



# DETECTION OF POLITICAL IDEOLOGY IN TEXT

## A COMPARATIVE STUDY

Aditya Srinivasan  
MS, Computer Science

Nithyashri Govindarajan  
MS, Computer Science

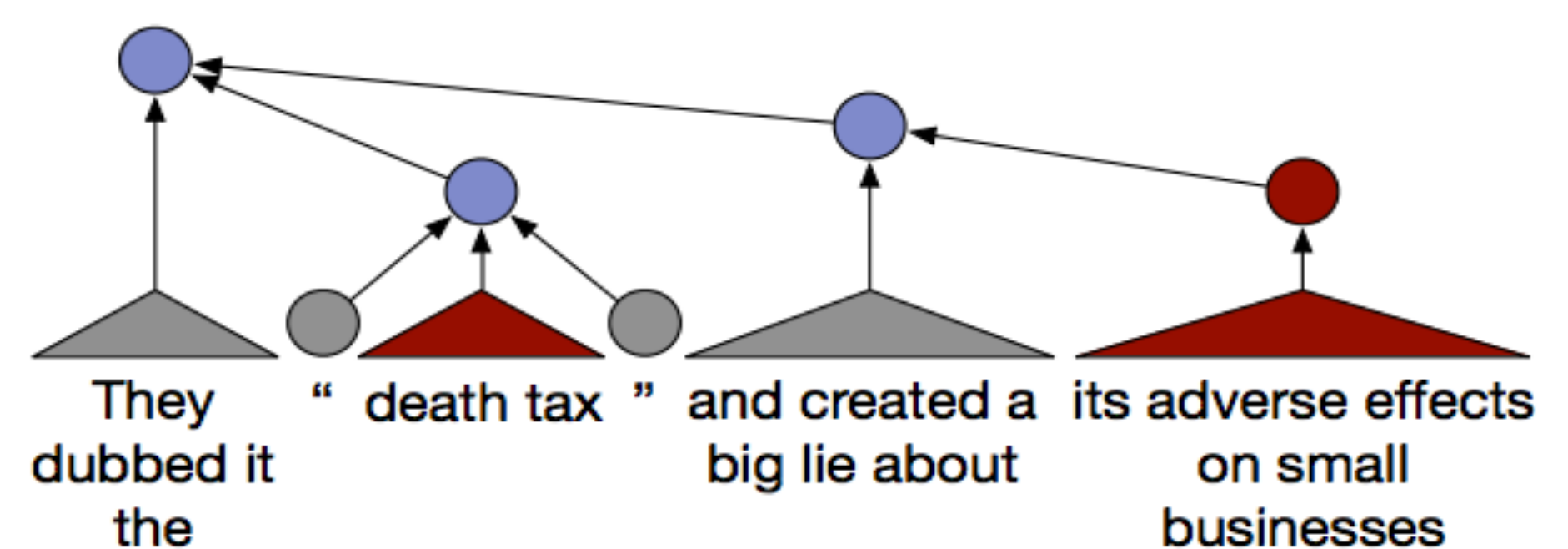
### PROBLEM STATEMENT:

- Identify the political leaning of a given sentence by classifying it as Liberal or Conservative.
- Compare different models for their accuracy and suitability.

### DATASET:

- Ideological Book Corpus - 4,062 sentences.
- Annotated at sentence and phrase level with three types of tags – Liberal, Conservative and Neutral.
- Represented by an annotated parse tree - 13,640 annotated nodes.

LABEL	NUMBER OF SENTENCES
Liberal	2025
Conservative	1701
Neutral	600



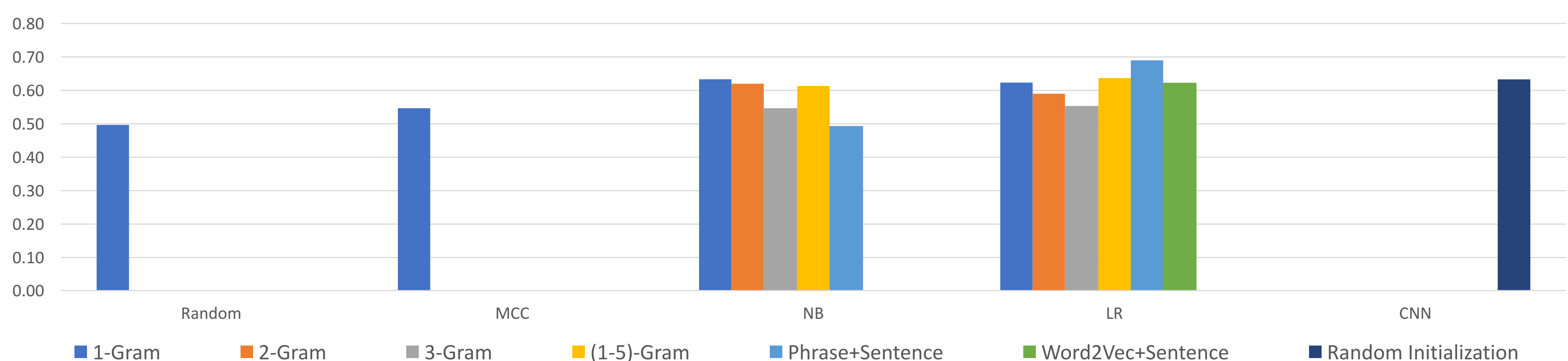
*Liberal Discriminators* - “tax”, “system”, “workers”  
*Conservative Discriminators* - “free”, “American”, “freedom”

### ANALYSIS OF MODELS:

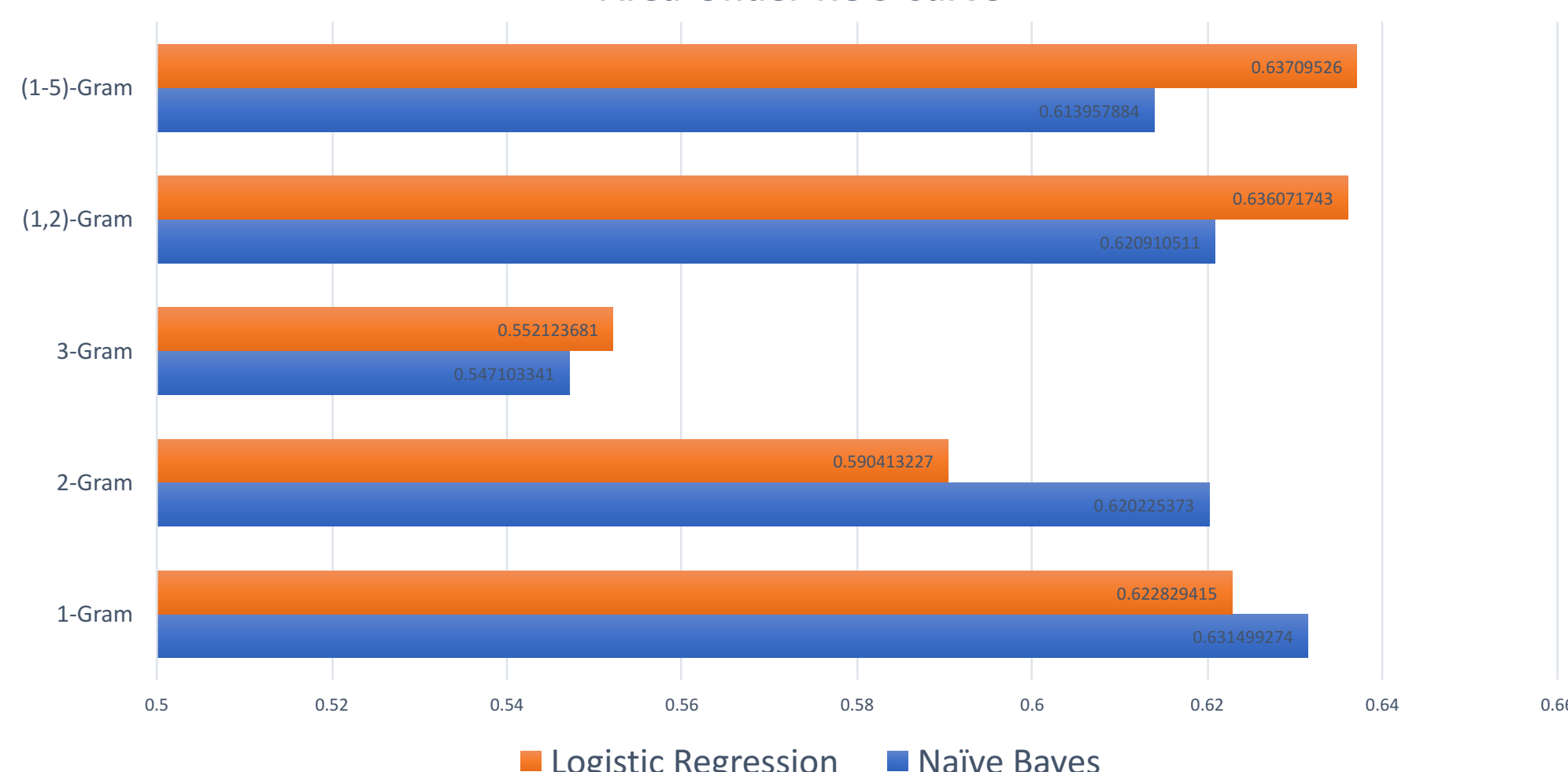
Comparison of Performance of “Traditional” NLP classifiers and Deep Learning classifiers:

- Random, Most Common Class (MCC), Naïve Bayes (NB), Logistic Regression (LR).
- Convolutional Neural Networks (CNN).

Models and Features vs Accuracy



Area-Under-ROC-Curve



*Logistic Regression has lower accuracy than Naïve Bayes when number of data points is low.*

*Phrasal Annotations are important features for both kinds of models.*

*Expected: Neural Networks to outperform traditional models.*

*Reason: Semantic Compositionality of text!*

*Logistic Regression with Phrasal Features and Sentence Features has comparable performance.*

*Next Step: Neural Network with Phrasal and Sentence Features*

