FAKE REVIEWS DETECTION

ProblemSet

In order to promote their products, people write unworthy positive reviews. In some instances, maliciously negative reviews of other (competitive) products are written to hurt their reputations. The main test is that a word can be both positive and negative in various settings. For instance, a positive opinion is expressed when battery life of a product is described as "long," while a negative opinion is expressed when the start time is described as "long." Another challenge is that people don't always express opinions the same way.
Lastly, there are occasions when individuals make statements that are in opposition to one another, making it hard to understand the nature of opinion. A negative review may conceal a positive impression. Additionally, there may be mixed reviews of the product at times. A person's words and actions can be significantly influenced by their emotions. combining a positive and negative comment with the same emoji. Finding reviews that are not genuine or that are used to influence consumers' opinions becomes even more difficult after these difficulties.

Proposed Solution

This project proposes classifying spammed fake reviews into fake and genuine to address the major issue that opinion spamming poses to online websites. Using the Naive Bayes , SVM, Random forest, Decision Trees algorithm, and Logistic Regression as classifying models, we attempt to label the reviews present in the crawled Amazon dataset. To improve accuracy, additional features like a comparison of the review's sentiment, verified purchases, ratings, product category, and overall score are utilized in addition to the review details. Later on, we will be using advanced classification techniques based on NLP like LSTMs, transformers, recurrent neural networks.

Literature Review

FakeTech: Identifying fake reviews using Collective-Positive Unlabeled Learning →This paper proposed a novel approach towards fake review detection that relies on identifying suspicious patterns in the distribution of positive ratings.

Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews →This paper describes features used in the model including sentiment analysis, part - of - speech tagging. And user behaviour analysis. The author concludes with the approach that it can automatically detect and remove fake reviews, improving the overall quality.

Fake Reviews Detection using Supervised Machine Learning →This paper proposes a feature-based approach to fake review detection that uses ML to analyse various features of reviews which include language used and the reviewer’s history.

Detection of fake reviews using NLP &Sentiment Analysis | IEEE Conference Publication→This paper uses CNN and DT. This system extracts features from the text of reviews using CNN and DT to classify whether they are genuine or fake.

[Detecting Fake Reviews Utilizing Semantic and Emotion Model | IEEE Conference Publication]() → This paper proposes a heuristic-based approach for detecting opinion spam. The approach uses the sentiment of the review and the frequency of specific words to identify spam.

[A Study on Identification of Important Features for Efficient Detection of Fake Reviews | IEEE Conference Publication]() → This paper discusses the impact of the fake review on business, and the different techniques. Used to detect them, including ML, NLP, DL and data mining.

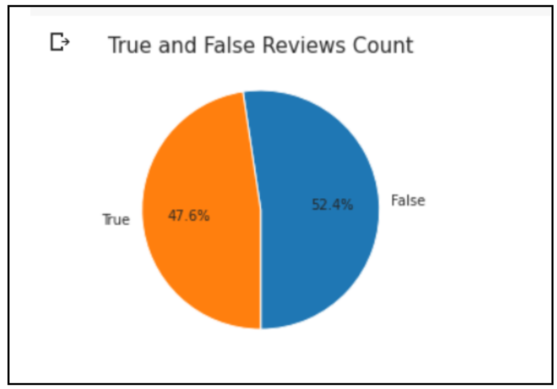
Dataset Description

Data columns (total 32 columns):

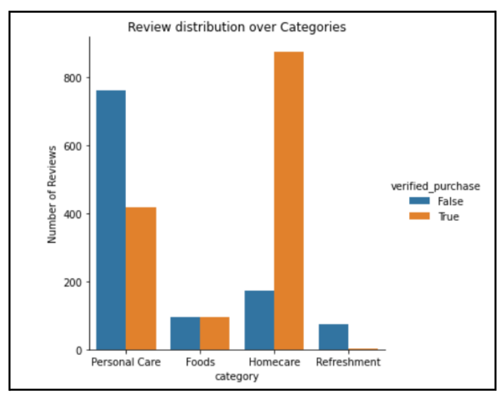
#	Column	Non-Null Count	Dtype
0	report_date	2501 non-null	object
1	online_store	2501 non-null	object
2	upc	2501 non-null	float64
3	retailer_product_code	2501 non-null	object
4	brand	2501 non-null	object
5	category	2501 non-null	object
6	sub_category	2501 non-null	object
7	product_description	2501 non-null	object
8	review_date	2501 non-null	object
9	review_rating	2501 non-null	int64
10	review_title	2403 non-null	object
11	review_text	2501 non-null	object
12	is_competitor	2501 non-null	int64
13	manufacturer	2501 non-null	object
14	market	2501 non-null	object
15	matched_keywords	0 non-null	float64
16	time_of_publication	0 non-null	float64
17	url	1654 non-null	object
18	review_type	2501 non-null	object
19	parent_review	2501 non-null	object
20	manufacturers_response	0 non-null	float64
21	dimension1	2501 non-null	object
22	dimension2	2501 non-null	object
23	dimension3	2310 non-null	object
24	dimension4	0 non-null	float64
25	dimension5	0 non-null	float64
26	dimension6	0 non-null	float64
27	dimension7	2499 non-null	object
28	dimension8	2501 non-null	object
29	verified_purchase	2501 non-null	bool
30	helpful_review_count	2501 non-null	int64
31	review_hash_id	2501 non-null	object

The dataset taken for our model is the Amazon Reviews dataset. It is focused on the products sold on Amazon website and the reviews that are added by the customers about the products. The dataset contains 2500 rows and 32 columns.

Exploratory Data Analysis



Percentage of True and False Reviews in the dataset



No of True and False Reviews per category

WordCloud showing common words in the reviews

Data Preprocessing

> Few columns contain many null values, so replacing these values with the mean values of the columns.
> Duplicate values were also removed.
> Feature Transformation: The data present is categorical, so the string values have been scaled to Integer for model prediction.
> Feature selection: On the basis of Information Gain, some of the attributes like dimensions, time-of-publication, category and few more columns which are product descriptive and which provide low information gain for review (dependent variable) are dropped.

After these steps, text processing was done like spelling correction: using TextBlob()
removing punctuations: using regex,
removing stopwords: from NLTK SW library ,
lemmatisation and tokenisation.

Baseline Models

After preprocessing the data, we have used several Machine Learning models to classify reviews based on the features in the sample.
We have used the following classifiers:

1. Logistic Regression: It is a type of regression used in case of classification problems. It learns a linear relationship from the given dataset and then introduces a non-linearity in the form of the Sigmoid function.

2. Gaussian Naive Bayes: It is a type of classification model which uses Bayes algorithm. It is easy and fast in multiclass classification as it needs less training data. It is used to determine the benchmark performance of the models.

3. Random Forests Classifier: It is ensemble learning of Decision Trees(which provides interpretability and is non-parametric in nature) where some weak classifiers are combined and the prediction is done by majority voting for classification problems.

4. Decision Tree Classifier: A decision tree is a non-parametric supervised learning algorithm which provides interpretability while doing classification. At each level, a feature is chosen as per its information gain or entropy for classifying data and final classification is obtained at the leaf level.

5. SVM : A support vector machine (SVM) is a supervised learning algorithm to classify or predict data groups. The goal of the SVM is to determine the unique decision boundary known as Optimum Separating Hyperplane (OSH) that can segregate n-dimensional space into the required number of regions for classification.

Results and Analysis

We have used Count-vectorizer and Tf-Idf vectorizer as evaluation metrics. Word vectorization is a process of converting words or phrases from a vocabulary into a corresponding vector of real numbers, which can be used to analyze word predictions and semantics. Some of the results are shown below.

Accuracies of various models

	MNB	SVM	LR	DT	RN
Count Vectorizer	80	80	82	76	78
Tfidf Vectorizer	79	82	82	76	76

Precisions of various models

	MNB	SVM	LR	DT	RN
Count Vectorizer	79	74	75	69	71
Tfidf Vectorizer	81	77	81	69	69

Recalls of various models

	MNB	SVM	LR	DT	RN
Count Vectorizer	77	89	91	85	87
Tfidf Vectorizer	72	86	81	85	85

F1-scores of various models

	MNB	SVM	LR	DT	RN
Count Vectorizer	78	81	83	85	87
Tfidf Vectorizer	76	81	81	85	85

From the above results , the overall accuracy were close to 80% for all the models. Also the precision is higher for some models while recall is higher for other models.

So the best method to evaluate is F1-score. According to which, strong classifiers like, Decision Trees , Random Forests and SVM gave better results as compared to weak classifier like Naive Bayes. It is because decision trees works uses nodes to classify the data and works well on categorical data . RF being ensemble learning technique on decision trees also gave good results.

Conclusion

The purpose of the fake review detection is to filter out fake reviews. Random Forests outperformed other classifiers when it came to classifying our dataset in this study. It shows that absolute classifiers are better for this errand. Exploring the dataset was made easier by the data visualization, and the features that were found improved the classification's accuracy. The accuracy of the various algorithms used demonstrates how well they have performed in relation to their accuracy factors.

In addition, the method gives the user the ability to recommend the most honest reviews so that the customer can make decisions about the product.

Future Work

> To make use of real-time or time-based datasets, which will enable us to compare the timestamps of the user's reviews in order to determine whether a particular user is posting an excessive number of reviews in a short amount of time.

> To apply and compare NLP-based models like Lstms and Transformers in order to broaden the scope of the study to deep learning methods.

> To create a method that is comparable for unsupervised learning of unlabeled data in order to identify fake reviews