

UNIVERSITÀ DELLA CALABRIA

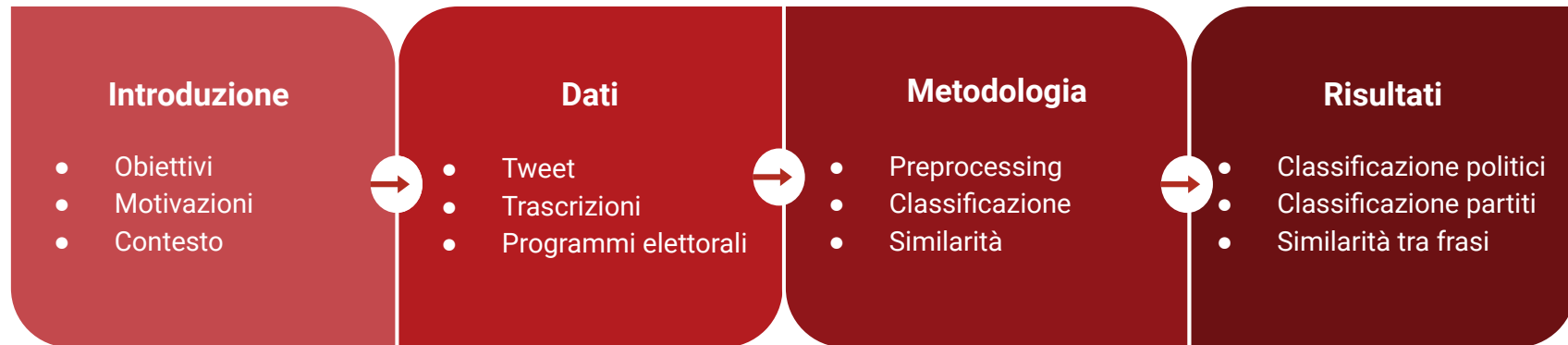
Dipartimento di Ingegneria Informatica,
Modellistica, Elettronica e
Sistemistica

Transformer-based Political Language Modeling
for the Italian landscape through the lens of
Twitter and YouTube



ANNO ACCADEMICO 2021/2022

Agenda



Introduzione

Introduzione - Obiettivi

1.

Coerenza contenuti social politici italiani

Valutare la coerenza dei contenuti social (Tweet e Speech) dei politici

2.

Coerenza tra programmi elettorali e contenuti social

Valutare coerenza e correlazione tra contenuti social dei politici e i rispettivi programmi elettorali

3.

Similarità tra programmi elettorali

Valutare quali programmi elettorali sono più simili tra di loro e su quali argomenti

Introduzione - **Motivazioni**

**Comunicazione
Politica**

Superficiale e semplicistica

Social Network

Contenuti brevi e immediati

**Panorama Politico
Italiano**

Fortemente frammentato e diviso

Introduzione - Politici e partiti



Enrico Letta
Partito Democratico



Silvio Berlusconi
Forza Italia



Giuseppe Conte
Movimento 5 Stelle



Matteo Salvini
Lega



Matteo Renzi - Carlo Calenda
ItaliaViva - Azione



Giorgia Meloni
Fratelli d'Italia

Introduzione - Contesto politico



Introduzione - Letteratura

Approccio inedito

Language-model poco utilizzati in questo contesto

Classificazione originale

Pochi esempi d'analisi che tengono conto dei discorsi politici

Similarità tra programmi

Analisi originale e poco approfondita in letteratura

Dati

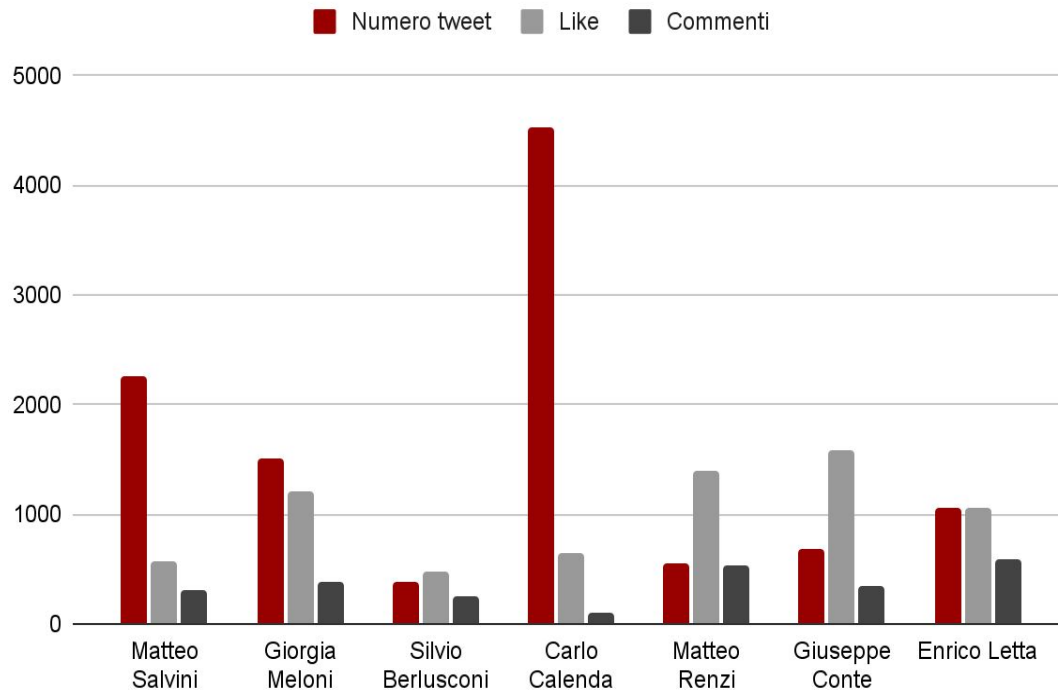
Dati - Tweet

Dettagli estrazione:

- Twitter API
- Periodo di osservazione

Osservazioni:

- Destra molto più presente della sinistra
- **Conte** molto *apprezzato*
- **Berlusconi** poco presente
- **Calenda** ha poche interazioni
- **Letta** molto commentato



Dati - Speech (Trascrizioni video YouTube)

Dettagli estrazione:

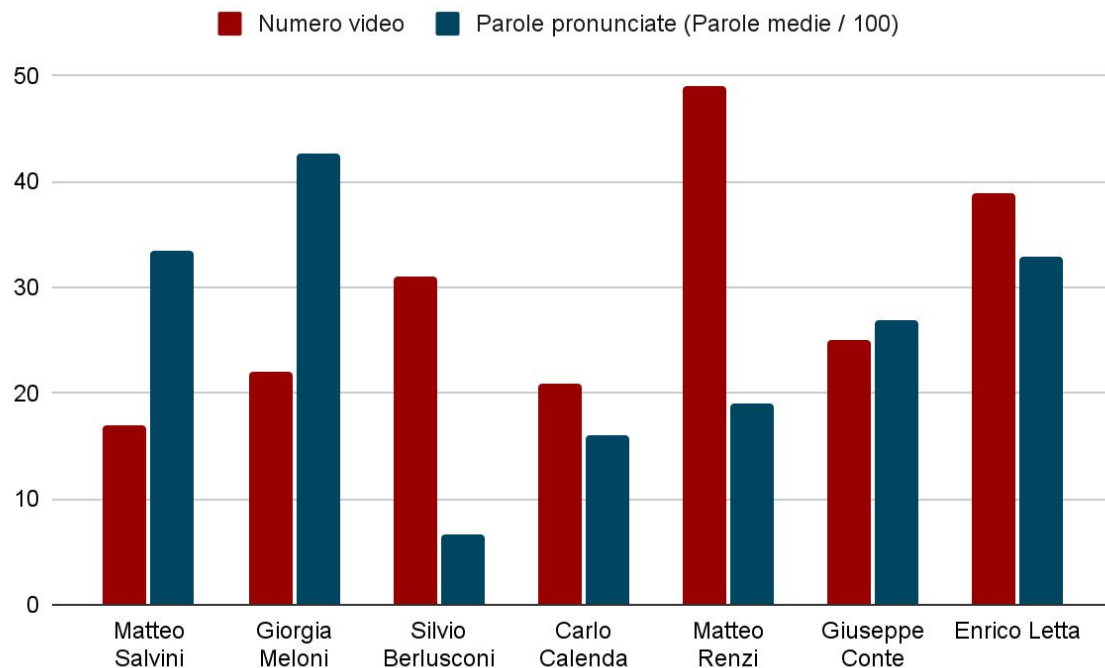
→ Tipologia di video:

- ◆ comizi
- ◆ conferenze
- ◆ riunioni
- ◆ dirette online

→ Estrazione trascrizioni

Osservazioni:

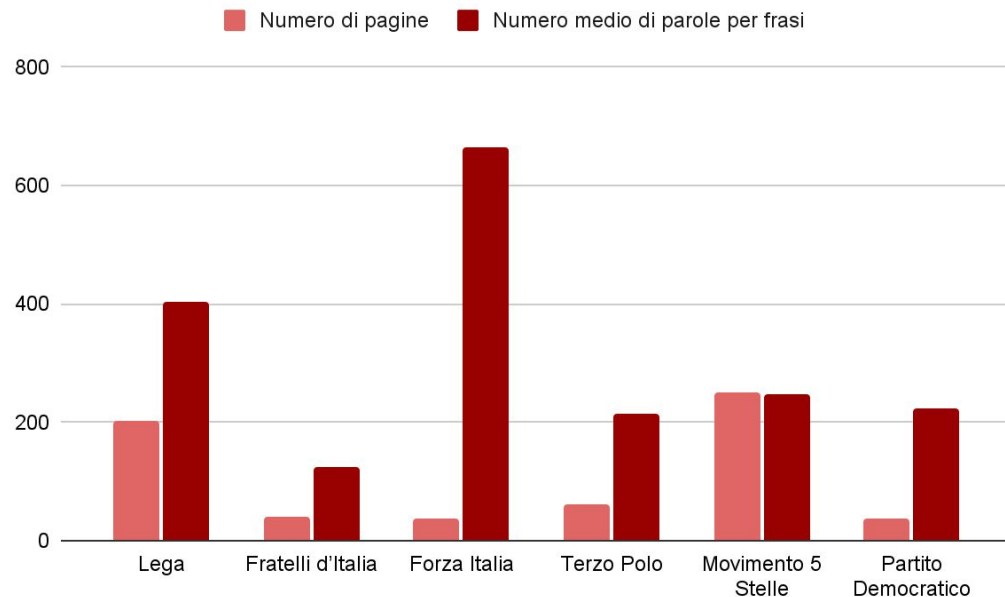
- **Meloni:** politico con più visualizzazioni e like
- **Letta:** poche visualizzazioni e bassa interazione



Dati - Programmi Elettorali

Osservazioni:

- **Lega** e **M5S** presentano programmi molto lunghi
- Il programma di **Forza Italia** è un elenco di proposte
- Programma **PD** organizzato per proposte chiave
- **Fdl** e **Terzo Polo** propongono un programma ben strutturato



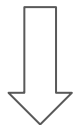
Metodologie

Metodologie - Pre-processing e data modelling

1.

**Classificazione politici
su contenuti social**

Coerenza contenuti social
politici italiani



Train

Val

Test

2.

**Classificazione partiti su
programmi elettorali**

Coerenza tra contenuti social
e programmi elettorali



Train

Val

Prog.
Elect.

3.

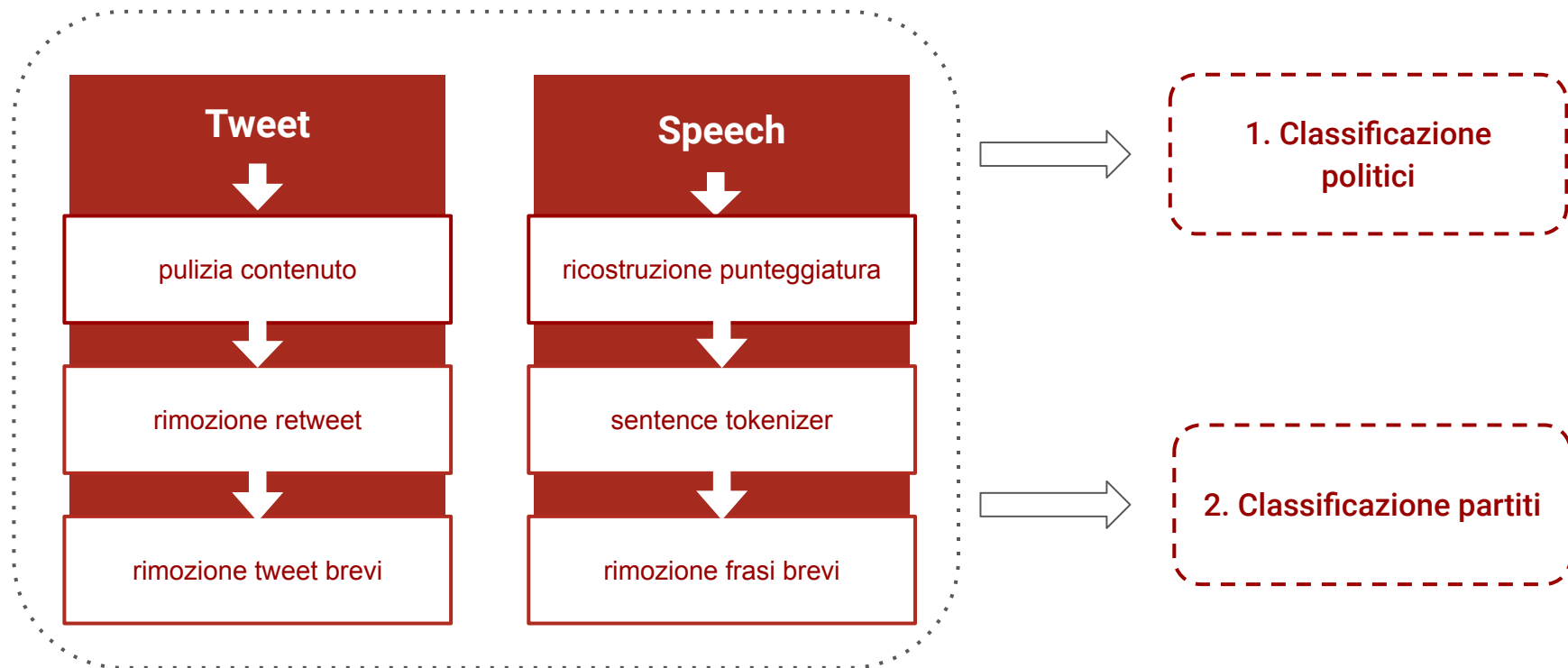
**Similarità tra programmi
elettorali**

Valutare programmi più simili
e su quali argomenti

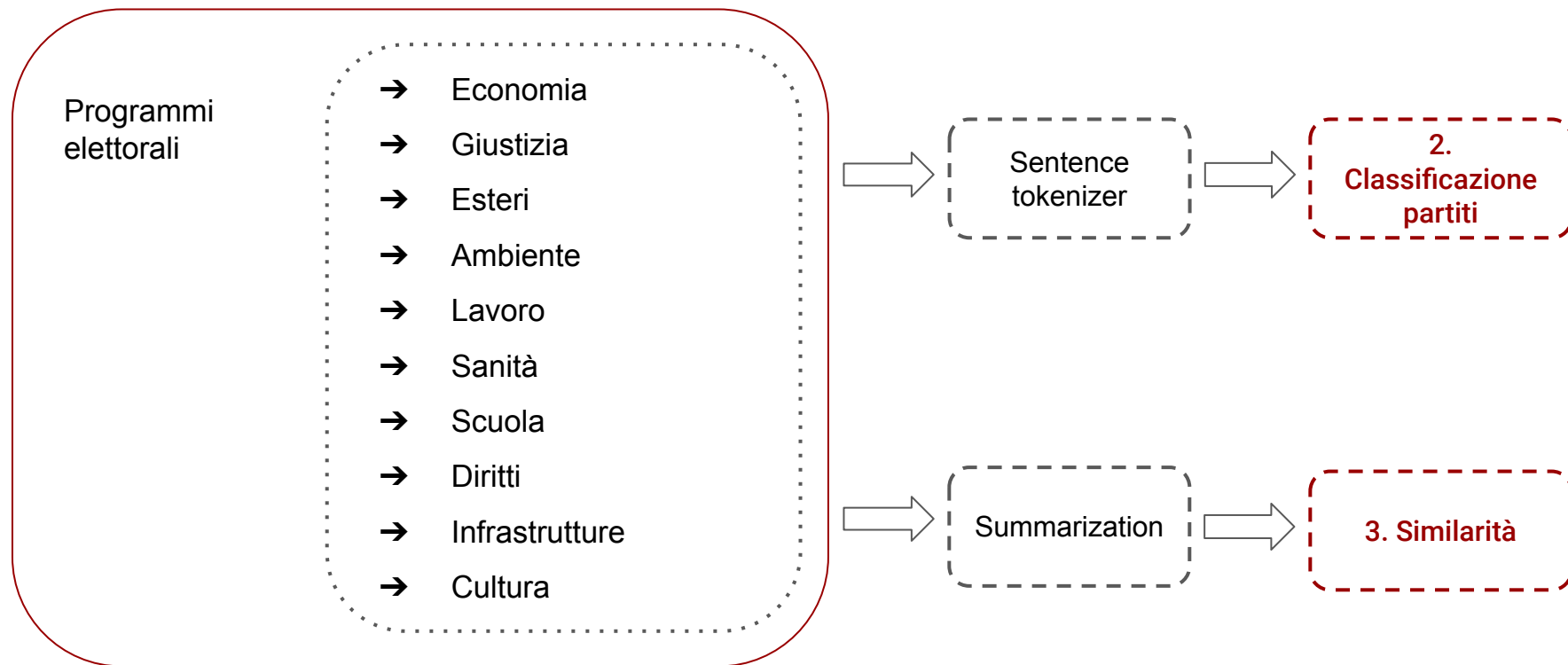


Sum.
Prog.
Elect.

Metodologie - Pre-processing e data modelling - **Tweet e Speech**



Metodologie - Pre-processing e data modelling - Programmi elettorali



Metodologie - Transformers BERT-based

→ **Caratteristiche:**

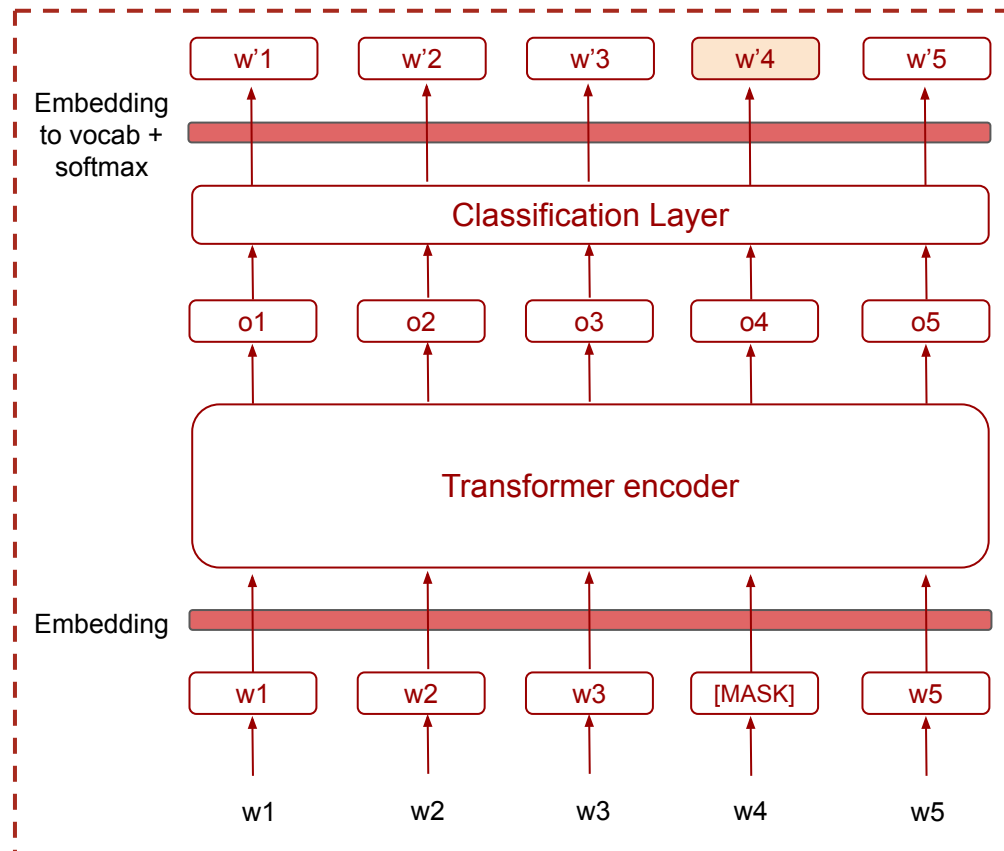
- ◆ considerano l'intera sequenza in ingresso
- ◆ parallelizzazione del training
- ◆ *Attention layer*

→ **Architettura:**

- ◆ *Encoder*

→ **Modelli:**

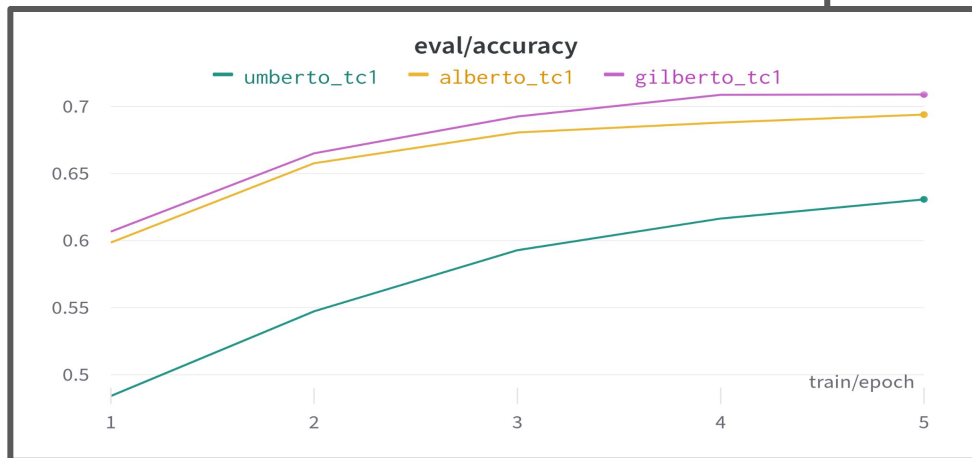
- ◆ BERT-based Italian language models



Metodologie - Classificazione politici

Osservazioni:

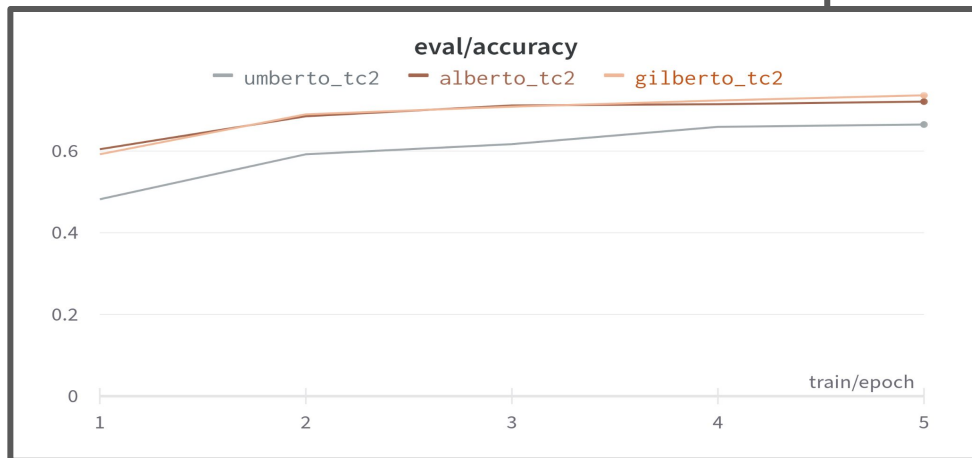
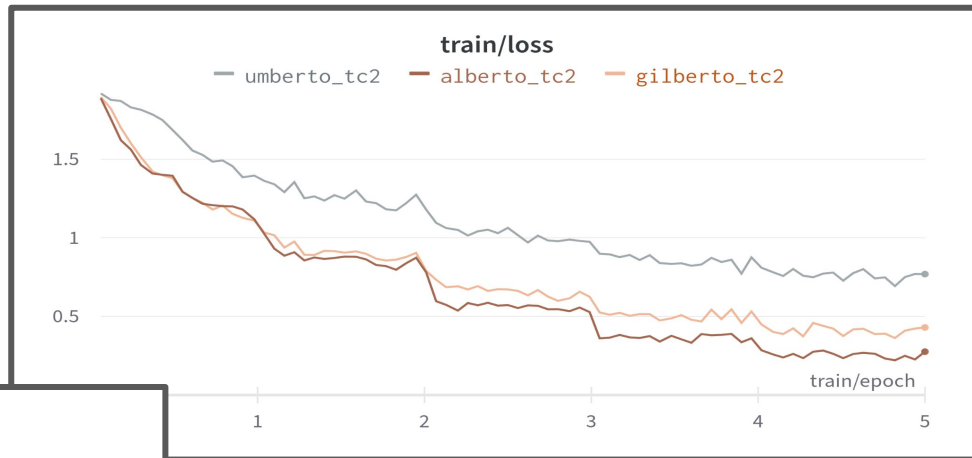
- Training su tweet e speech
- Convergenza dopo poche epoche
- **Gilberto e Alberto** hanno performance migliori
 - ◆ valori di *loss* bassi
 - ◆ *accuracy* più alta



Metodologie - Classificazione partiti

Osservazioni:

- Training su tweet e speech
- Convergenza veloce
- **Gilberto e Alberto** hanno performance migliori
 - ◆ valori di *loss* più bassi
 - ◆ *accuracy* più alta

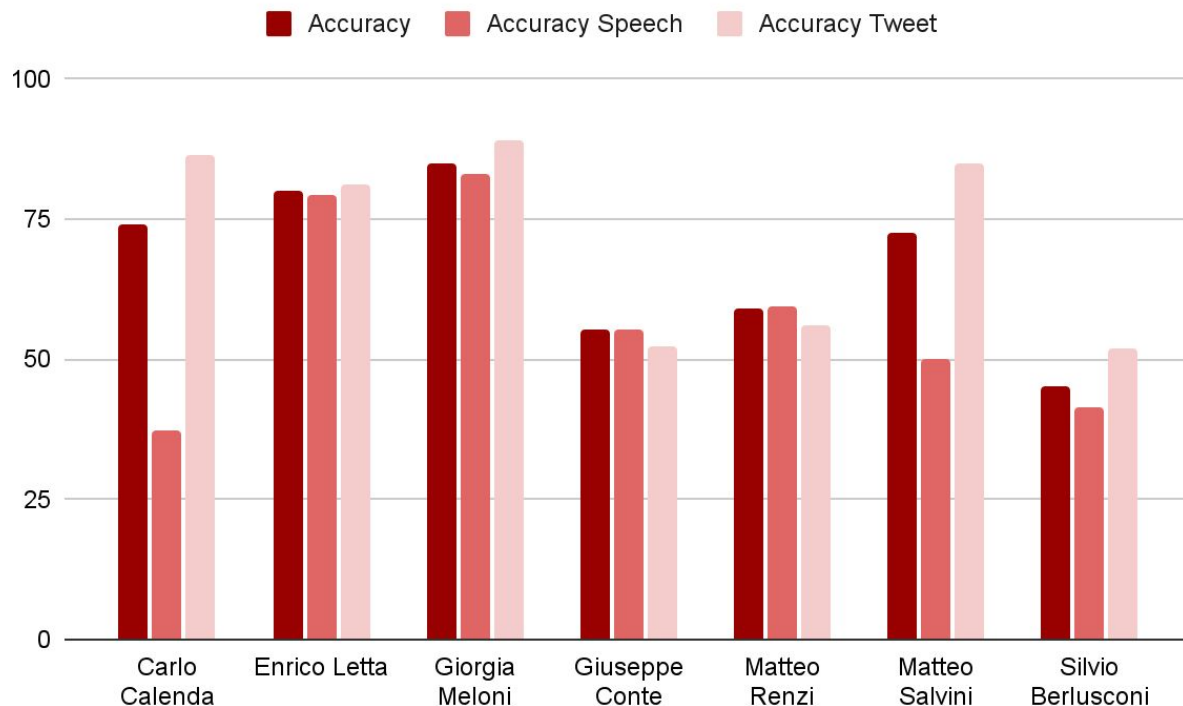


Risultati

Risultati - Classificazione politici su contenuti social

Osservazioni:

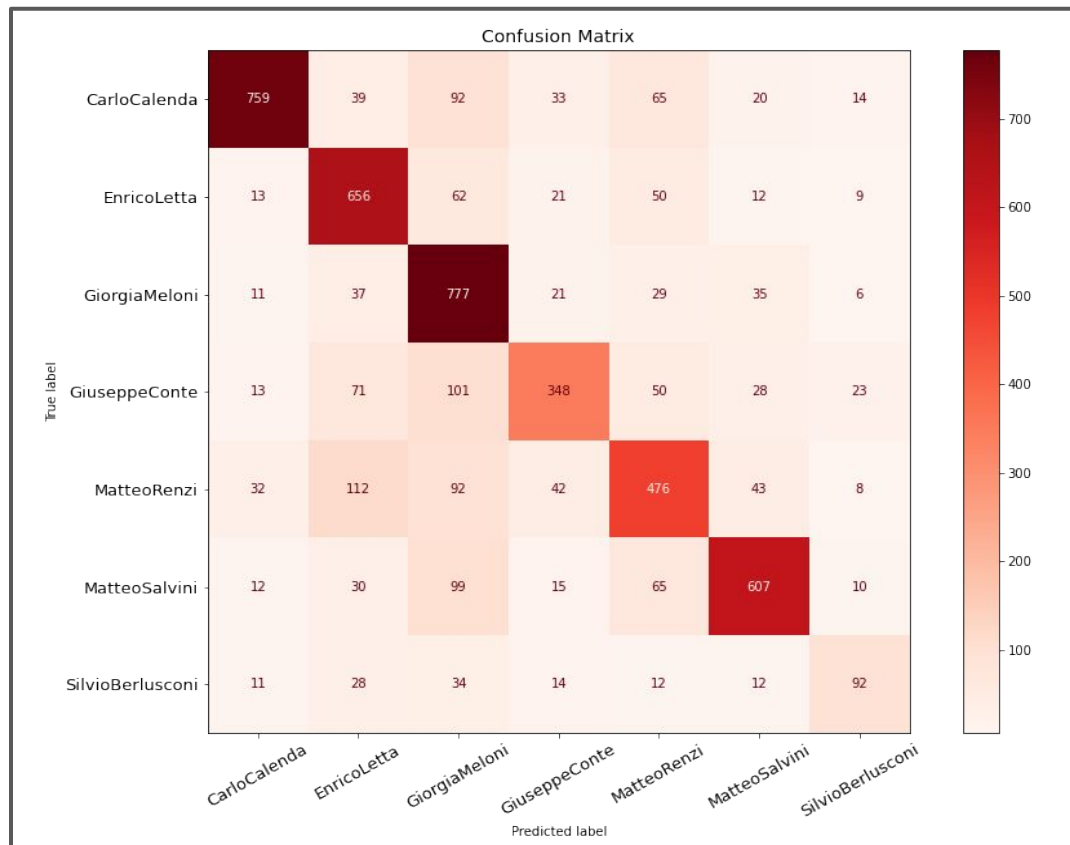
- Criteri di valutazione:
 - ◆ Accuracy: 70%
 - ◆ F1: 70%
 - ◆ Precision: 70%
 - ◆ Recall: 71%
- Accuracy alta: **Meloni** e **Letta**
- Accuracy bassa: **Conte** e **Berlusconi**
- Tweet generalmente meglio riconosciuti
- Accuracy variabile per tipo di dato



Risultati - Classificazione politica su contenuti social

Osservazioni:

- 42% degli errori per **Salvini** sono associati alla **Meloni**
- **Conte** 15% delle volte confuso con **Meloni**
- **Renzi** il 25% delle volte associato a **Letta** o **Meloni**
- **Berlusconi** poco riconosciuto

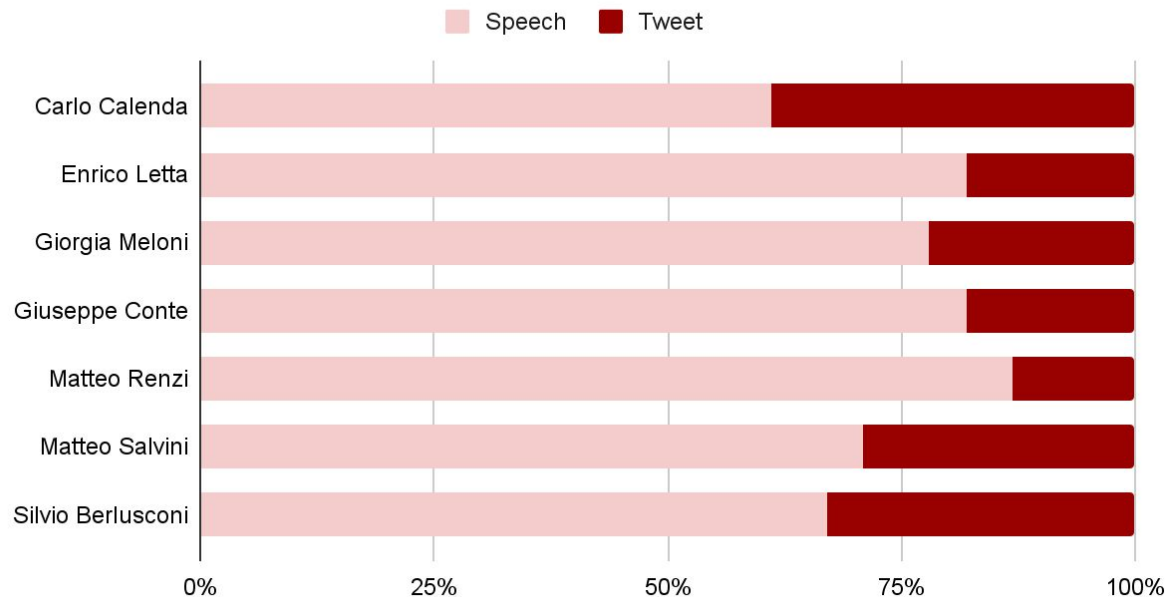


Risultati - Classificazione politici su contenuti social

Osservazioni:

- score di assegnamento molto alto per errori (60%)
- score costante per tutti i politici
- Errori più frequenti sugli speech
 - ◆ frasi decontestualizzate
 - ◆ contenuto poco significativo
- Frequenza maggiore anche in politici con meno speech
 - ◆ Calenda
 - ◆ Salvini

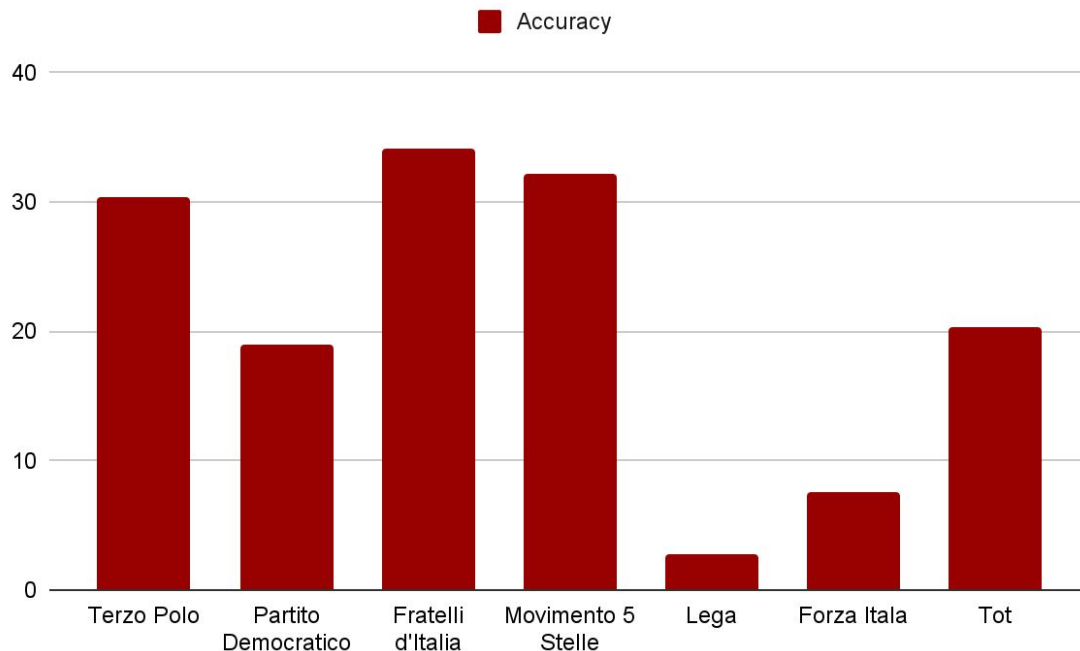
Percentuale di errori per tipo



Risultati - Classificazione partiti su programmi elettorali

Osservazioni:

- Criteri di valutazione:
 - ◆ *Accuracy*: 20%
 - ◆ *F1*: 19%
 - ◆ *Precision*: 24%
 - ◆ *Recall*: 20%
- Accuracy alta per **Terzo Polo**, **Fdl**, **M5S**
- Scarse performance per **Lega** e **FI**
 - ◆ programma **Lega** confuso con **Terzo Polo** e **Fdl**
- **PD** e **M5S** spesso confusi



Risultati - Classificazione partiti su programmi elettorali

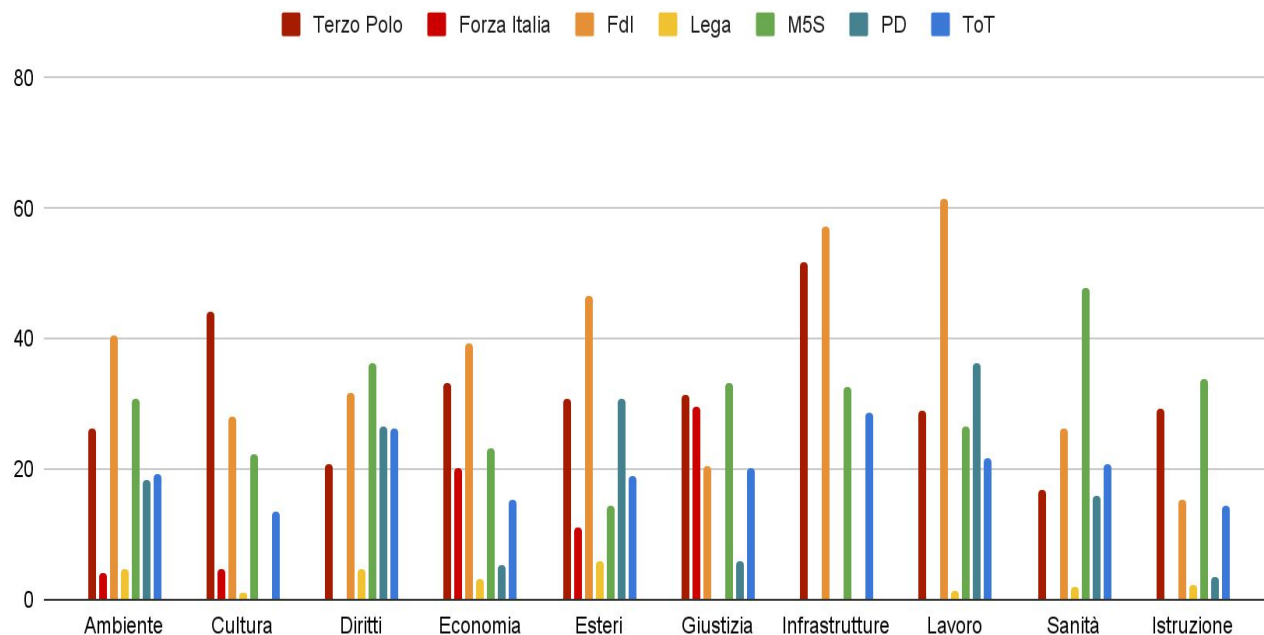


Risultati - Classificazione partiti su programmi elettorali

Osservazioni:

- *Lavoro, Esteri e Ambiente* sono le categorie meglio riconosciute
- Accuracys basse per *Cultura e Istruzione*
- Tendenza a valori migliori per alcuni temi:
 - ◆ **Terzo Polo** accuracy alta in *Cultura*
 - ◆ **Fdl** accuracy alta in *Lavoro*
 - ◆ **PD** accuracy alta in *Diritti*

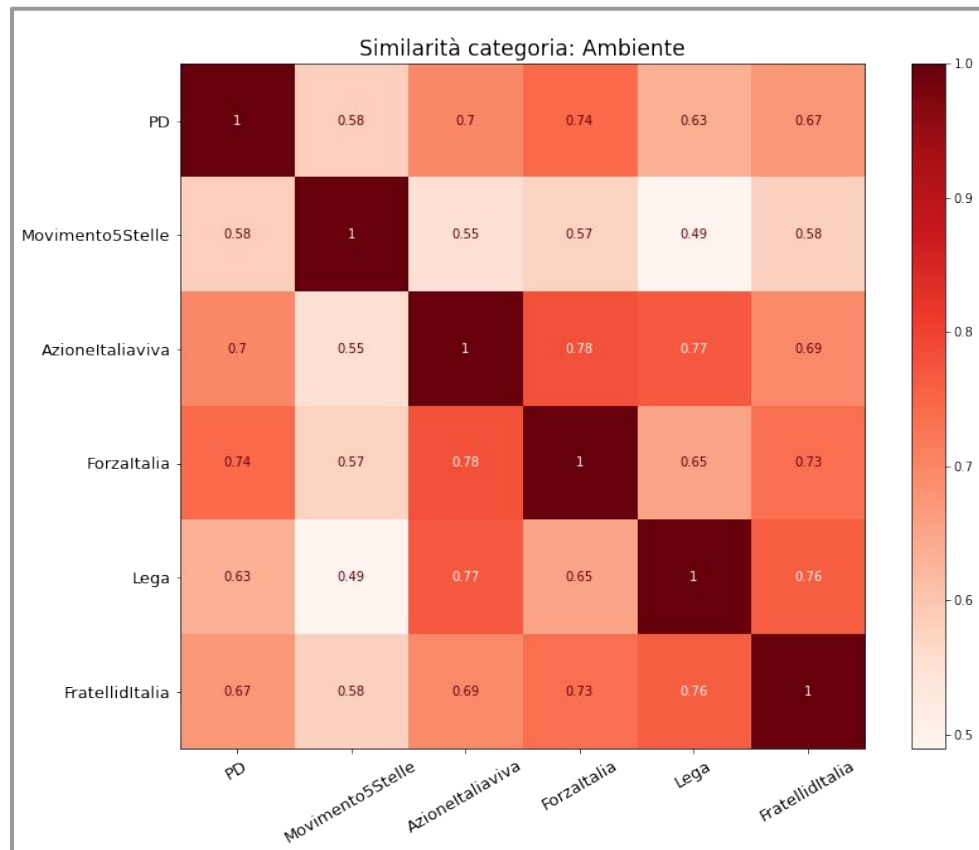
Accuracy per macro-area



Risultati - Similarità tra programmi elettorali

Osservazioni:

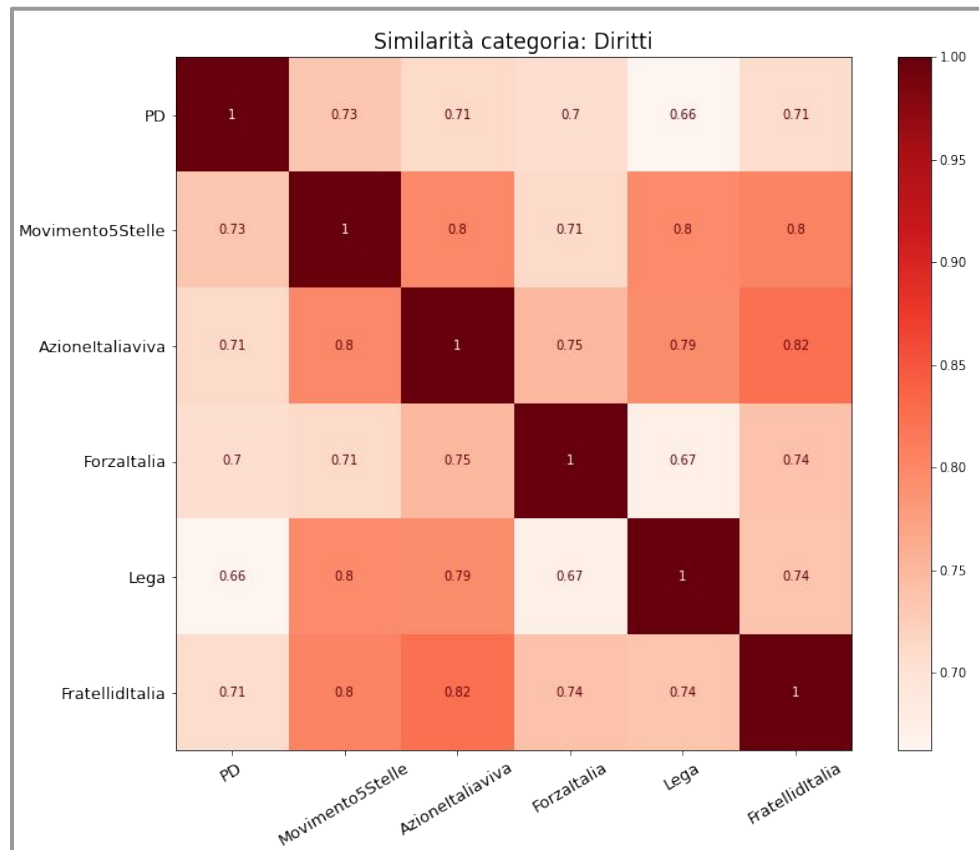
- Alta similarità tra i partiti di *centrodestra*:
 - ◆ centralità tema del gas
 - ◆ centralità tema dei rifiuti
 - ◆ centralità tema dei costi dell'energia
- **PD e FI** usano frasi simili
 - ◆ concetti simili proposti
 - ◆ sotto testo diverso
 - ◆ concetto d'innovazione centrale
- **M5S e PD** hanno bassa similarità
 - ◆ **M5S**: grande importanza alla battaglia contro gli allevamenti intensivi
 - ◆ **PD**: centralità del tema della transizione ecologica



Risultati - Similarità tra programmi elettorali

Osservazioni:

- Tema dei *diritti* affrontato in modo generale
 - ◆ diritto al lavoro, diritto alla pensione e “diritto al futuro”
- frasi più importanti non sono quelle più divisive
 - ◆ non compaiono diritti delle minoranze
 - ◆ diritti comunità *LGBTQ+*
- **Lega** e **Fdi** sono molto simili
 - ◆ centralità sostegno famiglie e lavoro
 - ◆ diritti delle minoranze poco o per nulla menzionati
- **Terzo Polo** molto simile a coalizione di *destra*



Conclusioni

Conclusioni - Criticità

Testo politico di difficile interpretazione	frasi o argomenti di difficile interpretazione e molto specifici
Materiale limitato	contenuto informativo debole in alcuni tweet o speech
Contenuti decontestualizzati	contenuti decontestualizzati sono interpretati diversamente
Modelli non allenati su contenuti politici	concetti politici non comprensibili dai modelli

Conclusioni - Considerazioni Generali

1.

Classificazione politici su contenuti social



- Contenuti social abbastanza riconosciuti
- Alcuni politici risultano difficili da riconoscere
- Gli *speech* sono difficili da classificare

2.

Classificazione partiti su programmi elettorali



- Programmi elettorali molto diversi dai contenuti social
- Programmi elettorali spesso molto specifici
- Pochi elementi presenti sono discussi sui social

3.

Similarità tra programmi elettorali



- Concetti distintivi spesso non evidenti
- Coalizione di *destra* molto compatta
- Similarità bassa tra i partiti di *sinistra*

Fine
