

# Text Analysis and NLP: Project Proposal

## r/Futurology

r/Futurology is a popular Reddit community dedicated to discussions about future trends, emerging technologies, and societal evolution. Discussions cover various topics, including advancements in AI, space exploration, robotics etc, and often feature realistic and speculative ideas of what the world might look like in the future.

## Project Overview

The goal of this project is to analyse text extracted from posts and comments on r/Futurology, mainly through sentiment analysis techniques. Depicted below is a flowchart describing the project.

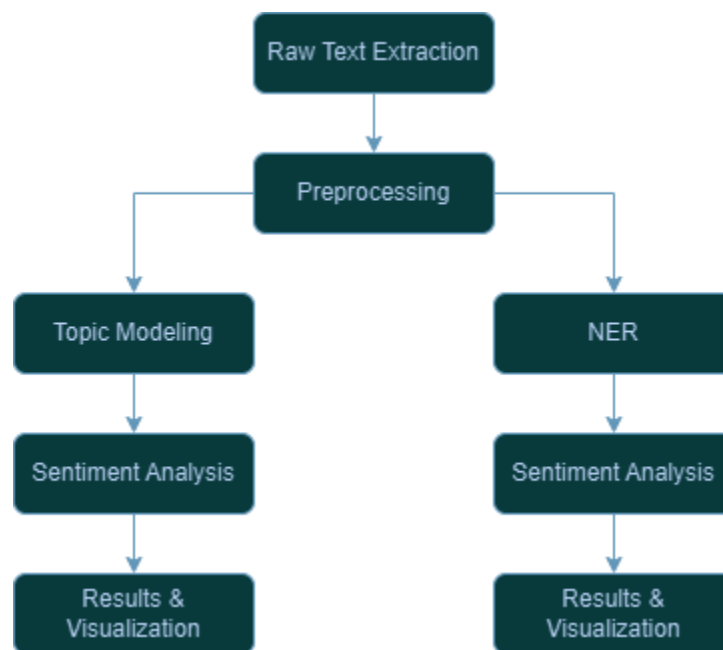


Figure 1: Project Summary

The project will consist of two parts.

In the first part, text is classified into various topics using **Topic Modeling** and overall sentiment is analysed within each topic and across all topics to make comparisons.

In the second part, important entities (for example, people and companies) are recognized with **Named Entity Recognition**. Similar to what will be done in the first part, sentiments toward individual entities are analysed by considering pieces of text where the entities appear. These pieces of text could be of levels of granularity ranging from sentences to posts/comments, and various such levels will be tried.

Where appropriate, visualizations will be used to illustrate results.

## Tools

The project will be done mostly in R.

The following table shows the tools to be used for different parts of the project.

Scraping	Pushshift API
Preprocessing	tidytext
Topic Modelling	topicmodels
Sentiment Analysis	syuzhet, sentimentr
Named Entity Recognition	spaCyR
Visualization	ggplot2

*Table 1: Tools to be used in the project*