FiscalNote Capstone Project Problem Statement

1. **Purpose of the Study**

FiscalNote/CQRC collects and maintains one of the largest collections of unstructured data coming out of government activity. Legislators and other policymakers often announce their intention or positions on a bill, regulation, or issues through speeches mad in Congress, tweets, press releases, and other forms of public statements.

The purpose of the study is to use the natural language processing and machine learning technique to build a program which can automatically get the insight of policymakers’ opinion on an issue from various data sources.

1. **Practical Objectives**

The practical objectives of this project are to identify in one or more of the data above when a policymaker is expressing their opinion on an issue, what that issue is, and what the policymaker opinion is.

The output of this project should be a program that takes text data as input, and outputs a tuple of policymaker’s name, the name of the issue, and the policymaker’s opinion on the issue.

1. **Research Questions**

How to identify the issue and policymaker’s opinion of it in any text data?

1. **Data Source**

The data source will be provided by FiscalNote, which is unstructured and collected from government activity.

1. **Potential Analysis Method**

This project can be categorized into two sub tasks of natural language processing.

Identifying the issue from text is a Named entity recognition (NER) task, and identifying the opinion is a Sentiment analysis problem.

There are two general approaches in solving these problems, one is linguistic grammar-based techniques, and another is statistical models. This research will implement both techniques on the data and evaluate the performance of these methods.