

Model Development Phase Template

Date	03 October 2024
Team ID	LTVIP2024TMID24974
Project Title	Analysis Of Amazon Cell Phone Reviews
Maximum Marks	6 Marks

Model Selection Report

For analyzing Amazon cell phone reviews, model selection involves choosing algorithms that accurately capture sentiment, aspects, and topics. Popular choices include Naive Bayes, Support Vector Machines and Random Forest. Considering factors like dataset size, complexity, and interpretability ensures optimal model selection, leveraging tools like NLTK, scikit-learn, and TensorFlow to achieve accurate and informative insights. In Model Selection multiple machine learning models are evaluated to identify the best approach for classifying amazon cell phone reviews. The models are compared based on performance, accuracy, complexity, and computational requirements. The aim is to select optimized model for Amazon cell phone reviews.

Model	Description
Random Forest	Random Forest is an ensemble learning algorithm that builds multiple decision trees and merges them to get more accurate and stable predictions. It is well-suited for handling overfitting and works well with complex data structures.
Logistic Regression	Logistic Regression is a linear classification model used to predict the probability of a binary outcome (positive or negative sentiment) by applying a logistic function to a linear combination of the input features. It works well for simple text classification tasks.

Support Vector Machine
(SVM)

SVM is a robust classification algorithm that works by finding the optimal hyperplane that best separates the data points into different classes (positive/negative sentiment). It is especially effective in Working with high-dimensional text data.

Conclusion

After analyzing all these models, we should use **Random Forest** over other models because of its robustness and ability to handle complex datasets without overfitting. Unlike linear models like Logistic Regression and SVM, Random Forest can model nonlinear relationships and interactions between features effectively. It is less sensitive to noisy data and automatically handles missing values, making it more adaptable for diverse text data.