Journal of Data Mining and Digital Humanities

The effect of Facebook behaviors on the prediction of review helpfulness

Emna Ben Abdallah[1] **and Khouloud Boukadi**[1]

[1]Mir@cl Laboratory, Sfax University, Sfax, Tunisia

Corresponding author: Emna Ben Abdallah , emnabenabdallah@ymail.com

Abstract

Facebook reviews contain reviews and reviewers' information and include a set of likes, comments, sharing, and reactions called Facebook Behaviors (FBs). We extend existing research on review helpfulness to fit Facebook reviews by demonstrating that Facebook behaviors can impact review helpfulness. This study proposes a theoretical model that explains reviews' helpfulness based on FBs and baseline features. The model is empirically validated using a real Facebook data set and different feature selection methods (FS) to determine the importance level of such features to maximize the helpfulness prediction. Consequently, a combination of the impactful features is identified based on a robust and effective model. In this context, the like and love behaviors deliver the best predictive performance. On the other hand, among the baseline features, review linguistic and subjectivity exhibit better predictive performance. Furthermore, we employ different classification techniques and a set of influencer features. The results showed the performance of the proposed model by 0.925 of accuracy. The outcomes of the current study can be applied to develop a smart review ranking system for Facebook product pages.

Keywords

Facebook review; Facebook behaviors; helpfulness; feature selection; machine learning

I INTRODUCTION

Nowadays, online reviews are an essential source of information to evaluate products and services. In particular, 80% of customers are impacted by Facebook reviews to make their purchasing decision according to Shaw [2018]. There are many ways that potential leads and current customers can share their experiences about a product or a service with others using Facebook reviews. The Facebook review is a form of feedback where a customer can rate a particular product/service (since August 2018, instead of reviews being posted on a scale of 1 to 5 stars, users can choose whether to recommend the business) or leave a comment. Users can react by likes comments, sharing one another's reviews, and openly talking about a provider. This review system, which appears on the provider's official page, can be viewed by anyone who consults this page.

However, the emergence and the growth of useless Facebook reviews prevent the customer from filtering the relevant information. Putting the most helpful reviews on the top can help users to get the relevant information quickly. In general, Facebook sorts reviews based on newness or connectivity. Meanwhile, previous studies have proved that helpfulness is a multi-faceted concept Zhou and Guo [2017], Li et al. [2022]. In other words, it is affected by many features, considered as cues of helpfulness, including the review quality, the review subjectivity, and the reviewer's characteristic Ghose and Ipeirotis [2011], Korfiatis et al. [2012]. The review quality denotes the review length, readability, and the writing style Malik and Hussain [2020], Ghose and Ipeirotis [2011], Zhou and Guo [2017]. On the other hand, the review subjectivity is explored using the sentiment analysis technique, which has several aspects, such as feature, opinion, and polarity Duan et al. [2012], Verma and Davis [2021]. As for the reviewer's characteristics, they reflect the reviewer's personal information, such as the name, location, and reputation Ngo-Ye and Sinha [2014], Mauro et al. [2021].

Besides, like posts and comments, a Facebook review contains a set of Facebook behaviors (likes, comments, sharing, and reactions) made by other users who read and react to the review. Moreover, in February 2016, Facebook rolled out the reaction feature consisting of five pre-defined reactions, namely "love", "haha", "wow", "sad" and "angry", which enable users to express their emotions wordlessly Smieško [2016]. Several studies have considered these components to analyze posts and comments in many fields Van Hooijdonk and Van Charldorp [2019], Kaur et al. [2019], Mohammad [2016]. These studies revealed how the number of Facebook behaviors helps not only to know more about its users but also to use this piece of information for the benefit of a targeted advertising Van Hooijdonk and Van Charldorp [2019] and improve health care Kaur et al. [2019], education Mohammad [2016] and Political affairs Del Vicario et al. [2017], Lin [2017], Alashri et al. [2018]. However, no study integrating such features has been carried out to improve Facebook review classification (i.e., helpful/not helpful) accuracy. For this reason, our study addresses the following three research questions:

- Is there any relationship between the Facebook helpful review and Facebook behaviors?
- What is the most relevant set of features that leads to the most effective helpfulness prediction?
- Which Machine Learning (ML) method provides better performance for helpfulness prediction on Facebook?

This research aims to investigate the effect of standard features and Facebook behaviors together on Facebook reviews to assess their relative importance. One objective is to compare the combination set (standard features + Facebook behaviors) to the art baseline state. Another objective is to examine the most appropriate set of features that leads to a better helpfulness prediction. There is also a need for an outperforming machine learning model.

To fulfill the above objectives, we collected Facebook reviews from 17 well-known cloud providers' official pages. Three feature selection methods are used and evaluated based on their robustness and performance to analyze helpfulness features. Besides, five ML methods are examined to build the helpfulness prediction model on the best effective set of features and the outperformed classifier. Theoretically, we make three contributions to the literature. First, this paper suggests a new model that ensures, for the first time, the assessment of Facebook review helpfulness. The findings indicate the impact of review linguistic and subjectivity feature categories among the other baseline features. Second, we explore the possibility of integrating the number of likes, comments, sharing, and reactions as helpfulness features to improve helpfulness prediction performance. The findings highlight the significant impact of love and like behaviors and the lesser impact of sharing behavior. Third, this study also provides practical implications. The proposed model can be the building block of a smart review ranking system for Facebook product pages to easily get the relevant information.

The remainder of this paper is organized as follows: Section II presents the related work followed by Section III, which details the used methodology. Subsequently, Section IV discusses preliminary evaluation results. Section V presents discussions and highlight the implications of this research before concluding remarks and future work in Section VI.

II LITERATURE REVIEW

2.1 Review helpfulness

Review helpfulness assesses the informativeness of reviews in terms of understanding and evaluating a product or a service Cao et al. [2011], Filieri [2015] In the past two decades, scholars have assessed review helpfulness according to different features considered clues to evaluate the review helpfulness. These features are deduced either from review content, such as readability, visibility, verbs, etc., or from the review/reviewer characteristic, such as productivity score and expertise. More details about helpfulness features are presented in Figure 1 and detailed in Section 3.3.1.

To analyze the impact of the proposed features on the review helpfulness, several studies Ren and Hong [2019], Craciun et al. [2020], Zhou et al. [2020] applied Tobit regression. Eslami et al. [2018] relied on ANOVA analysis to reveal the impact of each of the features on the helpfulness of online consumer's reviews. On the other hand, in the study of Malik and Hussain [2018], the authors apply the Pearson correlation method to examine the relationship between each feature and review helpfulness. As for the study in Ngo-Ye et al. [2017], the authors used correlation-based feature selection to identify a subset of features that have a high correlation with the review helpfulness.

For their part, Ghose and Ipeirotis [2011] analyzed several features of the reviewed text, such as spelling errors, readability, subjectivity, etc. and examined its effect on sales as well. It was discovered that linguistic correctness is a critical factor in affecting sales. On the other hand, compared to very short or very long reviews and getting spelling errors, there is an intuition that the concise and medium reviews and have fewer spelling errors are more useful to the customers. Moreover, Ghose and Ipeirotis [2011] developed three taxonomies for the reviewed text's characteristics, such as the ease of reading a review, spelling errors in the review, and the degree of subjectivity, while taking these points into account. As for Malik and Hussain [2018], they demonstrated the importance of helpfulness per day and syllables and auxiliary verbs as helpfulness features. They explained the high significance of helpfulness per day because a review of a reviewer with large helpfulness per day property attracts more readerships and receives more helpful votes. The syllable variable's great importance indicates that the reviews that contain more syllable words attract more customers and facilitate their purchasing decision making. Besides, customers prefer reviews that use more auxiliary verbs in the text. The authors also proved the high importance of space and productivity score variables. Lee et al. [2018] described the reviewer's features as important predicting factors of review helpfulness since they are the most performing in the classifications. In contrast, the review quality and review sentiment are poor predictors of review helpfulness. Meanwhile, the study of Ngo-Ye et al. [2017] investigated cognitive scripts for review helpfulness. Eslami et al. [2018] used the review length, score, and subjectivity (argument frame) and found that the most helpful reviews are those that are associated with a medium length, lower review scores, and a negative or neutral argument frame. On the other hand, the review title is considered as one of the helpfulness predictors. For instance, the study in Zhou et al. [2020] examined the similarity and the sentiment consistency between the review content and its title. The results revealed that the title-content similarity positively impacts the review helpfulness. Moreover, the authors in Ren and Hong [2019], Craciun et al. [2020] figure out that the emotional content is an essential factor in the review perceived helpfulness. Then, the authors in Ren and Hong [2019] indicated that the product type moderates the impact of three discrete emotions (sadness, fear, anger) on review helpfulness. Notably, they revealed the negative effect of the sadness emotion on the review helpfulness, contrary to the fear emotion that positively impacts. The anger emotion negatively impacts the review helpfulness for an experienced product than for a search product. Meanwhile, the study in Craciun et al. [2020] examined the correlation between the reviewer's gender and contextual emotional tone for the review helpfulness. The authors of this study demonstrated that the reviewer gender manages the impact of emotional content on the review helpfulness and that the readers' perceptions of the reviewer's credibility explain this effect. Their findings also revealed the relationship between female-expressed anger and review helpfulness.

On the other side, prediction models are more accurate and can thus be implemented further in practice Lee et al. [2018]. Notably, classification and regression techniques have been used to build prediction models (see Figure 1). Ngo-Ye et al. (2017) stated that the text regression model is used to estimate the helpfulness of the customer review. Consequently, the results showed that the proposed model offers higher accuracy with low training and testing times. As for Malik and Hussain [2018], they analyzed six standard machine learning techniques (NC-PQR, CART, MAR, NNET, RandF, and Stochastic GB). They then revealed that the stochastic gradient boosting ML model is the most effective method and that the proposed hybrid determinants have shown the best performance. Meanwhile, Lee et al. [2018] demonstrated that the RF classification-based algorithm performs the most efficient in predicting helpful reviews from all datasets of TripAdvisor reviews. In Ghose and Ipeirotis [2011], the classifiers SVM and Random Forest (RF) were used to predict the helpful reviews, where RF exceeded SVM in all cases. As for the study of Eslami et al. [2018], the authors used an artificial neural network approach to predict the review helpfulness. A. Lopez and R. Garza performed a topic modeling analysis to extract the main topics that consumers express in their reviews. Thereafter, the topics were used as regressors to predict the number of consumers who found the review helpful to test the serial mediation effect. Moreover, recently, the authors in Li et al. [2022] have empirically examined the effects of the numerical features used in online review comments on perceived review helpfulness and the underlying psychological mechanisms. The main findings of this study highlight the positive correlation between the numerical features in online review comments and perceived review helpfulness across different product categories. This relationship is mediated by two psychological responses of consumers: cognitive elaboration and credibility perception. On the other hand, Lee et al. Lee et al. [2021] proposed several prediction models for the helpfulness of Yelp business reviews using a variety of machine learning techniques, namely multivariate linear regression, random forest, support vector machine regression, and extreme gradient boosting (XGBoost). The results highlighted the outperformance of XGBoost for predicting review helpfulness among selected popular ML algorithms. Results revealed that the reviewer's credibility is an important feature to assess the review's helpfulness.

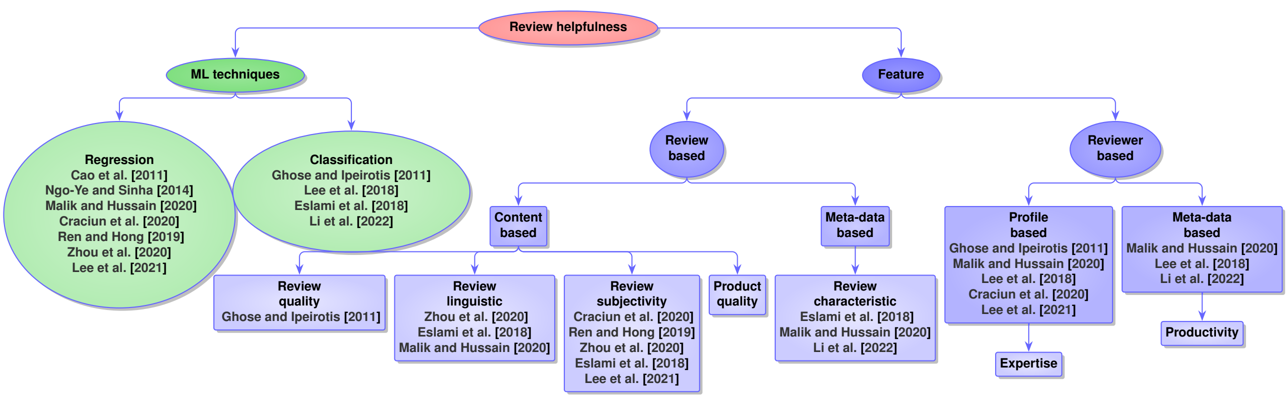


Figure 1: Helpful review prediction studies in the literature

2.2 Facebook behaviors

There are two types of interaction on Facebook: active one such as liking, sharing, commenting, reacting, and passive interaction such as clicking, watching, or viewing Ekström and Östman [2015], Kaur et al. [2019]. This research focuses on active interactions as they are the available publicly responses. Furthermore, the literature has demonstrated that this interaction type can help understand human behavior on social media Ding et al. [2017], Ross et al. [2018], Kaur et al. [2019].

In this context, Zell and Moeller [2018] showed that the like behavior is the fastest way for users to communicate on Facebook. It enables them to present their agreement on specific comments, pictures, wall posts, statuses, or fan pages. As for Sumner et al. [2018], they revealed that individuals often use the like button to appreciate the content of a post and to get closer with the poster. In the marketing areas, Ding et al. [2017] demonstrated the strong correlation between the like behavior and the box office of movies. Similarly, Pelletier and Horky [2015] revealed a positive impact of the like behavior on the product and service brands as it occurs between the company and the consumer. They also proved that the users were eight times more preferred to hit the like button than the sharing or comment button, with 44% of them liking content posted at least once a day and 29% did so several times a day. The sharing behavior, Rui and Stefanone [2013] proved that there is a strong link between sharing an item and the users' self-presentation. For instance, when presenting themselves on Facebook, the users carefully examine public self-evaluation and check whether online self-presentation is compatible with offline self-presentation.

On the other hand, comments are content that users can write under posted items. In the literature, many studies have proved the importance of considering Facebook comments for many purposes. In the field of crisis communication, Hong and Cameron [2018] confirmed the paramount role of the comments in influencing readers' perception of meaningful discussions. In the political field, many studies Lin [2017], Alashri et al. [2018] showed the effect of comments on altering voters' opinions in electoral elections. Moreover, The authors in de León and Trilling [2021] explored the relationship between political news valence, Facebook reactions, and news sharing during the 2018 Mexican elections. They used Negative Binomial Regressions Predicting to analyze the relationship between Reactions and news sharing. The findings revealed a strong relationship between negative news and the sad Reaction, as well as between negative news and the Wow Reaction. Besides like, comment and sharing, Facebook extended the like button by adding five reactions (love, haha, wow, sad and angry) to enable consumers to indicate how the content of a post/review makes them feel emotional in one simple click. According to Turnbull and Jenkins [2016], Facebook reactions provide an opportunity for marketers to gain a better understanding of how consumers emotionally engage with social media content. Moreover, many studies used Facebook reactions to improve sentiment classification Kaur et al. [2019], Smieško [2016].

III METHODOLOGY

The present study's methodology for predicting review helpfulness in Facebook consists of six primary phases: data collection, data cleaning, feature engineering, feature selection, helpful Facebook review prediction, and result comparison. Figure 2 depicts the overall workflow adopted in this study.

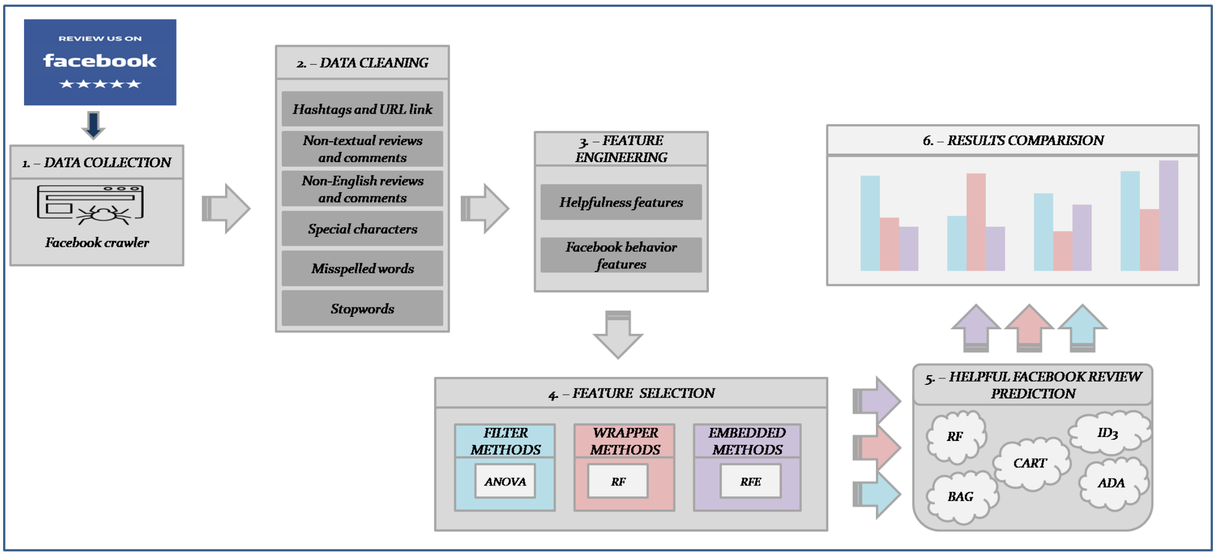


Figure 2: The workflow adopted in this study

3.1 Data collection

Since 2013, Facebook business pages allow for user-generated ratings or recommendations and reviews on the page. Although most people use Facebook primarily as a way to connect with their social network, rating businesses and reading reviews within such a context have great potential, given that Facebook is the most popular social media platform and the one that most users visit daily Kaur et al. [2019].
Unfortunately, Facebook Graph API [1] does not support the collection of reviews. Indeed, we propose a new crawler that allows the gathering of reviews posted on Facebook using Node.js and MongoDB as a database.
The developed crawler collects data from Facebook official pages firstly according to the Mongo model depicted in Figure 3, related to the review as well as to the reviewer such as the review content, date of the review, rating/ recommendation, reviewer (User) profile which contains reviewer information as job position and education level, etc. Moreover, as depicted in Figure 3, the number of likes, reactions, comments, and sharing were also extracted.
We extracted Facebook reviews from 17 official pages of cloud providers to evaluate our approach. These pages have an average of 4 reviews per day. The collected reviews were published between September 2019 and February 2022.

3.2 Data cleaning

The importance of "quality versus quantity" of data in social media scrapping and analysis cannot be overlooked. Of course, unstructured textual data can be very noisy (i.e., dirty). Therefore, data cleaning (or cleaning, scrubbing) is an important area for social media analytics. The process of cleaning the data, in this paper, consists of removing useless reviews and comments such as:

a) Hashtags and URL links
b) Non-textual reviews and comments (photo, video, GIF file, etc.)
c) Reviews and comments that were less than three words long.
d) Reviews and comments that were written in a language other than English.
e) Reviews and comments that had five or more misspelled words. Misspelled words were

[1] http://developers.facebook.com/docs/reference/api/

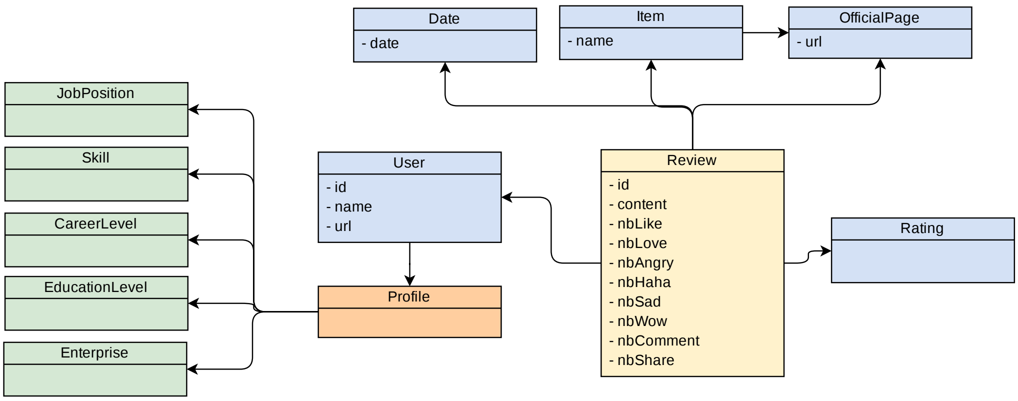


Figure 3: Structure of the collected data

considered as words that have been wrongly spelled either due to human mistakes or typos.

f) Special characters (, #, $, etc.).

Besides, this phase aims to correct misspelled words using the wordnik API dictionary [2] or to discard the word. We also removed reviews that do not contain any Facebook behavior.
In this phase, spam reviews and spam comments are also detected and filtered using Ben-Abdallah et al. [2018]. The proposed approach considers well-known spam features taken from the literature to them we add two new ones: the user profile authenticity to allow the detection of spam review from Facebook and opinion deviation to verify the opinion truthfulness.
The output of the data cleaning is a final *numberOfReviews* of raw data. Of these, approximately *20% (i.e. 5000)* were randomly extracted for this study. The *5000* reviews then went through the standard pre-processing phase of POS tagging, tokenization, stop word removal, and stemming.

Data annotation

To annotate the Facebook reviews as helpful or not helpful, we rely on seven cloud instructors from the IT department of the University of Sfax (considered as experts). This annotation is important for the helpfulness prediction of this study. An inter-rater reliability (IRR) analysis using subsamples was conducted where the Krippendorff's alpha Krippendorff [2004] was calculated. The generated IRR sets with 89% agreement.

3.3 Feature engineering

This section introduces the features used as clues to identify helpful cloud reviews. Table 1 presents the list of the features used in this study. We define two types of features: independent features and Facebook dependent features. The first type is used to assess any review extracted from any social media platform. Meanwhile, the second one is to evaluate only Facebook reviews, named Facebook Behavioral features.

3.3.1 Helpfulness features

As depicted in Figure 1, the independent features for the review helpfulness prediction process can be either review based or reviewer based. The review based category contains two feature

[2]https://developer.wordnik.com/

types: content-based feature and meta-data based features. The reviewer based feature category has, in turn, two feature types: profile-based feature and meta-data based feature. In this study, we consider only the review based features because the reviewer based features are not available on Facebook. Three types of review based features are investigated in this research: review quality, review subjectivity, and review characteristic. As for review quality, we consider readability, linguistic, and visibility features.

Category	Type	Feature	Description	Source
Review quality	Readability	ARI	Automated Readability Index	Ghose and Ipeirotis [2011]
		CLI	Coleman-Liau Index	Ghose and Ipeirotis [2011]
		FKGL	Flesch-Kincaid Grade Level	Ghose and Ipeirotis [2011]
		GFI	Gunning Fog Index	Ghose and Ipeirotis [2011]
		FRE	Flesch Reading Ease	Ghose and Ipeirotis [2011]
		SMOG	Simple Measure of Gobbledygook	Ghose and Ipeirotis [2011]
	Linguistic	W3SoM	Words with 3 syllables or more	Malik and Hussain [2020]
		Vrb	# of verbs	Hu and Chen [2016]
		Adv	# of adverbs	Hu and Chen [2016]
		Adj	# of adjectives	Hu and Chen [2016]
		Noun	# of nouns	Hu and Chen [2016]

	Visibility	Char	Length of a review in characters	Ghose and Ipeirotis [2011]
		Syllab	Average # of syllables per word	Lee et al. [2018]
		Word	Length of a review in words	Ghose and Ipeirotis [2011]
		Sent	Length of a review in sentences	Ghose and Ipeirotis [2011]
	Uncertainty	nb_if	Ratio of 'If'-terms according to the GI related to the entire number of words	Siering et al. [2018]
Review subjectivity	Compare	comp	The ratio of comparison words such as "bigger", "best" and "smaller"	Luo and Xu [2019]
	sentiment	polarity	The degree of emotions embedded in the wording of review mes-sages.	Malik and Hussain [2020]
Review	Age	Age	Age of the Review	Siering et al. [2018]
characteristic	Extremity	Extrm	Absolute value of Star Rating minus Mean Rating	Siering et al. [2018]
	like	nb_like	# of like in the review	this study
	love	nb_love	# of love in the review	this study
Facebook behavior	wow	nb_wow	# of wow in the review	this study
	haha	nb_haha	# of haha in the review	this study
	sad	nb_sad	# of sad in the review	this study
	angry	nb_angry	# of angry in the review	this study
	comment	nb_cmt	# of comment in the review	this study
	sharing	nb_sharing	# of sharing in the review	this study

Table 1: List of helpfulness features used in this study

1. Readability features: Existing studies Ghose and Ipeirotis [2011], Hu and Chen [2016] demonstrated that the review readability is one of the essential features for helpfulness prediction. A review with high readability tends to be read and perceived more votes from users. We rely on a set of well-used readability index to determine the review readability: Automated Readability Index (ARI), Coleman-Liau Index (CLI), Flesch-Kincaid Grade Level (FKGL), Gunning FOG Index (GFI), Flesch Reading Ease (FRE), and Simple Measure of Gobbledygook (SMOG).
2. Linguistic features: Linguistic features of the review text are other significant predictors that can impact the review helpfulness. Nouns, Verbs, Adverbs, Adjectives and words with three syllables or more features are confirmed to be determinant predictors in the literature Hu and Chen [2016], Chauhan et al. [2020], Malik and Hussain [2020]. These features are identified from the review text by counting their number in the review.
3. Visibility: According to Hu and Chen [2016], the longer a review has been posted, the higher possibility it will receive votes on helpfulness. Based on this hypothesis, four features are considered in this study: length of the review in characters, an average of the number of syllables per word, length of the review in words, and length of the review in sentences.

As for the review subjectivity features, many studies have proved their importance as determinants for review helpfulness prediction Luo and Xu [2019], Malik and Hussain [2020]. In this study, we consider the ratio of comparison and the polarity of the review. This study also considers review characteristic which includes review age and review extremity.

3.3.2 Facebook Behavioral features

Besides the features depicted above, we also exploit Facebook's different behaviors that exist around the review (likes, sharing, comments, and reactions). We had hypothesized that reviews receiving more likes, comments, and sharing would be associated with reporting higher helpfulness. Meanwhile, we will study in this paper the direction of the Facebook reaction effect.

3.4 Feature selection

Feature selection (FS) is an appropriate step in building machine learning methods as it enables the predictive model to achieve good, or even better, solutions with a restricted subset of features Saeys et al. [2008]. The FS techniques aim to remove irrelevant and/or redundant features and identify the most relevant features to understand better the subject of interest's mechanisms, instead of merely building a black box predictive model. In the context of the Facebook reviews' helpfulness, the feature importance must be studied to understand a reviewer's behavior that writes helpful reviews on Facebook and, second, to improve the performance prediction by neglecting the irrelevant and the disturbing features. For doing so, we choose to compare three FS methods belonging to different categories, i.e., wrapper method, embedded method, and filter. The use of these methods has led to analyze the most influencer features based on the robust and the performant FS method. The details of the used FS methods are depicted as follows:

(1) Random Forest (RF) is a wrapper feature selection method Genuer et al. [2010] that is often used for feature selection in a data science workflow. RF performs an implicit feature selection by creating multiple trees using regression trees, CART Genuer et al. [2010]. Typically, a collection of T decision trees are trained on T bootstrap samples of the data, respectively. A random subset of fixed size is selected from the features in each node of each tree and the one yielding the maximum decrease in Gini index is chosen for the split. The trees are fully grown and left unpruned. The class of a new sample

is determined by most of the votes of all trees in the RF. The test error of RF models is estimated on the out-of-bag (OOB) data. After each tree has been grown, the inputs that did not participate in the training bootstrap sample are used as a test set, then averaging over all trees gives the test error estimate. The Gini index uses the decrease of impurity after a node split as a measure of feature relevance. In general, the larger the reduction of impurity after a particular split, the more informative the corresponding input variable. The average decrease in Gini index over all trees in the RF defines the GI. The Gini index is closely related to entropy, both being measures of impurity.

(2) Recursive Feature Elimination (RFE) is an embedded feature selection method that evaluates multiple models by using different procedures. These procedures try to add and/or remove predictors until finding the optimal combination that maximizes the model performance Guyon and Elisseeff [2003].

(3) Analysis of Variance (ANOVA) is a filter-based feature selection method used to assess the means of two or more groups that are substantially different from each other Gueorguieva and Krystal [2004].

3.5 Helpfulness facebook review prediction

The techniques used to construct the prediction models were the models of Random Forest (RF), AdaBoost (ADA), Bagging (BAG), Classification And Regression Tree (CART), Iterative Dichotomiser 3 (ID3). These ML techniques are trained and tested for various sorts of experiments using baseline and Facebook behavior features.

On the other hand, the function matrix's size is $N * K$, where K is the number of the most impactful and efficient features, and N is the number of tuples. In the experiment, 100 training sets and 100 test sets were sampled from each of the available data sets in a ten-fold replicated cross-validation Demšar [2006]. Following the test process in Demšar [2006], the data set was sub-sampled to about ten groups: The different ML models were then trained with data from nine of the groups used to test the remaining group. Approximately 90 percent of the randomly selected findings from the total data set were considered for the training phase. In contrast, the remaining 10 percent were considered the test dataset for the comparative methods' output assessment. Moreover, the training and prediction process has been reproducing ten times. Then, the model prediction was validated against each of the ten rounds. By using the real values of review helpfulness, the predictive performance was measured with the ML performance metrics defined in Section 4.1.

All the above methods were implemented using Python 3.6.0, a high-level programming language. In particular, we used scikit-learn library Pedregosa et al. [2011] to create and fit our models under Google Colaboratory (also known as Colab) Bisong [2019]. Colab is a cloud service based on Jupyter Notebooks for the dissemination of machine learning and research. It provides a fully configured runtime for machine learning and free access to a robust GPU.

IV EXPERIMENTS AND RESULT COMPARISON

This section presents the series of conducted experiments that aims at assessing the impact of Facebook behavior features on the degree of review helpfulness and the performance of the proposed features compared with the state of the art baseline features. The experiments cover both the helpfulness prediction and feature-wise analyses.

4.1 Performance metrics

To evaluate the model performance, the metrics of accuracy, precision, recall and F-measure were considered in the present study. The confusion matrix in Table 2 was used to calculate these metrics as follows:

Table 2: Confusion matrix

		Predicted	
		Helpful	Not helpful
Actual	Helpful	TP	FN
	Not helpful	FP	TN

$$accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{1}$$

$$precision = \frac{TP}{FP + TP} \tag{2}$$

$$recall = \frac{TP}{FN + TP} \tag{3}$$

$$F - measure = 2 \times \frac{precision \bullet recall}{precision + recall} \tag{4}$$

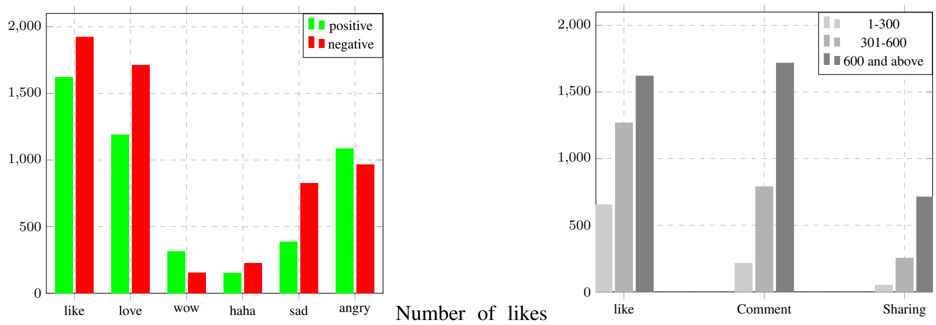
Besides, we rely on the receiver operating characteristic (ROC) to assess the area under the curve (AUC) Hand [2009].

4.2 Analysis of extracted reviews

We examined the collected reviews according to the review orientation (positive or negative) [3]. The results of likes and other Facebook reactions can be seen in Figure 4 where it can be observed that around 71% of reviews contain at least one like for both positive or negative reviews.

Perhaps due to the nature of the studied dataset, the usage of the reactions wow and haha do not exceed 9% of the time. In fact, people can agree by choosing like, love or sad if the review is negative or by choosing angry in case of disagreeing with the review opinion. For this reason, it can be observed in Figure 4 that 71% of reviews contain at least one like for both positive or negative reviews. Besides, 58% of reviews include at least one love. For the negative reviews, we found out that 41% of reviews contain at least one angry and 24% of reviews contain sad. Besides, it can also be observed from the figure that people react with negative reviews meaning that negative experiences affect people. Figure 5 also proves the previous hypothesis where 60% of reviews containing comments are negative ones. According to the collected reviews, we discovered that few people share reviews because sharing requires deeper cognitive processing Kaur et al. [2019], Kim and Yang [2017]. Among other fascinating discoveries made was that the longer the review's content material is, the extra likes, comments, and sharing it receives (Figure 4). More probably, it is the length of a review that pushes readers to pay greater attention when reading, which makes them think profoundly about the technique that eventually brings more likes and comments Kaur et al. [2019], Kim and Yang [2017].

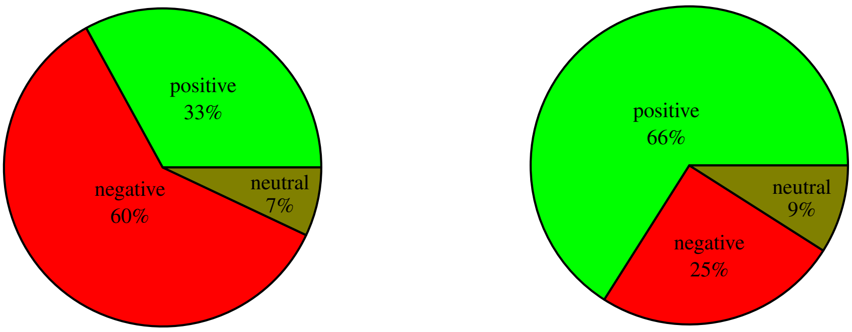
[3] Reviews having star rating > 3 or recommended reviews are considered as positive ones. Reviews having star ratings < 3 or not recommended reviews are regarded as negative ones. Reviews having star rating $= 3$ are considered as neutral ones.



Number of likes and reactions with respect to the review orientation

Number of Facebook Behaviors with respect to number of characters in a review

Figure 4: Facebook behavior analysis



Number of comments

Number of sharing

Figure 5: Number of comments and sharing of reviews according to the review orientation

4.3 Performance analysis of feature selection

As illustrated in Figure 2 and Section 3.4, three feature selection methods are applied to identify the best one in terms of performance. These methods' Feature importance is incorporated in Support Vector Machine (SVM) and distance-based k-Nearest Neighbors (KNN) classifiers. KNN and SVM are two simple and intuitive ones that belong to different families of ML Pathak et al. [2019]. Furthermore, there are often considered as powerful tools to assess the effectiveness of feature importance approaches Neumann et al. [2005]. This subsection is devoted to comparing the performance of the different feature selection methods as RF, ANOVA, and RFE on the Facebook review data set to select the best one.

Figure 6 plots the performance of KNN and SVM in terms of accuracy, AUC, and F-measure metrics averaged over 100 runs against different numbers of helpfulness features. It can be observed that RF outperforms ANOVA and RFE by generally achieving the best values overall metrics. This characteristic suggests that RF ranks the features properly. Using KNN the RF method, we can achieve outstanding performance (Accuracy = 0.87) with the top 16 features of the Facebook review data set, while SVM needs 23 features to achieve comparable results (Accuracy=0.85).

4.4 Robustness analysis of feature selection

The robustness aims at measuring the sensitivity to changes in the input data: a robust algorithm provides (almost) the same outcome when the original data set is disturbed to some extent, e.g., by adding or removing a given set of instances Saeys et al. [2008]. Besides the model performance, robustness is an essential task for the feature selection process. It verifies the FS algorithm's stability over an unstable one when only small changes are made to the data set. A robust FS algorithm would allow domain experts to have more confidence in the selected features, especially if subsequent analyses or validations of the selected feature subsets are

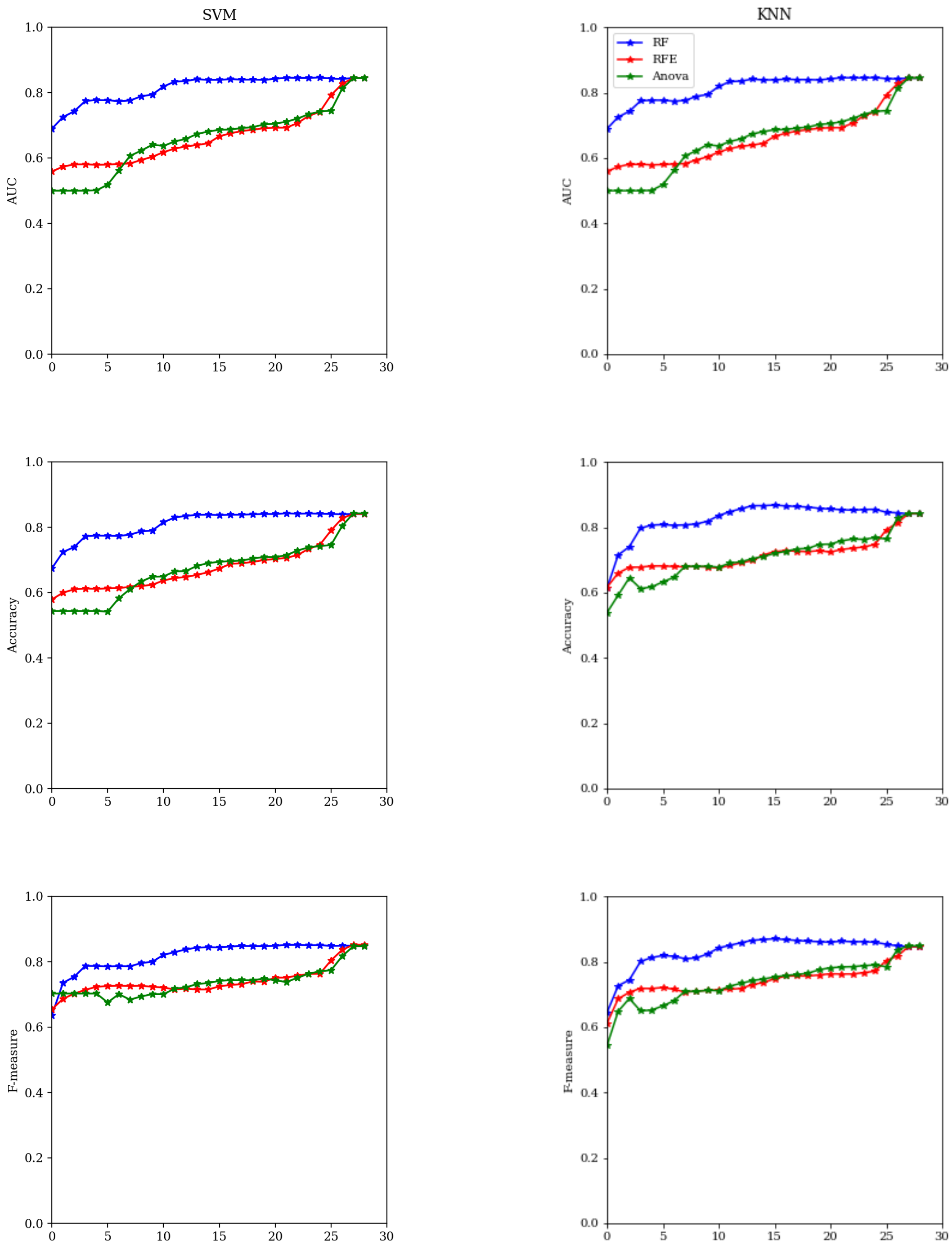


Figure 6: Feature selection methods' performance

costly.
Therefore, to identify the robustness of the various FS methods, we will focus on what follows on the comparison of feature rankings. The traditional Consistency Index (IC) was used for the top 10 percent of the best rankings over the 100 iterations Kuncheva [2007]. The Consistency Index for the two subsets Si and Sj, such that $|Si| = |Sj|$ is introduced by the Equation 5:

$$I_c(S_i, S_j) = \frac{rd - k^2}{r(d-k)} \tag{5}$$

where d is the number of features in the data set, $k = |Si| = |Sj|$ and r is the cardinality of the intersection of subsets Si and Sj.
The overall stability of a feature selection algorithm for a set of sequences of features $S_1, S_2, ..., S_K$ (K = 100 in our case) can be defined as the average overall pairwise consistency indices (Equation 6). The more similar all outputs are, the higher the stability measure will be.

$$Robustness = \frac{2 \sum_{i=1}^{K-1} \sum_{j=i+1}^{K} I_c(S_i, S_j)}{K(K-1)} \tag{6}$$

Table 3 summarizes the results of the robustness analysis across the Facebook review data set for the different feature ranking methods. ANOVA is the less stable algorithm. RF, on the other hand, proves to be a more common feature selection method. Thus, it seems that RF outperforms other feature selection methods regarding robustness.

Table 3: Robustness of the different feature selection methods across the Facebook review data set using the consistency index on the subset of 10% best features.

	RF	RFE	ANOVA
Robustness	0.984	0.939	0.875

Table 4: Harmonic mean of robustness and accuracy for the different feature selection algorithms using 10% of the features.

RF		RFE		ANOVA	
KNN	SVM	KNN	SVM	KNN	SVM
0.892	0.877	0.787	0.751	0.767	0.746

It should be noted that robustness should ideally be used in connection with the performance to enhance reliability and performance at the same time. Domain experts are not interested in a strategy that produces very robust features and a poorly performing model. For this reason, we rely on the soundness-performance trade-off (RPT) suggested by Saeys et al. [2008]. The RPT is a harmonious means of robustness and performance aimed at mutually assessing the trade-off bet between robustness and performance as introduced in the Equation 7.

$$RPT_\beta = \frac{(\beta^2 + 1) \times Robustness \times Performance}{\beta^2 \times Robustness + Performance} \tag{7}$$

The role of parameter β is to control the relative importance of robustness versus the performance and, therefore, can be used to exert more influence either on robustness or on the performance. On the other hand, the value of $\beta = 1$ is the standard formulation that treats both

essential robustness and performance.
Table 4 shows the results for the three FS ranking algorithms (ANOVA, RF, and RFE) where only 10 percent of the features are used. Therefore, the consistency index was used to measure robustness, while accuracy was used for performance. Moreover, it can be observed that with the use of the data collection, the RF results in a better RPT calculation compared to the other two feature ranking algorithms.

4.5 Feature-wise analysis

This section aims at investigating the importance of each feature for helpfulness prediction. The goal of the conducted experiments is to probe the contribution of each feature to the helpfulness of online reviews. The variables ranking applied to the Facebook review dataset using Random forest as a feature selection method is presented in Figure 7.

4.5.1 Impact of baseline features on Facebook review helpfulness

From Table 5 and Figure 7, it can be observed that some linguistic features (W3SoM, Word, Syllab), have a significant influence on helpfulness prediction. The different linguistic variables are appeared in the top 10 selected features using the three FS methods: RF, ANOVA (Adj, Adv), and RFE (Adj). The results echo a simple helpfulness evaluation assumption that "a longer review tends to be more helpful". It also indicates that an experienced customer tends to write a longer review, regardless of whether or not he or she likes the product as these types of reviews appear to be considered more than short reviews.
Table 5 illustrates the importance of the review subjectivity (Polarity, Comp) in the helpfulness evaluation. Comp takes the first place according to the RFE method (see Table 5), while polarity exists in the top 10 selected features using RF. The significance of the text review sentiment and polarity suggests that the reviews containing with more words of emotion and sentiment comparatively acquire more helpful votes.
Age and extremity are other important factors that caught our attention when evaluating reviews helpfulness. Extremity exists in the top 10 selected features according to the three FS methods. The high impact of extremity illustrates that reviews that have high/low ratings tend to be helpful. Besides, it is noticed that the readability variable (SMOG) also has an important impact on the helpfulness assessment (RF and ANOVA). In fact, this variable aims at estimating the years of education needed to understand the reviews. Table 5 lists the top 10 features according to the

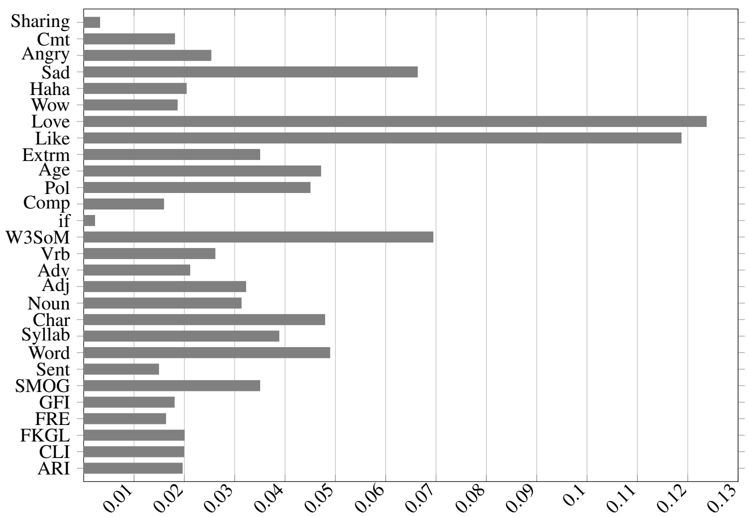


Figure 7: RF based importance of features

different FS methods.

Table 5: Top 10 selected features using RF, ANOVA and RFE

	RF	ANOVA	RFE
1	love	GFI	Comp
2	like	SMOG	love
3	W3SoM	haha	Extremity
4	sad	Adj	like
5	Word	Cmt	sad
6	Age	sharing	Cmt
7	Polarity	ARI	angry
8	Syllab	Adv	Adj
9	Extremity	sad	wow
10	SMOG	Extremity	haha

4.5.2 *Impact of Facebook behaviors on review helpfulness*

From the results depicted in Table 5 and Figure 7, it can be observed the highest importance of both likes and reactions, probably because these features are the most used by consumers when they want to act with a review. In particular, when considering the RF method, the love behavior gained the highest performance ($w = 0.1237$) compared to other Facebook reactions and the other features. It takes first place in the top 10 selected features (Table 5). Furthermore, as illustrated in Figure 6, the love reaction has achieved outstanding performance results using both classifiers (KNN and SVM) compared to the other features selected by the other FS methods. According to the RF method, the second top-ranked feature is the like behavior, which has a 0.1187 of RF importance weight. This finding demonstrates that the like and love behaviors have a strong relationship to review helpfulness since the difference between their RF importance weights and the other features' importance weights is meaningful. Meanwhile, the sad behavior ($w = 0.0663$) is the next important FB behavior since it takes fourth place. Moreover, it appears in the top 10 selected features using RF, ANOVA, and RFE methods. The angry, haha, and wow are the next four influencing behaviors that have lower rankings. Hence, these three reactions do not impact the review helpfulness prediction.

We cannot deny the slight importance of the commenting behavior since it appears in the top 10 selected features using ANOVA and RFE. It reveals that customers may prefer reviews that receive more comments. The sharing behavior has the least predictive performance according to the RF selection method as depicted in Figure 7. It may be explained by the fact that the sharing is only done after some form of cognitive persuasion Kaur et al. [2019], Kim and Yang [2017] (1000 reviews that have at least one sharing among 5000 reviews).

4.6 Model performance analysis

Throughout this section, a series of experiments are conducted to construct the helpfulness prediction models using five popular machine learning techniques: RF, ADA, BAG, CART, and ID3. According to the RF selection method, the five prediction models are trained and tested using the Facebook review data set and with 23 most important features. Table 6 presents the performance of the different classifiers in terms of AUC, accuracy, precision, recall, and F-measure metrics averaged over the 100 runs against the top 23 features.

It is evident from the results that the RF model yields the best performance compared to the other four ML methods with the 23 most important features. It illustrates the evidence of the powerful predictive capacity of the RF robust model.

Table 6: Performance results for the different classifiers

Method	AUC	Accuracy	Precision	Recall	F-measure
RF	**0.927**	**0.925**	0.939	0.926	**0.932**
ADA	0.895	0.896	0.912	0.893	0.898
BAG	0.924	0.917	0.958	0.883	0.924
CART	0.9	0.898	0.92	0.899	0.91
ID3	0.9	0.903	0.913	0.902	0.911

V DISCUSSION

Facebook has become the most popular way through which consumers recommend or review local businesses Van Hooijdonk and Van Charldorp [2019]. In turn, the enormous amount of reviews posted on Facebook pages makes the process of finding helpful reviews time-consuming and sometimes overwhelming. In the literature, many studies have proposed to examine review helpfulness from different social media platforms such as Amazon.com and Yelp.com using several features (considered as indicators of helpfulness). These features are based either on review content, review meta-data, or reviewer. This study evaluates Facebook review helpfulness using baseline features as well as Facebook behaviors.

Our proposed helpfulness prediction model is quite effective as it achieves the maximum accuracy of 0.925 and AUC of 0.927 using a real-life Facebook data set. The proposed review and Facebook behavior-based features are established to be useful predictors in improving helpfulness predictive performance based on results presented in Figures 6 and 7 and Table 5.

Each feature's importance is computed by relying on different feature selection methods such as Random Forest, ANOVA, and RFE to analyze the impact of baseline features on Facebook review helpfulness. The obtained findings indicate that the review linguistic, subjectivity, and characteristics (extremity and age) are the most significant parameters to determine the review helpfulness in the field of cloud computing as presented in Table 5. Similar findings are also supported by past studies Korfiatis et al. [2012], Wang et al. [2019], Lee and Choeh [2016], Ghose and Ipeirotis [2011]. Then, the next most influencing baseline feature for helpfulness is the extremity. A previous study Lee and Choeh [2016] also showed that products with extreme ratings received more helpful reviews. However, we surprisingly found that the standard readability index does not contribute to any critical feature, and consequently, does not help evaluate the Facebook review helpfulness. However, these results are contradictory with those of the literature Ghose and Ipeirotis [2011], which the used data set may explain (Facebook reviews *vs* Amazon.com).

Besides the baseline features, this study has investigated the relationship between Facebook behaviors, such as likes, comments, sharing, reactions, and the review helpfulness. The obtained results suggested that the likes and reactions are the most significant features to determine the review helpfulness (Table 5). In particular, the love reaction performs better than the other reactions. Moreover, Figure 4 indicates that love is the most used reaction. Indeed, we can endorse the hypothesis that Facebook users often use love reaction to express their agreements about a review. The next most impactful behavior is the like button. Still, even angry and sad reactions have great importance on the review helpfulness prediction as they positively impact prediction when the review is negative but an opposite (negative) impact when it is positive. This finding may be explained by the fact that users usually choose these two reactions to agree on the content in a negative review, and choose them to express their anger about a positive review to indicate that it is not true. Meanwhile, the findings showed that the wow and haha reactions do not affect review helpfulness. The next most significant FB feature is the commenting behavior, which indicates that users are more likely to comment on helpful reviews, which is in line with the previous studies Kaur et al. [2019], Kim and Yang [2017] that demonstrate that users are more likely to comment on posts that have logical information. Surprisingly, we found that the sharing feature does not contribute to any important variable, which implies that the sharing behavior does not significantly help evaluate review helpfulness. Therefore these results are different from those in the literature Kaur et al. [2019], Kim and Yang [2017] maybe because the type of the dataset is different since reviews are used in this research and posts were used in past studies, knowing that users tend to share posts more than reviews.

Implications

The research findings of this research have several significant implications. This study addresses the problem of Facebook review helpfulness assessment to build a practical predictive approach for Facebook review helpfulness. Baseline features are studied on Facebook reviews. Furthermore, we have extended the literature by adding new helpfulness indicators, Facebook behaviors (likes, comments, sharing, and reactions), to make Facebook a source of reviews. In terms of impact levels, the love and like have the highest level, the commenting behavior has a medium level, while and the sharing behavior has the lowest level.

The current study raises several practical implications with outcomes that can be applied to develop a smart review ranking system for Facebook product pages. Notably, when a consumer is looking for reviews of a particular product on its official page, the system can automatically identify helpful reviews as per the combination of helpfulness features of the target product. It is a highly desirable feature as product official pages can offer a deeper level of adaptive filtering. On the other hand, since online viewers usually have limited time to read many product reviews, this system can help users quickly grasp relevant information of the selected products and gain time during their online shopping process.

VI CONCLUSION

This study dealt with the problem of helpfulness prediction of Facebook reviews by building a useful model using five machine learning methods, including RF, BAG, CART, ADA, and ID3. The influential features are related to three identified feature categories, such as Facebook behavior, review quality, and review characteristics. In contrast, the hybrid set of features (Facebook behavior + review quality + review subjectivity) delivers the best predictive results. The RF's performance is better than that of the BAG, CART, ADA, and ID3 classification methods. Moreover, the proposed features' predictive performance was also compared to baseline features using the same data set and three FS methods: RF, ANOVA, and RFE. Experimental results showed that the proposed features (Facebook behaviors) outperform the baseline features in predicting review helpfulness using various evaluation metrics. Then, each feature's importance was also examined, and the list of influential features belonging to each category was highlighted. Variable importance measures revealed that love, like, and sad behaviors are the most significant features. Thus, our proposed features are useful indicators for the helpfulness prediction of Facebook reviews.

This study has theoretical as well as practical implications. This research contributes to the

body of knowledge by examining the effect of facebook behaviors on the helpfulness assessment. The results showed the considerable effect of this type of features to determine helpful reviews. This will help researchers and practitioners understand and explore the online market and the consumers' desire based on facebook behaviors. On the other hand, review platforms could integrate these results in order to encourage consumers to write helpful reviews for example by focusing the linguistic and subjectivity features.
However, the present study has some limitations. First, because of the lack of Facebook reviews in cloud computing, this research utilizes only 5000 Facebook reviews to explore and evaluate the proposed variables' contribution to Facebook review helpfulness. Therefore, future endeavors should include other types of products or different brands to enlarge the number of reviews and further enhance the findings presented in this paper. The second limitation of this study is that only English reviews were considered. Meanwhile, non-English reviews may also provide useful consumer opinion information, which should not be neglected. Hence future work could do experiments on other languages such as the Arabic language.

References

Saud Alashri, Srinivasa Srivatsav Kandala, Vikash Bajaj, Emily Parriott, Yukika Awazu, and Kevin C Desouza. The 2016 us presidential election on facebook: an exploratory analysis of sentiments. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.

Emna Ben-Abdallah, Khouloud Boukadi, and Mohamed Hammami. Spam detection approach for cloud service reviews based on probabilistic ontology. In *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, pages 534–551. Springer, 2018.

Ekaba Bisong. Google colaboratory. In *Building Machine Learning and Deep Learning Models on Google Cloud Platform*, pages 59–64. Springer, 2019.

Qing Cao, Wenjing Duan, and Qiwei Gan. Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems*, 50(2):511 – 521, 2011. ISSN 0167-9236.

Ummara Ahmed Chauhan, Muhammad Tanvir Afzal, Abdul Shahid, Moloud Abdar, Mohammad Ehsan Basiri, and Xujuan Zhou. A comprehensive analysis of adverb types for mining user sentiments on amazon product reviews. *World Wide Web*, pages 1–19, 2020.

Georgiana Craciun, Wenqi Zhou, and Zhe Shan. Discrete emotions effects on electronic word-of-mouth helpfulness: The moderating role of reviewer gender and contextual emotional tone. *Decision Support Systems*, 130: 113226, 2020.

Ernesto de León and Damian Trilling. A sadness bias in political news sharing? the role of discrete emotions in the engagement and dissemination of political news on facebook. *Social Media + Society*, 7(4): 20563051211059710, 2021. doi: 10.1177/20563051211059710.

Michela Del Vicario, Fabiana Zollo, Guido Caldarelli, Antonio Scala, and Walter Quattrociocchi. Mapping social dynamics on facebook: The brexit debate. *Social Networks*, 50:6–16, 2017.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.

Chao Ding, Hsing Kenneth Cheng, Yang Duan, and Yong Jin. The power of the "like" button: The impact of social media on box office. *Decision Support Systems*, 94:77 – 84, 2017. ISSN 0167-9236.

Dongxu Duan, Weihong Qian, Shimei Pan, Lei Shi, and Chuang Lin. Visa: A visual sentiment analysis system. In *Proceedings of the 5th International Symposium on Visual Information Communication and Interaction*, VINCI '12, pages 22–28, New York, NY, USA, 2012. ACM. ISBN 978-1-4503-1782-5.

Mats Ekström and Johan Östman. Information, interaction, and creative production: The effects of three forms of internet use on youth democratic engagement. *Communication Research*, 42(6):796–818, 2015.

Seyed Pouyan Eslami, Maryam Ghasemaghaei, and Khaled Hassanein. Which online reviews do consumers find most helpful? a multi-method investigation. *Decision Support Systems*, 113:32 – 42, 2018.

Raffaele Filieri. What makes online reviews helpful? a diagnosticity-adoption framework to explain informational and normative influences in e-wom. *Journal of Business Research*, 68(6):1261 – 1270, 2015. ISSN 0148-2963. doi: https://doi.org/10.1016/j.jbusres.2014.11.006. URL http://www.sciencedirect.com/science/article/pii/S014829631400349X.

Robin Genuer, Jean-Michel Poggi, and Christine Tuleau-Malot. Variable selection using random forests. *Pattern recognition letters*, 31(14):2225–2236, 2010.

A. Ghose and P. G. Ipeirotis. Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics. *IEEE Transactions on Knowledge and Data Engineering*, 23(10):1498–1512, Oct 2011.

Ralitza Gueorguieva and John H. Krystal. Move Over ANOVA: Progress in Analyzing Repeated-Measures Data and Its Reflection in Papers Published in the Archives of General Psychiatry. *Archives of General Psychiatry*, 61(3):310–317, 03 2004. ISSN 0003-990X.

Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of machine learning research*, 3(Mar):1157–1182, 2003.

David J Hand. Measuring classifier performance: a coherent alternative to the area under the roc curve. *Machine learning*, 77(1):103–123, 2009.

Seoyeon Hong and Glen T. Cameron. Will comments change your opinion? the persuasion effects of online comments and heuristic cues in crisis communication. *Journal of Contingencies and Crisis Management*, 26 (1):173–182, 2018.

Ya-Han Hu and Kuanchin Chen. Predicting hotel review helpfulness: The impact of review visibility, and interaction between hotel stars and review ratings. *International Journal of Information Management*, 36(6, Part A): 929 – 944, 2016.

Wandeep Kaur, Vimala Balakrishnan, Omer Rana, and Ajantha Sinniah. Liking, sharing, commenting and reacting on facebook: User behaviors' impact on sentiment intensity. *Telematics and Informatics*, 39:25 – 36, 2019. ISSN 0736-5853.

Cheonsoo Kim and Sung-Un Yang. Like, comment, and share on facebook: How each behavior differs from the other. *Public Relations Review*, 43(2):441 – 449, 2017. ISSN 0363-8111.

Nikolaos Korfiatis, Elena García-Bariocanal, and Salvador Sánchez-Alonso. Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content. *Electronic Commerce Research and Applications*, 11(3):205 – 217, 2012. ISSN 1567-4223.

Klaus Krippendorff. Reliability in content analysis. *Human communication research*, 30(3):411–433, 2004.

Ludmila I. Kuncheva. A stability index for feature selection. In *Proceedings of the 25th IASTED International Multi-Conference: Artificial Intelligence and Applications*, AIAP'07, page 390–395, USA, 2007. ACTA Press.

Minwoo Lee, Wooseok Kwon, and Ki-Joon Back. Artificial intelligence for hospitality big data analytics: developing a prediction model of restaurant review helpfulness for customer decision-making. *International Journal of Contemporary Hospitality Management*, ahead-of-print, 06 2021. doi: 10.1108/IJCHM-06-2020-0587.

Pei-Ju Lee, Ya-Han Hu, and Kuan-Ting Lu. Assessing the helpfulness of online hotel reviews: A classification-based approach. *Telematics and Informatics*, 35(2):436 – 445, 2018. ISSN 0736-5853.

Sangjae Lee and Joon Yeon Choeh. The determinants of helpfulness of online reviews. *Behaviour & Information Technology*, 35(10):853–863, 2016.

Hongliu Li, Xingyuan Wang, Shuyang Wang, Wenkai Zhou, and Zhilin Yang. The power of numbers: an examination of the relationship between numerical cues in online review comments and perceived review helpfulness. *Journal of Research in Interactive Marketing*, 2022.

Hsin-Chen Lin. How political candidates' use of facebook relates to the election outcomes. *International Journal of Market Research*, 59(1):77–96, 2017.

Yi Luo and Xiaowei Xu. Predicting the Helpfulness of Online Restaurant Reviews Using Different Machine Learning Algorithms: A Case Study of Yelp. *Sustainability*, 11(19):1–17, 2019.

MSI Malik and Ayyaz Hussain. An analysis of review content and reviewer variables that contribute to review helpfulness. *Information Processing & Management*, 54(1):88 – 104, 2018.

MSI Malik and Ayyaz Hussain. Exploring the influential reviewer, review and product determinants for review helpfulness. *Artificial Intelligence Review*, 53(1):407–427, 2020.

Noemi Mauro, Liliana Ardissono, and Giovanna Petrone. User and item-aware estimation of review helpfulness. *Information Processing & Management*, 58(1):102434, 2021.

Saif M Mohammad. Sentiment analysis: Detecting valence, emotions, and other affectual states from text. In *Emotion measurement*, pages 201–237. Elsevier, 2016.

Julia Neumann, Christoph Schnörr, and Gabriele Steidl. Combined svm-based feature selection and classification. *Machine learning*, 61(1-3):129–150, 2005.

Thomas L. Ngo-Ye and Atish P. Sinha. The influence of reviewer engagement characteristics on online review helpfulness: A text regression model. *Decision Support Systems*, 61:47 – 58, 2014.

Thomas L. Ngo-Ye, Atish P. Sinha, and Arun Sen. Predicting the helpfulness of online reviews using a scripts-enriched text regression model. *Expert Systems with Applications*, 71:98 – 110, 2017. ISSN 0957-4174.

Yadunath Pathak, KV Arya, and Shailendra Tiwari. Feature selection for image steganalysis using levy flight-based grey wolf optimization. *Multimedia Tools and Applications*, 78(2):1473–1494, 2019.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python.

Journal of machine learning research, 12(Oct):2825–2830, 2011.
Mark J. Pelletier and Alisha Blakeney Horky. Exploring the facebook like: a product and service perspective. *Journal of Research in Interactive Marketing*, 9(4):337–354, 2015.
Gang Ren and Taeho Hong. Examining the relationship between specific negative emotions and the perceived helpfulness of online reviews. *Information Processing & Management*, 56(4):1425 – 1438, 2019.
Björn Ross, Tobias Potthoff, Tim A Majchrzak, Narayan Ranjan Chakraborty, Mehdi Ben Lazreg, and Stefan Stieglitz. The diffusion of crisis-related communication on social media: an empirical analysis of facebook reactions. In *Proceedings of the 51st Hawaii International Conference on System Sciences*, 2018.
Jian Raymond Rui and Michael A. Stefanone. Strategic image management online. *Information, Communication & Society*, 16(8):1286–1305, 2013.
Yvan Saeys, Thomas Abeel, and Yves Van de Peer. Robust feature selection using ensemble feature selection techniques. In Walter Daelemans, Bart Goethals, and Katharina Morik, editors, *Machine Learning and Knowledge Discovery in Databases*, pages 313–325, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.
Atanu Shaw. Council post: Do people really look at facebook reviews? here's why it matters. Available in: https://www.forbes.com/sites/forbescommunicationscouncil/2018/04/27/do-people-really-look-at-facebook-reviews-heres-why-it-matters/#242dde8963b7, April 2018. (Accessed on 05/16/2020).
Michael Siering, Jan Muntermann, and Balaji Rajagopalan. Explaining and predicting online review helpfulness: The role of content and reviewer-related signals. *Decision Support Systems*, 108:1 – 12, 2018. ISSN 0167-9236.
Ivan Smieško. Criminal liability of facebook reaction buttons in environment of the slovak republic as a form of hate speech. *Societas Et Iurisprudentia*, 4:213, 2016.
Erin M Sumner, Luisa Ruge-Jones, and Davis Alcorn. A functional approach to the facebook like button: An exploration of meaning, interpersonal functionality, and potential alternative response buttons. *New Media & Society*, 20(4):1451–1469, 2018.
Sarah Turnbull and Simon Jenkins. Why facebook reactions are good news for evaluating social media campaigns. *Journal of Direct, Data and Digital Marketing Practice*, 17(3):156–158, Feb 2016.
Charlotte Van Hooijdonk and Tessa Van Charldorp. Sparking conversations on facebook brand pages: Investigating fans' reactions to rhetorical brand posts. *Journal of Pragmatics*, 151:30 – 44, 2019.
Kanishk Verma and Brian Davis. Implicit aspect-based opinion mining and analysis of airline industry based on user-generated reviews. *SN Computer Science*, 2(4):1–9, 2021.
Yani Wang, Jun Wang, and Tang Yao. What makes a helpful online review? a meta-analysis of review characteristics. *Electronic Commerce Research*, 19(2):257–284, Jun 2019.
Anne L. Zell and Lisa Moeller. Are you happy for me ... on facebook? the potential importance of "likes" and comments. *Computers in Human Behavior*, 78:26 – 33, 2018. ISSN 0747-5632.
Shasha Zhou and Bin Guo. The order effect on online review helpfulness: A social influence perspective. *Decision Support Systems*, 93:77 – 87, 2017. ISSN 0167-9236.
Yusheng Zhou, Shuiqing Yang, yixiao li, Yuangao chen, Jianrong Yao, and Atika Qazi. Does the review deserve more helpfulness when its title resembles the content? locating helpful reviews by text mining. *Information Processing & Management*, 57(2):102179, 2020.