Project Title: Amazon Cell Phone Review Analysis Using NLP Techniques

Milestone 1: Project Initialization and Planning Phase

GitHub Link: Project Initialization and Planning Phase
GitHub Link: Project Initialization and Planning Phase

Activity 1: Define Problem Statement

Problem Statement:
A significant challenge exists in understanding customer sentiment from Amazon cell phone reviews, particularly differentiating between positive and negative feedback. This analysis aims to predict customer sentiment based on reviews, providing insights to improve product offerings and customer satisfaction.

Activity 2: Project Proposal (Proposed Solution)

Proposed Solution:
The project, "Enhancing Product Insights with Amazon Review Analysis," will leverage NLP techniques to analyze customer reviews. By employing sentiment analysis on a dataset containing review text, star ratings, and reviewer demographics, the goal is to develop a predictive model that can classify reviews as positive, negative, or neutral. This initiative will help stakeholders make data-driven decisions to enhance product features and marketing strategies.

Activity 3: Initial Project Planning

Initial Planning:

- **Objectives:**
  - Analyze review sentiment.
  - Identify key factors influencing customer satisfaction.
- **Scope:**
  - Focus on text analysis and sentiment classification.
- **Stakeholders:**

    - Data scientists, product managers, marketing team.
- **Timeline:**
    - 4 months from data collection to model deployment.
- **Resources:**
    - Access to datasets, NLP libraries (e.g., NLTK, SpaCy), and machine learning frameworks (e.g., Scikit-learn).

Milestone 2: Data Collection and Preprocessing Phase

GitHub Link: Data Collection and Preprocessing Phase
GitHub Link: Data Collection and Preprocessing Phase

Activity 1: Data Collection Plan

Data Sources:
The dataset will be sourced from Kaggle, specifically the "Amazon Cell Phone Reviews" dataset. This dataset includes user reviews, ratings, and various product attributes.

Activity 2: Data Quality Report

Data Quality Assurance:

- **Verification:** Ensure all reviews are complete and relevant.
- **Missing Values:** Address missing values through imputation or removal.
- **Ethical Guidelines:** Maintain data privacy and adhere to ethical standards in analysis.

Activity 3: Data Exploration and Preprocessing

Data Exploration:

- Analyze review distributions, word frequencies, and sentiment distributions.
- Visualize trends over time and product categories.

Preprocessing Steps:

- Text cleaning (removing punctuation, lowercasing).
- Tokenization and lemmatization.
- Encoding categorical variables (e.g., ratings).
- Handling missing data.

Milestone 3: Model Development Phase

GitHub Link: Model Development Phase
GitHub Link: Model Development Phase

Activity 1: Feature Selection Report

Feature Selection:
Identify relevant features such as:

- Review text (main feature).
- Star rating (as a target variable).
- Review date (for time-based trends).
- Reviewer demographics (if available).

Activity 2: Model Selection Report

Model Selection:
Consider models such as:

- Logistic Regression for binary classification.
- Random Forest for capturing non-linear relationships.
- XGBoost for optimized performance.
- Support Vector Machines (SVM) for robust classification.

Activity 3: Initial Model Training Code, Validation, and Evaluation

Initial Model Training:
Develop code to train selected models on the preprocessed dataset. Use cross-validation to ensure model robustness.

Validation and Evaluation Metrics:

- Accuracy, precision, recall, and F1-score.
- Confusion matrix visualization.

Milestone 4: Model Optimization and Tuning Phase

GitHub Link: Model Optimization and Tuning Phase
GitHub Link: Model Optimization and Tuning Phase

Activity 1: Hyperparameter Tuning Documentation

Tuning Process:
Utilize techniques such as Grid Search or Random Search to optimize model parameters, focusing on:

- Learning rate for boosting models.
- Maximum depth for decision trees.
- Regularization parameters.

Activity 2: Performance Metrics Comparison Report

Performance Comparison:
Generate a report comparing model performances before and after tuning, highlighting improvements in metrics.

Activity 3: Final Model Selection Justification

Final Model Justification:
Choose the best-performing model based on metrics and alignment with project objectives. Document the rationale for the final selection.

Milestone 5: Project Files Submission and Documentation

GitHub Link: Project Files Submission and Documentation
GitHub Link: Project Files Submission and Documentation

Documentation:
Prepare comprehensive documentation detailing methodologies, findings, and code. Submit project files on GitHub.

Milestone 6: Project Demonstration

GitHub Link: Project Demonstration

GitHub Link: Project Demonstration

Demonstration:
Record a video presentation showcasing:

- An overview of the project.
- A live demonstration of model execution.
- Insights drawn from the analysis and future recommendations.