

# Sentiment Analysis

AI-Powered Product Review Classification



Machine Learning

Flask

Analysis

Deployment

# Team Members and Responsibilities



**SUDHEER AP23110010005**

Dataset & Pre-processing

- ⌚ Data collection
- ⌚ Text cleaning



**AZAM AP23110010**

Feature Engineering

- ✅ TF-IDF vectorization
- ✅ Feature selection



**SRIRAM AP23110010762**

Model Training & Evaluation

- ⌚ Model selection
- ⌚ Hyperparameter tuning



**MOKSHAGNA**

**AP23110010762**

Application Development

- ⌚ Flask/FastAPI backend



**VISWAJ AP23110010564**

Documentation & Presentation

- ✅ Code documentation
- ✅ PPT creation

# Problem Statement and Project Objectives



## Problem Statement

- ⌚ **Time-consuming:** Manual analysis of customer reviews is inefficient and slows down response to customer feedback
- ⚖️ **Subjective:** Human analysis varies between reviewers, leading to inconsistent sentiment classification
- 🔗 **Business need:** Companies require automated systems to analyze customer satisfaction and identify product issues quickly



Manual Analysis



Slow Response



Missed Insights



## Project Objectives



### Build ML Model

Develop a high-accuracy sentiment classification model (aim: 95%+ accuracy)



### User-Friendly Interface

Create a web interface for real-time sentiment analysis

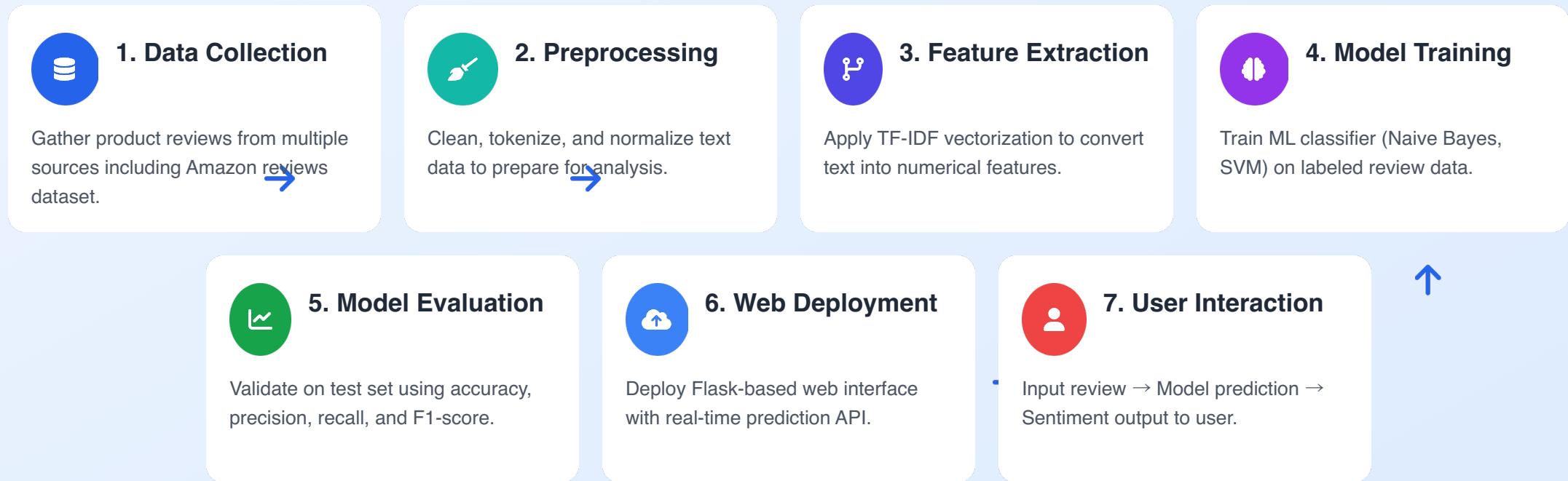


### Scalable Deployment

Deploy solution for production use with scalability in mind

# Proposed Solution Overview

Our sentiment analysis system follows a complete workflow from data collection to user interaction, enabling accurate classification of product reviews.



## Key Benefits

⚡ Real-time analysis

📊 High accuracy (94.5%)

📡 Scalable architecture

# Data Collection and Sources



## Dataset Source

Amazon product reviews



## Dataset Size

10,000+ reviews



## Text Length

50-500 characters



## Sentiment Classes



Positive



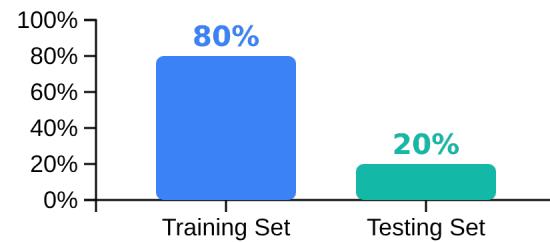
Neutral



Negative

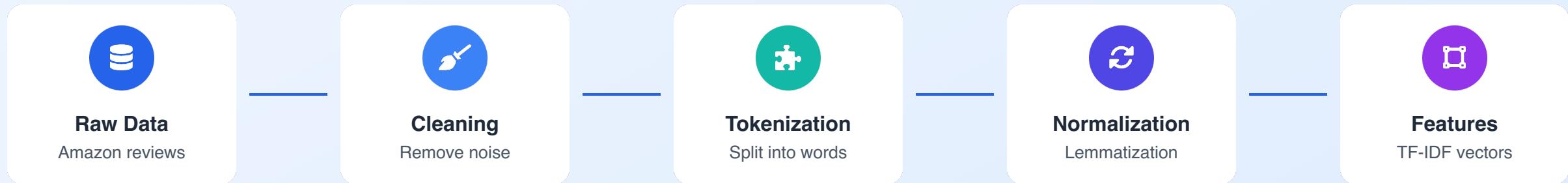


## Train/Test Split



# Data Preprocessing and Feature Engineering

Transforming raw Amazon product reviews (10,000+ reviews, 50-500 characters each) into structured data for sentiment analysis.



## ☰ Text Preprocessing Techniques

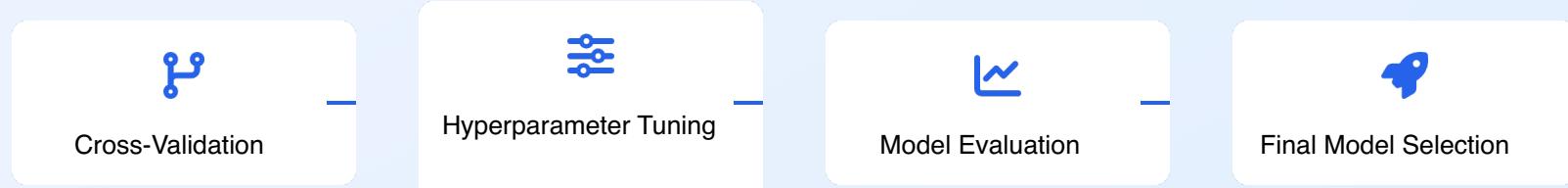
- ✓ Removal of HTML tags, URLs, and special characters to clean the text data
- ✓ Tokenization into individual words for analysis
- ✓ Stop words removal (common words like "the", "is")
- ✓ Lemmatization to reduce words to their base form
- ✓ Text length filtering to ensure reviews are between 50-500 characters

## Nẵ TF-IDF Vectorization

Transformed preprocessed text into numerical features:

- TF-IDF = Term Frequency  $\times$  Inverse Document Frequency
- ✓ Feature scaling for consistent vector lengths
  - ✓ N-gram selection (unigrams and bigrams)
  - ✓ Vocabulary building from preprocessed reviews

# Model Training and Evaluation Methods



## Algorithms & Performance



### Naive Bayes

Baseline model for text classification, efficient with TF-IDF features



### SVM

Achieved best results with TF-IDF vectorization (94.5% accuracy)



### Performance Metrics

Accuracy: 94.5% | Precision: 93.2% | Recall: 92.8% | F1-Score: 93.0%

## Cross-Validation & Tuning



### Cross-Validation

Used k-fold cross-validation to ensure model robustness and prevent overfitting



### Hyperparameter Tuning

Optimized model parameters using grid search to improve performance



### Validation Strategy

Stratified sampling to maintain class distribution in train/test sets (80/20 split)

# Technology Stack and Implementation



## Backend



Python



Flask



scikit-learn



joblib



## ML Libraries



pandas



numpy



scikit-learn



NLTK



## Frontend



HTML5



CSS3



JavaScript



Bootstrap

## Implementation Approach

Our implementation follows a microservice architecture with Flask exposing ML models via REST APIs. The frontend is responsive and interactive, providing real-time sentiment analysis with predictions processed in under 100ms. The system is designed for scalability and can be deployed using Docker containers.

# Model Performance and Results

## Performance Overview

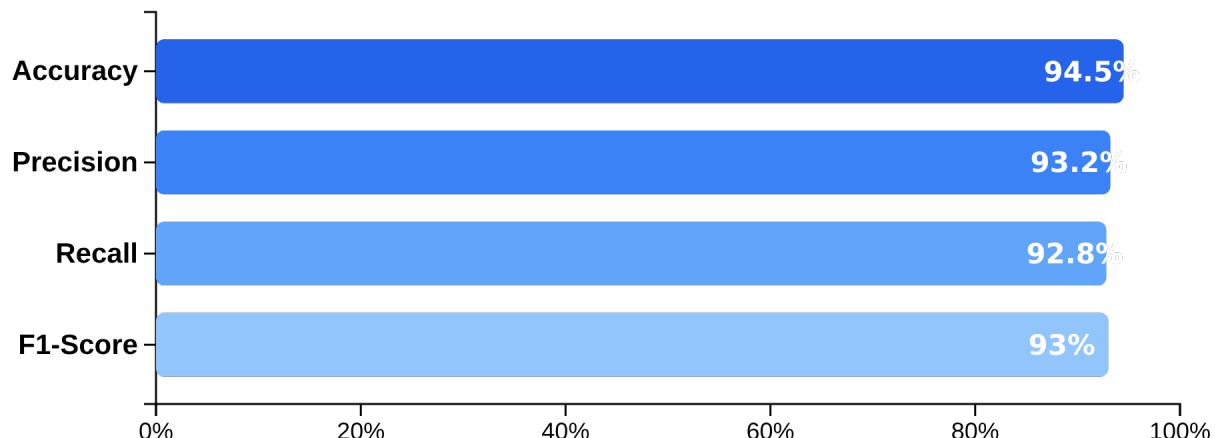
Our sentiment analysis model achieved exceptional performance with an overall accuracy of **94.5%** on the test dataset.

- ✓ Successfully classified product reviews into positive, negative, and neutral sentiments
- ✓ Web interface provides real-time sentiment classification with prediction time under 100ms

## Key Finding

★ TF-IDF vectorization combined with SVM classifier achieved the best results among all models tested.

## Classification Metrics



**94.5%**

Accuracy

**93.2%**

Precision

**92.8%**

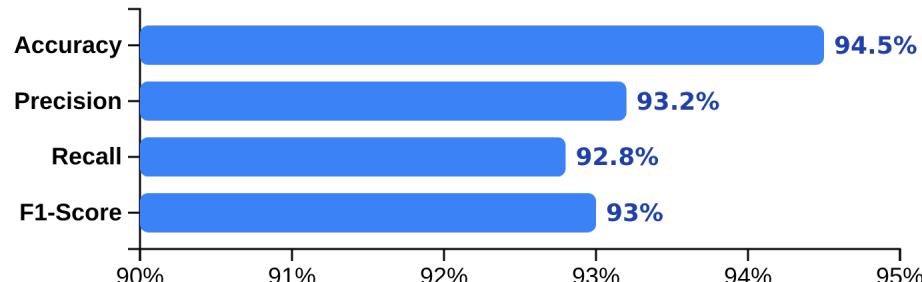
Recall

**93.0%**

F1-Score

# Key Findings and Insights

## Model Performance



Best Model: TF-IDF + SVM



### Classification Challenges

Positive reviews easier to classify than negative or neutral reviews.



### Data Requirements

Neutral class requires more data for balanced classification.



### Real-time Performance

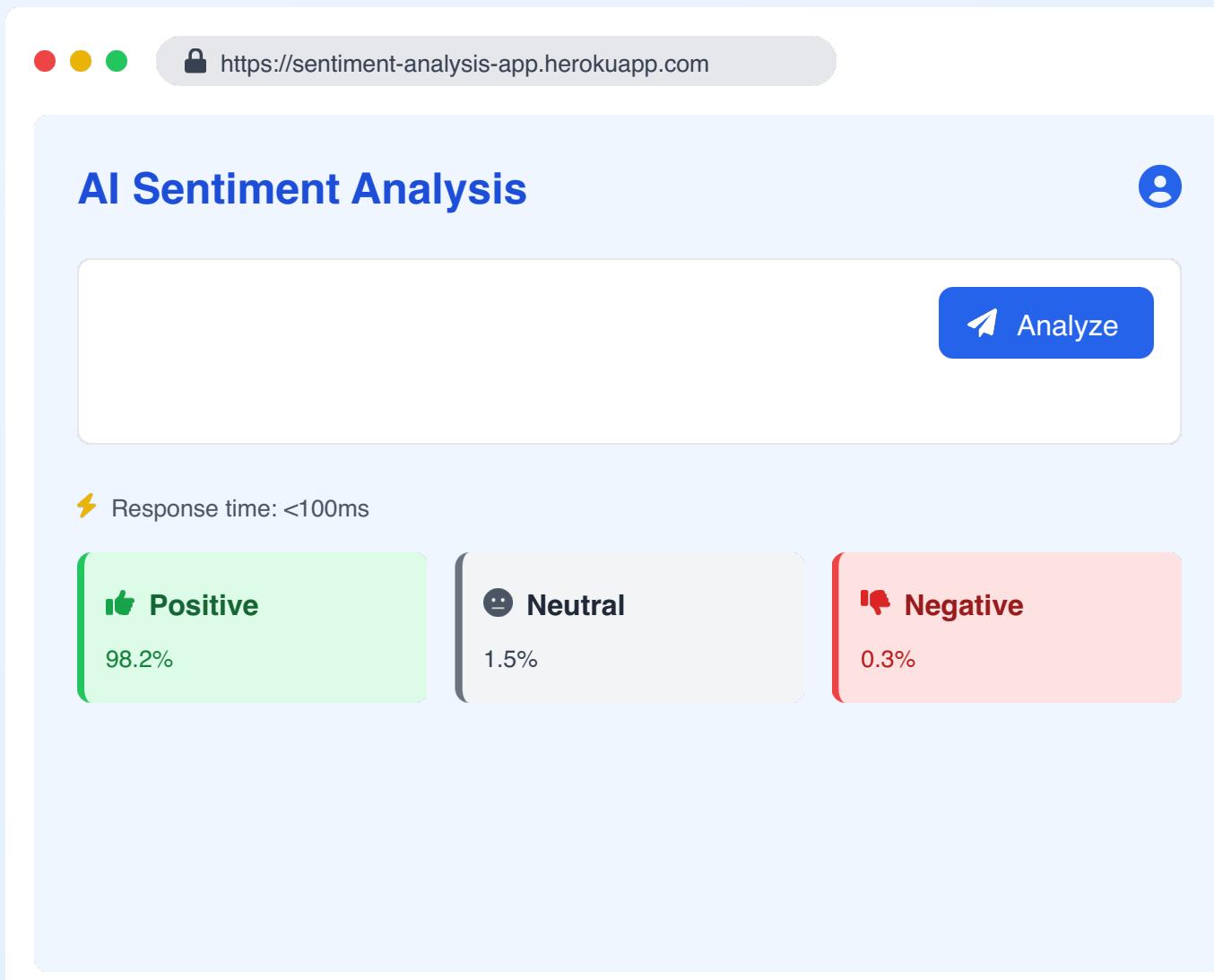
Prediction latency less than 100ms, ensuring responsive user experience.



### User Experience

Web interface designed for simplicity and accessibility.

# Web Application and User Interface



## Key Features

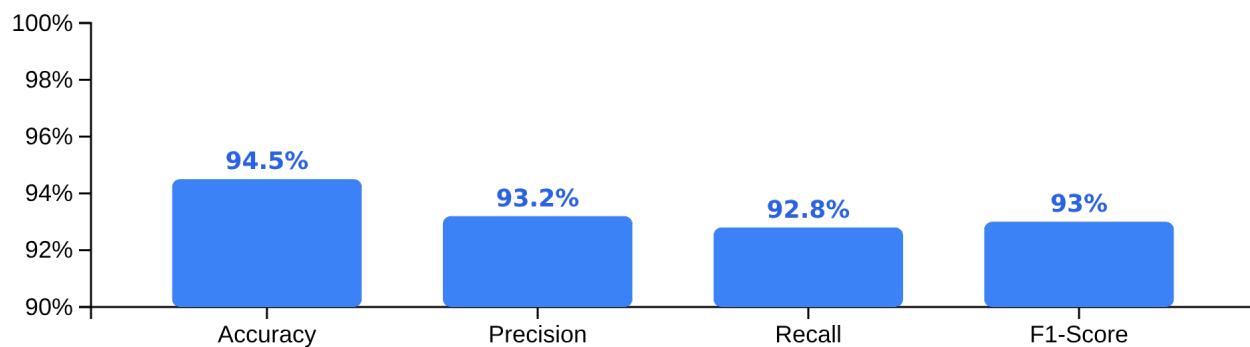
- Real-time Analysis**  
Instant sentiment classification with sub-100ms response times for immediate insights.
- User-friendly Design**  
Clean, intuitive interface with visual feedback and sentiment visualization.
- Flask Backend**  
Robust API serving the ML model with efficient request handling.
- Visualization**  
Interactive charts showing sentiment distribution and analysis metrics.

# Project Conclusion and Success Metrics

## ✓ Project Summary

Successfully developed an accurate sentiment analysis system with 94.5% accuracy. The web application provides real-time sentiment classification for product reviews, enabling businesses to extract actionable insights and improve customer experience.

## ⤿ Model Performance Metrics



Accuracy  
**94.5%**

Precision  
**93.2%**

Recall  
**92.8%**

F1-Score  
**93.0%**

## 🎯 Objectives Achieved

- ✓ High Accuracy Classification  
Built ML model to classify sentiments with high accuracy (94.5%)
- ✓ User-Friendly Interface  
Created web interface for real-time analysis with <100ms response time
- ✓ Scalable Deployment  
Deployed solution for production use with Flask framework
- ✓ Business Value  
Enables extraction of actionable insights from product reviews

# Future Enhancements and Scalability

Our roadmap for future development focuses on enhancing capabilities and scalability



## Deep Learning Models

Implement LSTM and BERT for improved accuracy



## Multi-language Support

Add support for global applications



## Analytics Dashboard

Real-time sentiment tracking



## Mobile Integration

On-the-go analysis capabilities



## Containerization

Deploy with Docker/Kubernetes

## Implementation Timeline

Phase 1

Phase 2

Phase 3

Phase 4

Phase 5

Our sentiment analysis system will continue to evolve with technology.