E-Commerce Product Review Analysis

Technology: Python, Scikit-Learn, TF- IDF,
SVM, Flask

Project Brief:
An AI-driven system that automatically analyzes customer product reviews, classifies sentiment in real time (Positive/Negative/Neutral), and identifies key product aspects to help e-commerce platforms make faster, data-driven decisions.

Team / Group Members

Presented By:

M. PRERANA REDDY----- AP23110010033
K. SRAVYA ---------------- AP23110010581
N. SRINITHYA ------------ AP23110010427
B. KARTHIKEYA----------- AP23110010584

Team Strength:
A technically driven group focused on building AI-powered solutions for real-time e-commerce analytics, combining machine learning, NLP, and practical deployment skills.

Problem Statement & Objectives

Problem Statement:

- E-commerce platforms receive massive volumes of product reviews.
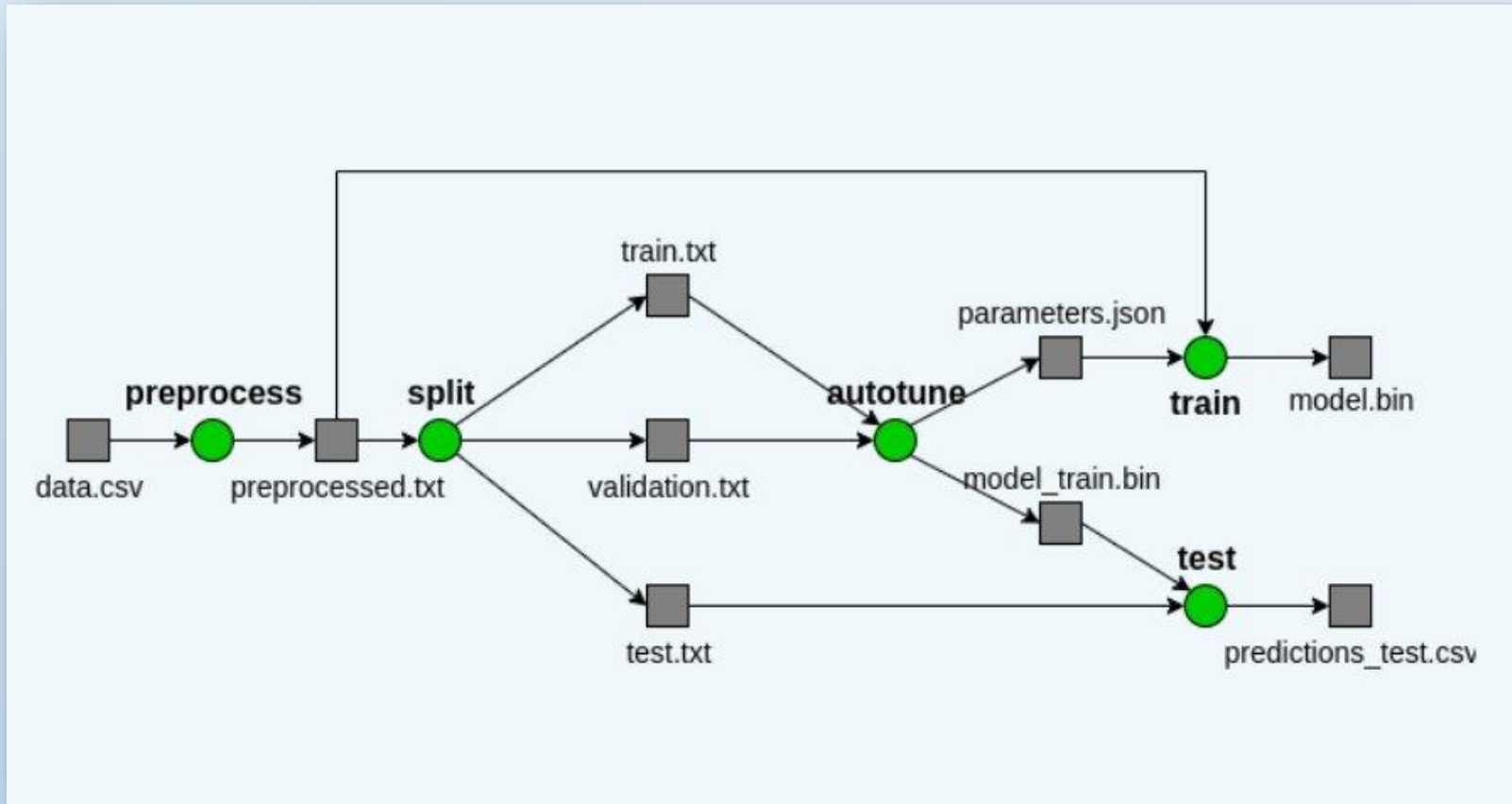- Manual review analysis is slow, inconsistent, and fails to extract actionable insights quickly.

Objectives:

- Build an automated system to classify product reviews as **Positive, Negative, or Neutral**.
- Enable **real-time feedback analysis** to improve product quality and customer satisfaction.
- Identify key **product aspects** (battery, design, performance, shipping, audio).
- Deploy a simple web interface for live predictions.

Proposed Solution & Workflow

Solution:

A machine-learning pipeline using **TF-IDF feature extraction + Linear SVM classifier** integrated into a **Flask web app** for real-time predictions.

Dataset & Methodology:

Dataset:

Primary Dataset: 20,063 reviews
Columns: title, review, rating (1=Negative, 2=Positive)
Sample Dataset: 60 labeled reviews for demo

Methodology:

Preprocessing: Tokenization, stopword removal, normalization
Feature Extraction: TF-IDF (unigram + bigram)
Classifier: Linear SVM
Evaluation: Accuracy, confusion matrix, classification report

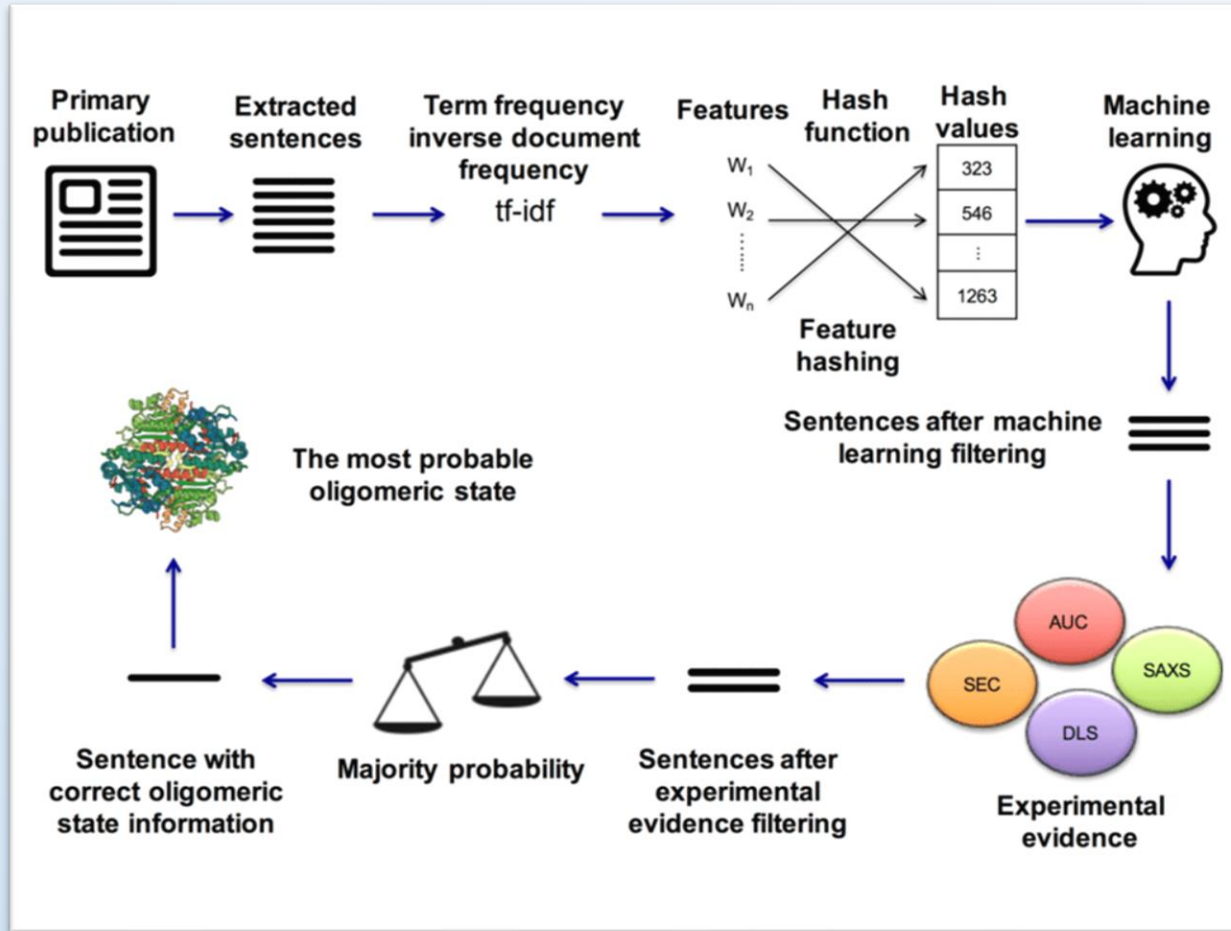# Technologies Used

Languages & Tools:

- Python
- Flask
- Jupyter Notebook
- Scikit-Learn
- NLTK
- Pandas, NumPy

Core Python Files:

ecommerce_sentiment_svm.py — Model training
flask_sentiment_app.py — Web interface
.ipynb notebook — Experimentation & visualization

Results & Findings

Model Performance:

- **Accuracy:** ~87.5%
- **Algorithm:** Linear SVM
- **Features:** TF-IDF bigram vectors
- **Confusion Matrix:** (Add screenshot if needed)

Key Insights:

- Positive reviews dominated electronics products
- Aspect detection helped pinpoint specific issues:
    - Shipping delays
    - Battery complaints
    - Design appreciation
- System handles real-time predictions efficiently

Conclusion & Future Scope

Conclusion:

The SVM-based sentiment analysis system automates review classification, cuts manual workload, and provides meaningful insights to improve product strategy and customer satisfaction.

Future Scope:

Integrate Deep Learning models (BERT, DistilBERT)
Expand aspect detection using NLP techniques
Add trend forecasting using time-series models
Deploy system as a REST API or microservice