A Hybrid Approach for Identifying Political Ideology and Power in Multilingual Parliamentary Speeches

HALE LAB AT NITK

Sevitha Simhadri, Mauli Mehulkumar Patel, Sowmya Kamath S.

Agenda

01 Problem Statement

02 Methodology

03 Results & Key Insights

04 Conlusion & Future Work

Problem Statement

- **Goal:** Identify political ideology and power (government vs. opposition) in parliamentary speeches.
- **Challenges**: Multilingual parliamentary speeches, introducing complexity in identifying political stances across varied languages.
- **Why this matters:** Parliamentary debates influence national policies and international relations.
- **Sub-tasks:**
    a. Identify ideology (left/right).
    b. Identify government/opposition stance

id	speaker	sex	text	text_en	label
lv00000	12e4d29ea4c1af0c8	M	Godājamais Prezidij	Honorable Bureau! D	1
lv00001	382e15bb9fedd997c	F	Labrīt, cienījamā Sa	Good morning, hono	1
lv00002	985899224d296176	M	Labdien, kolēģi! Vēl	Hello, colleagues! I'c	1
lv00003	e587d55787bf7611f	M	Cienītā Saeimas prie	President of the Sae	1
lv00004	6660d123d6afa6545	M	Labdien, cienījamā	Good afternoon, Mac	1
lv00005	f1bb423f1c0e74ebe	M	Augsti godājamā Sa	Highly honored Spea	1
lv00006	382e15bb9fedd997c	F	Cienījamā Saeimas	Dear Chairman of the	1
lv00007	508faeedcac361491	M	Cienījamie kolēģi, m	Ladies and gentleme	0
lv00008	12e4d29ea4c1af0c8	M	Cienījamā priekšsēdē	Dear President, Colle	1
lv00009	25e789d80c9636e5	M	Godātais sēdes vadīt	Honorary Head of the	1
lv00010	857ecf8abad953155	M	Cienījamie kolēģi, tā	Dear colleagues, this	1

Political Orientation

id	speaker	sex	text	text_en	label
lv02806	7af40836f6973e473	M	Labdien, dāmas un k	Hello, ladies and ger	1
lv02807	8444d76c9a547a92	F	Labrīt, godātie kolēģ	Good morning, dear	0
lv02808	68b92d492fe282a16	M	Godātais Prezidij! Kc	Dear Bureau! Collea	0
lv02809	084456182fd5359da	F	Cienījamā Saeimas	Dear President of the	0
lv02810	0a97cdc65b4cc276	F	Labdien, kolēģi! Es n	Hello, colleagues! I v	1
lv02811	85e3463cd1a85669	M	Labdien, cienījamā	Good afternoon, Mac	1
lv02812	e98e7c657fc992eee	M	Kolēģi! Šā gada 3.jū	Colleagues! On 3 Jul	0
lv02813	49c6c3fa99b3e38ef	M	Labrīt, cienījamie ko	Good morning, hono	0
lv02814	84ec9b59fa385d60e	F	Labdien, godātie kol	Good afternoon, dea	1
lv02815	0a97cdc65b4cc276	F	Labdien, kolēģi! Izgl	Hello, colleagues! Tl	1
lv02816	e587d55787bf7611f	M	Cienītā Saeimas prie	President of the Sae	0

Power Identification

Dataset - ParlaMint Corpus

- Speeches from 28 countries for ideology, 25 for power
- Multilingual text covering multiple political systems
- id (unique id for each text)
- speaker (unique id for each speaker, multiple speeches from the same speaker can exist)
- sex (gender of the speaker)
- text (transcribed text of the parliament speech)
- text_en (english translated speech)
- label (0 for left/government, 1 for right/opposition)
- **Real Challenge:** Handling diversity across languages while preserving semantic meaning.

Methodology

- **Data Preprocessing**
  - Tokenization, Stopword Removal, Lemmatization
- **Embedding**
  - TF-IDF
  - LASER
- **Classifier**
  - Traditional
    - SVC, Logistic Regression, Naive Bayes, Random Forest, Gradient Boosting, XGBoost,
  - Deep Learning
    - Simple Neural Network, LSTM
- **Evaluation**
  - F1 score for self evaluation

Results and Key Insights

Task - 1

Team	F_1-score																												
	Overall	AT	BA	BE	BG	CZ	DK	EE	ES	ES-CT	ES-GA	FI	FR	GB	GR	HR	HU	IS	IT	LV	NL	NO	PL	PT	RS	SE	SI	TR	UA
Policy Parsing Panthers	79	77	51	71	77	63	84	64	94	80	98	77	75	92	89	65	87	71	77	67	71	82	88	95	79	95	78	93	83
gerber	63	60	45	54	62	52	56	00	77	66	76	54	58	76	72	51	69	00	60	49	59	00	72	69	64	00	58	84	73
HALE Lab	61	56	44	59	60	52	56	52	76	69	84	52	48	74	71	43	67	57	60	49	53	61	62	67	55	77	49	83	60
Pixel Phantoms	59	58	49	56	56	47	56	54	72	64	75	59	58	72	71	55	68	57	57	54	60	54	59	54	51	61	47	78	56
Ssnites	59	50	53	55	53	50	61	52	61	58	64	55	56	64	59	53	60	58	53	51	56	66	71	64	64	75	58	79	53
Trojan Horses	59	61	25	57	61	51	60	57	72	67	00	33	60	73	74	53	71	55	66	00	60	61	68	63	00	74	00	80	68
INSA Passau	59	60	53	54	61	47	57	53	63	61	66	34	58	69	59	56	66	56	56	54	56	58	69	55	61	66	51	80	62
JU_NLP_DID	57	53	42	42	55	51	60	57	69	57	70	00	50	71	63	43	60	55	61	47	56	59	51	67	48	73	46	77	57
Baseline	56	52	42	45	53	52	56	47	72	65	67	54	43	74	74	43	57	39	56	45	51	62	46	63	53	75	39	84	58

- **Objective:** Classify speeches as government-oriented or opposition-oriented based on their ideological stance.
- **Baseline Performance:** F1 score of 0.56.
- **Our System Performance:** We achieved an F1 score of 0.61, surpassing the baseline by a significant margin.
- **Key Improvements:** Our hybrid approach, combining traditional ML (TF-IDF + Logistic Regression) with deep learning (LASER embeddings + LSTM), enabled us to better capture nuanced ideological stances across multiple languages.
- **Challenges:** Variability in ideological rhetoric across languages required the robust handling of multilingual texts.

Results and Key Insights

Task - 2

Team	F_1-score																								
	Overall	AT	BA	BE	BG	CZ	DK	ES	ES-CT	ES-GA	ES-PV	FI	FR	GB	GR	HR	HU	IT	LV	NL	PL	PT	RS	SI	TR	UA
Policy Parsing Panthers	83	88	56	74	81	78	87	88	91	98	90	80	82	83	95	75	97	78	75	74	90	85	84	81	94	65
HALE Lab	70	69	46	61	68	69	70	65	85	88	78	65	67	75	82	68	88	69	62	64	78	65	69	61	84	49
Trojan Horses	69	72	57	63	67	63	68	69	82	85	74	39	66	72	83	67	86	72	64	64	74	65	75	62	83	56
gerber	68	68	51	60	66	64	63	72	80	86	74	60	71	72	68	63	87	52	63	64	77	66	73	58	84	48
Vayam Solve Kurmaha	68	48	48	65	69	68	69	72	83	87	76	35	66	47	85	67	88	72	62	68	75	67	75	63	85	48
Pixel Phantoms	66	70	50	59	63	65	69	65	64	77	69	61	64	73	72	57	80	69	58	62	70	66	69	60	80	52
Baseline	64	66	45	61	68	64	56	65	78	83	71	56	66	71	63	60	86	43	51	62	76	62	65	53	83	46
JU_NLP_DID	63	68	47	55	58	57	67	60	78	55	72	00	59	00	77	65	83	71	47	63	70	63	54	56	78	43
INSA Passau	62	67	45	60	66	65	54	65	00	00	00	56	66	72	56	61	85	45	52	64	77	62	63	54	84	47
Ssnites	60	66	45	58	60	61	61	62	58	62	60	60	65	60	69	65	79	62	54	57	62	58	60	57	61	46

- **Objective:** Classify whether a speaker belongs to the government or opposition in parliamentary debates.
- **Baseline Performance:** F1 score of 0.64.
- **Our System Performance**: We outperformed the baseline with an F1 score of 0.70, marking a clear improvement.
- **Key Factors for Success:** The combination of deep learning methods like LASER embeddings, which handle multilingual data effectively, alongside traditional ML models, provided a better distinction between government and opposition speeches.
- **Challenges:** Multilingual variability and unclear speaker affiliations posed challenges. Additionally, capturing the influence of parliamentary context on power status added complexity to classification.

Conclusion & Future Work

Conclusion:

- Successfully identified political ideologies and power structures in parliamentary debates using diverse features and advanced models.
- Addressed challenges posed by multilingual and heterogeneous datasets.

Future Work:

- Investigate speech relationships and conversational flow using dialogue act recognition.
- Expand to more languages and legislative contexts to enhance model generalizability.

Thank you!

We are happy to take any questions or feedback!
You can also mail us at maulipatel03@gmail.com

Meet the Team

Get to know the faces behind the work

Simhadri Sevitha

Mauli Mehulkumar Patel

Sowmya Kamath S.