Kaggle competition: Modeling Pro-Government Votes

Due date: **May 6, 2025, 11:55 PM**

Deliverables:

1. Submission to Kaggle competition.

2. Python notebook includes your code and model results (ipynb format)

3. Lab Report (pdf format)

Guidelines:

1. **You can work alone or in a group of two on this project.**
2. Please fill out this document to write your Kaggle competition name:

   Google Form
3. You are not allowed to post your solutions online during and after the competition.
4. Your submission will be evaluated using Accuracy.
5. You can make at most **20 submissions** per day.
6. The grade will be based on **100** points.
7. Total Grading will be based on
   a. **Part I: Kaggle competition (70 %)**
      i. Ranking in the Kaggle competition.
      ii. If your submission is above 55 % accuracy (which is the accuracy if you submit all 0's)
   b. **Part II: Lab Report (30 %)**
      i. Abstract, Introduction, Data, Methods, Results
      ii. Flow, readability, level of detail, quality of visuals/tables, adherence to the guideline

Part I: Kaggle Competition (70 % of the grade)

Overview

The European Court of Human Rights (ECHR) adjudicates cases related to alleged human rights violations under the European Convention on Human Rights. It spans 46 member states in Europe, addressing cases brought by individuals or states. Member states independently appoint judges on the ECHR, but their voting behavior may reflect:

Home State Bias: Favoring their home state.

Geopolitical or Cultural Affinity: Aligning with countries sharing similar political or cultural backgrounds.

Strategic Behavior: Voting in line with broader legal or political trends.

This competition challenges participants to develop machine learning models that predict whether a judge will cast a pro-government vote in a case. The dataset offers a rich collection of features that span case-specific attributes, judge-specific characteristics, and country-specific variables, creating a multidimensional prediction problem at the intersection of law, political science, and data science.

Goals

Develop and evaluate a variety of machine learning approaches to predict whether a judge will cast a **pro-government vote** in court cases. The ultimate objective is to achieve the highest possible prediction accuracy score

Evaluation

Submissions will be assessed based on accuracy, which means the percentage of correctly predicted votes.

Submission File

For each ids in the test set, you must predict a probability for the **progovernment_vote** variable. The file should contain a header and have the following format:

```
ids,progovernment_vote
1,0
2,0
3,0
etc.
```

Dataset Description

Files

train.csv - the training set

test.csv - the test set

solutions_template.csv - a sample submission file in the correct format

Dataset

The dataset contains over 10,000 voting records and 68 features across judges, cases, and countries. The target variable is **progovernment_vote**:

1: Pro-government vote.

0: Non-pro-government vote.

Features

Case Details

• ids: Unique identifier for each case.

• case_title_list: Title of the case being judged.

• app_no_list: Application number for the case.

• date_list: The date on which the case was decided.

• year: The year the case was decided.

• issue: Number of issue in the specific court case.

• case_violation: Binary indicator for whether a violation occurred in the case.

• violations: description of the violations in the case.

• no_violations: description of non-violations in the case.

• conclusion: Outcome of the case (specific conclusion details).

• importance_level: Importance level of the case (e.g., key case, level 1 or level 2).

• originating_body: Body or chamber handling the case.

• women_issue: Whether the case involves women's rights (binary).

• number_of_women: Total number of female judges on the panel.

• number_of_men: Total number of male judges on the panel.

• number_of_judges: Total number of judges on the panel.

• fraction_female: Proportion of female judges on the panel.

Judge Characteristics

- judge_full_name: Full name of the judge involved in the case.
- female: Whether the judge is female (binary).
- recruited_from_position_type: The professional background of the judge.
- academic: Indicator for whether the judge has an academic background.
- diplomat: Indicator for whether the judge has a diplomatic background.
- judge: Indicator for whether the judge has judicial experience.
- lawyer: Indicator for whether the judge has a legal practice background.
- politician: Indicator for whether the judge has a political background.
- public_official: Indicator for whether the judge was a public official.
- omni: Indicator for other combined professional backgrounds.
- judgecnt: Home country of judge.
- home: If the judge is from the respondent country. (1=yes, 0=no)
- term: The term number of the judge.
- term_start: Start date of the judge's term.
- term_end: End date of the judge's term.
- biryear: Year of birth of the judge.
- age_at_election: Judge's age at the time of election to the court.
- age_at_retirement: Judge's age at retirement from the court.
- judge_age: Judge’s age at the time of the case.
- president_in_case: Whether the judge was the president for the case (binary).
- term_year: Year in the judge's term.
- term_length: Length of the judge’s term in years.
- last_two_years: Whether the judge is in the final two years of their term.

Voting and Opinions

- dissent_vote: Whether the judge dissented in their vote (binary).
- dissent_alone: Whether the judge dissented alone (binary).
- concur_vote: Whether the judge voted to concur (binary).
- concur_alone: Whether the judge concurred alone (binary).
- adhoc_case: Whether the case involved an ad hoc judge (binary).
- violation_vote: Whether the judge voted that a violation occurred (binary).
- case_separate_opinion: Whether the case included a separate opinion (binary).

• number_of_separate_opinions: Total number of separate opinions in the case.

• number_of_concurring_opinions: Total number of concurring opinions in the case.

• number_of_dissenting_opinions: Total number of dissenting opinions in the case.

• number_of_all_opinions: Total number of opinions in the case.

• concur_case: Indicator for a concurring case.

• dissent_case: Indicator for a dissenting case.

Country Attributes

• respondent_country: Country alleged to have violated rights.

• respondent_country_2, respondent_country_3, respondent_country_4, respondent_country_5: Alternative or additional respondent countries involved in the case.

• region: Geopolitical region of the judge's country.

• southern_europe: Whether the judge's country is in Southern Europe (binary).

• eastern_europe: Whether the judge's country is in Eastern Europe (binary).

• v2x_libdem: Liberal democracy index of the judge's country.

• v2x_gender: Gender equality index in the judge's country.

• eu: Binary indicator for EU membership of the judge's country.

• EU_membership_period: Duration of EU membership of judge's country (if applicable).

• eu_years: Total number of years judge's country has been a member of the EU.

• separate_opinion_tradition: Indicator for a tradition of issuing separate opinions in the judge's country

Part II: Lab Report (30 % of the grade)

- 2-page maximum in IEEE two-column format. Use the IEEE manuscript template, available at: **IEEE Templates**. Include you name (or your name and your group member's name), all of your findings and the visuals that you created.
- Reports expected to have following parts:

  1. Abstract: A short summary of your report (This part should include a very brief summary of your methods and analysis and the answer to why you think what you have done is important)

  2. Introduction: A summary of what you expected and did, and two-three of your most significant findings (please use some numerical results here)

3. Data Analysis: Overview of variables and descriptive findings

The Data Analysis section should include the following components:

3.a. Create a summary table showing the **minimum, average, median, and maximum values** for key variables: ***Progovernment_vote, v2x_libdem, female, judge_age,*** and ***home***

3.b. Use the training data to analyze the distribution of pro-government votes and identify patterns related to judge and country characteristics. Include **at least two visualizations** to illustrate key trends and comparisons.

Suggested analyses include:

- Proportion of pro-government votes by judge, region, or issue type
- Respondent country trends by year
- Voting trends by gender, professional background of the judge (e.g., academic, diplomat, lawyer)
- Visualization of judges from EU vs. non-EU voting patterns

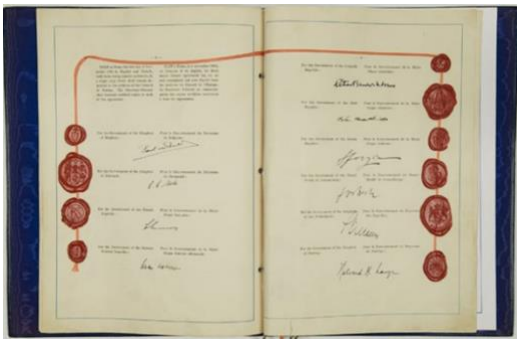
4. Methods: Provide a description of your strategy and the steps you took to improve your prediction model (this includes the steps you followed for data-preprocessing, setting up the model, and checking the strength of the model).

5. Results: A detailed discussion on the results you obtained. What is your Accuracy value? Evaluate and criticize yourself / your team. You may include a **confusion matrix** or **performance summary**.

Name the file **kaggle_project_report.pdf.**

Submit **kaggle_project_report. pdf** and **kaggle_project. ipynb** to **MyCourses→Assignments→Final Project**

Some Interesting Facts about the European Court of Human Rights



A Court for Human Rights in Europe

The European Court of Human Rights (ECHR) was established in 1959 based on the European Convention on Human Rights, a treaty adopted by the Council of Europe to protect fundamental rights and freedoms across its member states.

Not an EU Court!

Despite popular belief, the ECHR is not part of the European Union. It belongs to the Council of Europe, which has 46 member states, including many non-EU countries like Turkey and Ukraine.

A Judge for Every Country

The ECHR has one judge from each member state. Judges do not represent their country's interests—they serve independently. Whether a country is large like Germany or small like Liechtenstein, it gets an equal voice at the Court.

A Futuristic Fortress for Justice

Located in Strasbourg, France, the ECHR building is an architectural icon. Its twin cylindrical towers and glass center symbolize transparency and fairness in European human rights law.

What do judges wear?

ECHR judges wear black robes with a white band and the blue European flag badge on the left shoulder. For more judge fashion: https://x.com/siir_ezgi/status/1702191772699336928

Female judges

Among 46 judges currently there 17 female judges which is 37 %.