AI-Powered Customer Review Insights - Detailed Documentation

1. Introduction

Customer reviews are one of the richest sources of feedback for businesses. However, analyzing them manually is time-consuming and inefficient. This project leverages AI and NLP techniques to automate the analysis of customer reviews. It enables businesses to gain actionable insights from feedback at scale. The system provides sentiment analysis, problem detection, and suggestion extraction from reviews. It visualizes insights through an interactive Streamlit dashboard and allows exporting summaries as CSV or PDF reports.

2. Workflow & System Design

The system follows a modular pipeline for review analysis: 1. Raw Review Data (JSON format) Reviews are collected in JSON structure. 2. Data Ingestion (DataIngester) - Reads the JSON file and normalizes fields such as review_id, date, rating, and text. 3. Preprocessing & NLP (ReviewAnalyzer) - Performs sentiment analysis (positive, negative, neutral). - Extracts problems and customer suggestions using keyword/lexicon-based methods. - Identifies key topics from reviews. - Calculates metadata such as word count, character count, and numeric rating. 4. Structured Data (Pandas DataFrame) - Reviews are converted into structured tabular format for analysis and visualization. 5. Storage & Export - Processed data can be exported as CSV or PDF reports. 6. Visualization (Streamlit Dashboard) - Displays metrics, charts, insights, and review browser.

3. Key Features

- Upload reviews from JSON or CSV file
- Paste JSON text directly for quick testing
- Sentiment analysis (positive, negative, neutral) using lexicon and lightweight models
- Extraction of problems and customer suggestions
- Interactive charts for sentiment and rating distribution
- Insights on top problems and top suggestions
- Review browser for exploring individual reviews
- Export report in CSV and PDF format

4. Example Input and Output

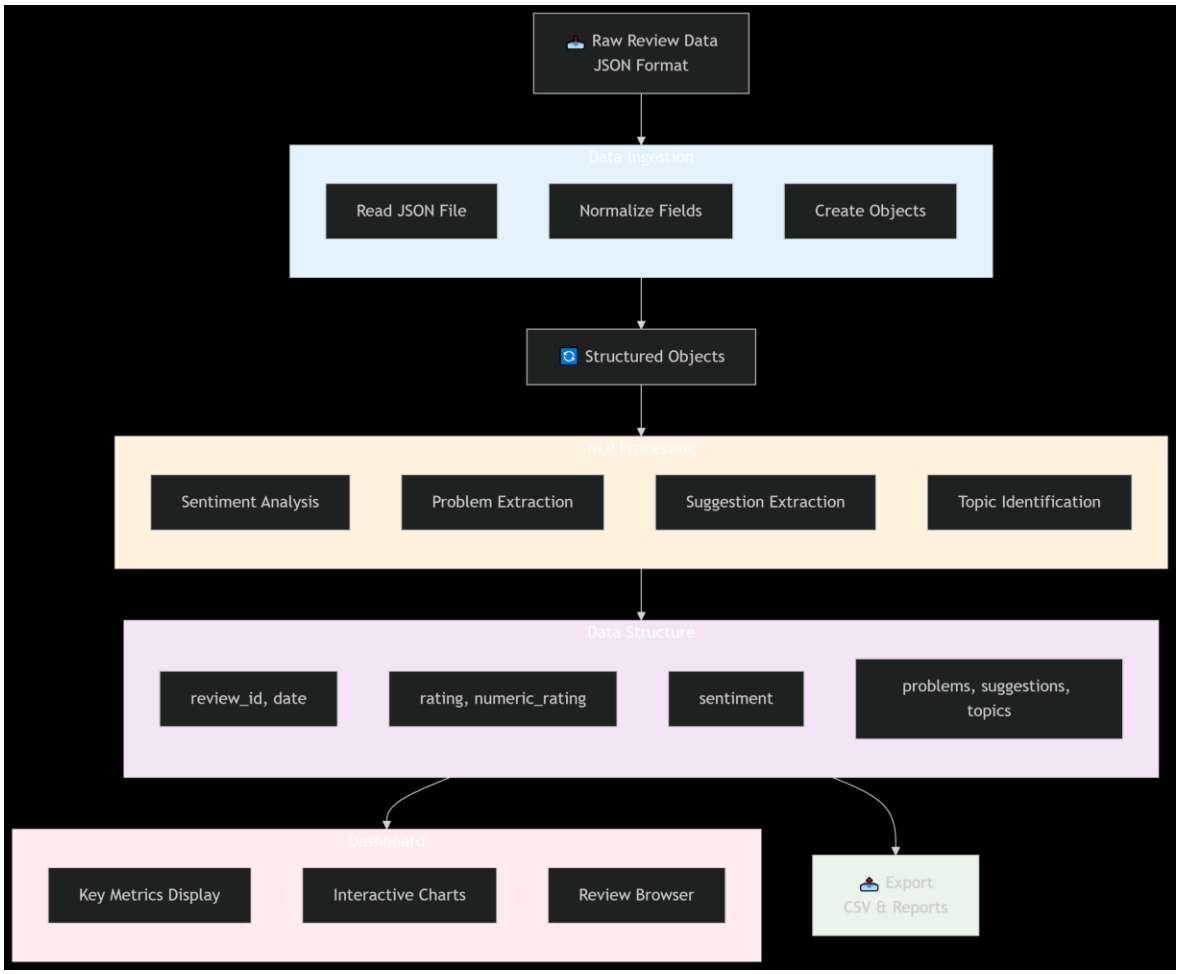
Example Input (JSON):
[
{
"review_id": "R67890",
"date": "2025-01-05",
"rating": "★★★★★ (5 stars)",
"text": "Delivery was super fast and the rider was polite. The food was hot and fresh." }
]

Example Output (Structured DataFrame row):
review_id: R67890

date: 2025-01-05 numeric_rating: 5
sentiment: positive problems: []
suggestions: [] topics: ["delivery",
"food"] word_count: 15 char_count: 78

5. System Architecture

The following diagram shows the flow of data through the system:



6. How to Run

1. Install dependencies from requirements.txt: pip install -r requirements.txt

2. Run the Streamlit app: streamlit run app.py

3. In the sidebar, upload a JSON/CSV file or paste JSON reviews.

4. Navigate through the following tabs in the dashboard:
- Overview: sentiment and rating charts
- Sentiment Analysis: detailed review-level insights
- Insights: aggregated problems and suggestions
- Review Browser: explore reviews one by one

- Export Report: download summary as CSV or PDF

5. Use export options to save structured analysis for reporting.

7. Export Options

The Export Report tab provides the following options: - Download Report Summary (CSV): Includes review-level and aggregated insights. - Download Report Summary (PDF): Includes overview, metrics, top problems, and suggestions.

8. Future Enhancements

Possible improvements for future versions include: - Integration with live APIs (e.g., Google Reviews, Trustpilot) - Multi-language review analysis using pre-trained transformers - Advanced topic modeling using BERTopic or LDA - Dashboard enhancements with drill-down insights and trend analysis