Action Plan for Sentiment Analysis of Product Reviews Project

Project Goal Definition

To develop a machine learning model capable of classifying product review sentiments as positive, negative, or neutral.

Data Acquisition

Use datasets such as Amazon Product Reviews or Yelp Dataset for training and evaluation.

Environment Setup

Python with Pandas, NumPy, Scikit-learn, NLTK or SpaCy for NLP, and Matplotlib/Seaborn for visualization.

EDA and Text Preprocessing

Analyze sentiment distribution, review length, and vocabulary size. Preprocess text through tokenization, lowercasing, stop word removal, and stemming/lemmatization.

Feature Extraction

Use TF-IDF and optionally word embeddings to convert text into numerical features.

Model Selection

Implement Naive Bayes, SVM, and Logistic Regression models for sentiment classification.

Model Training and Evaluation

Split data into training and testing sets, evaluate using accuracy, precision, recall, and F1-score, and compare results.

Model Optimization

Perform hyperparameter tuning to enhance performance of the best model.

Documentation and Reporting

Document the process, findings, and conclusions in a final project report.

Timeline

Week 1: Data acquisition and environment setup
Week 2: EDA and preprocessing
Week 3: Model training and evaluation
Week 4: Optimization and reporting