# Larry on Ballots

Digging into Swiss Referendum Voting Patterns

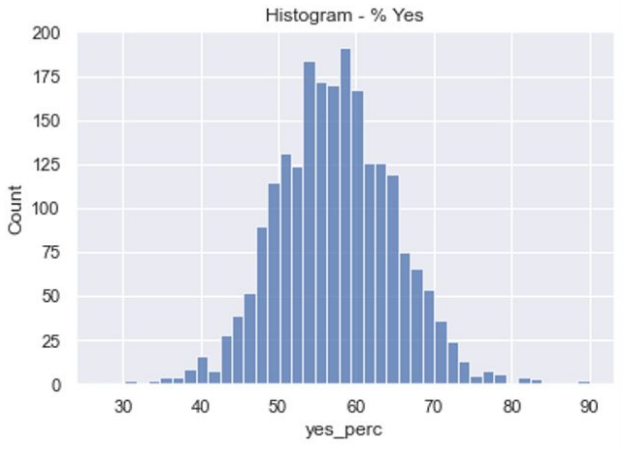
# Project Introduction

- Switzerland is a direct democracy
- Voted on banning face-covering in 2021
- Accepted by the popular vote with 51.42%


- Who votes for or against it?
  - Urban-rural divide
  - Language divide
- Which model works the best?
- Significant accuracy improvements?

# Datasets

- Swiss Federal Office of Statics
- Two datasets:
  - Socio-economic data
  - Voting outcomes
- Population, land use, local economy, local politics
- 2179 municipalities, 83.1% yes



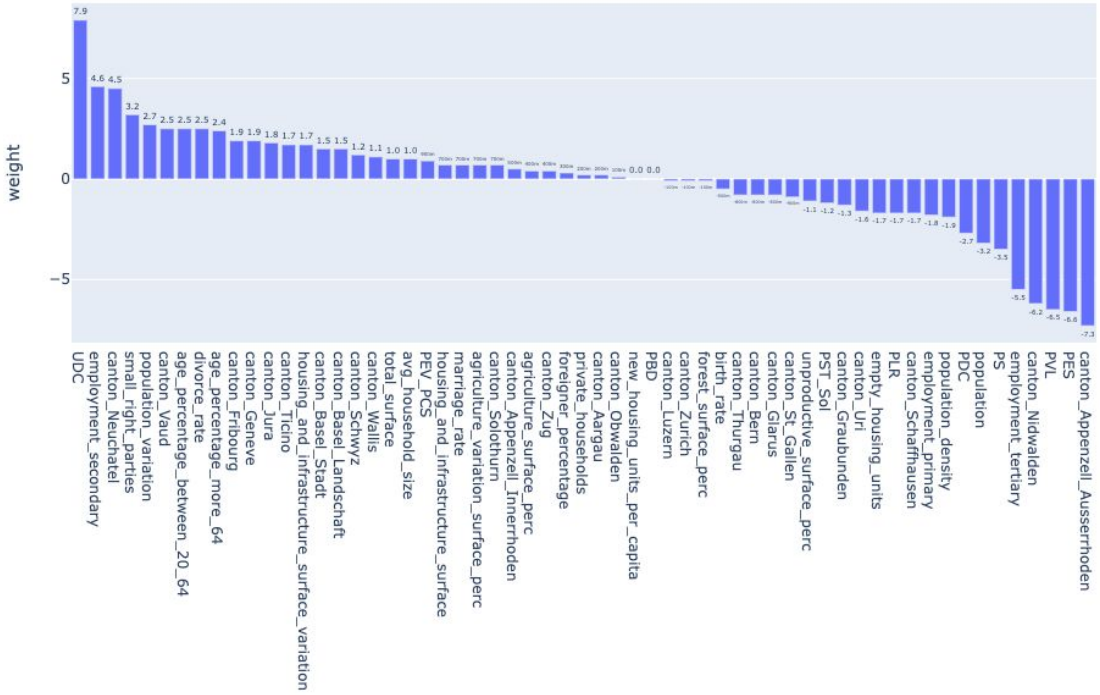
# Experiment Setup

Binary classification problem

1. Merge datasets on municipalities
2. Remove irrelevant features (collinear, redundant, irrelevant outcomes)
3. 80/10/10 train/dev/test split
4. Min-Max scaler
5. Train the models
    a. Logistic Regression
    b. K-Nearest Neighbors
    c. Decision Tree
    d. Random Forest
    e. Neural Networks
6. Hyperparameter tuning
7. Test the models with and without political features

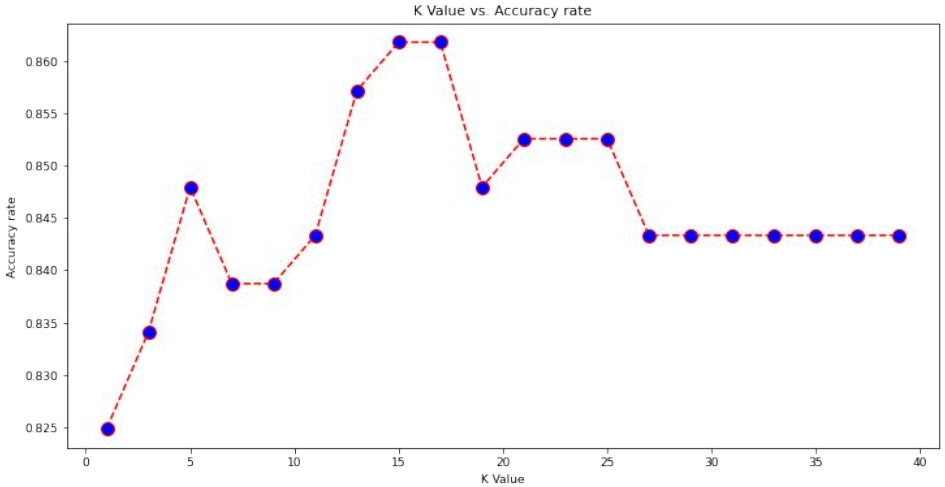
# Logistic Regression

Accuracy = 92.17%
Hyperparameters: L2 penalty, LibLinear solver, C=10

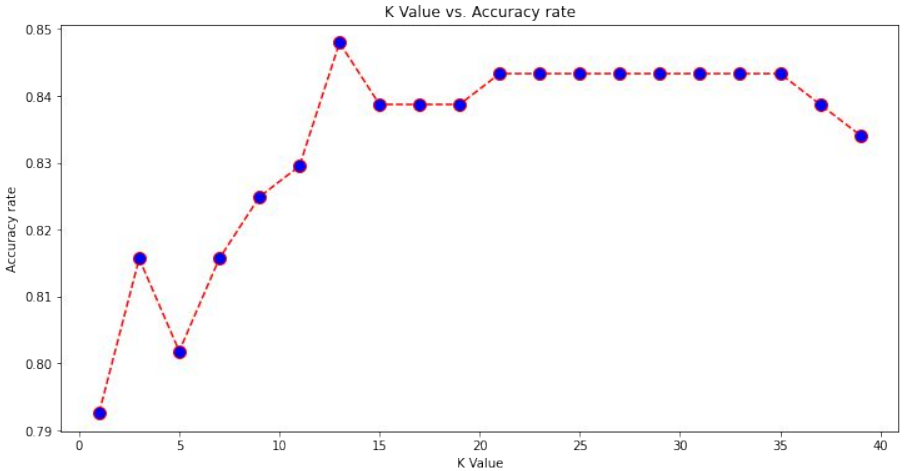


# K-Nearest Neighbors

**With political parties:**




**Without political parties:**




**With political parties:**

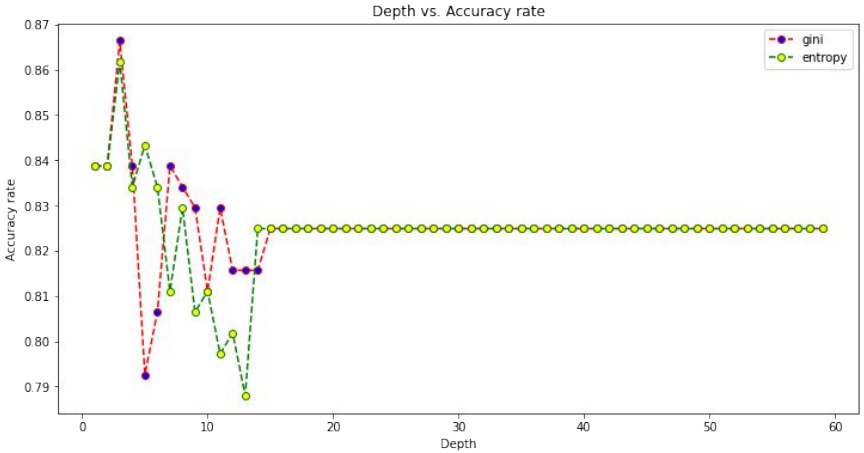
**Optimal k**: 15 - 17
**Accuracy**: 89.86%

**With political parties:**

**Optimal k**: 13
**Accuracy**: 88.01

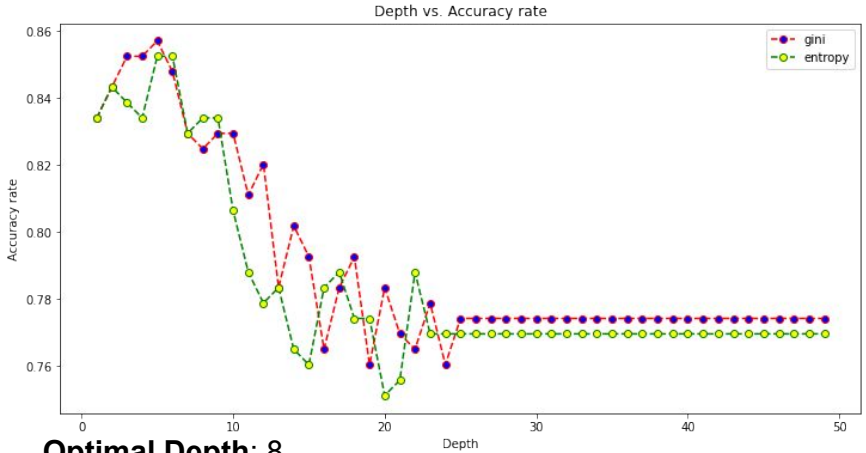
# Decision Tree

**With Political Parties**




**Optimal Depth**: 3
**Accuracy**: 89.4%
**Key Splitting Attributes**: The Swiss People's Party, The Green Liberal Party, and The Liberals

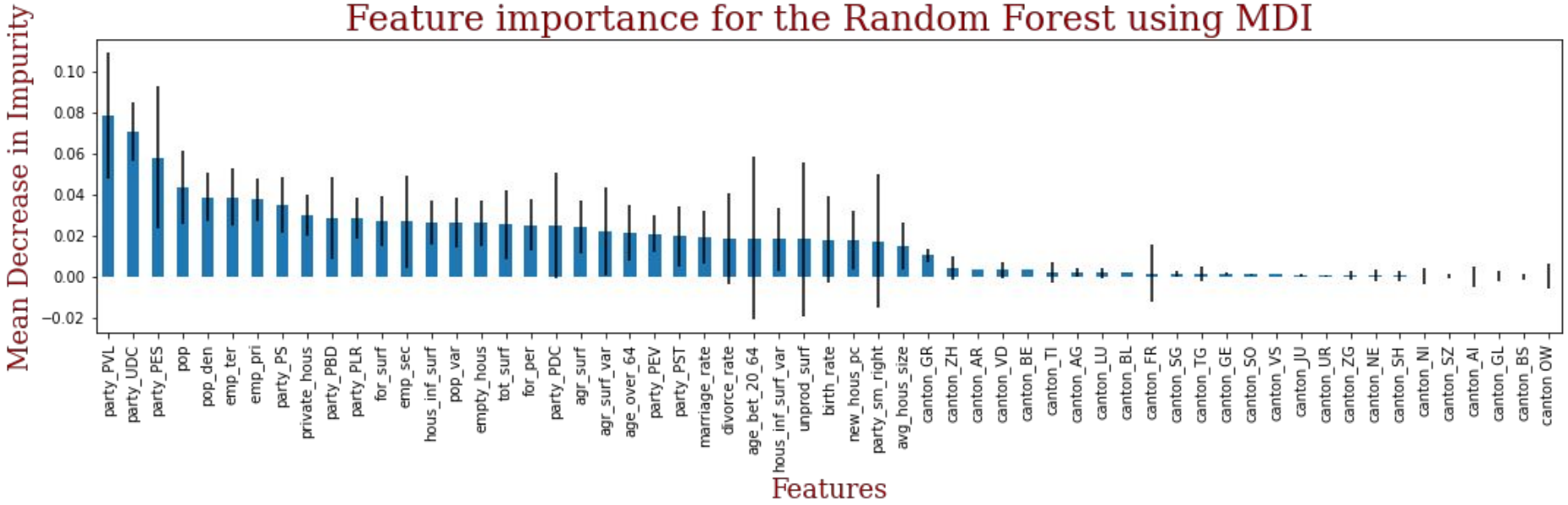
**Without Political Parties**




**Optimal Depth**: 8
**Accuracy**: 87.10%
**Key Splitting Attributes**: Population, Private Households, and Graubunden canton

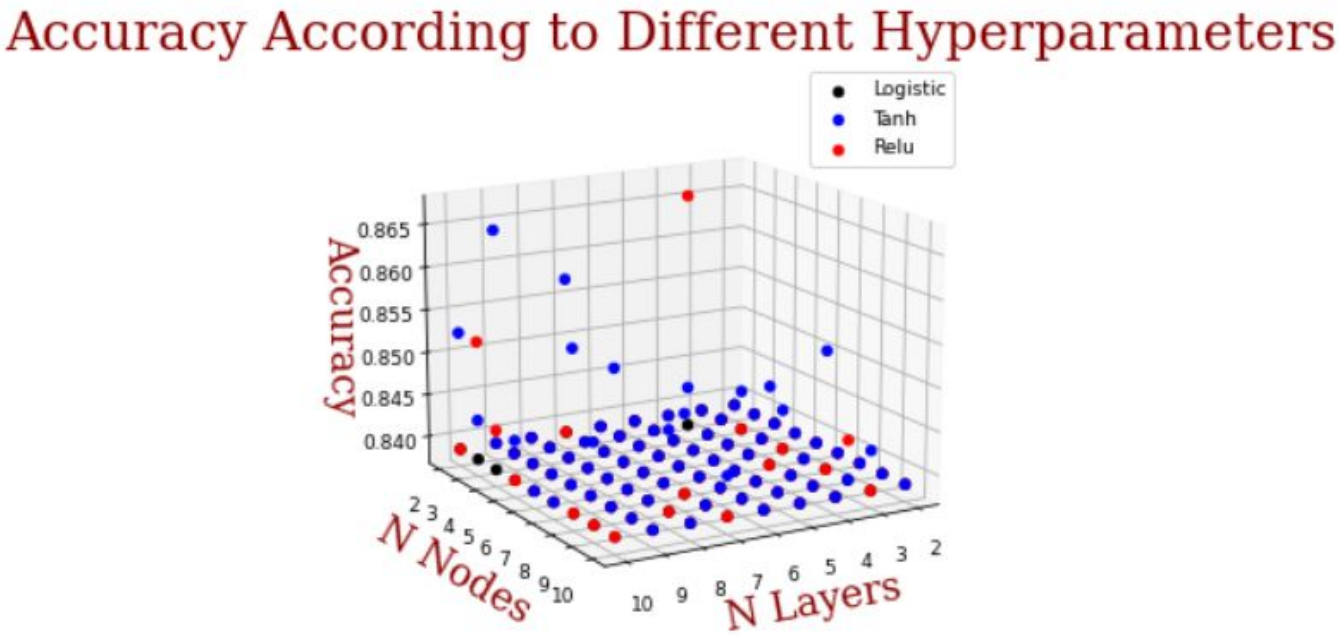
# Random Forest

Accuracy = 88.48%
Hyperparameters: Gini, 40 trees, max depth of 40



# Neural Networks

Development Accuracy = 86.63
Hyperparameters: Tanh, 10 layers, 4 Nodes



# Results

**With political parties:**

| | Model | Accuracy | Specificity | recall |
|---|---|---|---|---|
| 0 | Logistic Regression | 0.921659 | 0.500000 | 0.979058 |
| 1 | KNN | 0.898618 | 0.230769 | 0.989529 |
| 2 | Decision Trees | 0.894009 | 0.230769 | 0.984293 |
| 3 | Random_Forest | 0.884793 | 0.307692 | 0.963351 |
| 4 | Neural Networks | 0.898618 | 0.153846 | 1.000000 |

**Without political parties:**

| | Model | Accuracy | Specificity | recall |
|---|---|---|---|---|
| 0 | Logistic Regression | 0.884793 | 0.192308 | 0.979058 |
| 1 | KNN | 0.880184 | 0.115385 | 0.984293 |
| 2 | Decision Trees | 0.870968 | 0.230769 | 0.958115 |
| 3 | Random_Forest | 0.889401 | 0.307692 | 0.968586 |
| 4 | Neural Networks | 0.880184 | 0.000000 | 1.000000 |

Poor performance on no-municipalities

# Conclusions

Main findings

- Best features: local politics, population, employment sectors
- Patterns show an urban-rural divide, but not across languages
- More complex models were not the most performant, indicating overfitting

Future Steps:

- Weigh no-municipalities
- Other referendums, countries, types of votes