Project Report: Myntra Review Scraper

Abstract

The Myntra Review Scraper is a web-based project designed to extract, analyze, and visualize product reviews from Myntra and potentially other clothing e-commerce platforms. By leveraging web scraping techniques, machine learning algorithms, and data visualization tools, this project provides users with insightful information regarding product reviews, pricing trends, and ratings. The system enhances online clothing business's decision-making by presenting detailed analytics and comparisons of products. The implementation supports real-time data fetching, classification, clustering, regression, and boosting algorithms for deeper insights. The cross-platform web-based interface ensures usability across different devices.

Introduction

The Myntra Review Scraper project aims to build an intelligent review analysis system that helps online shoppers make informed purchase decisions. By automatically collecting and analyzing reviews, this system delivers real-time insights into product quality, pricing trends, and user satisfaction. The project is an excellent opportunity to test and refine web scraping, data analysis, and machine learning (ML) skills, making it beneficial for both end-users and developers interested in e-commerce analytics.

Literature Survey

Several existing platforms and research studies discuss sentiment analysis, price tracking, and review aggregation. However, most systems focus either on static data or limited filtering options, making it difficult for users to gain in-depth insights into a product's long-term performance.

Relevant studies that support this project include:

- **Sentiment Analysis for E-Commerce Reviews:** Discusses how NLP and ML techniques can be applied to analyze customer sentiments.
- **Web Scraping for Competitive Price Analysis:** Explores how scraping technologies are used in retail to track product prices and trends.
- **Machine Learning for Customer Feedback Classification:** Examines how supervised learning techniques help classify user feedback.

By integrating real-time data scraping, machine learning, and interactive visualization dashboards, this project aims to overcome the limitations of existing review-based analytics tools.

Objectives

The primary objectives of this project are:

- **Review Scraping & Analysis:** Extract product reviews from Myntra and other fashion e-commerce platforms to analyze trends and customer sentiments.
- **Price and Rating Comparison:** Provide users with insights into product pricing fluctuations and average ratings.
- **Machine Learning Integration:** Implement classification, clustering, regression, decision trees, and boosting algorithms for advanced analysis.
- **Data Visualization Dashboards**: Present insights using interactive graphs and dashboards for better decision-making.
- **Real-Time Data Updates:** Ensure that data is continuously updated to reflect the latest market trends.
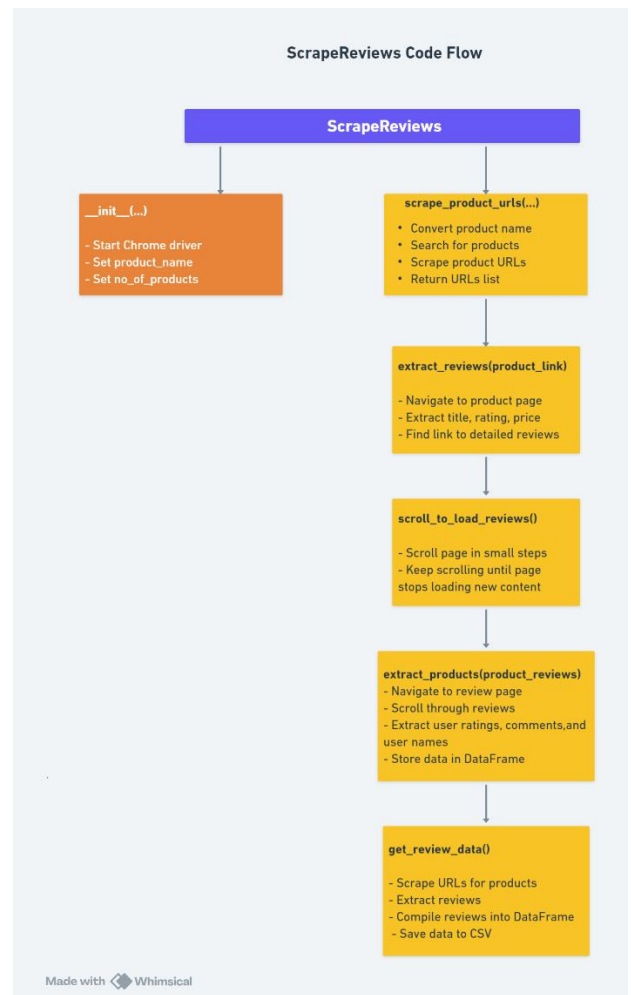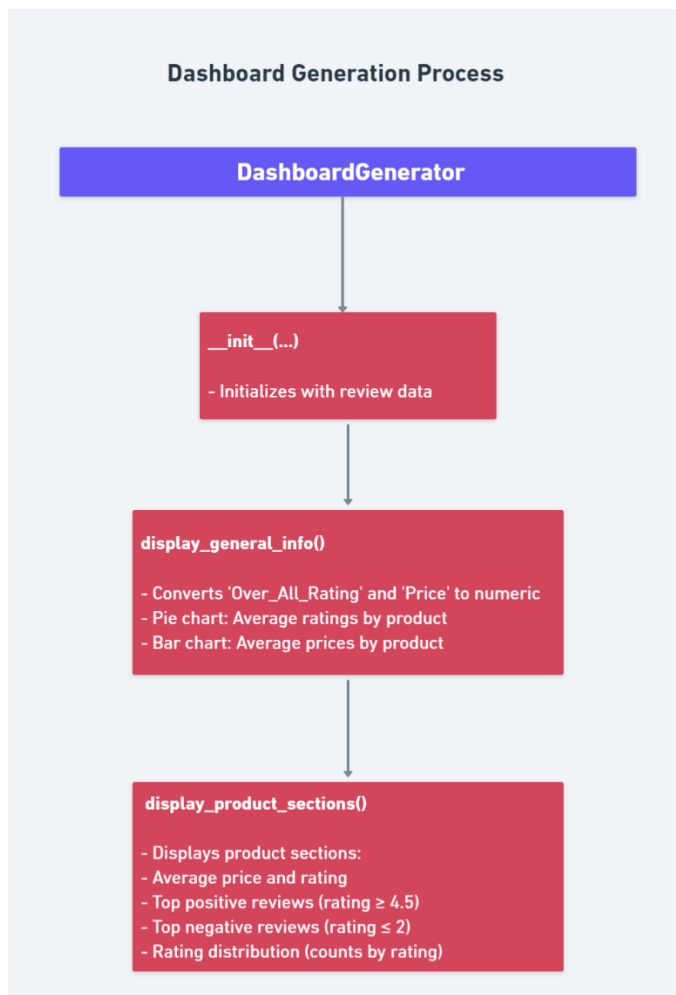
Methodology

The development process follows a structured approach:

1. **Requirement Analysis:** Identify and document detailed project requirements.
2. **Data Collection:** Implement web scraping using BeautifulSoup, Selenium, and related libraries to extract reviews, prices, and ratings.
3. **Database Management:** Store collected data in MongoDB for structured analysis and retrieval.
4. **Machine Learning Implementation:** Apply classification, clustering, and regression algorithms for trend analysis and predictive insights.
5. **Dashboard Development:** Use Streamlit and Plotly to create an interactive visualization platform.
6. **Testing and Deployment:** Conduct rigorous testing across different products and categories, ensuring data accuracy and system stability.

Dashboard Generation Process

The dashboard visualization follows a structured pipeline:

1. Data Initialization
2. Loads review data into the system.
3. General Information Display
4. Converts textual ratings and prices into numeric values.
5. Generates pie charts for average ratings.
6. Generates bar charts for price variations.
7. Product-Specific Analysis
8. Shows average prices and ratings per product.
9. Highlights top positive and negative reviews.
10. Displays rating distribution trends.

Dashboard Generation Process

DashboardGenerator

__init__(...)
- Initializes with review data

display_general_info()
- Converts 'Over_All_Rating' and 'Price' to numeric
- Pie chart: Average ratings by product
- Bar chart: Average prices by product

display_product_sections()
- Displays product sections:
- Average price and rating
- Top positive reviews (rating ≥ 4.5)
- Top negative reviews (rating ≤ 2)
- Rating distribution (counts by rating)

ScrapeReviews Code Flow

ScrapeReviews

__init__(...)
- Start Chrome driver
- Set product_name
- Set no_of_products

scrape_product_urls(...)
• Convert product name
• Search for products
• Scrape product URLs
• Return URLs list

extract_reviews(product_link)
- Navigate to product page
- Extract title, rating, price
- Find link to detailed reviews

scroll_to_load_reviews()
- Scroll page in small steps
- Keep scrolling until page stops loading new content

extract_products(product_reviews)
- Navigate to review page
- Scroll through reviews
- Extract user ratings, comments, and user names
- Store data in DataFrame

get_review_data()
- Scrape URLs for products
- Extract reviews
- Compile reviews into DataFrame
- Save data to CSV

Made with Whimsical

Technologies & Libraries Used

The project utilizes the following technologies:

1. **Web Scraping:** BeautifulSoup, Selenium
2. **Backend & API:** Flask, Flask-CORS, Gunicorn
3. **Database:** MongoDB
4. **Machine Learning:** Scikit-learn, NLP techniques, Deep Learning
5. **Data Visualization:** Streamlit, Plotly

Gantt Chart

A timeline-based representation will outline different phases of development, testing, and deployment.
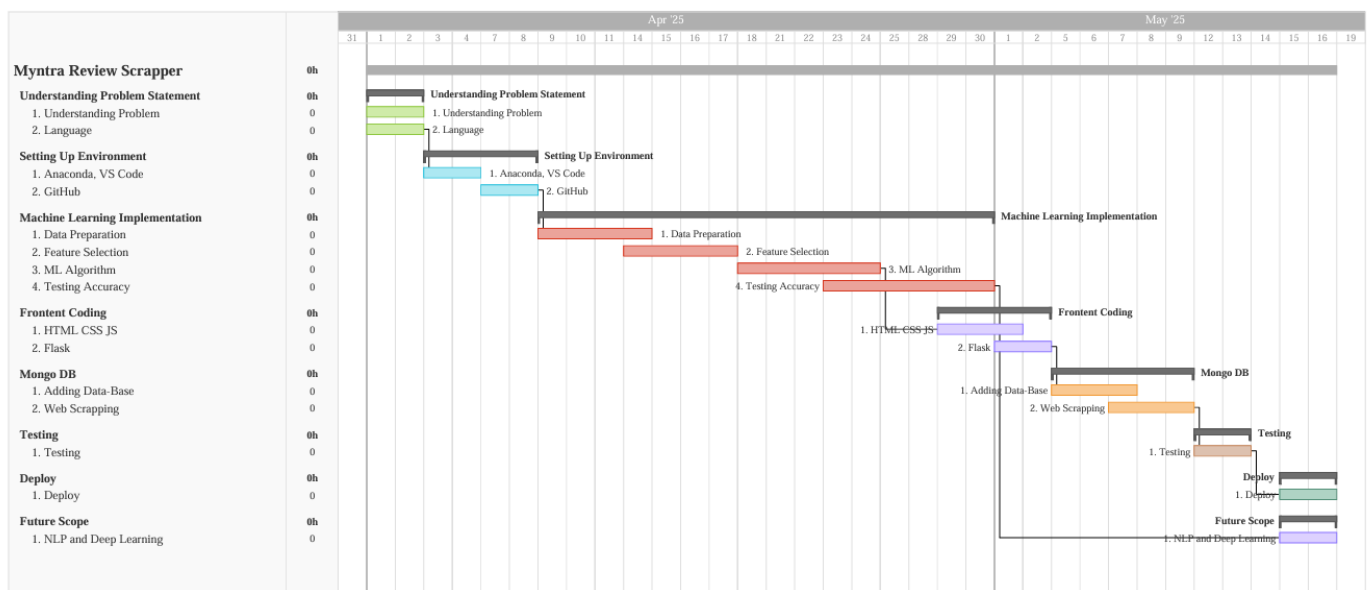
Conclusion

The Myntra Review Scraper project offers a comprehensive solution for online shoppers to make better purchasing decisions by analyzing real-time reviews, prices, and ratings. By incorporating machine learning models and interactive dashboards, the system enhances data-driven decision-making. Future enhancements may include support for multiple e-commerce platforms, improved NLP-based sentiment analysis, and deep learning models for even more accurate predictions. The project stands as a significant step toward intelligent e-commerce analytics, helping users find the best products with data-backed confidence.