**Enhancing the Instagram user experience through Advanced Analytics**

A PROJECT REPORT

in partial fulfillment of the award of the Diploma

of

POST GRADUATE DIPLOMA IN MANAGEMENT

AT

INTERNATIONAL MANAGEMENT INSTITUTE, KOLKATA

Submitted By: -

1. DEBANGSHU RATHI 22PGDM016
2. AMAN AGARWAL 22PGDM081
3. KIRAN KUMARI DAS 22PGDM099
4. PRATIK GHOSH 22PGDM112
5. SAMIKSHA JAISWAL 22PGDM198

Submitted to: -

Dr. Arghya Ray

**INTERNATIONAL MANAGEMENT INSTITUTE, KOLKATA**

**BONAFIDE CERTIFICATE**

Certified that this project report on “**Enhancing the Instagram user experience through Advanced Analytics”** is the bonafide work of “Debangshu Rathi, Ritankar Maity, Aman Agarwal, Kiran Kumari Das, Pratik Ghosh, Samiksha Jaiswal” who carried out the project work under my supervision.

**Supervisor**

**Dr. Arghya Ray**

**TABLE OF CONTENT**

|  |  |  |
| --- | --- | --- |
| Serial no | Contents | Page no. |
| 1 | Problem Statement | 4 |
| 2 | Interpretations | 5-9 |
| 3 | Solutions | 10-12 |

**PROBLEM STATEMENT**

As an IT analyst at the company overseeing Instagram, there's a pressing need for an automated system to streamline the analysis of incoming reviews on the platform. Manually reviewing and understanding these reviews is highly time-consuming and inefficient for senior management.

To address this issue, the proposed solution involves implementing machine learning algorithms, namely Naive Bayes for sentiment analysis and topic modeling for identifying key discussion points, on pre-processed and cleaned review data. This system aims to achieve several objectives:

* Predicting Review Ratings: By assessing both sentiment and content, the system will predict the rating (ranging from 1 to 5) of new reviews. This prediction will help in gauging customer satisfaction levels and prioritizing actions based on the anticipated negativity in reviews.
* Sentiment Classification: In addition to predicting ratings, the system will classify the sentiment of reviews as positive, negative, or neutral. This classification will provide clearer insights into the emotional tone of user feedback, guiding efforts to address concerns effectively.
* Identifying Key Topics: Utilizing topic modeling, the system will extract and highlight major themes discussed in reviews. This analysis will unveil crucial areas of customer focus, enabling the prioritization of development efforts and communication strategies.
* Automated Summarization: The system will automatically generate summaries of lengthy reviews. These summaries will offer top management a concise overview of customer feedback, saving time and enhancing their understanding of user sentiments.

This comprehensive system will yield valuable insights into customer sentiments, preferences, and pain points. Ultimately, it will empower us to enhance the Instagram app, address user concerns, and foster a loyal user base.

**THESE ARE THE WAYS DIFFERENT SEGMENTS OF THE CODE HELPED IN PREPARING THE FINAL OUTPUT**

1. Performing appropriate cleansing on reviews in text analytics is crucial for several reasons specific to the field:

* Improves accuracy by removing noise, ensuring more reliable sentiment analysis and topic modeling.
* Mitigates biases by eliminating fake reviews and irrelevant content.
* Enhances the effectiveness of sentiment analysis, topic modeling, and feature extraction.
* Reduces data noise, ensuring more precise and meaningful analysis.
* Maintains consistency in analysis, providing reliable insights over time.

1. Performing filtering on cleansed reviews in text analytics is vital because it:

* Enhances accuracy by ensuring only the most relevant and valuable reviews are considered.
* Improves efficiency by focusing on high-quality data, saving computational resources and time.
* Enables better decision-making by providing more reliable insights from the filtered reviews.
* Refines data for specific analyses, improving the effectiveness of text analytics techniques.
* Enhances user experience by presenting users with the most helpful and pertinent information.

1. Naive Bayes Classification is employed in text analytics for the following reasons:-

* Efficiency: It's computationally efficient and quick in processing large volumes of text data due to its simplicity and independence assumptions.
* Effectiveness: Despite its "naive" assumptions of independence between features, Naive Bayes often performs well in text classification tasks, especially with limited training data.
* Scalability: It scales well with high-dimensional data (such as word frequencies in text), making it suitable for text analytics tasks where the feature space can be extensive.
* Baseline Model: Naive Bayes serves as a good baseline model for text classification. It provides a starting point for more complex models and helps in understanding the problem space.
* Applicability to Text Data: It's particularly suited for text data due to its ability to handle categorical features and is commonly used in tasks like sentiment analysis, spam filtering, and document classification.

1. Performing text analytics based on content words is essential for several reasons:

* Relevance: Content words carry the most meaning in a text, including nouns, verbs, adjectives, and adverbs. Analyzing these words helps focus on the core message or information conveyed in the text.
* Semantic Understanding: Content words provide insight into the semantics and context of the text. Analyzing them helps in understanding the underlying concepts and topics within the text data.
* Reduced Noise: Focusing on content words helps filter out less informative or irrelevant elements like stop words (e.g., "the," "and," "is") or punctuation, reducing noise and improving the quality of analysis.
* Improved Feature Extraction: Content words serve as crucial features in text analytics tasks such as sentiment analysis, topic modeling, and document classification. By focusing on these words, more relevant and meaningful features can be extracted for analysis.
* Enhanced Accuracy: Analyzing content words leads to more accurate results in various text analytics applications, ensuring that the analysis is based on the most informative and meaningful elements of the text.

1. Performing text analytics based on Words

* Comprehensive insights: Analyzing all words ensures a holistic understanding of the text, capturing nuanced meanings, sentiments, and context.
* Minimize bias: Examining all words mitigates cherry-picking or overlooking crucial information, reducing potential biases in the analysis.
* Accurate patterns: Inclusive analysis of all words enables the discovery of intricate patterns, improving the accuracy of findings in text analytics.

1. Performing text analytics based on function words is beneficial for several reasons:

* Understanding Sentiment and Tone: Function words (like articles, prepositions, and pronouns) can reveal the emotional tone and sentiment of a text. Analyzing their frequency and context helps in gauging whether the text is positive, negative, or neutral.
* Author Profiling and Style Analysis: Function words contribute to an author's unique writing style and linguistic fingerprint. Analyzing these words assists in author profiling, identifying writing patterns, and determining attributes such as gender, age, or education level of the writer.
* Content Summarization and Topic Modeling: Function words aid in identifying grammatical structures and sentence formations. Analyzing these helps in content summarization and topic modeling, allowing for the extraction of key themes and topics within a body of text.

1. Performing text analytics based on content words and function words is beneficial for several reasons:

* Understanding Context: Analyzing both content and function words helps in grasping the context of a text, enabling a deeper comprehension of its meaning and nuances.
* Semantic Understanding: Incorporating both types of words aids in extracting the semantic relationships within the text, offering a more comprehensive analysis of its content.
* Improved Accuracy: By considering both content and function words, text analytics algorithms can achieve higher accuracy in tasks like sentiment analysis, topic modeling, and information extraction.

1. Performing text analytics based on n-grams is beneficial for several reasons:

* Language Patterns: N-grams help identify sequential word patterns, aiding in language modeling and understanding context in text analytics.
* Feature Extraction: By breaking text into n-word sequences, they enable efficient feature extraction for tasks like sentiment analysis or machine translation.
* Improved Accuracy: Incorporating n-grams enhances the accuracy of language models by capturing more nuanced relationships between words and phrases in textual data.

1. CONDUCTING SENTIMENT ANALYSIS on the reviews is because:

* Insight Generation: Sentiment analysis helps extract emotions and opinions from text data, enabling businesses to understand customer feelings, preferences, and trends.
* Decision Making: It aids in making informed decisions by gauging public perception, allowing organizations to adjust strategies, improve products, or address issues promptly.
* Customer Engagement: Analyzing sentiment allows for personalized customer interactions, fostering better engagement and tailored services based on individual sentiments and needs.

1. Sentiment Predictor is used for:

* Understanding Public Opinion: Sentiment analysis helps gauge public sentiment towards products, services, or topics, providing businesses and policymakers with insights into public perception.
* Decision Making: It aids in decision-making processes by swiftly identifying positive, negative, or neutral sentiments, enabling timely actions or adjustments.
* Enhancing Customer Experience: By analyzing sentiments in customer feedback, companies can improve their products/services based on customer preferences and address issues promptly.

1. Here Topic Modelling is used for:

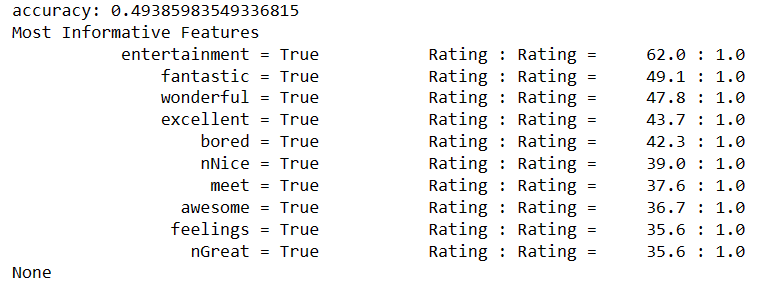
* Identifying Themes: Topic modeling helps in uncovering underlying themes or subjects within a large corpus of text, facilitating the extraction of valuable insights from unstructured data.
* Content Organization: It aids in organizing and categorizing textual data automatically, enabling better content management, retrieval, and understanding.
* Information Retrieval and Summarization: Through topic modeling, text analytics systems can efficiently retrieve relevant information and generate summaries, enhancing decision-making processes and content understanding.

1. Text Summarization is used in the analysis for:

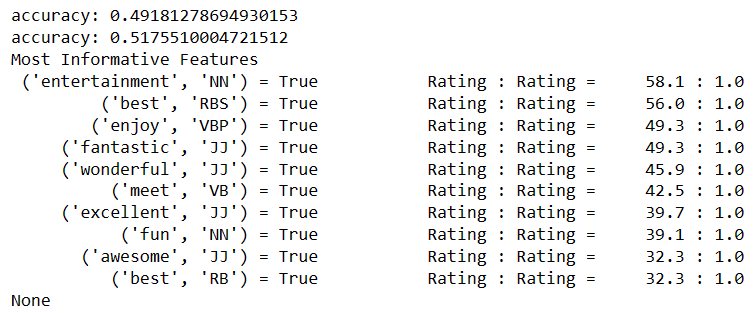
* Concise Information Retrieval: Text summarization condenses lengthy text into a shorter version, aiding quick comprehension and information retrieval.
* Time-Efficient Processing: It helps in swiftly extracting key points, reducing the time needed to comprehend large volumes of text.
* Enhanced Understanding: Summaries offer a clear overview, assisting in grasping the main ideas and important details from complex documents or articles.

**INTERPRETATIONS**

* There are around 2340 reviews out of 50000 reviews that have more than 20 words in it.
* The rating given to the new review is Rating 1 which means very poor.
* There is a strong unfavorable feeling expressed in the latest review. Phrases like "does not work properly," "often fails," "most pathetic app," and the declaration that you would never suggest the app to anybody are examples of key indicators. The user also asks the providers to value the time spent with their customers and to take customer feedback carefully. In general, the review expresses an extremely negative sentiment.
* The most important features according to Function and Content Words: -



* Based on POST tags, these are the outputs:

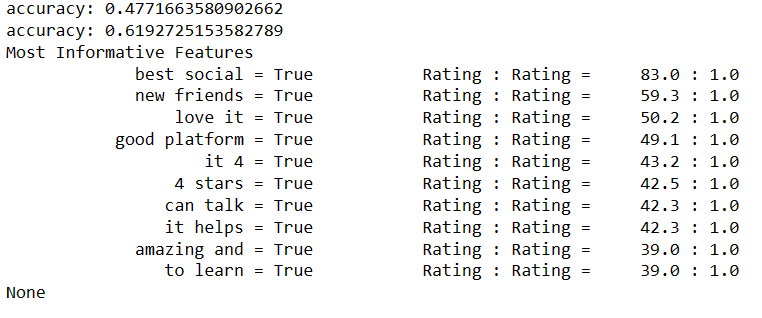


The model's accuracy ranges from approximately 49.18% to 51.76%. The list displays word pairs, their POS tags, and the degree to which they are associated with a certain rating. In contrast to being connected with the general 'Rating' class, the word 'entertainment' is 58.1 times more likely to be associated with a specific rating category when it is classified as a noun ('NN').

Similarly, the most helpful qualities for predicting the rating category are terms like "best," "enjoy," "fantastic," "wonderful," "meet," "excellent," "fun," "awesome," and "best." It appears that the model gives weight to adjectives that are upbeat and enthusiastic, such as "best," "enjoy," "fantastic," etc.

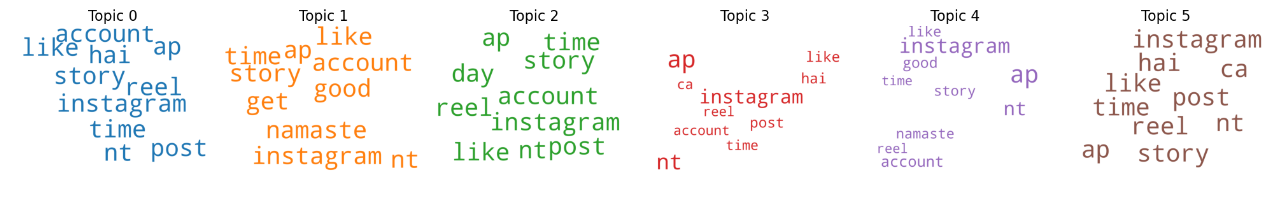
According to the POS tags, several word categories—such as verbs ('VB', 'VBP') and adjectives ('JJ')—have an impact on rating prediction.

* Based on n-grams, these are the outputs:



The model, which used n-grams as features, showed varying accuracy in several runs (0.4771663580902662 and 0.6192725153582789). The list of most informative features provides insight into important terms and phrases that affect how ratings are categorized. Higher scores are noteworthy when it comes to positive sentiments like "best social," "new friends," "love it," and "good platform." Furthermore, as indicated by the corresponding likelihood ratios of 83.0:1.0 and 59.3:1.0, certain n-grams such as "it 4," "4 stars," "can talk," and "it helps" favorably contribute to the prediction of higher ratings. This implies a high correlation between these traits' presence and particular rating classes. Positive experiences, contentment, and social interactions are found to be significant features, highlighting the significance of user sentiment in predicting higher ratings.

* Out of 2340 reviews most of them are positive reviews about the app and neutrals are the least.
* Major topics of all the ratings are-



* The organizational impacts are as follows: -
* Quality Improvement: The unfavorable tone of the most recent review draws attention to particular problems with the app. Enhancing customer happiness and loyalty can be achieved through giving priority to functional enhancements, resolving technical issues, and respecting the time of customers.
* Enhancement of Features: By identifying the features that most closely match favorable evaluations, the company can give these a higher priority and make improvements, which may draw in additional consumers.
* User Engagement: By identifying key themes in reviews, the company may better understand what features customers find appealing or objectionable. This knowledge can direct activities related to user involvement in general, feature updates, and communication tactics.
* Customer Relationship Management: It's critical to actively listen to customer feedback and make the required adjustments after acknowledging the opinions and concerns voiced by users, particularly in the most recent review.