Text Preprocessing: The approach involved removing non-alphanumeric characters, converting to lowercase, removing URLs and links, tokenizing, lemmatizing, and filtering out stopwords using NLTK. These techniques help to standardize the text, reduce noise, and remove irrelevant information, improving the model's ability to extract meaningful features.

Feature Extraction: The TF-IDF vectorizer was used to extract features from the preprocessed text. TF-IDF assigns weights to words based on their frequency in the document and across the corpus. This helps the model to focus on important words and down-weight common words, leading to better representation and improved model performance.

Machine Learning Algorithms: Logistic Regression, Naive Bayes, and SVM were experimented with. SVM was chosen as the final model due to its superior recall values for both classes. Model performance was evaluated using accuracy, precision, recall, F1-score, and cross-validation.

Challenges: The class imbalance was a challenge in the dataset. SMOTE was used to address this issue by oversampling the minority class.

Insights: The dataset was imbalanced with more non-disaster tweets. The choice of model and feature extraction techniques significantly impacted recall values, which are crucial for this problem statement to correctly identify disaster tweets.

Class Imbalance Handling: SMOTE (Synthetic Minority Over-sampling Technique) was used to balance the class distribution by creating synthetic samples of the minority class.

Model Selection: SVM was chosen as the final model due to its superior recall values for both classes. Recall is essential in disaster prediction as it ensures timely response and assistance. The model aligns with the problem statement and focuses on correctly predicting disaster tweets.