

Action Plan for Sentiment Analysis of Product Reviews

1. Project Title

Sentiment Analysis of Product Reviews using Natural Language Processing and Machine Learning

2. Project Objective

The main objective of this project is to build a machine learning model that can automatically classify product reviews into **positive**, **negative**, or **neutral** sentiments. This will help in understanding customer opinions and extracting meaningful insights from textual review data using Data Analytics, Data Science, and Artificial Intelligence techniques.

3. Dataset Selection and Data Acquisition

- **Dataset:** Amazon Product Reviews
- **Source:** Publicly available datasets from Kaggle or official dataset repositories
- **Data Type:** Text data (customer reviews) along with sentiment labels

Steps:

- Identify and download a suitable sentiment-labeled dataset
 - Load the dataset into the Python environment using Pandas
 - Perform an initial inspection to understand the structure and size of the dataset
-

4. Environment Setup

Programming Language:

- Python

Libraries and Tools:

- Pandas, NumPy – data handling
- NLTK / SpaCy – text preprocessing
- Matplotlib, Seaborn – data visualization

- Scikit-learn – feature extraction and machine learning models
 - Jupyter Notebook – code implementation and documentation
 - GitHub – version control and project management
-

5. Exploratory Data Analysis (EDA)

Planned Activities:

- Inspect dataset structure, data types, and missing values
- Analyze the distribution of sentiment labels (positive, negative, neutral)
- Visualize sentiment distribution using bar charts or pie charts
- Explore review text characteristics such as:
 - Review length
 - Commonly used words
 - Vocabulary size

Outcome:

- Better understanding of the dataset and sentiment distribution
-

6. Text Preprocessing

Steps to be Performed:

- Convert text to lowercase
- Remove punctuation, special characters, and numbers
- Tokenize the text into words
- Remove stop words
- Apply stemming or lemmatization
- Clean and normalize text data

Outcome:

- Cleaned and standardized text suitable for feature extraction
-

7. Feature Extraction

Techniques:

- **TF-IDF (Term Frequency–Inverse Document Frequency)** to convert text into numerical features
- (Optional) Word Embeddings such as Word2Vec or GloVe

Outcome:

- Numerical representation of text data for model training
-

8. Model Building

Machine Learning Models:

- Multinomial Naive Bayes
- Logistic Regression
- Support Vector Machine (SVM)

Steps:

- Split data into training and testing sets (e.g., 80% training, 20% testing)
 - Train each model using the training dataset
 - Generate predictions on the testing dataset
-

9. Model Evaluation

Evaluation Metrics:

- Accuracy
- Precision
- Recall
- F1-score
- Confusion Matrix

Outcome:

- Compare model performance and identify the best-performing model for sentiment classification
-

10. Model Optimization (Optional)

- Perform hyperparameter tuning using GridSearchCV or RandomizedSearchCV
 - Improve model performance if required
-

11. Web Application Development (Optional)

Framework Options:

- Streamlit or Flask

Features:

- Text input box for user reviews
- Button to predict sentiment
- Display predicted sentiment result

Outcome:

- Simple user-friendly interface for sentiment prediction
-

12. Documentation and Reporting

- Prepare a research report explaining theoretical concepts
 - Document all code with comments and markdown explanations
 - Save notebooks and scripts clearly labeled
-

13. GitHub Repository Management

Planned Actions:

- Create a GitHub repository for the project
 - Upload all project files (reports, notebooks, scripts)
 - Maintain proper folder structure
 - Write a detailed README file explaining:
 - Project overview
 - Setup instructions
 - How to run the code
-

14. Project Timeline

Task	Duration
Dataset collection & setup	1 day
Exploratory Data Analysis	1–2 days
Text preprocessing	1 day
Feature extraction	1 day
Model building & evaluation	2 days
Optional web app	1 day
Documentation & GitHub	1 day

15. Expected Outcome

By the end of this project, a complete sentiment analysis system will be developed that can accurately classify product reviews into sentiment categories. The project will demonstrate practical skills in NLP, machine learning, and data analytics while following real-world data science workflows.
