Fake Review Classification and Topic Modeling

Final Project

S.Vinodha

Objectives

Classifying reviews	Classifying reviews (fake or real) using both traditional machine learning , deep learning models, Transformers
Clustering Reviews	Clustering similar reviews to group related feedback together.
Topic Modelling	Identifying underlying topics in the reviews to understand customer sentiments and issues.

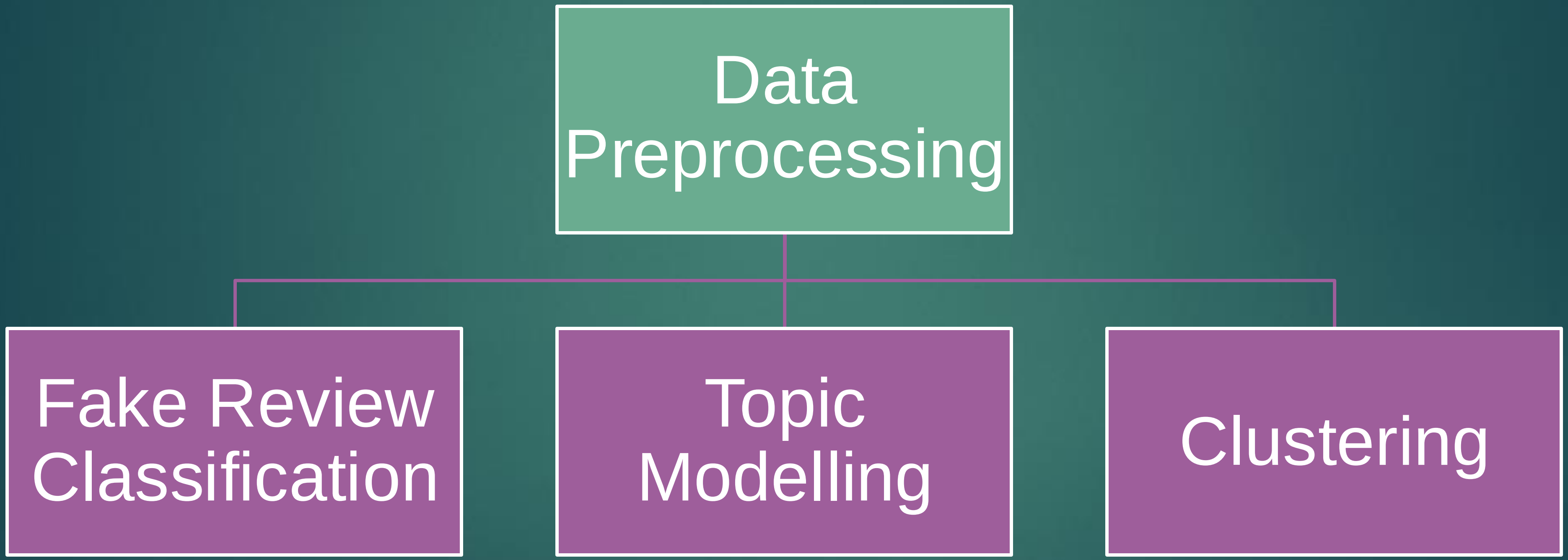
Business Use Cases

CUSTOMER TRUST: DETECTS FAKE REVIEWS TO PROTECT CUSTOMERS FROM MISLEADING INFORMATION.

PRODUCT FEEDBACK ANALYSIS: GROUP SIMILAR REVIEWS TO HIGHLIGHT COMMON PRODUCT ISSUES OR FEATURES.

CONTENT MODERATION: AUTOMATICALLY FILTER OUT FAKE OR HARMFUL REVIEWS FROM PRODUCT PAGES.

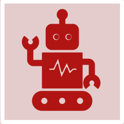
Approach


Data Preprocessing
Fake Review Classification
Topic Modelling
Clustering

Text Preprocessing

Tokenization	• Splits the review text into individual words or subword tokens.
Stopword Removal	• Removes common, unimportant words (e.g., "and", "the").
Lemmatization	• Converts words to their base or dictionary form (e.g., "running" to "run").
TF-IDF Vectorization	• Converts the review text into a matrix of numerical features, capturing the importance of words in each review.

FAKE REVIEW CLASSIFICATION (SUPERVISED LEARNING)



Traditional Machine Learning

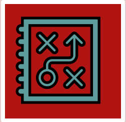
Logistic Regression

Random Forest

Support Vector Machine (SVM)

Deep Learning

LSTM (Long Short-Term Memory)



Transformers (Hugging Face)

BERT (Pre-trained model) for sentiment and text classification.

Traditional Machine Learning

Text Preprocessing

- Remove special characters and numbers.
- Convert to lowercase.
- Tokenize text.
- Remove stopwords

Word Embedding (Word2Vec)

- Words- numerical vectors

Model Training

- Random Forest
- Logistic Regression
- Support Vector Machine

Model Evaluation

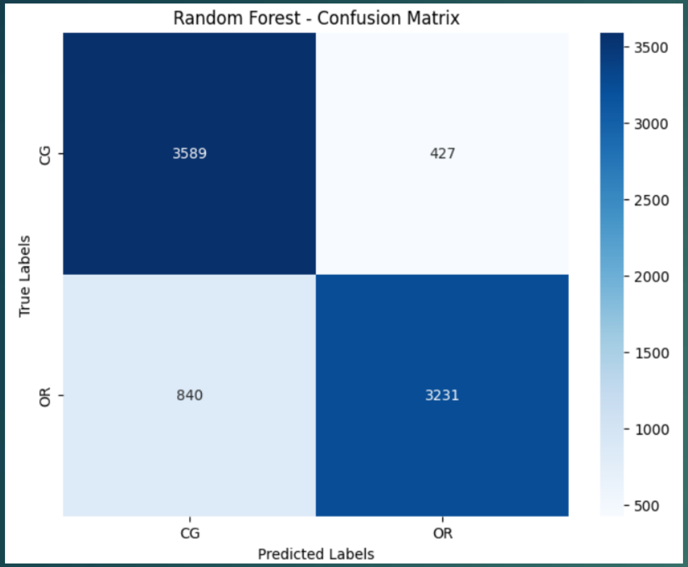
- Accuracy
- Precision, Recall, F1-Score
- Confusion Matrix (Visual Example)

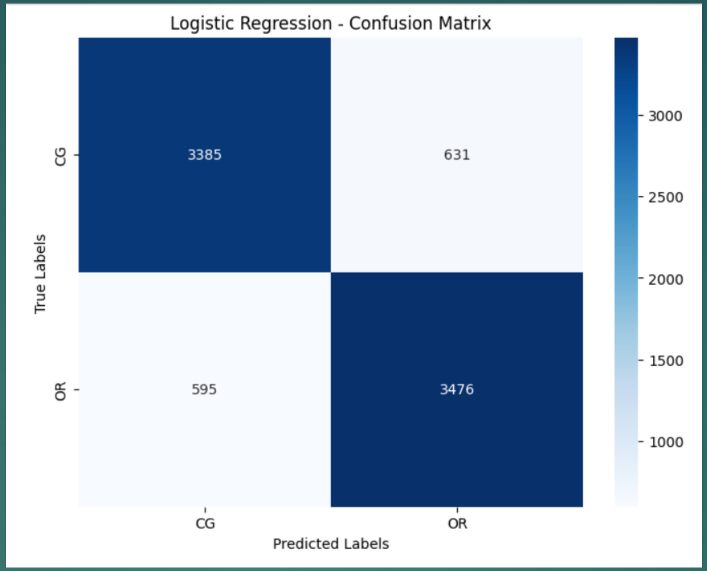
Hyperparameter Tuning

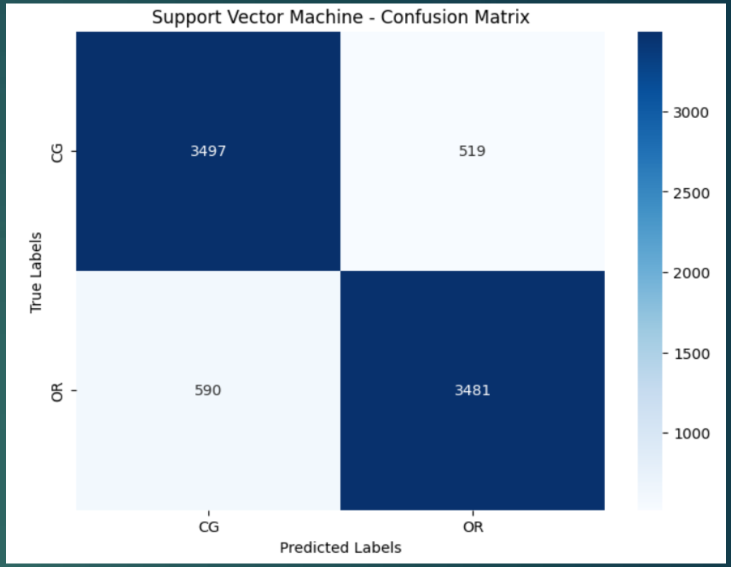
- To optimize model performance.

Model	Precision (CG)	Recall (CG)	F1-Score (CG)	Precision (OR)	Recall (OR)	F1-Score (OR)	Accuracy
Random Forest	0.82	0.89	0.85	0.88	0.80	0.84	0.846
Logistic Regression	0.85	0.84	0.85	0.85	0.85	0.85	0.848
Support Vector Machine	0.86	0.87	0.86	0.87	0.86	0.86	0.864

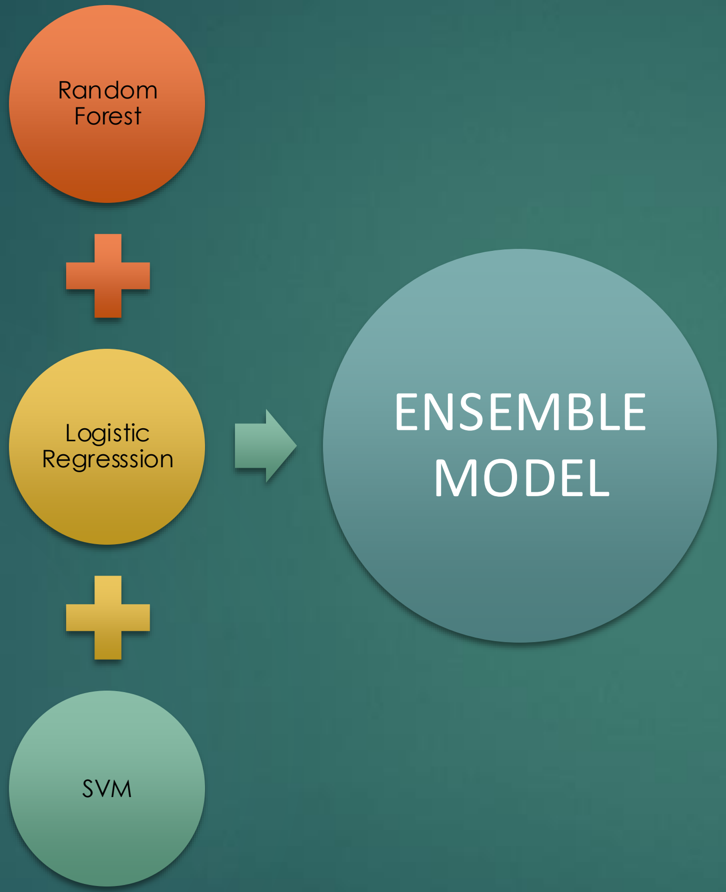
SVM IS THE BEST CHOICE FOR THIS CLASSIFICATION PROBLEM BASED ON OVERALL METRICS.







Random Forest	Logistic Regression	Support Vector Machine
• TP (OR correctly classified as OR): 3231 • TN (CG correctly classified as CG): 3589 • FP (CG misclassified as OR): 427 • FN (OR misclassified as CG): 840	• True Positives: 3476 • True Negatives: 3385 • False Positives: 631 • False Negatives: 595	• True Positives: 3481 • True Negatives: 3497 • False Positives: 519 • False Negatives: 590
Lower recall for OR (79%) shows more OR instances are missed.	Balanced precision and recall for both classes.	High recall for both CG (87%) and OR (86%).



Ensemble Methods:

Combine the outputs from different models to improve overall performance.

Using a weighted average of the predictions from different classifiers.

- Voting Classifier combines the strengths of multiple models to improve performance.
- It uses majority voting ('hard') to decide the final prediction.
- Performance evaluation includes accuracy, classification report, and confusion matrix visualization.
- The model is saved for later use, avoiding the need for re-training.

The model seems to be **balanced** between the two classes.

The **recall for CG (0.88)** is better than that for OR (0.85), which suggests that the model is better at identifying fake reviews than original reviews.

Deep learning model

- **Data Preparation**
  - **Reshaping Data for LSTM**
  - **Encoding**
- **LSTM Model Structure Defining**
  - **First LSTM Layer**
  - **Second LSTM Layer**
  - **Dense Layers**- ReLU activation.
  - **Final layer** - softmax activation
- **Model Compilation**
  - **Optimizer:** Adam
  - **Loss Function**: Categorical cross entropy
  - **Metric:** Accuracy.
- **Model Training**
  - **Evaluation Accuracy: 86.7**
  - **Saving**

Text classification model combining BERT embeddings and LSTM

Model Architecture → BERT Layer: Converts text into embeddings. → LSTM Layer (128 units): Captures sequential dependencies. → LSTM Layer (64 units): Further refines the model's understanding. → Dense Layer (32 units): Fully connected layer to learn complex features. → Output Layer (Softmax): Multi-class classification output.

Model Training →

Batch Processing:
- Text data is processed in batches to manage memory efficiently.

→ **Loss Function:**
- Categorical Cross entropy (since it's a multi-class classification problem).

→ **Optimizer:**
- Adam optimizer used for efficient training.

→ **Training Metrics:**
- Accuracy is tracked during training. 82.96%

Transformers (Hugging Face): Pre-trained models for sentiment and text classification.

Installing Required Libraries

transformers Provides pre-trained models like BERT.

Datasets Offers easy dataset handling.

Pandas For loading and manipulating CSV data.

Loading and Preparing the Dataset

Load CSV Dataset

Convert to Hugging Face Dataset

Mapping Labels/Map Sentiment Labels

- "CG" to 0 (Fake Review)."OR" to 1 (Original Review)./"negative" to 0 and "positive" to 1.

Tokenization and Split Dataset

Load BERT Tokenizer

Split the Dataset into Train and Test Sets (80% Train, 20% Test)

Training the Model

Initialize Trainer: Set Parameters for the Training Process

Pass the model, datasets, and training arguments.

Fine-Tune the Model

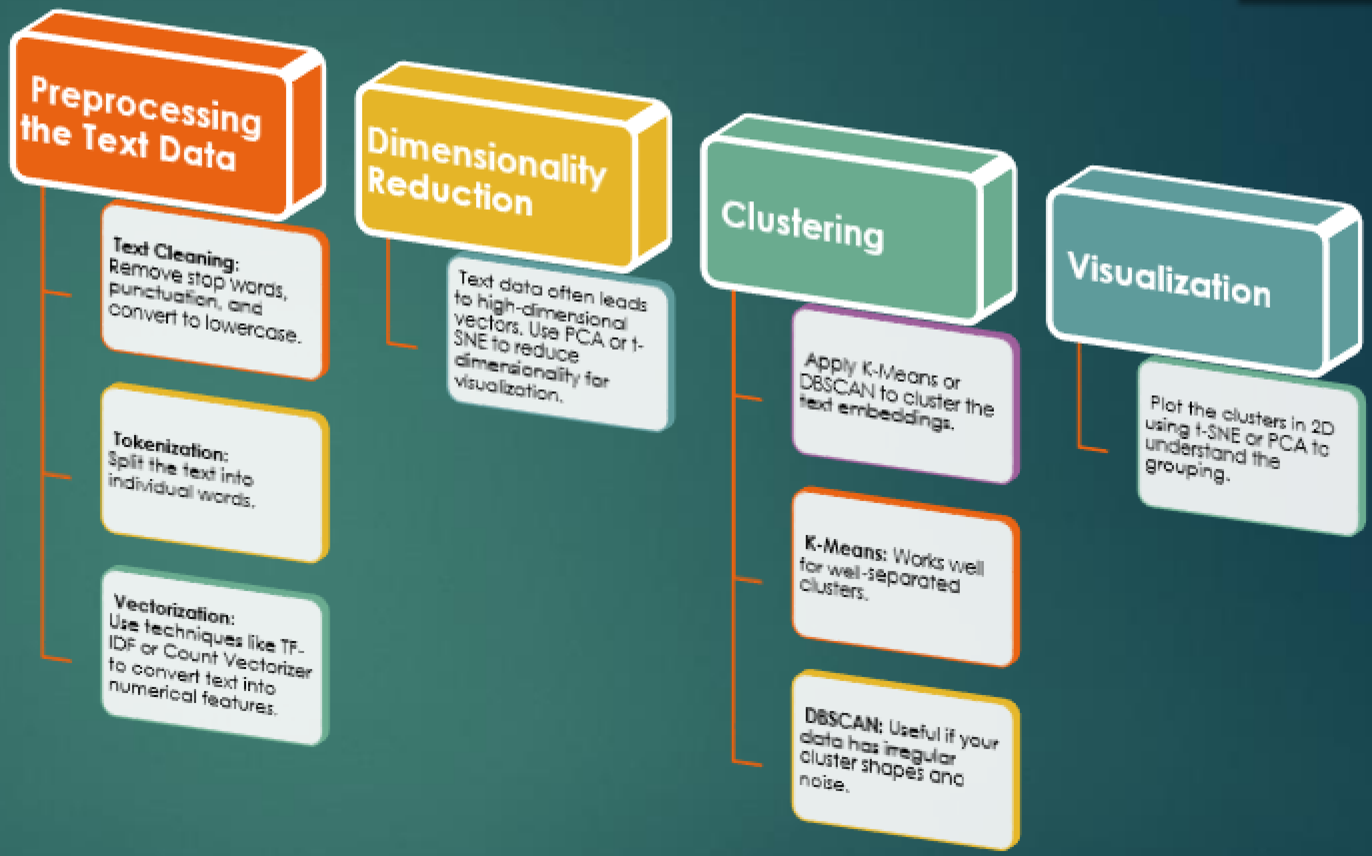
Model Evaluation and Saving

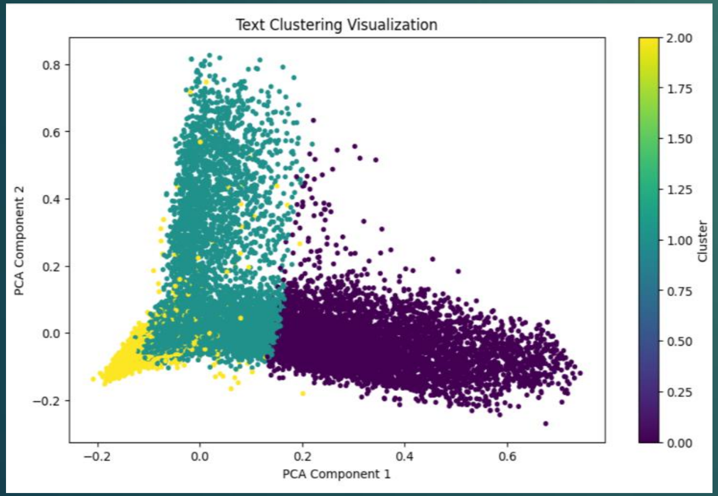
After training, use the Trainer to evaluate F1-score, etc., on the test dataset.

Accuracy=**97%,80%**

Save the trained model for future use.

Clustering (Unsupervised Learning)





Cluster Analysis

Cluster Composition Analysis

- **Cluster 0 (Books/Stories)** - book, read, story, characters, series, author, enjoyed, good, reading, books
- **Cluster 1 (Movies/Entertainment)** - good, movie, love, just, like, little, use, nice, really, time
- **Cluster 2 (Products/Items)** - great, loves, bought, product, dog, quality, size, son, price, little

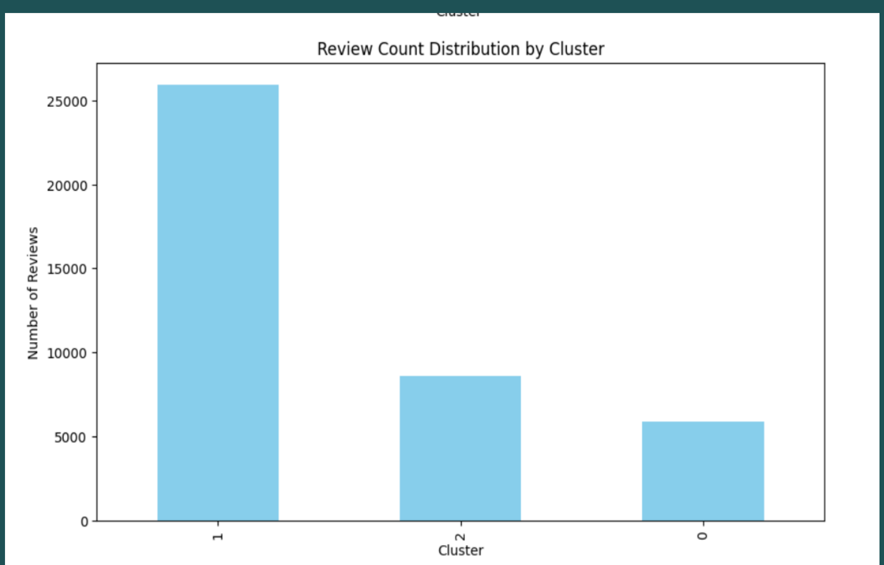
Text Length Distribution

- Cluster 0 – long review- more narrative-style reviews,
- Clusters 1 and 2 – shorter reviews.

Cluster Characteristics

Sentiment Analysis

Review Count Distribution



Topic Modeling (Unsupervised Learning)

Preprocessing:
- It preprocesses the text by removing stopwords using the stopwords.words from NLTK.
- Transforms text data into a document-term matrix using CountVectorizer with a limit on the number of features (max_features=1000).

Topic Modeling:
- Applies LDA to extract topics from the document-term matrix.
- Defines the number of topics (num_topics=5) and uses a random state for reproducibility.

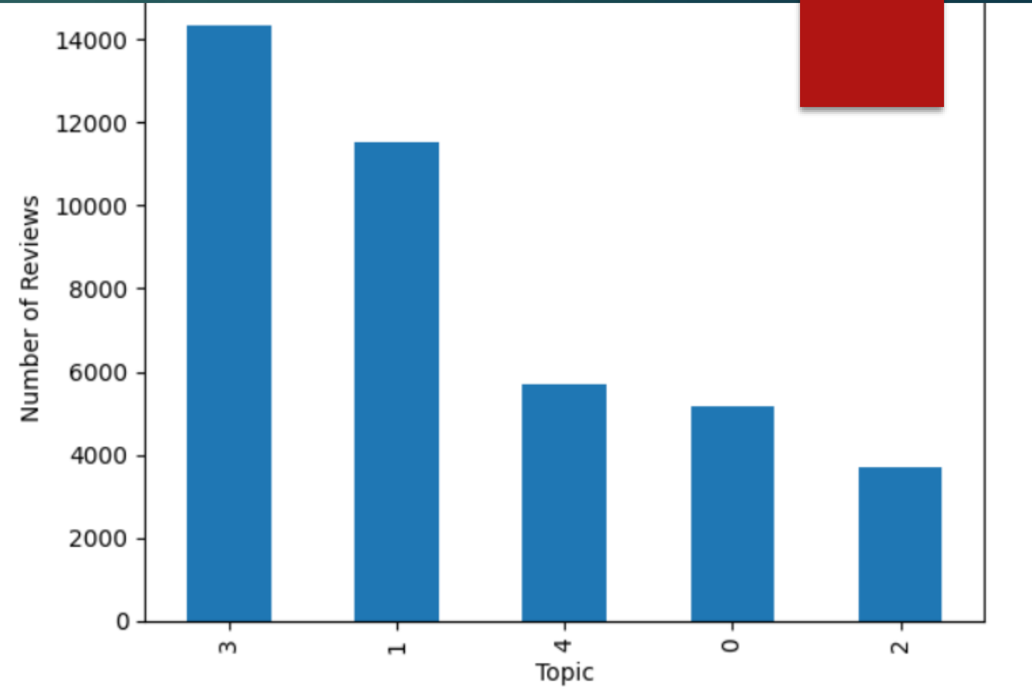
Topic Interpretation:
- Extracts and displays the top n_words=10 most significant words for each topic.

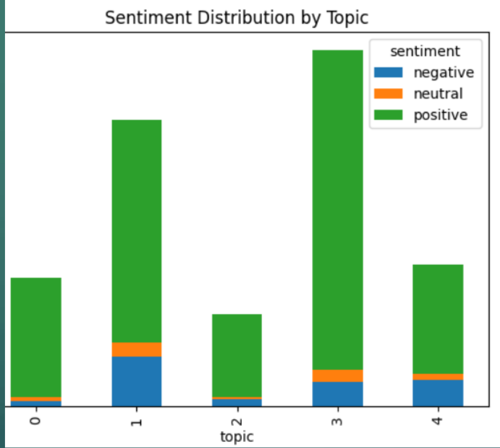
Topic Assignment:
- Uses the LDA output to assign each review the most likely topic.

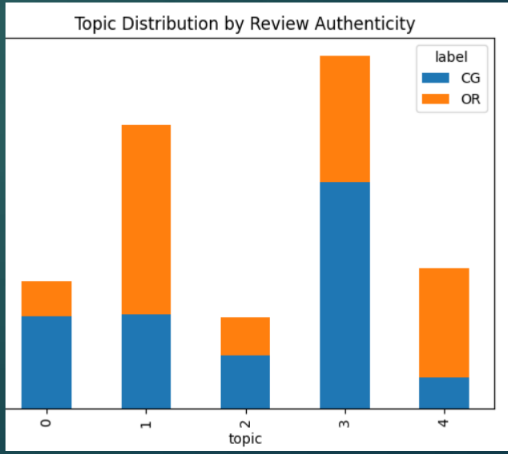
Result Output:
- Displays a preview of the DataFrame with the original text and assigned topic.

Topic Modeling (Unsupervised Learning)

- Topic 1: Reviews about books or stories ("book", "read", "story", "characters", "great").
- Topic 2: Reviews about use of a product ("use", "great", "good", and "light").
- Topic 3: Reviews related to movies ("movie", "acting", "fun", "watch", and "film").
- Topic 4: Reviews on small products ("great", "love", "little", "size", "small").
- Topic 5: Reviews about personal stories ("book", "story", "read", "life", "first").







Insights

Topic 0: Product Quality

Example words: "book", "read", "story", "characters", "good", "quality".

Associated Theme: This topic seems to be related to books or narratives, but also reviews that mention the quality of a product.

Positive sentiment associated with this topic suggests that customers are generally happy with the product's quality.

Topic 1: Usability and User Experience

Example words: "use", "great", "good", "comfortable", "light".

Associated Theme: This topic likely focuses on the usability of products. Reviews in this topic mention how easy or comfortable

a product is to use, and issues like setup or comfort, which can be crucial for user experience.

Topic 2: Product Type or Age Group Target

Example words: "2-year-old", "game", "good", "top-rated".

Associated Theme: This topic revolves around products targeted toward children or specific age groups, focusing on reviews for toys,

games, and other child-friendly items. Positive sentiment suggests these products are well-received.

Topic 3: Shipping and Delivery

Example words: "delivery", "great", "love", "bought", "wide-angle".

Associated Theme: This topic could be related to products that are commonly delivered or shipped.

Reviews may highlight the speed of delivery, packaging quality, and customer satisfaction related to the shipping process.

Topic 4: Price and Value

Example words: "Dracula", "growing up", "first", "life".

Associated Theme: While this topic seems to focus on books or nostalgic themes, it could also involve reviews discussing the value of the product relative to its price. It might include thoughts on whether the product is worth its cost, especially if the reviews mention it in a comparative context.

Insights and Conclusions

Sentiment: Fake reviews (CG) appear more positive on average, likely because they are designed to highlight only positive aspects of a product.

Review Length: Fake reviews are shorter, while original reviews tend to provide more thorough feedback.

Unique Words: The higher count of unique words in original reviews suggests that genuine reviews are more descriptive and personalized.

These differences suggest that fake reviews might focus on generating positive sentiment in a concise format, while original reviews offer a broader and more detailed perspective.

Thank You