 

**Project Title: Sentiment Analysis on Amazon Product Reviews**

**1. Dataset Overview:**

- Provide a brief overview of the Amazon product review dataset.

- Describe the columns: `reviewText` (textual content of the review) and `Positive` (binary label, 1 for positive, 0 for negative).

**2. Data Preprocessing:**

- Handle missing values, if any.

- Perform text preprocessing (lowercasing, removing stop words, punctuation, etc.) on the `reviewText` column.

- Split the dataset into training and testing sets.

**3. Model Selection:**

- Choose at least three different machine learning models for sentiment classification. Suggested models include:

- Logistic Regression

- Random Forest

- Support Vector Machine (SVM)

- Naïve Bayes

- Gradient Boosting (e.g., XGBoost)

**4. Model Training:**

- Train each selected model on the training dataset.

- Utilize appropriate vectorization techniques (e.g., TF-IDF, word embeddings) for the text data.

www.aiquest.org

**5. Formal Evaluation:**

- Evaluate the performance of each model on the testing set using the following metrics:

- Accuracy

- Precision

- Recall

- F1 Score

- Confusion Matrix

**6. Hyperparameter Tuning:**

- Conduct hyperparameter tuning for one or more selected models using techniques like Grid Search or Random Search.

- Explain the chosen hyperparameters and the reasoning behind them. **7. Comparative Analysis:**

- Compare the performance of different models based on the evaluation metrics. - Identify the strengths and weaknesses of each model.

**8. Conclusion:**

- Summarize the findings of the project.

- Provide insights into the challenges faced and lessons learned.

**9. Comments:**

- Create well-organized comments on each step of the project. - Highlighting key results, visualizations, and model comparisons.

www.aiquest.org

**10. Code Submission:**

- Include well-commented code for each step in the project. - Submit the code along with the documentation.

This project will assess their understanding of classification models and hyperparameter tuning and their ability to communicate results effectively.

www.aiquest.org