifiers/links/0fcfd50d0c1d1f41ad000000/k-Nearest-neighbour-classifiers.pdf (accessed 20 December 2020)

Datafiniti (2019). *“Hotel Reviews: A list of 1,000 hotels and their online reviews.”* Kaggle. [online]https://www.kaggle.com/datafiniti/hotel-reviews (accessed 15 June 2019)

Domingos, P. (2012). “A few useful things to know about machine learning”, *Communications of the ACM*, Vol. 55, No. 10, pp.78-87. doi:10.1145/2347736.2347755

Dondokova, A., Aich, S., Kim, H.C., and Huh, G.H. (2019). *“A Text Mining Approach to Study Individuals’ Food Choices and Eating Behavior Using Twitter Feeds.”* In: Hung J., Yen N., Hui L. (eds) Frontier Computing. FC 2018. Lecture Notes in Electrical Engineering, vol. 542. Springer, Singapore. doi:10.1007/978-981-13-3648-5_60

Doo, I.C., Shin, H.D., and Park, M.H. (2021). "Automated product review collection and opinion analysis methods for efficient business analysis." *International Journal of Computing and Digital Systems*, Vol. 10 No. 1, pp.37-45.

Dooms, S., De Pessemier, T., and Martens, L. (2013). *Movie Tweetings: A Movie Rating Dataset Collected From Twitter*. ResearchGate.net.

Dooms, S., De Pessemier, T., and Martens, L. (2014). *“Mining cross-domain rating datasets from structured data on Twitter.”* Proceedings of the 23rd International Conference on World Wide Web - WWW ’14 Companion. pp.621-624. doi:10.1145/2567948.2579232

D’Souza, A., Bathla, R., and Giri, N. (2013). *"Twitter based TV rating system."* Third International Conference on Computational Intelligence and Information Technology (CIIT 2013). doi:10.1049/cp.2013.2586

Duan, W., Gu, B., and Whinston, A.B. (2008). "Do online reviews matter? — An empirical investigation of panel data." *Decision Support Systems*, Vol. 45, No. 4, pp.1007–1016. doi:10.1016/j.dss.2008.04.001

Engler, T.H., Winter, P., and Schulz, M. (2015). "Understanding online product ratings: A customer satisfaction model." *Journal of Retailing and Consumer Services*, Vol. 27, pp.113–120. doi:10.1016/j.jretconser.2015.07.010

Gavilan, D., Avello, M., and Martinez-Navarro, G. (2018). “The influence of online ratings and reviews on hotel booking consideration.” *Tourism Management*, Vol. 66, pp.53–61. doi:10.1016/j.tourman.2017.10.018

Geetha, M., Singha, P., and Sinha, S. (2017). "Relationship between customer sentiment and online customer ratings for hotels - An empirical analysis." *Tourism Management*, Vol. 61, pp.43–54. doi:10.1016/j.tourman.2016.12.022

Gu, Y., Yang, X., Peng, M., & Lin, G. (2020). Robust weighted SVD-type latent factor models for rating prediction. *Expert Systems with Applications*, Vol. 141, pp.112885. doi:10.1016/j.eswa.2019.112885

Guyon, I., and Elisseeff, A. (2003). “An introduction to variable and feature selection”, *The Journal of Machine Learning Research*, Vol. 3, pp.1157-1182.

Haque, R., Hasanuzzaman, M., and Way, A. (2019). “Mining Purchase Intent in Twitter.” *Computación y Sistemas*, Vol. 23, No. 3, pp.871–881. doi:10.13053/CyS-23-3-3254

Jitpakdee, P., and Uyyanonvara, B. (2017). “Computer-aided detection and quantification in glistenings on intra-ocular lenses”. *Multimedia Tools and Applications*, Vol. 76 No. 18, pp.18915-18928.

Jockers, M. (2017a). “*Syuzhet: Extract sentiment and plot arcs from text*”. [online]https://github.com/mjockers/syuzhet (accessed 10 September 2020)

Jockers, M. (2017b). *“Introduction to the Syuzhet Package”*. [online]https://cran.r-project.org/web/packages/syuzhet/vignettes/syuzhet-vignette.html (accessed 10 September 2020)

Kayahan, D., Sergin, A., and Diri, B. (2013). *“Determination of TV programme ratings by Twitter”*. 2013 21st Signal Processing and Communications Applications Conference (SIU). doi:10.1109/SIU.2013.6531205

Kesharwani, A., and Bharti, R. (2017). “Movie rating prediction based on twitter sentiment analysis.” *Journal of Advanced Computing and Communication Technologies*. Vol. 5, No. 1, pp.6-10.

Khan, Z.Y., & Niu, Z. (2021). CNN with depthwise separable convolutions and combined kernels for rating prediction. *Expert Systems with Applications*, Vol. 170, pp.114528. doi:10.1016/j.eswa.2020.114528

Klammer, T., Schulz, M., and Della V. A. (2000). *“Analyzing English Grammar”*, 6/e, Pearson Education.

Koppel, M. (2002). “Automatically categorizing written texts by author gender”, *Literary and Linguistic Computing*, Vol. 17, No. 4, pp.401-412. doi:10.1093/llc/17.4.401

Kostyra, D.S., Reiner, J., Natter, M., and Klapper, D. (2016). “Decomposing the effects of online customer reviews on brand, price, and product attributes.” *International Journal of Research in Marketing*, Vol. 33, No. 1, pp.11–26. doi:10.1016/j.ijresmar.2014.12.004

Kumari, S., Agarwal, B., and Mittal, M. (2021). "A Deep Neural Network Model for Cross-Domain Sentiment Analysis." *International Journal of Information System Modeling and Design (IJISMD)*, Vol. 12 No. 2, pp.1-16.

Kumar, A., and Goyal, D. (2018). “Performance evaluation of naïve bayes and maximum entropy model through opinion mining”. *International Journal of Engineering, Science and Mathematics*, Vol. 7 No. 5, pp.207-212.

Kunimoto, R., and Saga, R. (2014). "Purchase Factor Expression for Game Software Using Structural Equation Modeling with Topic Model in User's Review Texts", *International Journal of Innovation, Management and Technology*, Vol. 5, No. 6, pp.417-421.

Lerman, K., and Ghosh, R. (2010). *"Information Contagion: An Empirical Study of the Spread of News on Digg and Twitter Social Networks"*. Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media, (pp.90–97).

Li, K., Chen, Y., & Zhang, L. (2020). Exploring the influence of online reviews and motivating factors on sales: A meta-analytic study and the moderating role of product category. *Journal of Retailing and Consumer Services*. Vol. 55, pp.102-107.

Mahadevan, A., and Arock, M. (2020). "Review rating prediction using combined latent topics and associated sentiments: an empirical review." *Service Oriented Computing and Applications*. Vol. 14, pp.19–34. https://doi.org/10.1007/s11761-019-00278-6

Mehta, S. (2019). *"Zomato Restaurants Data: Analyzing the best restaurants of the major cities."* Kaggle. [online] https://www.kaggle.com/shrutimehta/zomato-restaurants-data/downloads.

Mhaisgawali, A., and Giri, N. (2014). "Detailed Descriptive and Predictive Analytics with Twitter Based TV Ratings". *International Journal of Computing and Technology*, Vol. 1, No. 4, pp.125-130.

Mohammad, S.M. & Turney, P.D. (2013). *"Crowdsourcing a Word–Emotion Association Lexicon"*. [online]https://arxiv.org/pdf/1308.6297.pdf. (accessed 26 January 2020)

Mosley, M. (2018). *"Why Is Social Media Important for Businesses?"* [online]https://www.business2community.com/social media/why-is-social media-important-for-businesses-02142048 (accessed 26 January 2020)

Mostafa, M.M. (2018). "Clustering halal food consumers: A Twitter sentiment analysis." *International Journal of Market Research*. Vol. 61, No. 3, pp.1-18. doi:10.1177/1470785318771451

Mukherjee, S., and Bala, P.K. (2017). "Detecting sarcasm in customer tweets: an NLP based approach", *Industrial Management & Data Systems*, Vol. 117, No. 6, pp.1109-1126, doi:10.1108/IMDS-06-2016-0207Naqvi, U., Majid, A., and Abbas, S.A. (2021). "UTSA: Urdu text sentiment analysis using deep learning methods." *IEEE Access*, Vol. 9, pp.114085-114094.

Ng, A.Y., and Jordan, M.I. (2002). "On discriminative vs. generative classifiers: a comparison of logistic regression and Naive Bayes", *Advances in Neural Information Processing Systems*, Vol. 2, pp.841-848.

NRC Emotion library (n.d.). *"NRC Word-Emotion Association Lexicon"*. [online]http://saifmohammad.com/WebPages/NRC-Emotion-Lexicon.htm (accessed 10 September 2020)

Oheix, J. (2018). *"Detecting bad customer reviews with NLP"*. [online]https://towardsdatascience.com/detecting-bad-customer-reviews-with-nlp-d8b36134dc7e (Accessed 26 January 2020)

Patel, N. (2016). *"How to Make Visitors Read Your Entire Article"*. [online]https://contentmarketinginstitute.com/2016/01/visitors-read-article (accessed 26 January 2020)

Pennacchiotti, M. and Popescu, A.-M. (2011). "A machine learning approach to Twitter user classification", *ICWSM*, Vol. 11, No. 1, pp.281-288.

Prasetyo, V. R., and Winarko, E. (2016). *"Rating of Indonesian Sinetron based on public opinion in Twitter using Cosine similarity."* 2016 2nd International Conference on Science and Technology-Computer (ICST). doi:10.1109/icstc.2016.7877374

Prasanna, V.L., Chowdary, E.D., Sistla, V., and Kishore, K.V.K. (2019). "An experimental evaluation on utility prediction of integrated child development services". *Advances in Modelling and Analysis B*, Vol. 62, No. 1, pp. 24-30. https://doi.org/10.18280/ama_b.620104

Ray, A., Bala, P.K., and Dasgupta, S. (2019a). "Role of authenticity and perceived benefits of online courses on technology based career choice in India: A modified technology adoption model based on career theory," *International Journal of Information Management*, Vol. 47, pp.140-151. doi:10.1016/j.ijinfomgt.2019.01.015

Ray, A., Dhir, A., Bala, P. K., and Kaur, P. (2019b). "Why do people use food delivery apps (FDA)? A uses and gratification theory perspective." *Journal of Retailing and Consumer Services*, Vol. 51, pp. 221–230. doi:10.1016/j.jretconser.2019.05.025

Ray, A., Bala, P.K. and Jain, R. (2020a), "Utilizing emotion scores for improving classifier performance for predicting customer's intended ratings from social media posts", *Benchmarking: An International Journal*, Vol. 28 No. 2, pp.438-464. doi:https://doi.org/10.1108/BIJ-01-2020-0004

Ray, A., Bala, P.K., and Dwivedi, Y.K. (2020b), "Exploring values affecting e-Learning adoption from the user-generated-content: A consumption-value-theory perspective," *Journal of Strategic Marketing*, Vol. 29, No. 5, pp.430-452. doi:https://doi.org/10.1080/0965254X.2020.1749875

Razis, G., and Anagnostopoulos, I. (2014). *"InfluenceTracker: Rating the Impact of a Twitter Account."* In: Iliadis L., Maglogiannis I., Papadopoulos H., Sioutas S., Makris C. (eds) Artificial Intelligence Applications and Innovations. IFIP Advances in Information and Communication Technology, vol. 437. Springer, Berlin, Heidelberg. doi:10.1007/978-3-662-44722-2_20

Read, A. (2016). *"55% of Visitors Read Your Articles For 15 Seconds or Less: Why We Should Focus on Attention Not Clicks."* [online]https://buffer.com/resources/55-visitors-read-articles-15-seconds-less-focus-attention-not-clicks (accessed 26 January 2020)

Rentpath (2017). *The importance of online ratings and reviews*. [online]https://www.rentpath.com/blog/the-importance-of-online-ratings-and-reviews (accessed 26 January 2020)

Saleh, K. (2015). *"The Importance of Online Customer Reviews."* [online]https://www.invespcro.com/blog/the-importance-of-online-customer-reviews-infographic (accessed 26 January 2020)

Schmit, W., and Wubben, S. (2015). *"Predicting Ratings for New Movie Releases from Twitter Content."* Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015), pp.122-126.

Schürer, S.C. and Muskal, S.M. (2013). "Kinome-wide activity modeling from diverse public high-quality data sets", *Journal of Chemical Information and Modeling*, Vol. 53, No. 1, pp.27-38.

Sentimentr_Document (2019). *Package 'sentimentr'*. [online]https://cran.r-project.org/web/packages/sentimentr/sentimentr.pdf (accessed 10 September 2020)

Siering, M., Deokar, A.V., and Janze, C. (2018). "Disentangling consumer recommendations: Explaining and predicting airline recommendations based on online reviews." *Decision Support Systems*, Vol. 107, pp.52-63. doi:10.1016/j.dss.2018.01.002

Siersdorfer, S., Chelaru, S., Nejdl, W., and San Pedro, J. (2010). *“How useful are your comments?”* Proceedings of the 19th International Conference on World Wide Web - WWW ’10. doi:10.1145/1772690.1772781

Simon, T., Goldbergb, A., and Adinia, B. (2015). “Socializing in emergencies—A review of the use of social media in emergency situations.” *International Journal of Information Management*, Vol. 35, No. 5, pp.609–619. doi:10.1016/j.ijinfomgt.2015.07.001

Sparks, B.A., and Browning, V. (2011). "The impact of online reviews on hotel booking intentions and perception of trust." *Tourism Management*, Vol. 32, No. 6, pp.1310–1323. doi:10.1016/j.tourman.2010.12.011

Srivastava, A., Dasgupta, S.A., Ray, A., Bala, P.K. and Chakraborty, S. (2021), "Relationships between the “Big Five” personality types and consumer attitudes in Indian students toward augmented reality advertising", *Aslib Journal of Information Management*, Vol. 73 No. 6, pp.967-991. doi:https://doi.org/10.1108/AJIM-02-2021-0046

Subroto, A., and Christianis, M. (2021). "Rating prediction of peer-to-peer accommodation through attributes and topics from customer review." *Journal of Big Data*. Vol. 8, e.9 . https://doi.org/10.1186/s40537-020-00395-6

Thelwall, M., Buckley, K., and Paltoglou, G. (2011). “Sentiment in Twitter events”, *Journal of the American Society for Information Science and Technology*, Vol. 62, No. 2, pp.406-418.

Turney, P.D. (2002). *“Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews”*, Proceedings of the 40th Annual Meeting on Association for Computational Linguistics, pp.417-424.

Valence-shifters (2019). “*hash_valence_shifters: Valence Shifters*”. [online]https://rdrr.io/cran/lexicon/man/hash_valence_shifters.html (accessed 10 September 2020)

Wakamiya, S., Lee, R., and Sumiya, K. (2011). *"Towards better TV viewing rates."* Proceedings of the 5th International Conference on Ubiquitous Information Management and Communication - ICUIMC ’11. doi:10.1145/1968613.1968661

Willas, S. (2018). “*7 Reasons Online Reviews are Essential for Your Brand.”* [online]https://mention.com/en/blog/online-reviews (accessed 26 January 2020)

Winkler, E. (2012). “*A Basic Course in Linguistics”*, Continuum International Publishing Group, London.

Yan, X., and Yan, L. (2006), “*Gender classification of weblog authors”*, AAAI Spring Symposium Series on computational Approaches to Analysing Weblogs, pp.228-230.

Zhou, J., Liu, F., and Zhou, H. (2018). “Understanding health food messages on Twitter for health literacy promotion.” *Perspectives in Public Health*, Vol. 138, No. 3, pp.173–179. doi:10.1177/1757913918760359

Table 1. The relevant themes which guide this study are discussed

Theme	Example Sources
Online reviews and ratings	Duan, Gu, and Whinston (2008); Saleh (2015); Engler, Winter, and Schulz (2015); Gavilan, Avello, and Martinez-Navarro (2018); Chatterjee (2019)
Social media feeds and rating prediction	Wakamiya, Lee, and Sumiya (2011); D'Souza, Bathla, and Giri, (2013); Akarsu and Diri (2016); Prasetyo and Winarko (2016); Dooms, De Pessemier, and Martens (2013, 2014); Kesharwani and Bharti (2017)
Sentiment and emotion analysis	Turney (2002); Alsmearat et al., (2015); Siering, Deokar and Janze (2018)
Linear and Non-linear classifiers	Argamon, Koppel, Fine, and Shimoni (2007); Cunningham and Delany (2007); Yan and Yan (2006); Ng and Jordan (2002)

Table 2. General literature review on the study of predicting ratings from social media data

Social media Channel	**Author(s)**	**Findings**
Twitter	Wakamiya et al. (2011)	Improvement was noted in the outcome of relevance computation.
	Dooms et al. (2013)	Dataset containing recent ratings was developed. But dataset was 99.93% sparse.
	D'Souza et al. (2013)	Findings showed the learning precision to be 75%.
	Kayahan et al. (2013)	Predicting TV ratings based on customer sentiments.
	Dooms et al. (2014)	Dataset containing recent ratings was developed in the domains (books, music, videos).
	Mhaisgawali and Giri (2014)	Predicting TV rating of Indian TV shows.
	Razis and Anagnostopoulos (2014)	Influence needs to be calculated from the number of followers, interaction, awareness and visibility.
	Schmit and Wubben (2015)	Predict ratings of movie releases based on public opinions.
	Akarsu and Diri (2016)	Predicting TV ratings of Turkish TV using tweets.
	Prasetyo and Winarko (2016)	Findings were different from what the Indonesian Broadcasting Commission stated.
	Kesharwani and Bharti (2017)	Findings show ratings in twitter is lesser than other website ratings.
	Ray, Bala, and Jain (2020a)	Results of this study show that using combining sentiment and emotion values shows an improvement in the accuracy of prediction.
Facebook	Cheng et al. (2016)	Evaluating the correlation between Facebook posts and TV ratings.
YouTube	Siersdorfer et al. (2010)	Predicting ratings of YouTube comments based on customer sentiments.

Table 3. The features used in this study

Features	**Description**
Content Words	Words containing independent meaning (Winkler, 2012)
Function Words	Words like, ‘the’, ‘and’, etc. that does not contain much lexical meaning but helps to identify the grammatical relationships between different words in a sentence and also sometimes specifies attitude/mood of the person (Klammer, Schulz and Della, 2000).
Parts-of-speech (POS) tags	Depicting words in a phrase based on different parts of speech. For e.g., These/PNN, are/VB, tweets/NN). Here, PNN=pronoun, VB=verb, ART=article, NN=noun. (Church, 1989).
Content and function words	Using them together helps to get both style- and topic-based features (Mukherjee and Bala, 2017).
n-grams	This helps to determine the next item in a sequence based on one (bigram), two (trigrams), or more preceding items. Researchers have generally stressed on using trigrams as its works better than other higher n-grams (Koppel, 2002; Mukherjee and Bala, 2017).

Table 4. Comparison of the classifier performance (accuracy measures)

	Naïve Bayes			**Max-entropy**			**kNN Classifier**		
	Simple*	Sentiment	Sentiment+Emotion	Simple*	Sentiment	Sentiment+Emotion	Simple*	Sentiment	Sentiment+Emotion
All Words	0.5866	0.6579	0.6346	0.5867	0.4668	0.6346	0.6671	0.6680	0.6616
Content Words	0.5983	0.6479	0.6454	0.5984	0.6133	0.6455	0.6624	0.6788	0.6596
Function Words	0.5683	0.6732	0.6182	0.5683	0.6309	0.6182	0.6572	0.6592	0.6536
POS	0.5811	0.6461	0.9503	0.5846	0.6310	0.6310	0.6604	0.6644	0.6808
Ngrams	0.5811	0.7960	0.6220	0.5775	0.495	0.6219	0.6756	0.6648	0.66
Content+Function Words	0.5823	0.6517	0.6341	0.5823	0.6124	0.6554	0.6910	0.6920	0.6616
Average Score	**0.5830**	**0.6788**	**0.6841**	**0.582967**	**0.5749**	**0.6344**	**0.6690**	**0.6712**	**0.6629**

*Simple = It is the case where the classifier performs without the influence of any sentiment or emotion scores

[Note: Using the various combinations of sentiment and emotion values in context of e-Learning]

Table 5. Comparing the classifier performance (accuracy measures)

	Naïve Bayes			Max-entropy			kNN Classifier		
	Simple*	Sentiment	Sentiment+Emotion	Simple*	Sentiment	Sentiment+Emotion	Simple*	Sentiment	Sentiment+Emotion
All Words	0.7101	0.4797	0.8285	0.5323	0.4667	0.4749	0.7228	0.7005	0.7084
Content Words	0.7105	0.5001	0.8281	0.6856	0.4998	0.7739	0.7016	0.6960	0.6944
Function Words	0.7610	0.4306	0.8288	0.4983	0.4530	0.7413	0.6868	0.6784	0.6132
POS	0.7867	0.7880	0.8879	0.7867	0.7880	0.8880	0.7376	0.7188	0.7172
Ngrams	0.6973	0.6964	0.8239	0.5861	0.6542	0.6727	0.7140	0.7136	0.7140
Content+Function Words	0.7046	0.4931	0.8269	0.6758	0.4914	0.7708	0.7228	0.7005	0.7033
Average Score	**0.7284**	**0.5647**	**0.8374**	**0.6275**	**0.5589**	**0.7203**	**0.7143**	**0.7013**	**0.6918**

*Simple = It is the case where the classifier performs without the influence of any sentiment or emotion scores.

[Note: Using the various combinations of sentiment and emotion scores in context of OTAs]

Table 6. Precision, recall, F-measure for e-Learning and OTA services

	e-Learning			OTAs		
	Precision	Recall	F-Score	Precision	Recall	F-Score
	Naïve Bayes			Naïve Bayes		
Simple	0.58	0.72	0.64	0.60	0.69	0.62
Sentiment	0.59	0.72	0.64	0.62	0.69	0.63
Sentiment+Emotion	0.60	0.68	0.63	0.62	0.62	0.62
	kNN Classifier			kNN Classifier		
Simple	0.59	0.68	0.63	0.58	0.69	0.62
Sentiment	0.58	0.67	0.62	0.59	0.68	0.62
Sentiment+Emotion	0.59	0.69	0.63	0.59	0.68	0.62

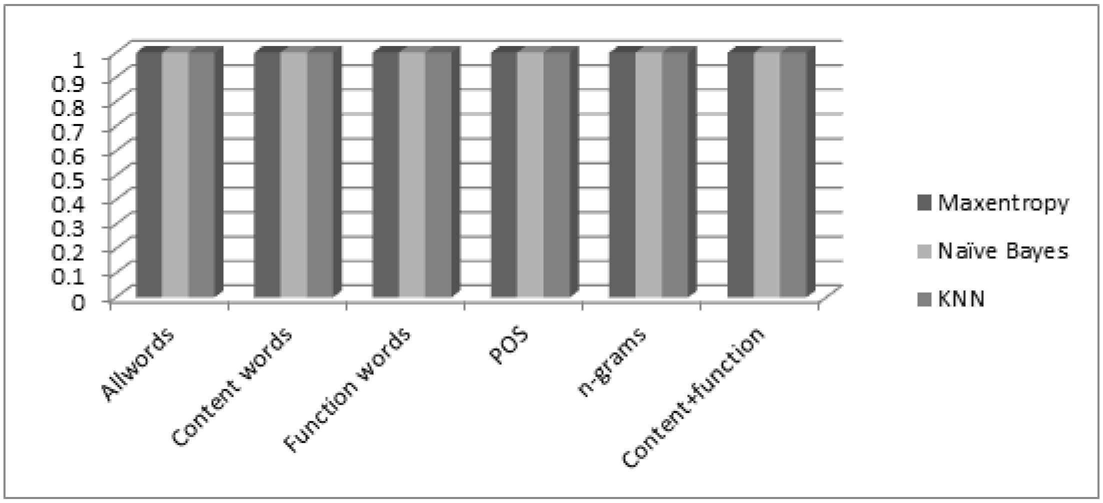

Figure 1. Accuracy of the various classifiers using different features in context of OFDs

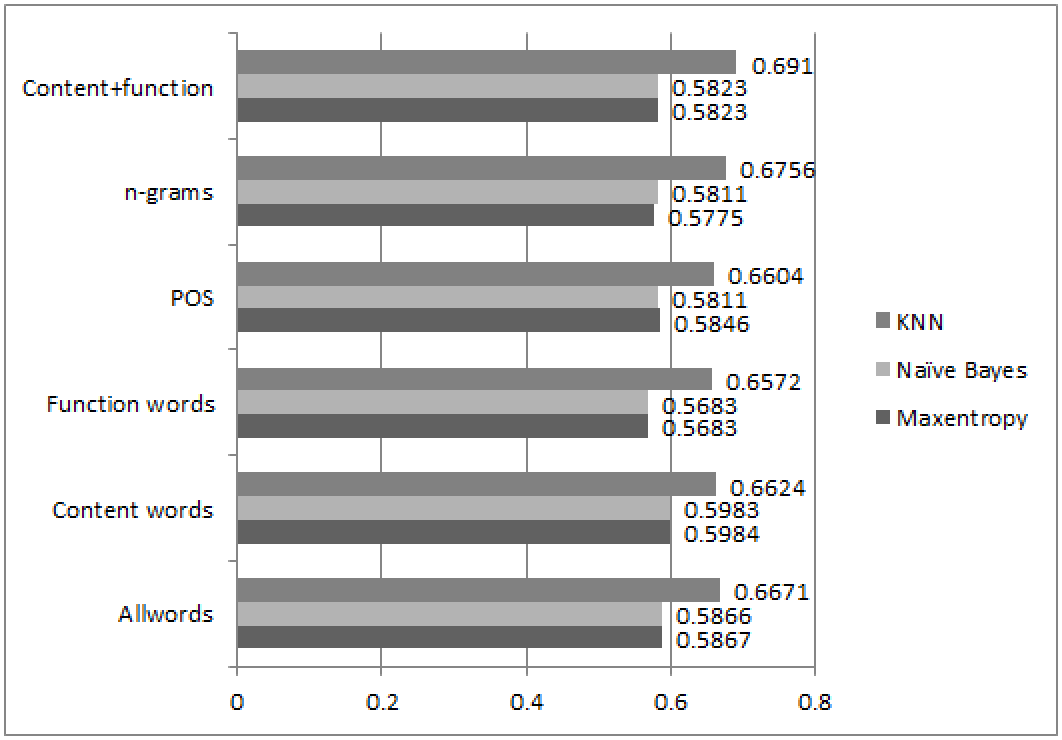

Figure 2. Classifier performance (accuracy) of e-Learning services

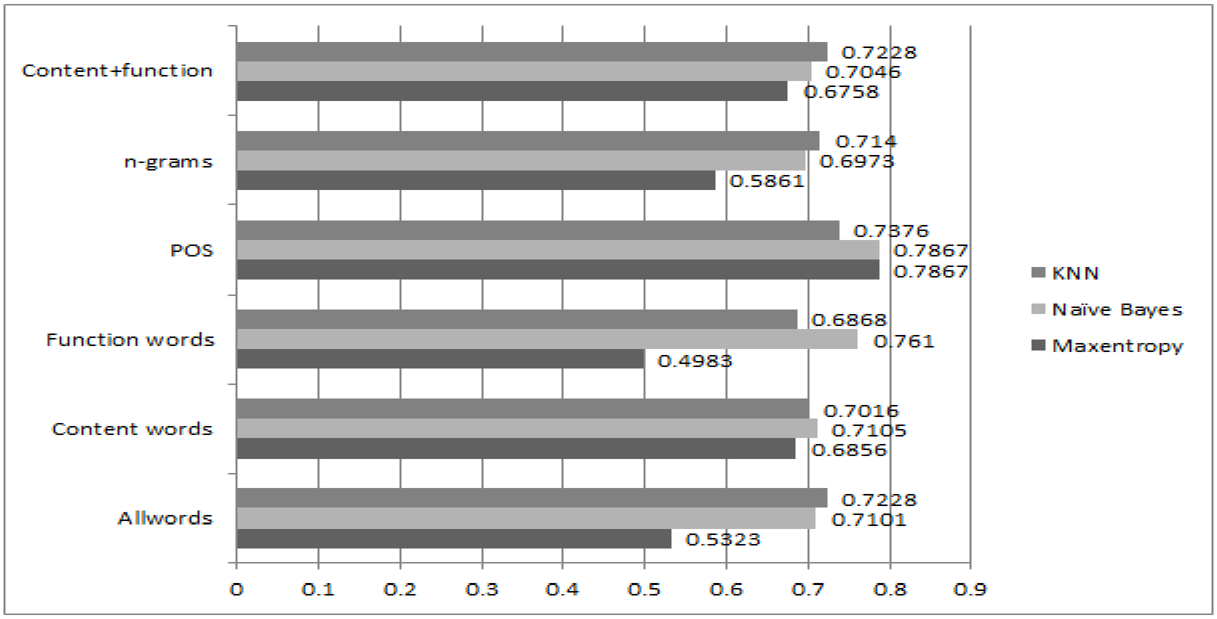

Figure 3. Classifier performance (accuracy) of OTAs

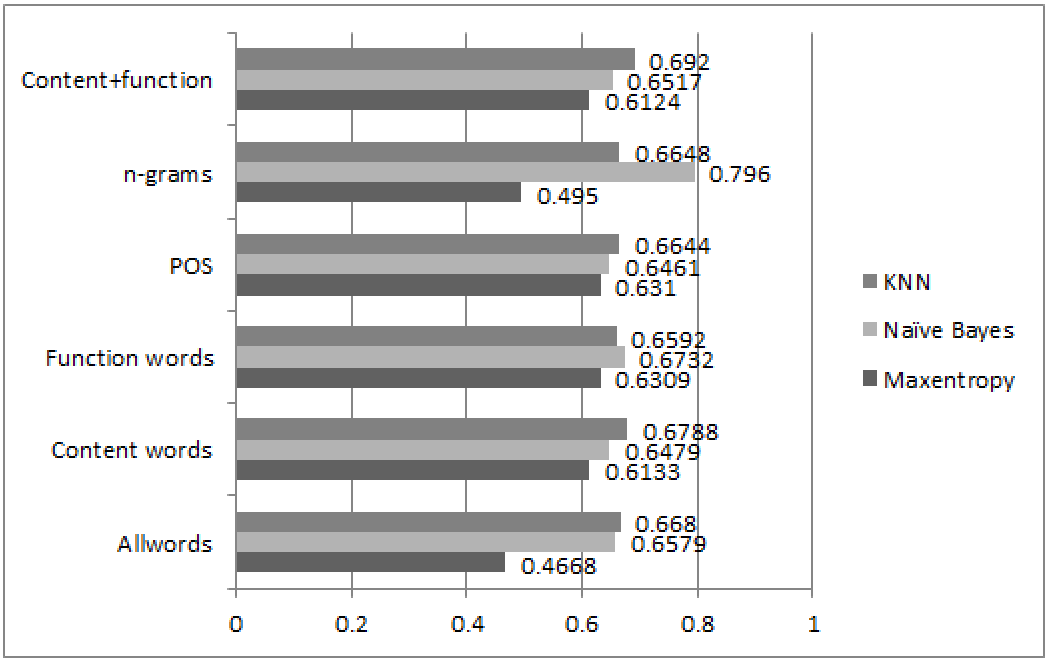

Figure 4. Classifier performance (accuracy) for combination of sentiment scores in e-Learning services

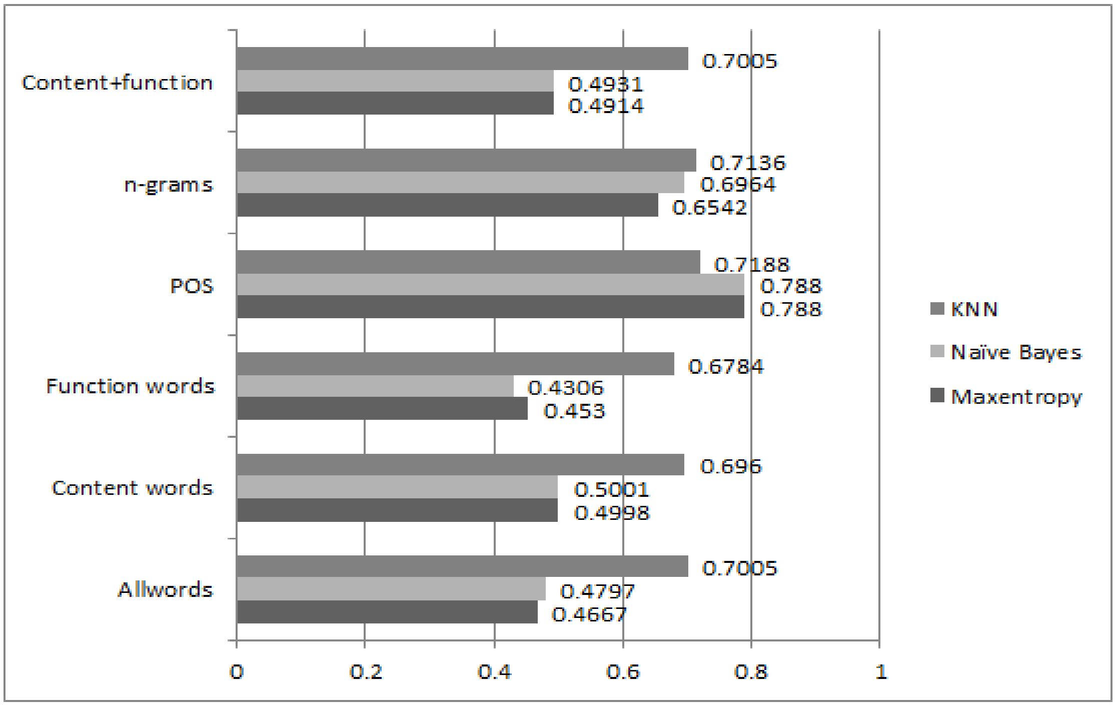

Figure 5. Classifier performance (accuracy) using combination of sentiment scores for OTA services

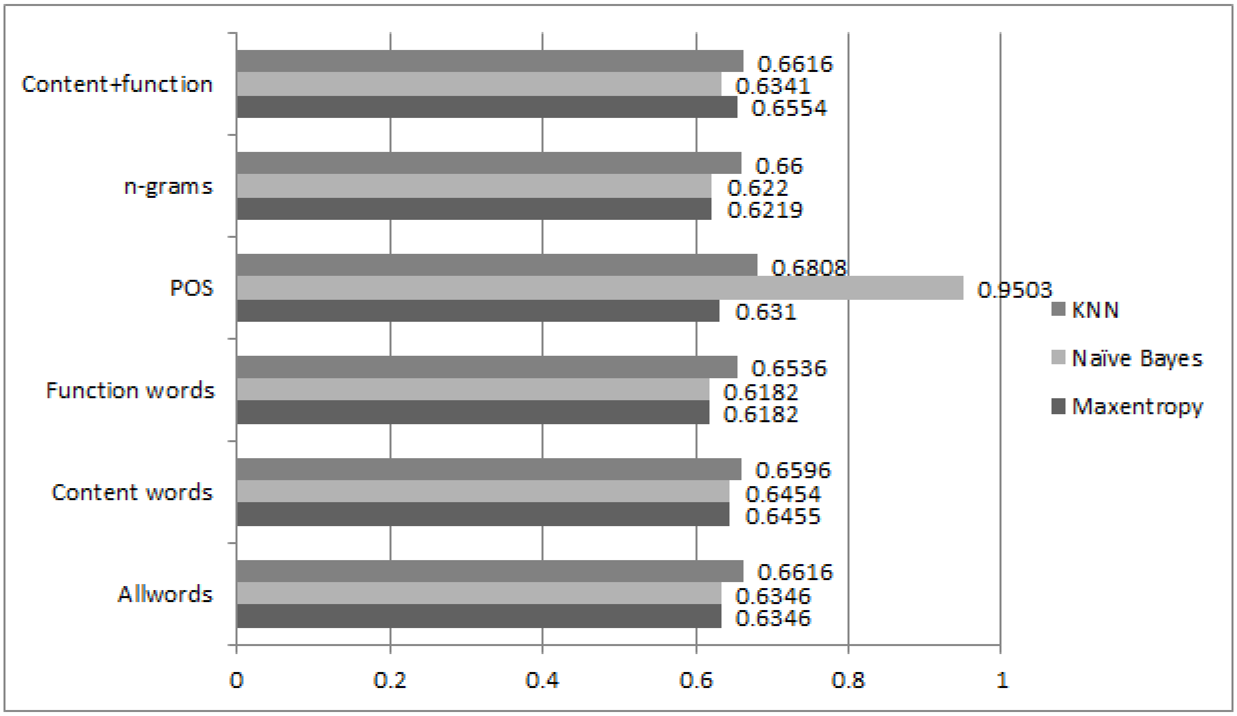

Figure 6. Classifier performance (accuracy) for sentiment scores and emotion scores for e-Learning services

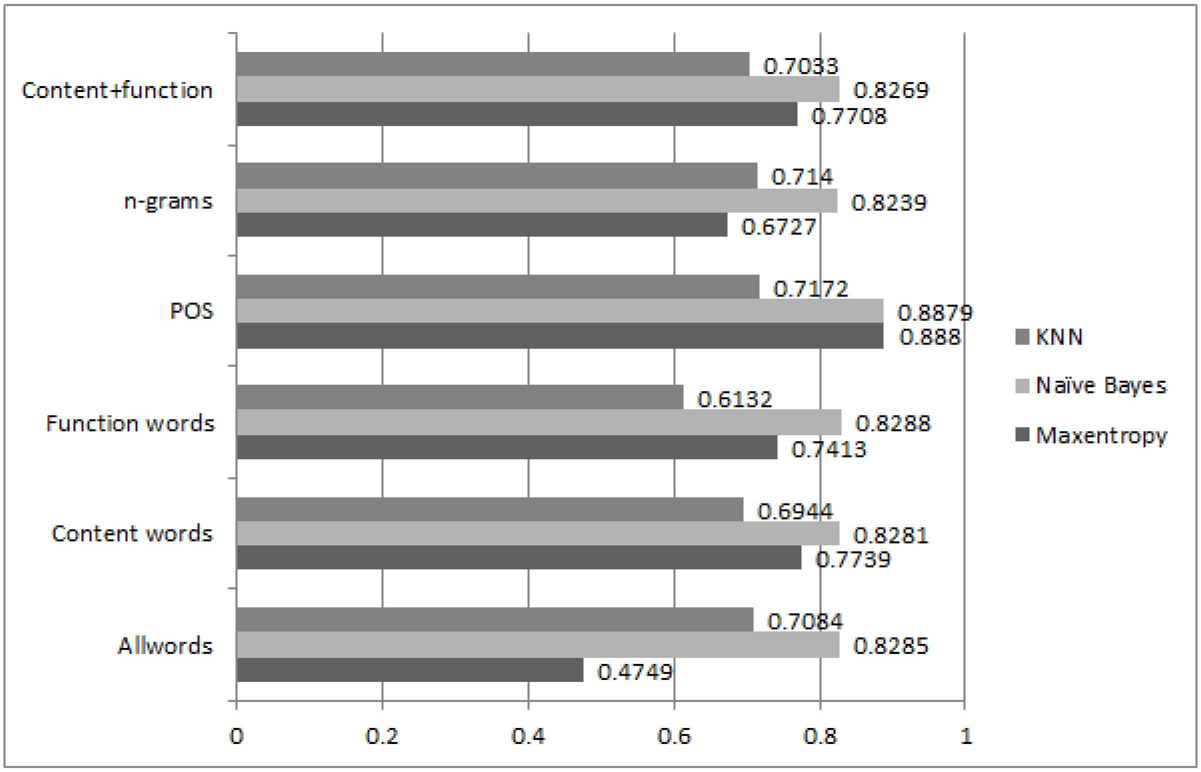

Figure 7. Classifier performance (accuracy) of sentiment scores and emotion scores for OTAs

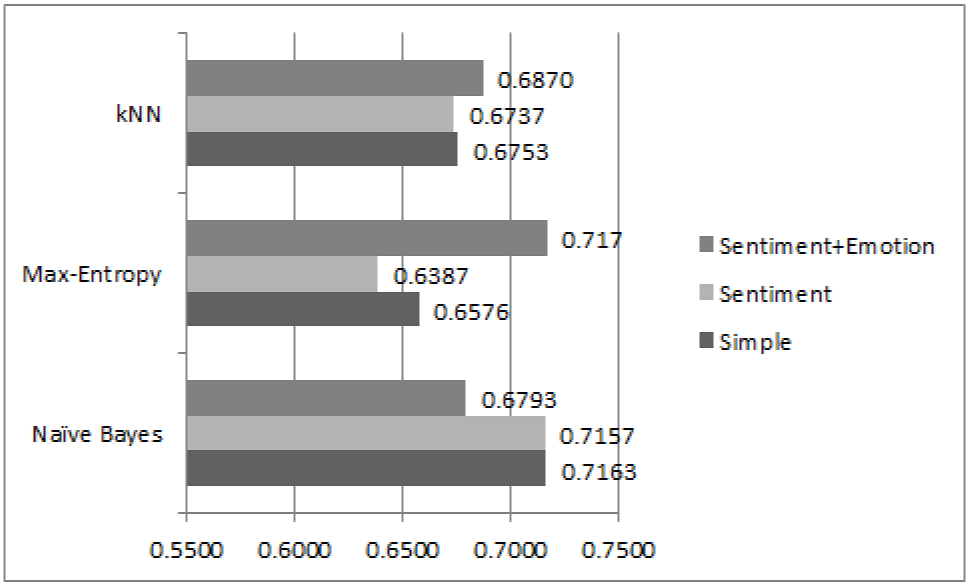

Figure 8. Accuracy of classifiers using the factors-of-importance as a feature for e-Learning services

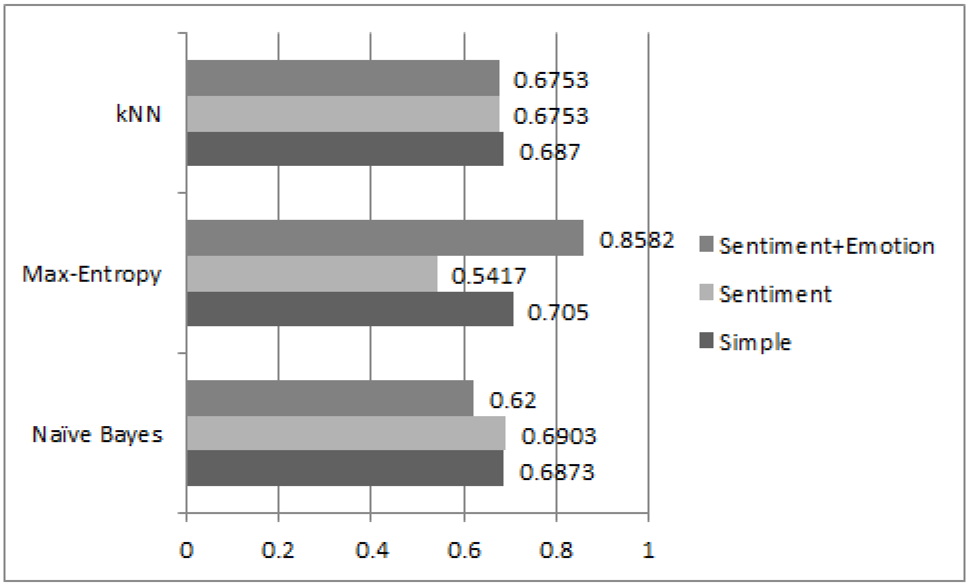

Figure 9. Accuracy of classifiers using the factors-of-importance as a feature for OTAs