



PROJECT

CLASSIFICATION OF POLITICAL TEXTS

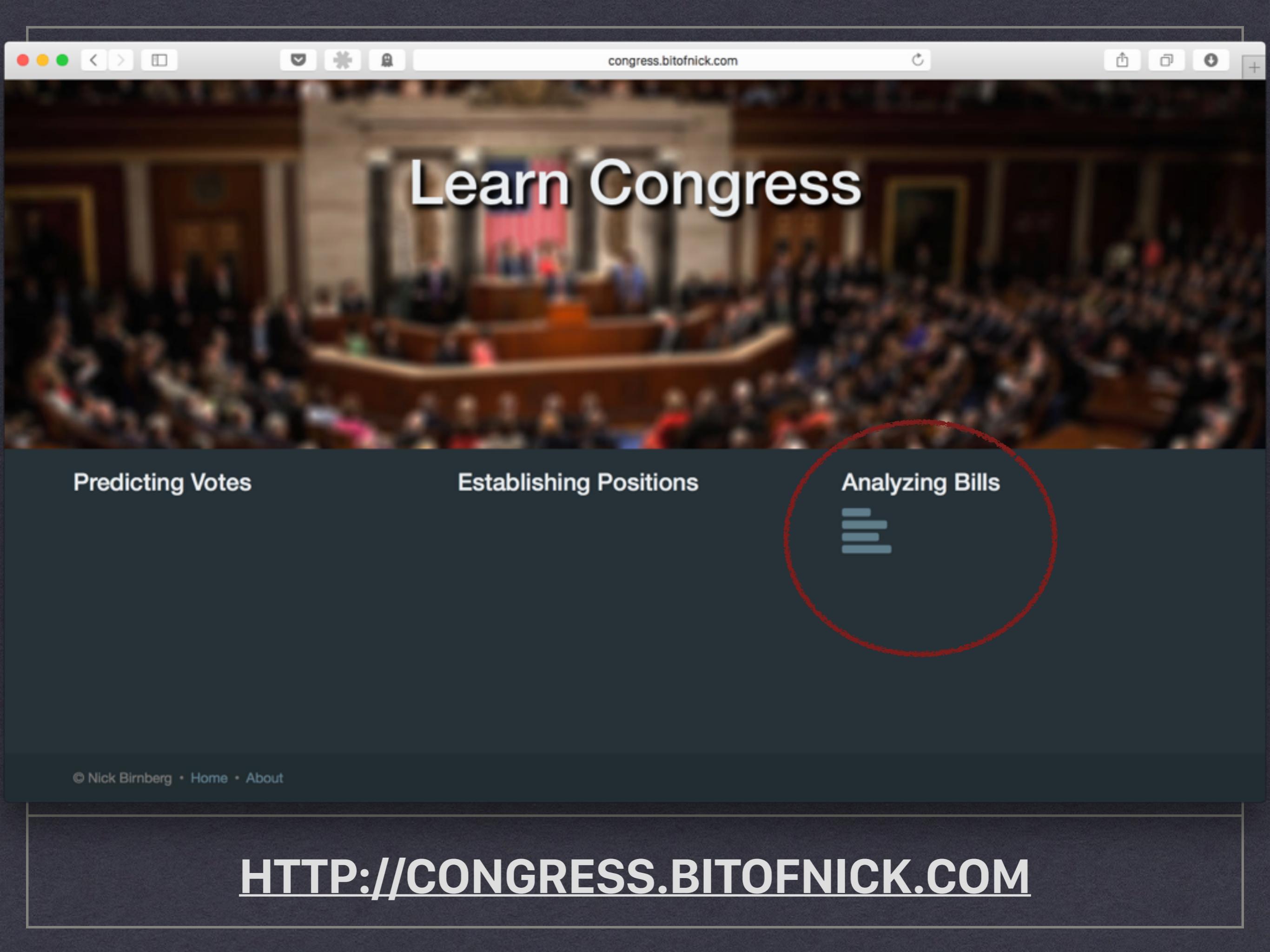
BY NICK BIRNBERG

DATE

26/5/15

MENTOR

PIERRE NUGUES



Predicting Votes

Establishing Positions

Analyzing Bills

A screenshot of a terminal window titled "data.json". The window displays a large block of JSON code representing a bill's metadata. The code includes fields such as "introduced_at", "number", "official_title", "popular_title", "related_bills", "short_title", "sponsor", "status", "status_at", and "subjects". The "subjects" field lists various topics like "Child safety and welfare", "Education", "Higher education", and "Teaching, teachers, curricula". The "titles" field contains two entries, both labeled "introduced" and "is_for_portion": false. The "summary" field provides a brief description of the bill's purpose. The bottom status bar shows "Line 192, Column 61", "Spaces: 2", and "JSON".

```
data.json UNREGISTERED
{
    "document_signature": false,
    "enacted": false,
    "vetoed": false
},
{
    "introduced_at": "2013-01-22",
    "number": "3",
    "official_title": "A bill to improve education and provide all students in the United States wit",
    "popular_title": null,
    "related_bills": [],
    "short_title": "Strengthen our Schools and Students Act",
    "sponsor": {
        "district": null,
        "name": "Reid, Harry",
        "state": "NV",
        "thomas_id": "00952",
        "title": "Sen",
        "type": "person"
    },
    "status": "REFERRED",
    "status_at": "2013-01-22",
    "subjects": [
        "Child safety and welfare",
        "Education",
        "Education programs funding",
        "Elementary and secondary education",
        "Higher education",
        "Preschool education",
        "Science and engineering education",
        "Student aid and college costs",
        "Teaching, teachers, curricula"
    ],
    "subjects_top_term": "Education",
    "summary": {
        "as": "Introduced",
        "date": "2013-01-22",
        "text": "Strengthen our Schools and Students Act - Expresses the sense of the Senate that Cong"
    },
    "titles": [
        {
            "as": "introduced",
            "is_for_portion": false,
            "title": "Strengthen our Schools and Students Act",
            "type": "short"
        },
        {
            "as": "introduced",
            "is_for_portion": false
        }
    ]
}
Line 192, Column 61
Spaces: 2
JSON
```

Data Collection

raw/s_113/s1/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s169/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s189/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s202/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s223/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s260/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s266/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s292/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s293/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s295/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s296/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s299/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s300/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s301/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s302/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s303/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s461/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s478/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s518/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s527/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s600/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s616/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s645/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s683/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s703/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s744/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s883/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s891/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s990/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s1059/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s1566/text-versions/cps/document.txt,Immigration,Liberal
raw/s_113/s1614/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s1817/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s1850/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s1943/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s1990/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2048/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s2195/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2218/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s2302/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s2315/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s2463/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2600/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2611/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2624/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s2628/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2631/text-versions/pcs/document.txt,Immigration,Conservative
raw/s_113/s2632/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2633/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2663/text-versions/is/document.txt,Immigration,Liberal
raw/s_113/s2666/text-versions/pcs/document.txt,Immigration,Conservative
raw/s_113/s2779/text-versions/pcs/document.txt,Immigration,Conservative
raw/s_113/s2781/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2869/text-versions/is/document.txt,Immigration,Conservative
raw/s_113/s2953/text-versions/is/document.txt,Immigration,Conservative

Extracting Features

- * bag of words
- * tf-idf



Classifier: Multinomial Naive Bayes

- * bag of words assumption
- * assumption of independence between every pair of features
- * can be used with a small amount of training data

Architecture

- * Front-End:
 - * node.js + React + Flux + Bootstrap
 - * JustGage
- * Back-End:
 - * Python (Flask + Natural Language Toolkit + scikit-learn)

RESULTS AND APPLICATION

Party

13

Immigration

46