Predicting the Outcome of Votes Taken by the Swiss National Council Applying Machine Learning Models

Daniel Müller [1,2] (*) | (1) stellus.ch, Liestal, Switzerland, (2) Alumni, EPFL Extension School, Lausanne, Switzerland

Aim

To explore the possibility to predict the outcome of votes taken by the Swiss National Council (SNC) applying Machine Learning Models (MLM).

Methods

From the website of the Swiss Parliament (1) 9004 voting records in German from 2011 up to 2019 were downloaded along with related documents and descriptions of the political affairs.

Votes about affairs of the Federal Council (n=5833) were excluded along with a subset of selective votes of other affairs (n=475). The excluded votes were essentially about incremental adaptations at the level of (sub) articles and not about entire political matters.

The final data set contained n=2696 records. The data selection process is shown in Figure 1 and the distribution of the categories of political affairs in Figure 2.

The votes of the winter session of 2019 (n=198) were used as a test set. N=33 records where retained after applying the above described selection criteria. The final test set contained 23 parliamentary initiatives and one request and nine motions.

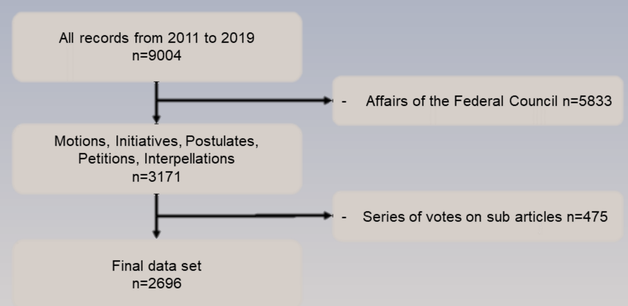

Figure 1: Data selection process

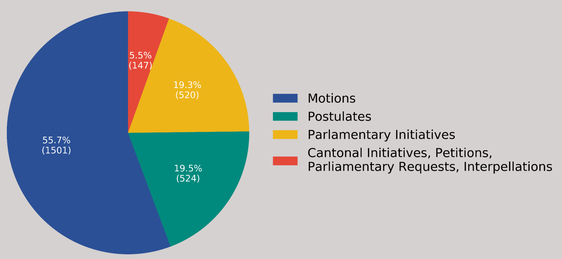

Figure 2: Distribution of the categories of political affairs contained in the final data set

(*) This study is motivated by personal interest and is politically independent.

The 2'696 voting records were randomly split by 90 to 10 into a training and validation set. In the training set the mean number of words per record was 1'572 (108 to 25'807). Scikit-learn (2) was used to preprocess the texts with a word tokenizer extracting 46'255 unique 1-grams.

Terms linked with the outcome of a vote (such as: "accepted", "rejected", "written off", etc.) were removed.

A Logistic Regression Classifier (LRC) was trained and validated. To avoid overfitting a least absolute shrinkage and selection operator (lasso) was used that retained 789 nonzero features.

Results

The accuracy of the model on the validation set was 97% and on the test set 88%. The accuracy of the majority classifier (MC) for voting "yes" in the validation set was 62% and in the test set 52% (see Figure 3).

The precision and recall of the model on the test set were 83% and 94%. The F1-score was 0.88.

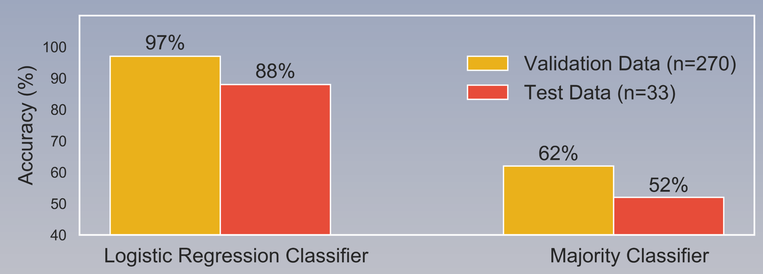

Figure 3: Accuracies of the Logistic Regression and Majority Classifier for the validation and test set

The 30 terms found to have the largest absolute weights to determine the outcome of votes are shown in Figure 4 (on blue background favoring "Yes" and on red favoring "No").

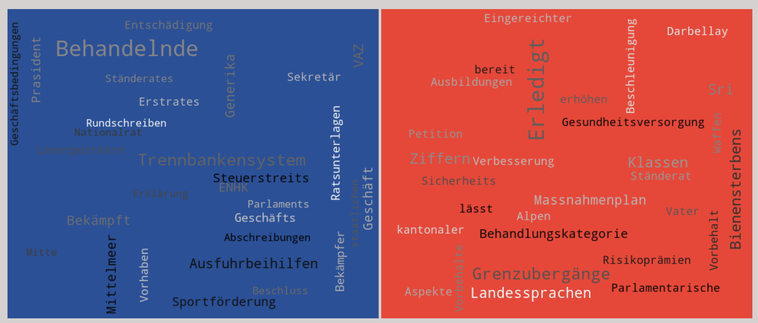

Figure 4: Word clouds based on the terms wth the largest absolute weights (blue: "Yes", red: "No")

Discussion

An LRC can predict the outcome of a subset of votes taken by the SNC with notable accuracy.

The LRC performed best among other models including artificial neural networks.

In a previous study predicting the outcome of Swiss federal votes based on voter information a similar accuracy of 92% was found (3).

Some keywords favoring a positive outcome of a vote are: "commercial banks", "tax dispute", "sports promotion" and "generics". Terms favoring a negative outcome: "border", "weapons", "risk premium" and "bee death".

Parliamentary group strengths were initially included as features but were eliminated during lasso regularization.

The model only applies to a subset of votes taken by the SNC and cannot be generalized to all categories of votes. Therefore the results should be considered with caution.

Voting records are only released after the vote took place and contain information related to the outcome. Such information has been removed but it is still possible that some evidence has leaked into the model representing a potential source of confounding.

The predictability of the outcome of votes can rise fears of manipulation in a democracy. But this study also shows that the complex legislative process cannot be easily reconstructed by means of machine learning.

Conclusions

The outcome of votes taken by the SNC can be predicted quite accurately by an LRC. The results should be considered with caution as only a subset of votes taken by the SNC was included.

Key parts of the complex and subtle legislative process (votes on incremental changes at the level of (sub) articles) could not be modeled by readily available means of machine learning.

More information can be found at: www.stellus.ch

References

1. The Federal Assembly - The Swiss Parliament. https://www.parlament.ch/en.

2. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

3. Müller, Daniel. Stellus.ch. https://www.stellus.ch