ENHANCING CUSTOMER FEEDBACK INTELLIGENCE ON FACEBOOK: AUTOMATED RATING PREDICTION, SENTIMENT ANALYSIS, TOPIC MODELLING, AND SUMMARIZATION FOR INFORMED DECISION MAKING

An End-Term Project Report

Submitted by:

Group - 02

CHITRAN MUKHERJEE(22PGDM090)

SAYON GHOSH(22PGDM127)

DANIYA NIZAM(22PGDM165)

SHIKSHA JHUNJHUNWALA (22PGDM203)

SIMRAN KHAITAN(22PGDM208)

In partial fulfilment for the award of the degree

Of

Post Graduate Diploma in Management (PGDM)

AT

INTERNATIONAL MANAGEMENT INSTITUTE (IMI), KOLKATA

DECEMBER 2023

TABLE OF CONTENTS

PROBLEM IDENTIFICATION 3
OBJECTIVES 4
UNIT CONTRIBUTION 5
MODEL PERFORMANCE ANALYSIS 8
INTERPRETATION OF THE OUTPUTS 12
RECOMMENDATION TO THE ORGANIZATION 14

PROBLEM IDENTIFICATION

An IT critic has linked the daunting challenge of rooting, analysing, and recapitulating client feedback on Facebook. The main focus is on automated standing predictors, sentiment predictors, content modelling ways for relating crucial discussion motifs, and developing automated textbook summaries.

Data mining involves reacquiring 100,000 Facebook prints, with a specific criterion of no further than 20 words. The thing is to train machine literacy algorithms to directly prognosticate standing sentiment. Sentiment analysis includes relating crucial conversational issues and furnishing perceptivity into client satisfaction. Thematic slice consists of relating six crucial themes from the exploration and relating associated terms.

The thing of creating an automated summary is to condense long analyses and make them easily understood by top operation. Interpretation and reporting bear analysis of results to give meaningful perceptivity for informed decision timber. Integration and robotization bear flawless integration of models and tools and ensures automated analysis of rearmost reviews. Meeting these challenges will enable the company to more assay and understand client feedback, enabling it to make further informed opinions and products for increased overall client satisfaction.

OBJECTIVES

The objectives of the project are as follows.

• Trained on the existing research and presentation to develop rating predictor algorithm using Naive Bayes Classifier by using 1,00,000 Facebook reviews.

• Building a sentiment prediction model for the latest research by classifying sentiment as positive, negative, or neutral.

• Use of thematic models to identify 6 key themes from the research and provide insight into common themes.

• To use the trained model to predict rates and sensitivity to new information.

• Create summaries to summarize longer proposals for senior management.

• Provide recommendations based on studies.

UNIT CONTRIBUTION

Unit 1 Code is about entering and preparing a data set for Facebook-related user review analysis. Notably, the contribution includes data filtering based on a minimum word count of 20 words to ensure that only relevant information is included.

```
# Filter reviews with more than 20 words
df_rev['word_count'] = df_rev['content'].apply(lambda x: len(str(x).split()))
df_rev = df_rev[df_rev['word_count'] > 20]

# Extracting 50,000 reviews
num_samples = min(50000, len(df_rev))
selected_reviews = df_rev.sample(num_samples, random_state=42)
```

In **Unit 2** A Naive Bayes classifier model is trained on a selection of searches with more than 20 terms and ranked from 1 to 5. Contribution depends on training a model that is able to quantify the prediction.

```
# Step 2: Automatic Rating Predictor (Naive Bayes Classifier)
X_train, X_test, y_train, y_test = train_test_split(
    selected_reviews['content'], selected_reviews['score'], test_size=0.2, random_state=42
)
```

Unit 3 Introduces a machine learning pipeline for sensitivity analysis, including Naive Bayes classification with TF-IDF vectorization. This integration enables the model to understand the meaning of words in any literature and enhances the ability to describe context and meaning in different words.

```
# Create a pipeline with TF-IDF vectorizer and Naive Bayes Classifier
model = make_pipeline(TfidfVectorizer(), MultinomialNB())

# Train the model
model.fit(X_train, y_train)
```

Unit 4 Evaluates the accuracy of the model and provides a quantitative indication of its effectiveness in correctly classifying information.

```
# Predict ratings on the test set
y_pred = model.predict(X_test)

# Evaluate the model
accuracy = metrics.accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
```

Unit 5 uses the trained model to predict the identification of some additional information, exemplified by user-supplied information. The contribution lies in the practical application of the model for analytical forecasting.

```
# Prediction for New Comment
new_comment = ""The app does not work properly. It often fails to function at the right time. Don't know howusers have given
new_comment_rating = model.predict([new_comment])[0]

print(f"Predicted Rating for the New Comment: {new_comment_rating}")
```

Unit 6 includes the integration of the Hugging Face Transformers library to perform sensitivity analysis using pre-trained models. This study contributes by providing a powerful tool for analysing the sensitivity of the data.

```
from transformers import pipeline

# Load the sentiment analysis model
sentiment_analyzer = pipeline("sentiment-analysis")

# Predict sentiment for the new comment
result = sentiment_analyzer(new_comment)

# Print the predicted sentiment
predicted_sentiment_label = result[0]['label']
predicted_sentiment_score = result[0]['score']

print(f"Predicted Sentiment: {predicted_sentiment_label} (Score: {predicted_sentiment_score})")
```

Unit 7 emphasizes the importance of pre-processing information in natural language processing and machine learning. Text data cleaning and tokenization help create more meaningful representations for subsequent modelling steps.

```
# Preprocess the text data
nltk.download('stopwords')
nltk.download('punkt')
stop_words = set(stopwords.words('english'))

def preprocess_text(text):
    words = word_tokenize(text)
    words = [word.lower() for word in words if word.isalpha() and word.lower() not in stop_words]
    return words

processed_docs = df_rev['content'].apply(preprocess_text)
```

Unit 8 focuses on the development of the dictionary and document word matrix using the Gensim corpora Dictionary and the Bag of Words model. This contribution is important for the transformation of unstructured text into a format suitable for machine learning algorithms.

```
# Create a dictionary representation of the documents.
dictionary = corpora.Dictionary(processed_docs)

# Create a document-term matrix.
corpus = [dictionary.doc2bow(doc) for doc in processed_docs]
```

Unit 9 uses the Latent Dirichlet Allocation (LDA) topic model to identify topics in text data. Contributions included pre-processing textual data, generating document term matrices, and extracting potential topics for machine learning models.

```
# Train the LDA model
lda_model = models.LdaModel(corpus, num_topics=6, id2word=dictionary, passes=15)

# Print the topics and associated terms
topics = lda_model.print_topics(num_words=5)
for topic in topics:
    print(topic)
```

Unit 10 demonstrates the use of the LexRank algorithm for extraction summarization, which contributes to feature engineering and text summarization. The brevity of the text contributes to its potential to improve general sampling by focusing on key points to capture valuable information, reduce dimensionality, and reduce noise too appropriately.

```
# Extractive Summarizer
def summarize_text(text, num_sentences=3):
    parser = PlaintextParser.from_string(text, Tokenizer('english'))
    summarizer = LexRankSummarizer()
    summary = summarizer(parser.document, sentences_count=num_sentences)  # Corrected argument name
    return ' '.join(str(sentence) for sentence in summary)

df_rev['summary'] = df_rev['content'].apply(lambda x: summarize_text(x))

# Print the summaries
for index, row in df_rev.iterrows():
    print(f"Original Content:\n{row['content']}\n")
    print(f"Summary:\n{row['summary']}\n{'='*50}\n")
```

MODEL PERFORMANCE ANALYSIS

This is the training part of the model on which we would perform analysis of the data.

The 68% accuracy indicates the average classification correctly by the machine learning model. In other words, 68% of all sampled data points were predicted. Accuracy metrics are often used in classification tasks, where the objective is to classify samples into classes or character sets.

```
# Evaluate the model
accuracy = metrics.accuracy_score(y_test, y_pred)
print(f"Accuracy: {accuracy}")
```

```
Accuracy: 0.6822489754098361
```

The number of predicted new assessments is 1, which means that, according to the model we used, the data are classified or the rating class is predicted to be 1. For sensitivity assessment or rating prediction, they are usually stopped exist for ratings as discrete categories, and 1 is one of these groups. We predict that users will indicate on a scale from 1 to 5, where 1 corresponds to the lowest rating, 5 corresponds to the highest rating, and a projected rating of 1 means the model considers the emotion or quality of the feedback to be very negative.

```
print(f"Predicted Rating for the New Comment: {new_comment_rating}")
```

```
Predicted Rating for the New Comment: 1
```

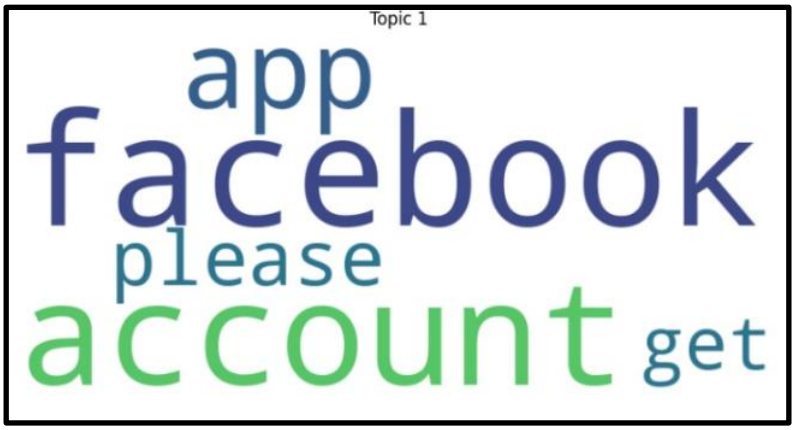
This section shows the sensitivity scores assigned to the inputs by the sensitivity analysis model. In this case, the model predicts that the sensitivity of the text is negative. Sentiment analysis generally involves classifying information into categories such as positive, negative, or neutral based on the emotional tone expressed in the text. The score represents the confidence or probability of the model's predictions. In this model, a high score of 0.999669075012207 indicates that the model is reliable in predicting negative emotions. Scores closer to 1 indicate greater confidence in the predicted category.

```
Predicted Sentiment: NEGATIVE (Score: 0.999669075012207)
```

The results show the identified topics with the most important terms associated with each topic, and their corresponding weights.

Topic 0: Facebook Account Issues: This topic seems to revolve around issues related to Facebook accounts, with significant terms like "Facebook," "account," and "app." Users may be discussing problems or concerns regarding their accounts on the platform.

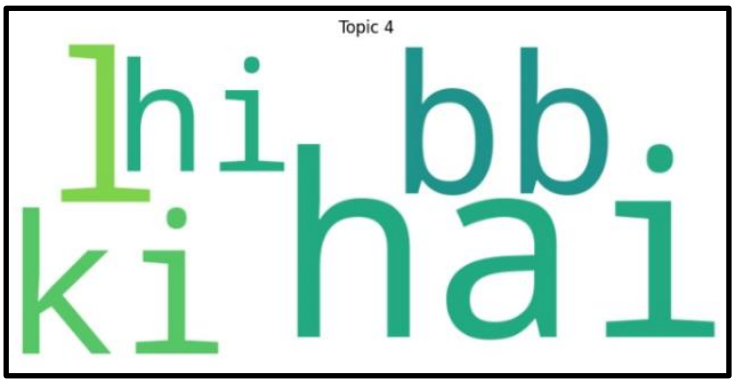
Topic 1: Short Phrases and Characters: This topic appears to contain short phrases and characters such as "u," "na," "r," "mm," and "n." It might represent a set of brief and informal expressions commonly found in social media conversations.



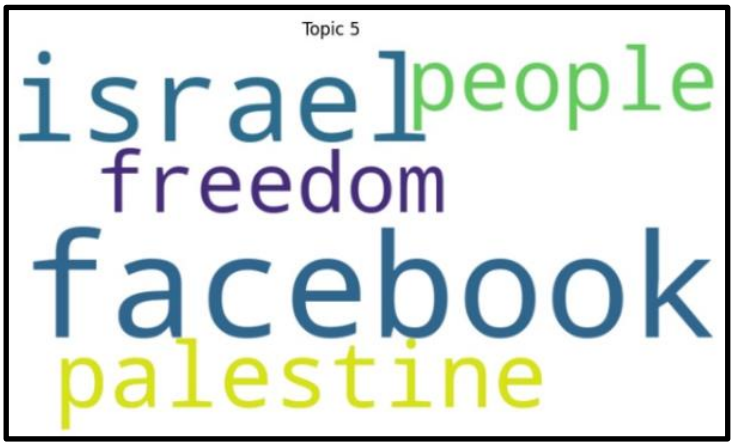
Topic 2: App Update and Functionality: This topic is likely related to discussions about the Facebook app, mentioning terms like "back," "app," "button," and "update." It suggests conversations regarding the functionality and user experience of the app.

Topic 3: Geopolitical Discussion (Israel, Palestine): This topic seems to involve geopolitical discussions, with terms like "Israel," "Palestine," "people," and "freedom." It suggests that some users are engaging in conversations related to political or social issues.

Topic 4: Multilingual Content: This topic appears to contain terms in different languages, such as "hai," "Facebook," "l," "p," and "se." It may represent multilingual content or discussions involving users who communicate in various languages.



Topic 5: Social and Political Issues (Zionist, Expression): This topic involves terms like "Zionist," "expression," "و" (Arabic for "and"), and "women." It suggests discussions related to social and political issues, potentially including topics like freedom of expression and women's rights.


Topic 5
israel
people
freedom
facebook
palestine

INTERPRETATION OF THE OUTPUTS

There have been many reviews regarding the performance of Facebook, the interpretation of which is as follows:

Positive Aspects of Facebook Reviews

- ***Maintaining Contact with Friends and Family:*** The users seem to value Facebook as a means of maintaining contact with friends and family, particularly those who live far away. This is evident from reviews like “Facebook is how I stay in touch with my college friends who live across the country.,” "Love seeing updates and photos from my kids and grandkids!", "Groups are great for finding like-minded people and sharing interests."
- ***Discovering New Things:*** Reviews like “Always find interesting articles and news stories on Facebook.", "Groups and pages help me learn about new hobbies and activities.", "Recommend groups for events and local businesses, helps me explore my city." attests to the platform's ability to offer users engaging content, news, and the chance to discover novel hobbies and interests through groups and pages.
- ***Expressing Oneself:*** The users appreciate that Facebook lets them share their artistic and musical creations. Reviews such as "Posting updates and pictures is a fun way to keep everyone in my life informed" and "Facebook events make it easy to organize and promote local community initiatives" further emphasize how simple it is to plan and publicize events on Facebook, which is evident by the review “Sharing my art and music on Facebook has helped me connect with potential fans."

Negative Aspects of Facebook Reviews

- ***Technical Issues:*** Reviews like "App crashes constantly, makes it frustrating to use," "Website is slow and buggy, often takes ages to load pages," and "Notifications and messages sometimes disappear or don't work" are common examples of how the users are expressing frustration with technical issues, which also include app crashes, slow loading times, and issues with notifications and messages.
- ***Privacy Concerns:*** Facebook's data collection and usage practices, as well as the lack of control over who can access personal information, are frequently cited as the causes for concern. Reviews such as "Concerned about Facebook's data collection and usage,"

"Lack of control over who sees my information is concerning," and "Targeted ads feel intrusive and creepy" reflect this.

- ***Information moderation:*** The users have bemoaned Facebook's apparent censorship, its ambiguous regulations regarding the removal of offensive and harmful information, and the fact that the platform contains content that needs to be filtered more skilfully. Reviews like "Content removal policies are unclear and inconsistently applied," "Censorship of certain viewpoints is unfair and biased," and "Seen offensive and harmful content that shouldn't be on the platform" serve as examples.
- ***Excessive Ads:*** The users have expressed a great deal of dissatisfaction with the platform due to the abundance of ads, particularly those that are repetitive, irrelevant, and intrusive. Reviews like "Ads are everywhere, interrupting my experience and taking up too much space," "Repetitive and irrelevant ads make me want to leave Facebook," and "Feel like Facebook prioritizes ad revenue over user experience" are indicative of this.
- ***Account Issues:*** Reviews like "Trouble logging in and recovering my hacked account, no support from Facebook.", "Customer service is slow and unhelpful, impossible to get assistance." and "Account locked for no reason, frustrating and time-consuming to resolve" indicate that users have frequently complained about issues with account security, login difficulties, and have received inadequate customer support. As a result, users feel frustrated and unsupported in resolving these issues.

While examining Facebook reviews, a prevailing negative feeling is apparent and is expressed through statements of dissatisfaction, annoyance, perceived unfairness, and worries about excessive components on the network. This sentiment is reflected in a considerable number of user reviews.

On the other hand, although they are less common, positive reviews do exist and they emphasize important aspects of gratitude. Facebook is valued by users for helping them connect with family and friends, find latest content, and express themselves. Nevertheless, compared to the more common unfavourable opinions voiced in the evaluations, these positive opinions are irregular.

RECOMMENDATION TO THE ORGANIZATION

- The study highlights the importance of addressing the many issues raised by users, particularly those related to account management, technical performance, privacy, branding, and advertising experiences If you watch on these issues can generally improve the platform user experience in the following ways.
- Technical improvements: To solve this problem and provide a better user experience, Facebook should provide resources to fix technical issues by investing in business improvements and bug fixes
- Facebook can focus on increasing user satisfaction, by improving the reliability and efficiency of its notification messaging system in addition to vague error messages a provided for troubleshooting.
- Increased privacy: To address privacy concerns, Facebook needs to increase transparency about its data collection and use policies.
- A safer and better user experience will come from establishing stronger data security measures and giving customers more control over their data, such as the right to opt out of targeted advertising the results of the building
- Improving data processing: In order to remain consistent and accurate, Facebook needs to analyze and revise its data processing.
- It is important that appeals are easily accessible and that disposal decisions are made in a transparent manner.
- Investing in human content management in sophisticated AI algorithms could improve Facebook's ability to detect and remove harmful content
- Reduce intrusive ads: Facebook should limit the number of ads displayed and increase the value of targeted ads to address user complaints about products the role of over-advertising is huge.
- In addition to emphasizing the value of advertising, providing hierarchical systems or ad-free ordering options will help users gain access to advertising a much better experience.
- Improve account management: According to the survey, Facebook should improve customer support, simplify login processes and strengthen account security measures.

- Facebook can use various strategies to provide a better experience for users, for example, it can increase users' trust in account management by providing better account recovery options, easy and logical to raise.
- Strengthen social engagement: Facebook needs to provide new tools and features that make it easier for users to interact and engage to maximize the quality.
- User experience can be enhanced by promoting content related to personal interactions that are easily recognized by the platform and by encouraging creative community building through group activities and initiatives.
- Curate content discovery: To improve users' experience of content discovery on Facebook, Facebook can create personalized suggestions for news stories and videos based on what users are happy This can improve research capabilities by collaborating with top producers.
- Facebook should conduct regular focus groups and surveys with users to learn more about their needs and wants.
- Facebook should also clearly and actively communicate with users about updates and changes to the platform.