Kirtika Iyer (Roll no.: 19)
Chirag Thanth (Roll no.: 50)
Abhay Shukla (Roll no.:57)
Dhanesh Yadav (Roll no.: 67)

Sentiment analysis on McDonald's Store Reviews

Synopsis:

Abstract:

Sentiment analysis has emerged as a crucial technique in natural language processing (NLP) to decipher customer emotions and opinions expressed in textual data. In this project, we conduct sentiment analysis on McDonald's store reviews obtained from the Kaggle dataset, aiming to gain valuable insights into customer perceptions and sentiments towards the renowned fast-food chain. The dataset comprises a diverse collection of customer reviews, encompassing positive, negative, and neutral sentiments, reflecting their experiences at various McDonald's locations. The results of the sentiment analysis unveil a comprehensive overview of customer sentiments towards McDonald's, highlighting aspects that resonate positively with customers and areas that warrant improvement. We gain valuable insights into the factors contributing to customer satisfaction and dissatisfaction, as well as the overall customer experience at McDonald's outlets.

Methodology:

Data Preparation:

i. Obtain the McDonald's store review dataset from Kaggle or other relevant sources.
ii. Perform data cleaning, removing irrelevant information, and handling missing values.
iii. Label the reviews with sentiment labels (e.g., positive, negative, neutral).

Text Preprocessing:

i. Tokenize the text data into words or subwords.
ii. Convert the text to lowercase to ensure uniformity during analysis.
iii. Remove stop words, punctuation, and special characters.
iv. Pad or truncate the reviews to ensure a consistent sequence length.

Data Split:

i. Divide the dataset into training, validation, and testing sets.
ii. The training set is used for model training, the validation set for hyperparameter tuning, and the testing set for final evaluation.

LSTM Model:

i. Build the LSTM model using TensorFlow or Keras.
ii. The model should include an embedding layer to process the word embeddings, one or more LSTM layers for sequence processing, and a final dense layer for classification.

Model Training:

i. Compile the LSTM model with an appropriate loss function (e.g., binary cross-entropy) and an optimizer (e.g., Adam).
ii. Train the model on the training data, using the validation data for monitoring the model's performance and preventing overfitting.
iii. Experiment with different hyperparameters like the number of LSTM units, batch size, learning rate, etc.

Evaluation:

i. Evaluate the LSTM model on the testing set to assess its performance.
ii. Calculate evaluation metrics such as accuracy, precision, recall, and F1-score to measure the model's effectiveness in sentiment classification.

Inference:

i. Use the trained LSTM model to predict sentiment labels for new, unseen reviews.

Interpretation and Visualization:

1. Analyze the LSTM model's predictions to gain insights into customer sentiments towards McDonald's.
2. Visualize the results using plots or charts to showcase the sentiment distribution

Technical Terms:

1. Sentiment Analysis: The process of using natural language processing and machine learning techniques to identify and extract sentiment or emotional information from textual data.
2. McDonald's Store Reviews Dataset: A collection of customer reviews specifically related to McDonald's restaurants, sourced from Kaggle or other data repositories.
3. LSTM (Long Short-Term Memory): A type of recurrent neural network (RNN) architecture designed to handle sequential data with long-range dependencies, making it well-suited for text classification tasks like sentiment analysis.
4. Data Cleaning: The process of removing irrelevant or noisy information from the dataset, handling missing values, and ensuring data consistency.

5. Tokenization: The process of breaking down text into smaller units (tokens), such as words or subwords, to facilitate further analysis.
6. Data Split: Dividing the dataset into training, validation, and testing sets to ensure a fair evaluation of the model's performance.
7. Model Training: The process of feeding the training data to the LSTM model and updating the model's parameters to learn from the data.
8. Hyperparameters: Parameters of the LSTM model that are set before training and affect the model's learning process, such as the number of LSTM units, batch size, learning rate, etc.
9. Evaluation Metrics: Performance metrics used to assess the model's effectiveness in sentiment analysis, including accuracy, precision, recall, and F1-score.
10. Fine-tuning: The process of adjusting hyperparameters or modifying the architecture of the LSTM model to improve its performance.
11. Inference: Using the trained LSTM model to predict sentiment labels for new, unseen reviews.
12. Visualization: Representing the results of sentiment analysis using plots, charts, or other visual representations to showcase the sentiment distribution.
13. Overfitting: When a model performs well on the training data but poorly on new, unseen data due to memorizing the training data instead of generalizing to new instances.
14. Imbalanced Data: When the number of instances belonging to different classes (sentiments) in the dataset is uneven, which can lead to biased model performance.
15. NLP (Natural Language Processing): The field of AI concerned with the interaction between computers and human language, encompassing tasks like text processing, sentiment analysis, and language understanding.
16. Preprocessing: The preparatory steps taken on the text data before feeding it into the LSTM model, including tokenization, padding, and word embedding.
17. Contextual Embeddings: Word embeddings that capture contextual information, representing words based on their meaning in the context of the surrounding words.
18. Deep Learning: A subset of machine learning techniques that use neural networks with multiple layers to learn patterns and representations from data.