# Community Pulse: An Automated Reddit Scraper for East Bay Community Sentiment Analysis

This project aims to develop an automated web scraping system that collects data from Oakland/East Bay community Reddit forums on a scheduled basis. The tool will create a comprehensive web dashboard featuring visualizations, sentiment analysis, and natural language processing capabilities to query the collected data. The system will identify, categorize, and track public sentiment, emerging community issues, concerns, and notable events, providing timely alerts to users. By leveraging NLP techniques learned in class, this project will help monitor and understand community discourse in the East Bay area, potentially serving as a valuable resource for community organizations, local government, and residents who want to stay informed about neighborhood developments and concerns.

# 

# Project - Phase 1 (Feasibility Study)

Start Assignment

* **Due** Tuesday by 11:59pm
* **Points** 100
* **Submitting** a text entry box, a website url, or a file upload

## Overview

Before diving into implementation, it's crucial to assess the feasibility of your proposed NLP project. This assignment requires you to thoroughly analyze your project's viability, identify potential challenges, and develop a strategic implementation plan. This feasibility study will serve as a roadmap for Phase 2 (Implementation) and help ensure project success.

## Deliverable

Submit a comprehensive feasibility analysis document (5-8 pages) following the structure outlined below.

## Assignment Structure

### 1. Project Overview (10 points)

* Refined project title and description
* Clear statement of project objectives and success criteria
* Target users/beneficiaries
* Expected impact and applications

### 2. Data Acquisition & Processing Analysis (20 points)

* Data sources identification (APIs, websites, datasets)
* Data access methods (APIs, web scraping, manual collection)
* Legal and ethical considerations for data collection
* Data storage requirements and approach
* Data preprocessing needs (cleaning, normalization, transformation)
* Sample data exploration results (if available)

### 3. Technical Feasibility Assessment (20 points)

* Required NLP techniques and algorithms
* Hardware and software requirements
* Libraries, frameworks, and tools selection with justification
* Technical skills required and team competency assessment
* Potential technical limitations and proposed solutions

### 4. Implementation Plan (15 points)

* Project timeline with key milestones
* Task breakdown and responsibility assignment
* Development methodology (Agile, Waterfall, etc.)
* Testing approach and validation methods
* Deployment strategy (if applicable)

### 5. Risk Assessment (15 points)

* Identification of potential technical challenges
* Data quality and availability risks
* Time management and scope creep concerns
* Mitigation strategies for each identified risk
* Contingency plans for significant risks

### 6. Resource Requirements (10 points)

* Computational resources needed
* Dataset sizes and storage requirements
* External services or APIs (including any costs)
* Time commitment estimation

### 7. Proof of Concept (10 points)

* Small-scale demonstration of core functionality
* Results from preliminary experiments
* Evidence that critical components are viable
* Code snippets or pseudocode for key algorithms

## Grading Rubric

| **Category** | **Excellent (90-100%)** | **Good (80-89%)** | **Satisfactory (70-79%)** | **Needs Improvement (<70%)** |
| --- | --- | --- | --- | --- |
| **Project Overview** | Comprehensive, clear objectives with measurable success criteria | Clear objectives but some success criteria not measurable | Basic objectives defined but lacks specificity | Vague objectives, unclear project scope |
| **Data Analysis** | Thorough data sources analysis with clear acquisition strategy and ethical considerations | Good data sources identified but some gaps in acquisition strategy | Basic data sources identified, limited consideration of access methods | Incomplete data source identification, major gaps in strategy |
| **Technical Feasibility** | Comprehensive analysis of all technical aspects with justified tool selection | Good technical analysis but some gaps in justification | Basic technical requirements identified but limited analysis | Inadequate technical analysis, poor tool selection |
| **Implementation Plan** | Detailed timeline with clear milestones and task assignments | Good timeline but some milestones lack specificity | Basic timeline present but lacks detail | Inadequate planning, unrealistic timeline |
| **Risk Assessment** | Thorough risk analysis with effective mitigation strategies | Good risk identification but some mitigation strategies lacking | Basic risks identified but inadequate mitigation plans | Poor risk identification, inadequate mitigation strategies |
| **Resource Requirements** | Comprehensive resource planning with detailed justification | Good resource identification but some justification lacking | Basic resource needs identified | Inadequate resource planning |
| **Proof of Concept** | Compelling evidence of core functionality viability | Good preliminary evidence but some gaps | Basic demonstration of concept but limited evidence | Inadequate proof of core concept viability |

## Submission Guidelines

* Document format: PDF or shared Google Doc
* Length: 5-8 pages (excluding code snippets or appendices)
* Due date: [Insert appropriate date - typically 2-3 weeks before Phase 2 begins]
* Submit via Canvas

## Additional Resources

* Sample feasibility analysis documents (available in Canvas)
* Office hours for consultation: [Insert your office hours]
* Recommended literature on NLP project planning

## Feedback Process

After submission, I will review your feasibility analysis and provide detailed feedback within [X] days. You'll have an opportunity to revise your analysis based on feedback before proceeding to Phase 2 implementation.

## Important Notes

* This feasibility analysis is a crucial checkpoint. Projects cannot proceed to Phase 2 without approval of this analysis.
* Be realistic in your assessment - identifying potential challenges demonstrates critical thinking.
* You may begin preliminary implementation (such as data collection) while working on this analysis, but major development should wait for Phase 1 approval.

Good luck with your feasibility analysis! Remember that thorough planning now will save significant time and effort during implementation.

[Previous](https://peralta.instructure.com/courses/78596/modules/items/4901286)

[Next](https://peralta.instructure.com/courses/78596/modules/items/4931665)