**Analyzing Amazon Product Reviews**

**Overview**

This project aims to analyze Amazon product reviews using various natural language processing (NLP) and machine learning techniques. By analyzing these reviews, we can gain insights into customer sentiments, identify trends, and predict product ratings based on textual data.

**Author**

Saron Yaya

**Initial work**

* [Portfolio Projects](https://github.com/Saron222/PortfolioProjects)

**Released on**

* GitHub

**My professional profile on LinkedIn**

* [My LinkedIn Profile](https://www.linkedin.com/in/saron-yaya/)

**Technologies Used**

* Python
* pandas
* numpy
* matplotlib
* seaborn
* scikit-learn
* imbalanced-learn
* nltk
* textblob
* wordcloud

**Usage**

To utilize this project, follow these steps:

1. **Data Preparation:** Ensure you have the dataset containing Amazon product reviews. In this project, the dataset used is "Datafiniti\_Amazon\_Consumer\_Reviews\_of\_Amazon\_Products\_May19.csv".
2. **Data Analysis:** The project involves various steps such as data cleaning, exploratory data analysis, sentiment analysis, classification, and visualization. Make sure you understand the purpose of each step and its implementation.
3. **Installation:** Clone the repository and install the required dependencies using the following commands:

* git clone <https://github.com/Saron222/PortfolioProjects.git>
* cd PortfolioProjects
* pip install -r requirements.txt

1. **Execution:** Run the Python script containing the project code:

* python Analyzing\_Amazon\_Product\_Reviews.py

1. **Interpretation:** Analyze the generated results, including visualizations, sentiment analysis, classification reports, and accuracy scores, to gain insights into Amazon product reviews.

**Features**

* **Data Cleaning:** Handle missing values, duplicates, and outliers in the dataset.
* **Exploratory Data Analysis (EDA):** Visualize distributions, summary statistics, and correlations in the data.
* **Sentiment Analysis:** Determine the sentiment (positive, negative, or neutral) of each review using TextBlob.
* **Text Classification:** Build machine learning models to classify reviews into rating categories using Support Vector Machines (SVM) and Random Forest classifiers.
* **Visualization:** Generate word clouds, histograms, box plots, and correlation heatmaps to visualize the data and analysis results.

**Contributions**

Contributions to this project are welcome. To contribute:

1. Fork the repository from [Saron222/PortfolioProjects](https://github.com/Saron222/PortfolioProjects).
2. Create your feature branch: **git checkout -b feature/your-feature-name**.
3. Commit your changes: **git commit -am 'Add your feature'**.
4. Push to the branch: **git push origin feature/your-feature-name**.
5. Create a new Pull Request in the original repository.