# **Community Pulse**

# **Product Requirements Document (PRD)**

## **1. Introduction & Background**

**Product Name:** Community Pulse

**Purpose:** Monitor and analyze public sentiment in East Bay communities (Oakland, Berkeley, etc.) through Reddit and other sources, providing visual dashboards, alerts, and insights.

**Primary Audience:**

* Local government agencies
* Neighborhood organizations and nonprofits
* Researchers and community advocates
* Concerned citizens

## **2. Product Vision**

Empower local stakeholders with real-time insights into community sentiment on social media, enabling proactive interventions and informed decision-making.

## **3. Objectives & Success Criteria**

* **Automated Data Pipeline:** Daily or hourly scraping from Reddit.
* **Accurate NLP Processing:** ≥85% sentiment classification accuracy.
* **Alerts & Visual Dashboard:** Configurable alerts and intuitive visual analytics.
* **Scalability & Reliability:** Stable performance, reliable storage.
* **Ease of Use:** User-friendly interface.

## **4. Scope & Feature Summary**

### **Core Features**

* Reddit data scraping (PRAW API)
* MongoDB storage and AWS S3 backup
* NLP: Sentiment analysis (DistilBERT), NER (spaCy), Topic modeling (BERTopic)
* Visualization Dashboard (Streamlit, Plotly, Folium)

### **Out of Scope (Current Phase)**

* Advanced sarcasm/irony detection
* Precise geolocation tagging
* Multilingual support

## **5. Functional Requirements**

### **5.1 Data Ingestion**

* Scheduled and on-demand scraping
* Customizable date range configuration

### **5.2 NLP Processing**

* Text preprocessing: URL removal, tokenization, lemmatization
* Transformer-based sentiment analysis
* Named Entity Recognition (NER)
* Topic modeling (BERTopic)

### **5.3 Dashboard & Visualization**

* Aggregated sentiment visualization (pie/bar charts)
* Trending keywords word cloud
* Topic distribution visualization
* Geospatial heat maps
* Recent posts overview

### **5.4 Alerts**

* Threshold-based SMS/email notifications

### **5.5 System Administration**

* Subreddit and data collection management
* Adjustable time filters

## **6. Non-Functional Requirements**

* **Performance:** Handle hundreds of posts daily; dashboard loads <5 sec.
* **Security & Compliance:** Reddit TOS compliance, data anonymization.
* **Reliability & Availability:** ≥99% dashboard uptime
* **Scalability:** Cloud-ready deployment options (AWS, Google Cloud)
* **Maintainability:** Modular Python code, containerization-ready

## **7. Use Cases**

* **Local Government Official:** Monitors daily sentiment and receives alerts.
* **Community Advocate:** Identifies and responds to trending local concerns.
* **Neighborhood Moderator:** Uses data for community engagement and reporting.

## **8. Technical Architecture & Implementation**

### **Data Pipeline**

* **Ingestion:** PRAW (Reddit API), Pushshift API
* **Processing:** NLP pipeline in Python
* **Storage:** MongoDB, AWS S3 backups

### **Application Server**

* Streamlit front-end visualization
* Python-based backend scripts

### **Alerts & Notifications**

* Twilio (SMS), SendGrid (email)

### **Visualization & Analytics**

* Plotly, WordCloud, Folium

## **9. Timeline & Milestones**

| **Week** | **Milestone** |
| --- | --- |
| 1 | Project setup, initial data scraping |
| 2 | NLP preprocessing, basic sentiment tagging |
| 3 | Dashboard integration |
| 4 | Advanced NLP features (NER, topics, alerts) |
| 5 | Testing, map visualization, UI refinements |
| 6 | Final presentation, documentation, feedback |

## **10. Risks & Mitigation**

* **API Constraints:** Use asynchronous requests, rotating tokens
* **Data Quality & Bias:** Local data fine-tuning
* **Security & Privacy:** Strict anonymization
* **Time Constraints:** Prioritize essential features

## **11. Acceptance Criteria**

* **Data Ingestion:** ≥100 weekly posts
* **Sentiment Accuracy:** ≥85% accuracy
* **Dashboard Functionality:** Accurate data visualizations
* **Alerts:** Notifications sent within 5 min
* **Performance:** Dashboard loads in under 5 seconds

*Document Prepared by Jesse Katz & Hung Lu | Date: April 10, 2025*