

**SCHOOL OF ENGINEERING****DEPARTMENT OF AI & ML (IIIrd Year II Semester)****Application Development- Web Application with Natural Language Processing & IOT Explore (MR22-1CS0264)****Date: 11-02-25****PROBLEM STATEMENT**

The ecommerce product reviews sentiment analysis project seeks to automate the extraction of meaningful insights from vast amounts of customer feedback available on online shopping platforms. This challenge involves processing unstructured textual data—often laden with noise, informal language, and diverse expressions of sentiment—to accurately gauge customer satisfaction. To address this, the project begins with data ingestion and exploration, where reviews that may include star ratings, review texts, and timestamps are carefully examined for missing values and noise.

The next step involves thorough data preprocessing: cleaning the text by removing irrelevant characters, punctuation, and stop words, and then normalizing it through tokenization, stemming, or lemmatization. Sentiment labeling is achieved by mapping existing ratings to sentiment classes (for example, categorizing 1-2 star reviews as negative, 3-star reviews as neutral, and 4-5 star reviews as positive) or by employing lexicon-based methods. Once the data is prepared, the text is transformed into numerical features using techniques such as Bag of Words, TF-IDF, or word embeddings, making it suitable for machine learning algorithms.

The core of the project is model building, where various machine learning techniques—from traditional models like logistic regression, Naive Bayes, and support vector machines to advanced deep learning models like LSTM are trained to classify the sentiment of new, unseen reviews. These models are rigorously evaluated using metrics such as accuracy, precision, recall, and F1-score to ensure their reliability.

Furthermore, visualizations play a crucial role in the project by illustrating sentiment distributions, identifying trends, and highlighting correlations between review sentiments and other factors like star ratings or review timestamps. The insights derived from these visualizations can drive actionable business decisions aimed at enhancing product offerings and customer service strategies. Overall, this project harnesses Python's extensive ecosystem of libraries—such as pandas, numpy, nltk, spaCy, scikit-learn, and visualization tools like matplotlib and seaborn—to deliver an end-to-end solution for sentiment analysis in the ecommerce context.