

# Fall 2022 Erdos Project: The Ideology of Speeches

Team *Lilac*

<sup>1</sup> Hai Le and <sup>2</sup> Enan Srivastava

<sup>1</sup> Postdoc, Department of Mathematics, University of Michigan, Ann Arbor

<sup>2</sup> Data Scientist, Aware

## INTRODUCTION

Given a political speech, will we be able to tell which side of the aisle it supports? We are interested in how the languages of the two major parties of the United States have changed over the years and are looking to develop an NLP model to classify the partisanship of the words that members of Congress use to assert their ideology. We use the dataset “*Congressional Record for the 43rd-114th Congresses*” from the Stanford’s Social Science Data Collection. The dataset includes speeches given in both chambers of Congress which we only analyze from 1981 to 2017. The goal is to determine the party, Democratic or Republican, the speech belongs to.

## METHODS

Given the forgetful nature of Congress, we decided that an LSTM classifier is suitable for the job. In our neural network, we feed the data through one LSTM unit, then a dropout layer to avoid overfitting, and finally a fully connected layer for classification.

## RESULTS AND DISCUSSION

Classification Report:

	precision	recall	f1-score	support
1	0.8934	0.0014	0.0029	309404
0	0.5253	0.9998	0.6888	342001
accuracy		0.5256		651405
macro avg	0.7094	0.5006	0.3458	651405
weighted avg	0.7001	0.5256	0.3630	651405

(1 is Republican, 0 is Democrat)

The LSTM was highly successful in correctly classifying speeches between Democrats and Republicans. Despite no hyperparameter tuning nor stopword elimination, the model still achieved mediocre statistics. Even though the model couldn’t manage to differentiate between the two parties, we remain hopeful with further work we can build on the results. Considerably improving the performance of the model can be done by increasing the number of epochs from 5 to 100, as well as adding longer speeches.