

# Finnish parliament elections 2015

- Are all parties the same?

## Scope

The aim of the analysis is to find out if one can deduct the party of a candidate, based on what the candidate says he/she wants to work for.

## Findings

The model could make predictions with approx. 44 % accuracy on the test data.

As seen on the classification report below, Vihreä liitto and Suomen Kristillisdemokraatit were the easiest ones to spot, while Suomen Keskusta was the most difficult one. While most parties have similar numbers in precision and recall, Vasemmistoliitto shows a larger dispersion with 56 % of the correct candidates being identified but only 34 % of the ones labelled as Vasemmistoliitto were correct. Perhaps an indication that environmental issues are emphasized in several parties?

|                                     | precision | recall | f1-score | support |
|-------------------------------------|-----------|--------|----------|---------|
| Kansallinen Kokoomus                | 0.431     | 0.4098 | 0.4202   | 61.0    |
| Perussuomalaiset                    | 0.4464    | 0.4237 | 0.4348   | 59.0    |
| Suomen Keskusta                     | 0.3385    | 0.3607 | 0.3492   | 61.0    |
| Suomen Kristillisdemokraatit (KD)   | 0.5306    | 0.4815 | 0.5049   | 54.0    |
| Suomen Sosialidemokraattinen Puolue | 0.3623    | 0.4098 | 0.3846   | 61.0    |
| Vasemmistoliitto                    | 0.5556    | 0.3448 | 0.4255   | 58.0    |
| Vihreä liitto                       | 0.4625    | 0.6271 | 0.5324   | 59.0    |
| accuracy                            | 0.4358    | 0.4358 | 0.4358   | 0.4358  |
| macro avg                           | 0.4467    | 0.4368 | 0.4359   | 413.0   |
| weighted avg                        | 0.4444    | 0.4358 | 0.4344   | 413.0   |

The confusion matrix below tells us that Vasemmistoliitto are often falsely classified as Vihreä liitto or Suomen Sosialidemokraattinen Puolue. Kansallinen Kokoomus seems to be easy to falsely classify as Suomen Keskusta and the other way around.

```

[[25  5 16  7  4  1  3]
 [ 7 25  8  5  8  3  3]
 [13  9 22  3  5  1  8]
 [ 4  7  5 26  8  0  4]
 [ 2  6  7  4 25  7 10]
 [ 4  1  3  1 14 20 15]
 [ 3  3  4  3  5  4 37]]

```

The six most typical words or word combinations used by candidates for each party (after stemming) are listed below. The wordings are quite well in line with what could be expected for each party.

| Party                               | 1           | 2                       | 3            | 4              | 5             | 6               |
|-------------------------------------|-------------|-------------------------|--------------|----------------|---------------|-----------------|
| Kansallinen Kokoomus                | tekemin     | työnteko                | työn tekemin | kannustav      | satakun       | työ on          |
| Perussuomalaiset                    | maahanmuuto | maahanmuuttopolitiik    | suome etu    | ulkom          | järk          | pakkoruots      |
| Suomen Keskusta                     | biotaloud   | luonnonvar              | koko suome   | asu            | ja työllisyys | halua kehity    |
| Suomen Kristillisdemokraatit (KD)   | kristillis  | kristillis arvo         | ¥            | miehe ja       | miehe         | naise           |
| Suomen Sosialidemokraattinen Puolue | tule tuke   | julkist palvelu         | jokais on    | reilu          | maks          | yhteiskun on    |
| Vasemmistoliitto                    | elvyttäv    | elvyttäv talouspolitiik | verotettav   | talouspolitiik | investoin     | tuloero         |
| Vihreä liitto                       | suoje       | ilmastonmuutos          | ilmasto      | luonto         | eläint        | ilmastonmuutoks |

## Material

The data used for the prediction comes from avoindata.fi and consists of parliament candidates answers to yle's election machine 2015:

<https://www.avoindata.fi/data/fi/dataset/eduskuntavaalien-2015-ylen-vaalikoneen-vastaukset-ja-ehdokkaiden-taustatiedot>

In this analysis, the columns “party” and “what will you work for in the upcoming parliamentary period ?” have been used. Rows containing NA:s in these columns, parties with less than 150 candidate answers and answers in another language have been omitted.

## Methods

Special signs were removed from the “work for column” and the words were stemmed.

Bag of Words was used to find word frequencies and Multinomial Naive Bayes to create a model for party prediction.

# Limitations

A party's or politician's viewpoint in 2015 is not necessarily the same as it is today and the algorithm would need to be retrained if it was to be used with current data . Both party agendas and hot topics in society are likely to vary with time.

All possible model variations have not been tested, and more invested time in this could result in more precise predictions.