PUBLIC OPINION ON NEWS

SENTIMENT ANALYSIS AND TEXT MINING PROJECT

Nicole Maria Formenti

DSE - Università degli Studi di Milano

AIMS OF THE ANALYSIS

1. Create a multimodal multitask neural network to predict controversy of comments.

2. Study the relationship between controversy and topics of articles.

3. Create unsupervised lexicon-based labels based on sentiment polarity and intensity of comments:

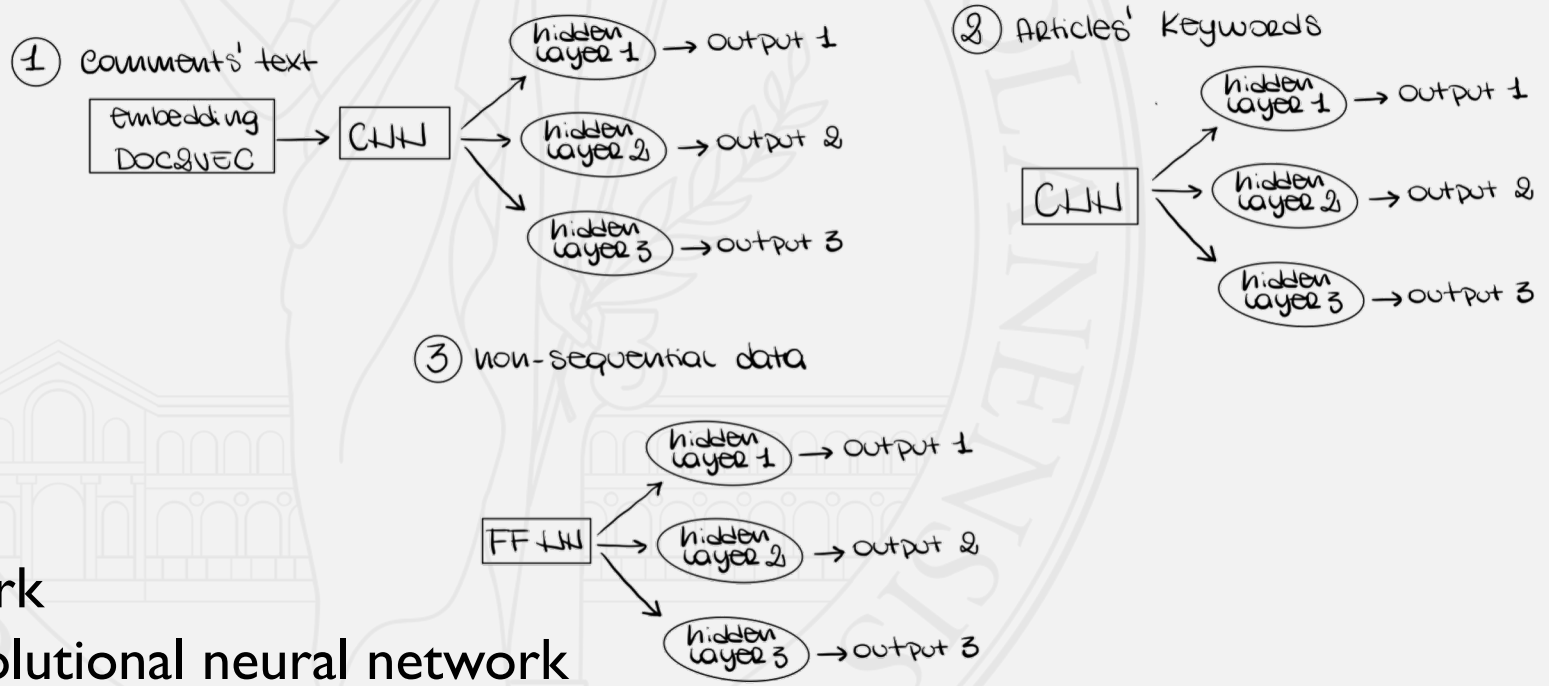
    I. Understand if they are a good predictor for controversy to use them as alternative labels

    II. Study the relationship between sentiment polarity and topics of articles

1. BUILDING MULTITASK AND MULTIMODAL NEURAL NETWORK

3 targets:

- `editorsSelection`
- `recommendations`
- `replyCount`

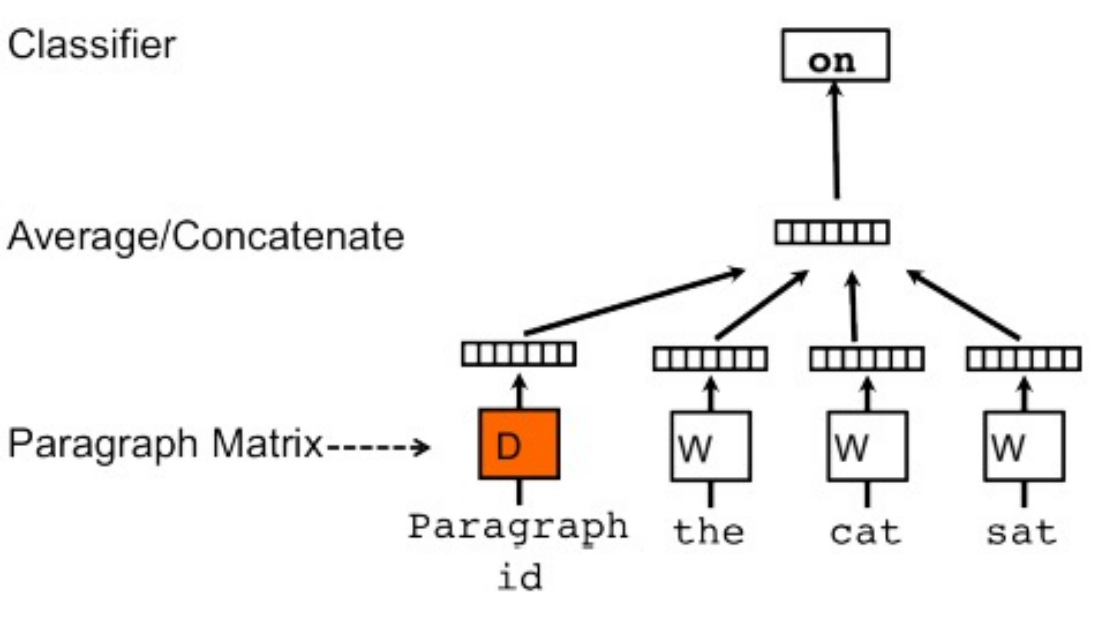


3 Neural Networks:

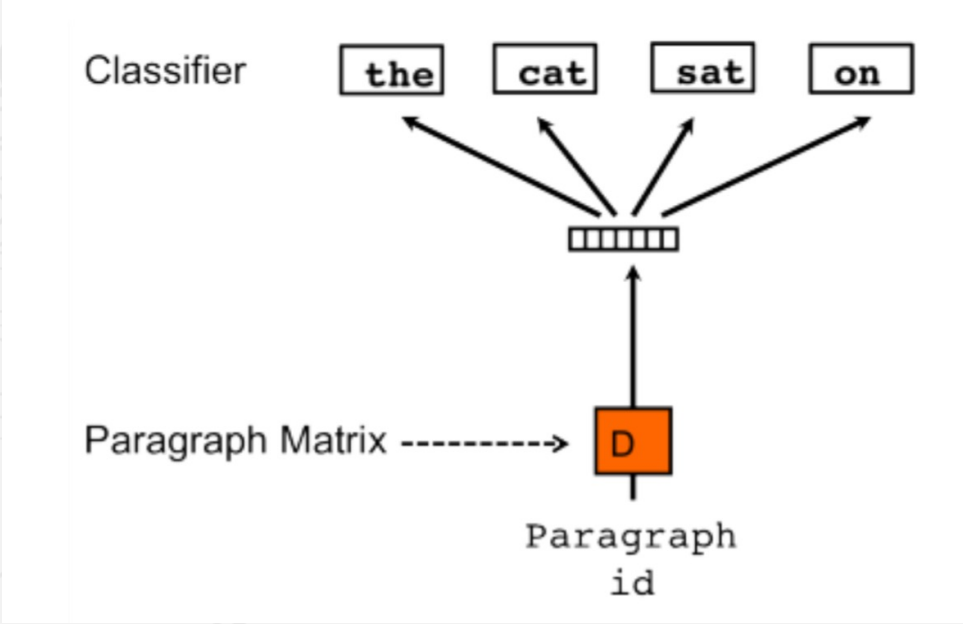
- Feed forward neural network
- TF-IDF embedding + Convolutional neural network
- Doc2vec embedding + Convolutional neural network

DOC2VEC EMBEDDING

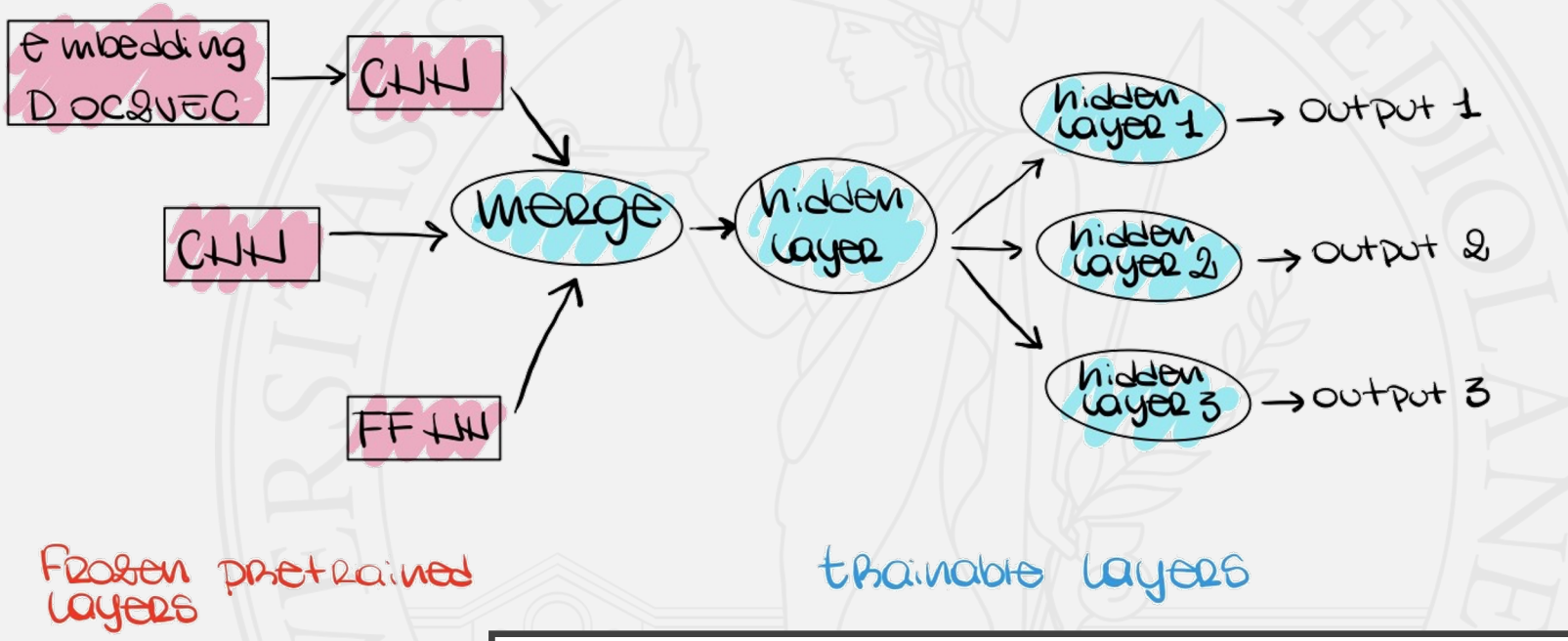
PV-DM



PV-DBOW



Multimodal Neural Network Architecture



Parameters:

- RMSprop optimiser
- Learning rate: 0.00319
- 20 epochs with early stop
- F1 score

FINAL PERFORMANCE:
TESTING F1 SCORE: 0.8539

Good, but might be biased due to imbalance and naive implementation of multimodal neural network

2. CONTROVERSY OF TOPICS

2 topics variables:

- `newDesk`
- `sectionName`

2 Summary target variables for controversy:

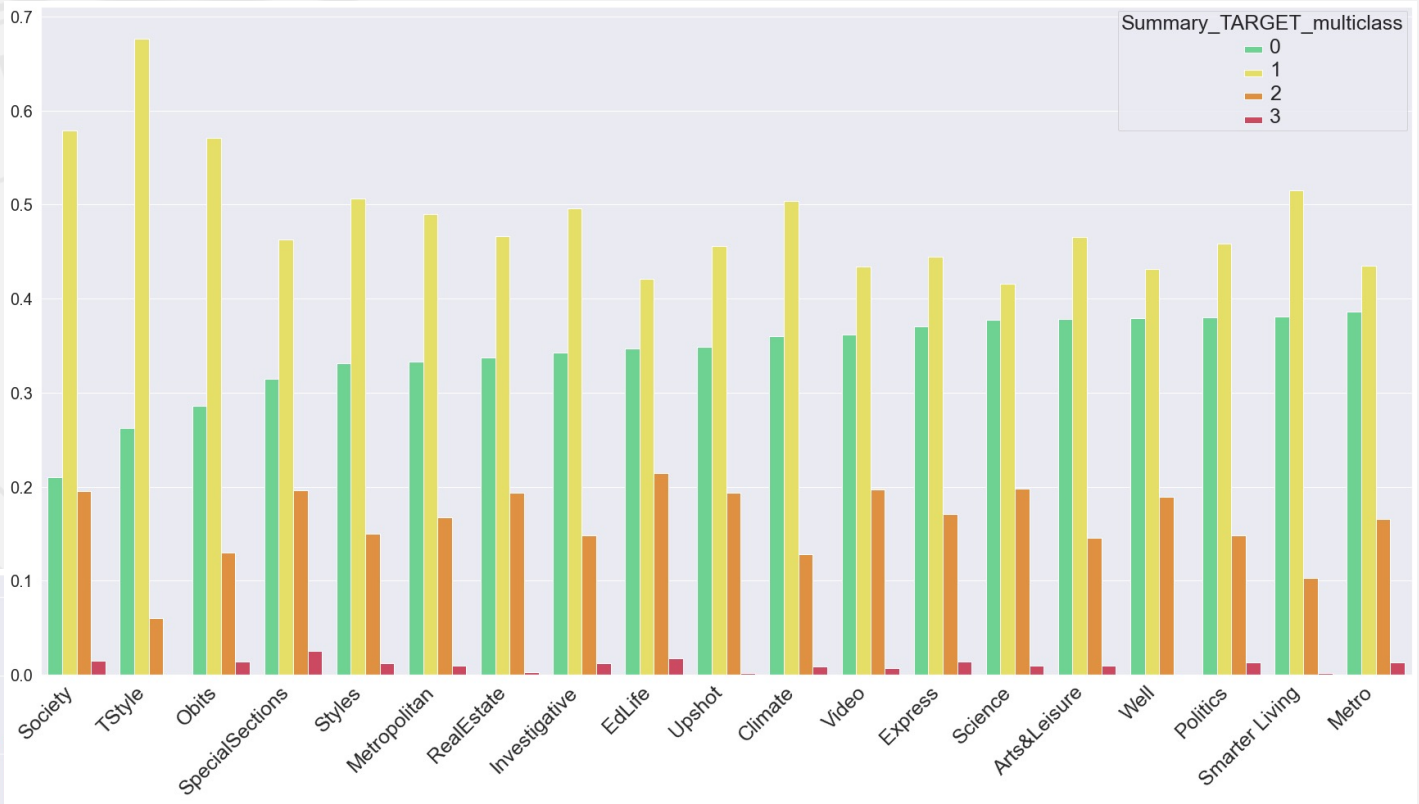
Binary target:

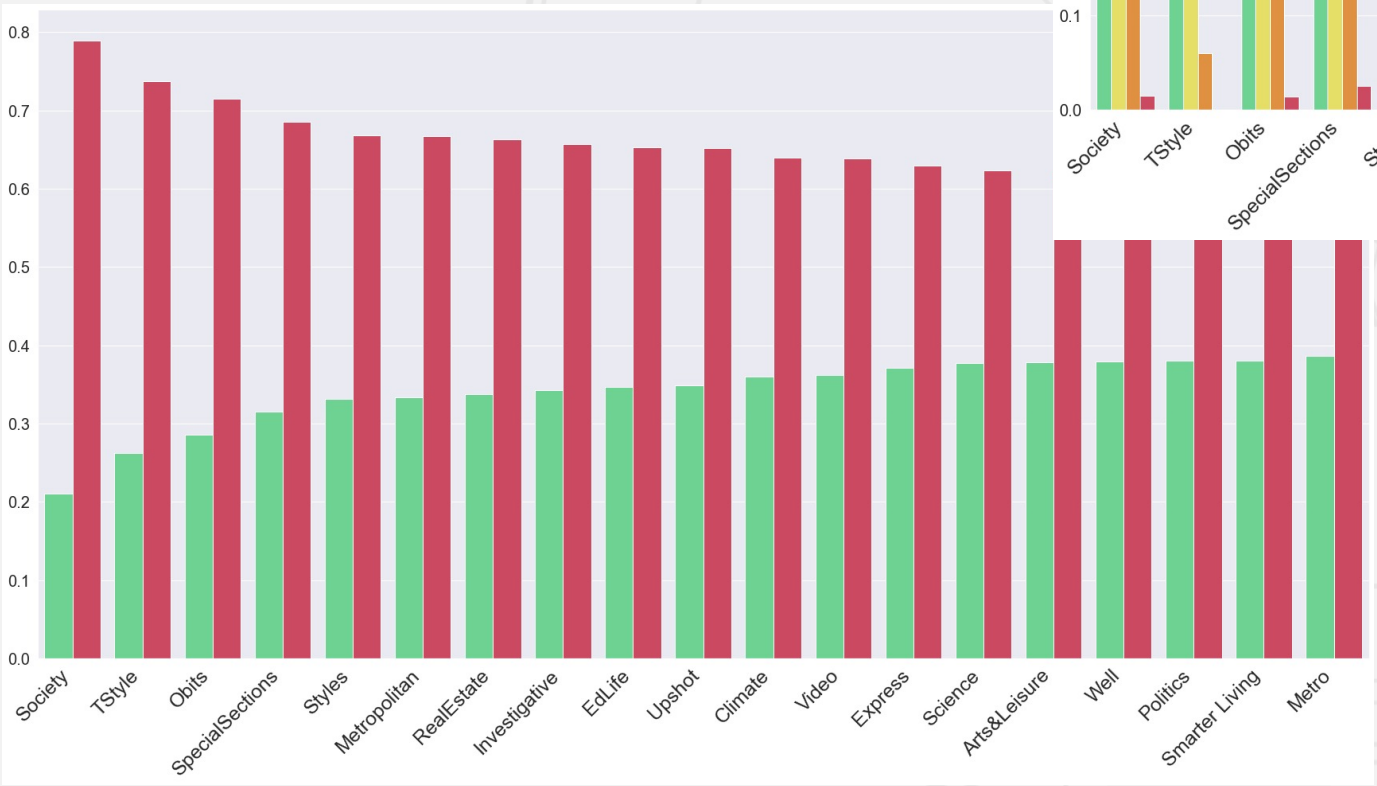
$$\begin{cases} 0 \; if \; no \; positive \; targets \\ 1 \; if \; at \; least \; 1 \; positive \; target \end{cases}$$

Multiclass target:

$$\begin{cases} 0 \; if \; no \; positive \; targets \\ n \; if \; n \; positive \; targets \end{cases}$$

Variable: newDesk

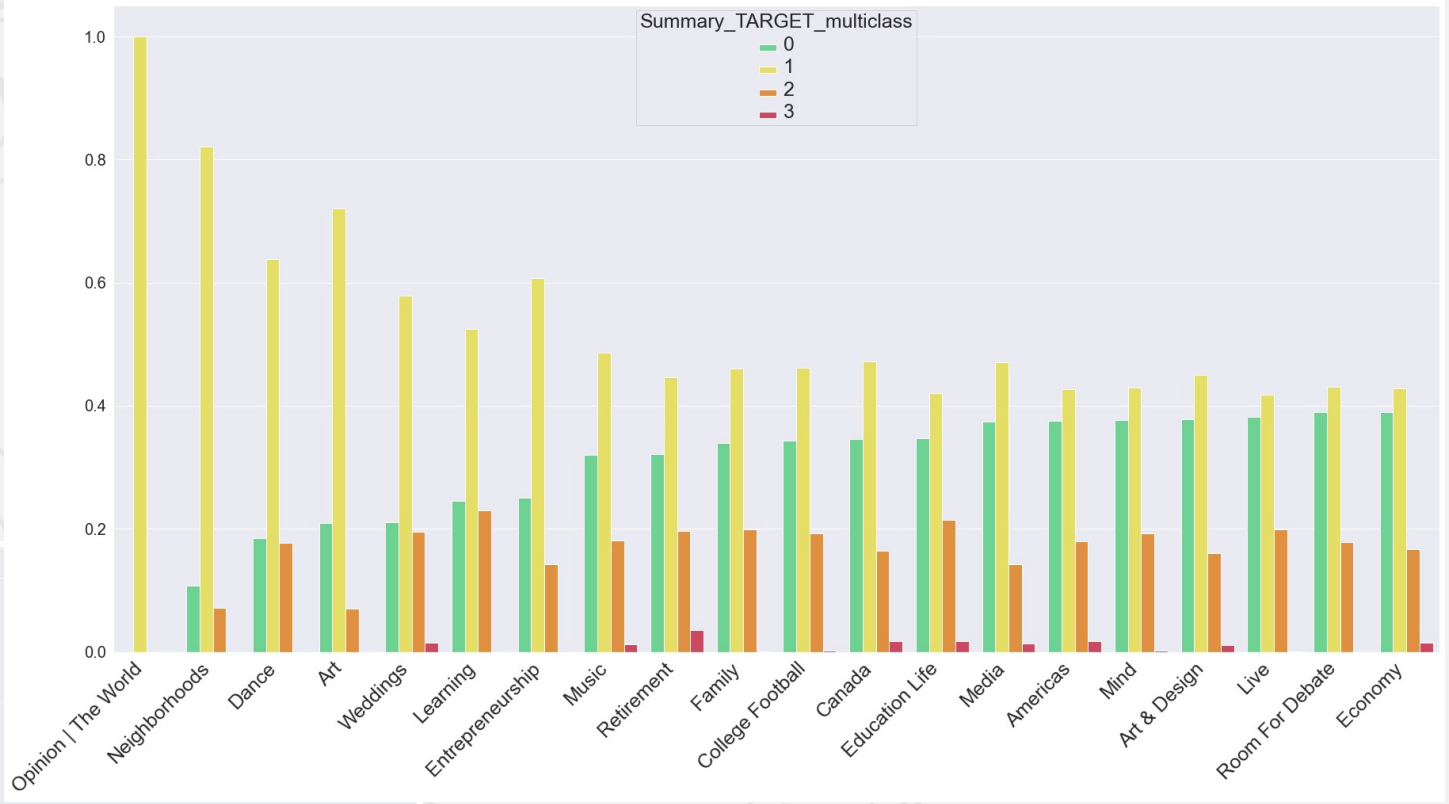


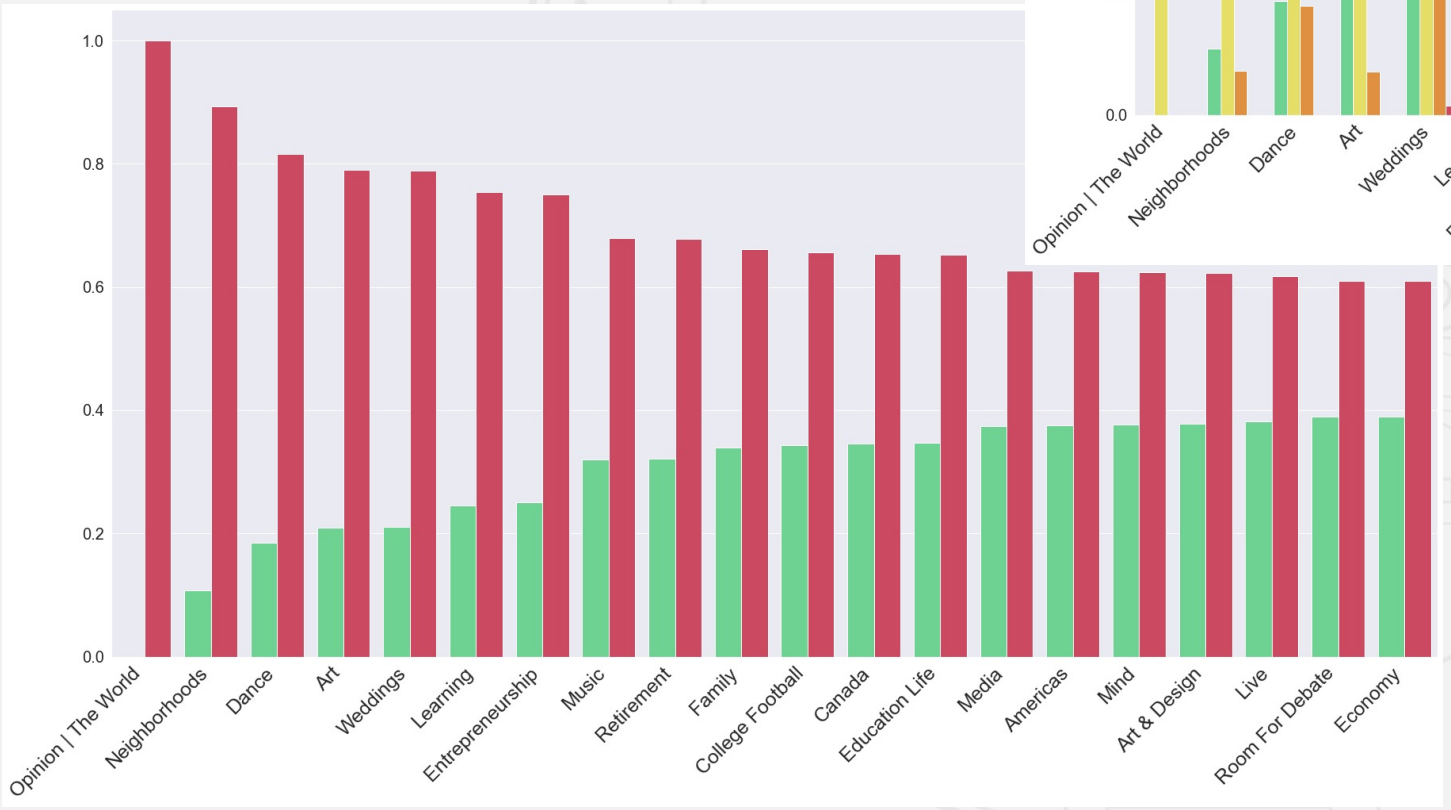


TOP 3 CONTROVERSIAL TOPICS:

1. Society
2. Tstyle
3. Obits

Variable: `sectionName`





TOP 3 CONTROVERSIAL TOPICS:

1. Opinion | The World
2. Neighborhoods
3. Dance

3. LEXICON-BASED SENTIMENT POLARITY OF COMMENTS

1. **Is it a good predictor for the controversy of comments?**

Statistical analysis

- Point Biserial Correlation
- Logistic Regression: AUPRC vs Baseline score

Visual exploration

- Comparison of density distributions for different values of target
- Comparison of whiskers and box plot for different values of target

2 libraries for sentiment labelling:

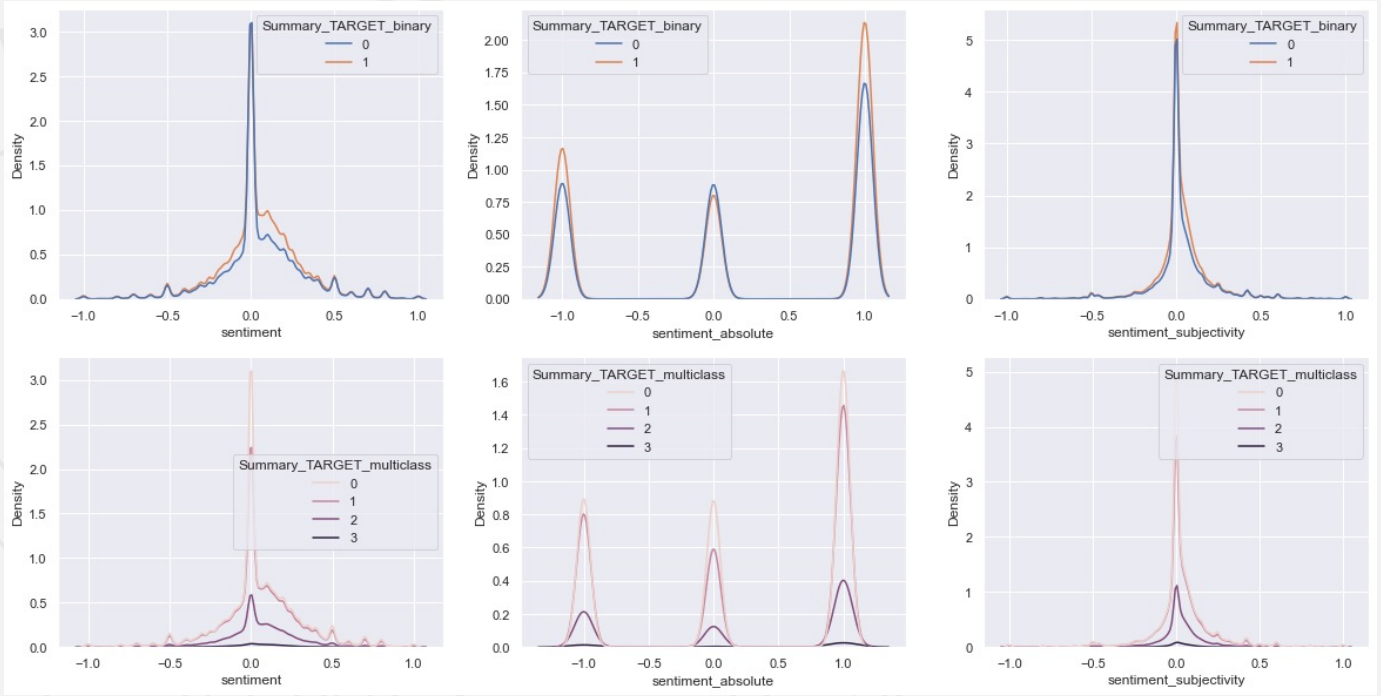
- TextBlob
- Vader

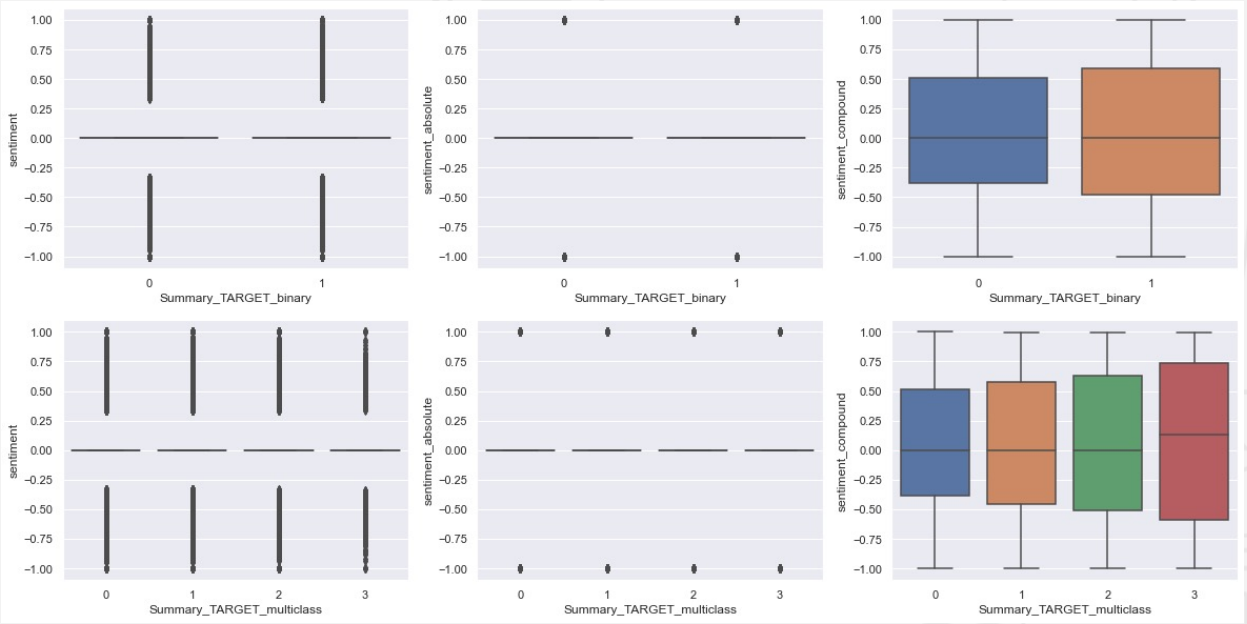
1. **Is it a good predictor for the controversy of comments?**

Statistical analysis → No correlation
Polarity not predictive of controversy

Visual exploration → No difference in distributions

Visual exploration





1. **Is it a good predictor for the controversy of comments?**

Not a good alternative labelling method for controversy of comments

3. LEXICON-BASED SENTIMENT POLARITY OF COMMENTS

2. **Analysis of the relationship between polarity and topics**

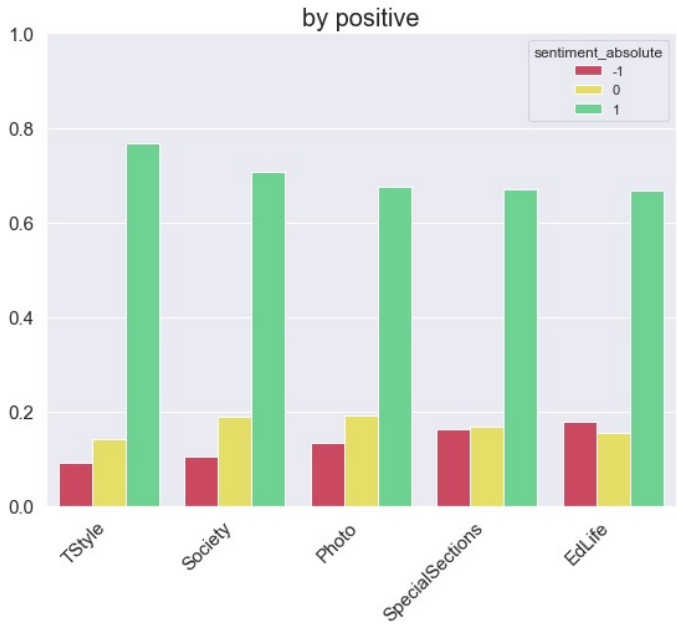
2 topics variables:

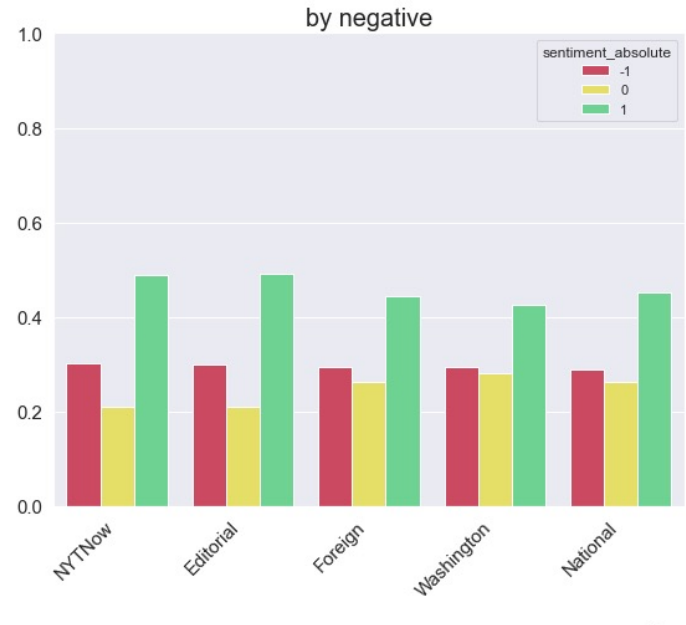
- `newDesk`
- `sectionName`

3 absolute sentiment polarities:

- Positive
- Neutral
- Negative

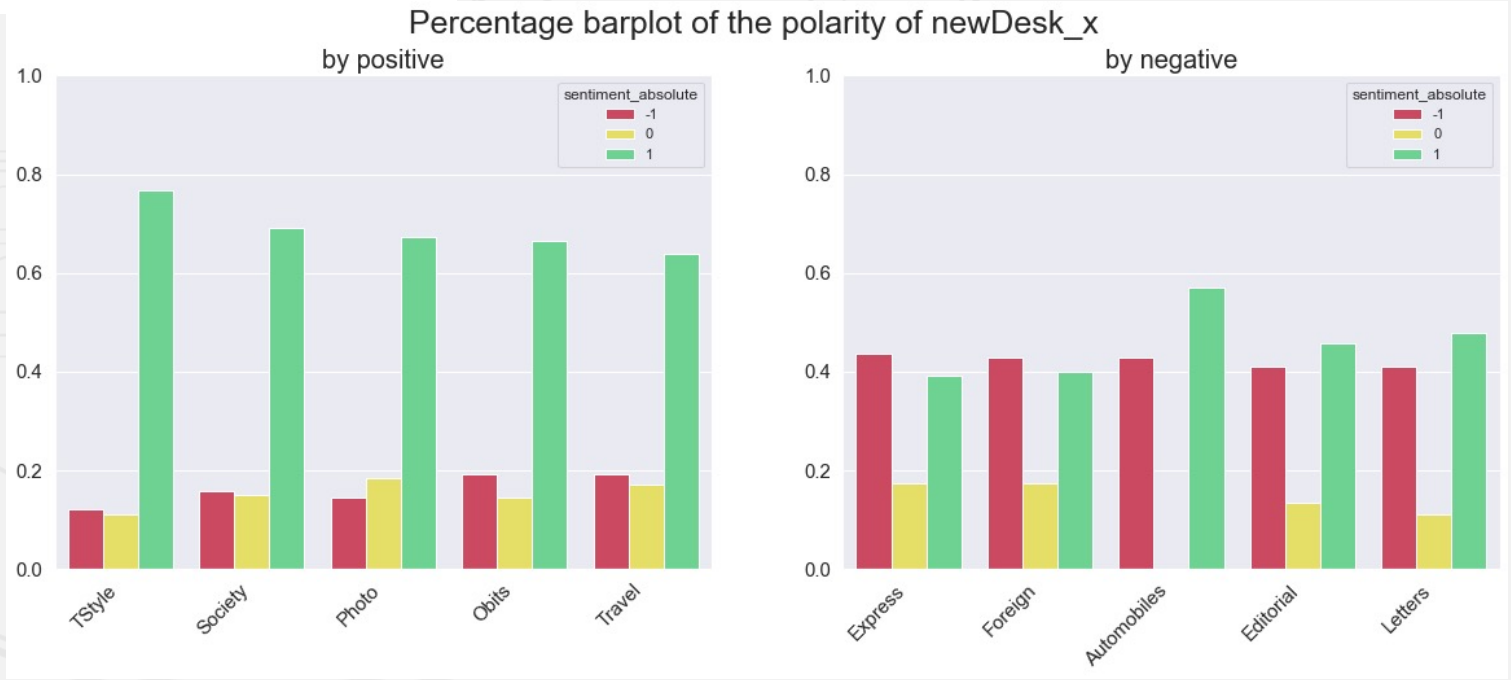
TextBlob





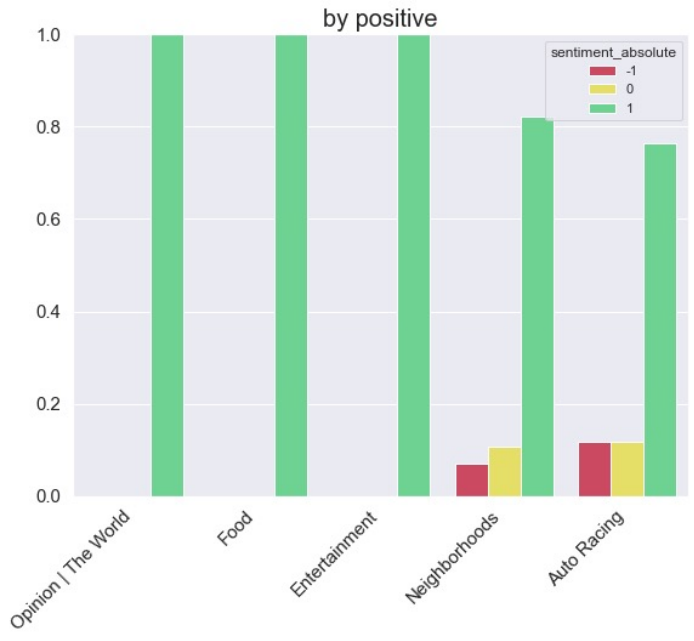
- Same result for top-3 positive topics
- Different result for top-3 negative topics
- Vader finds greater percentage of negative comments

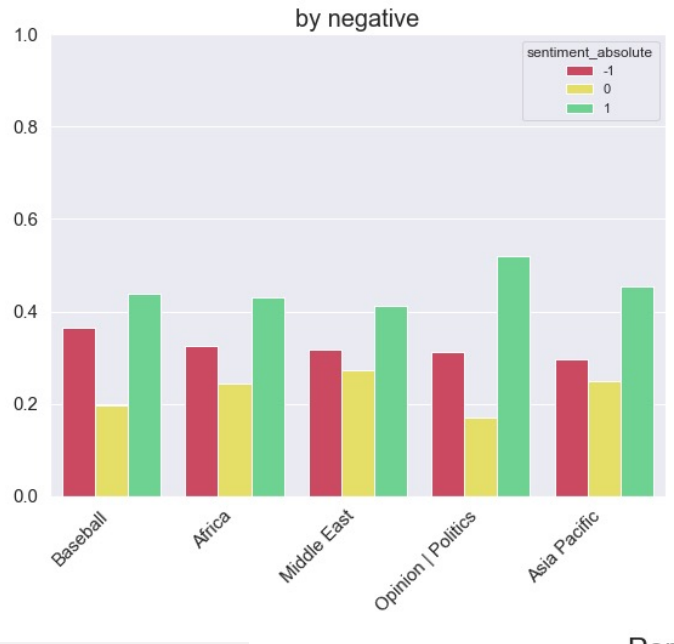
Vader



Variable: `newDesk`

TextBlob

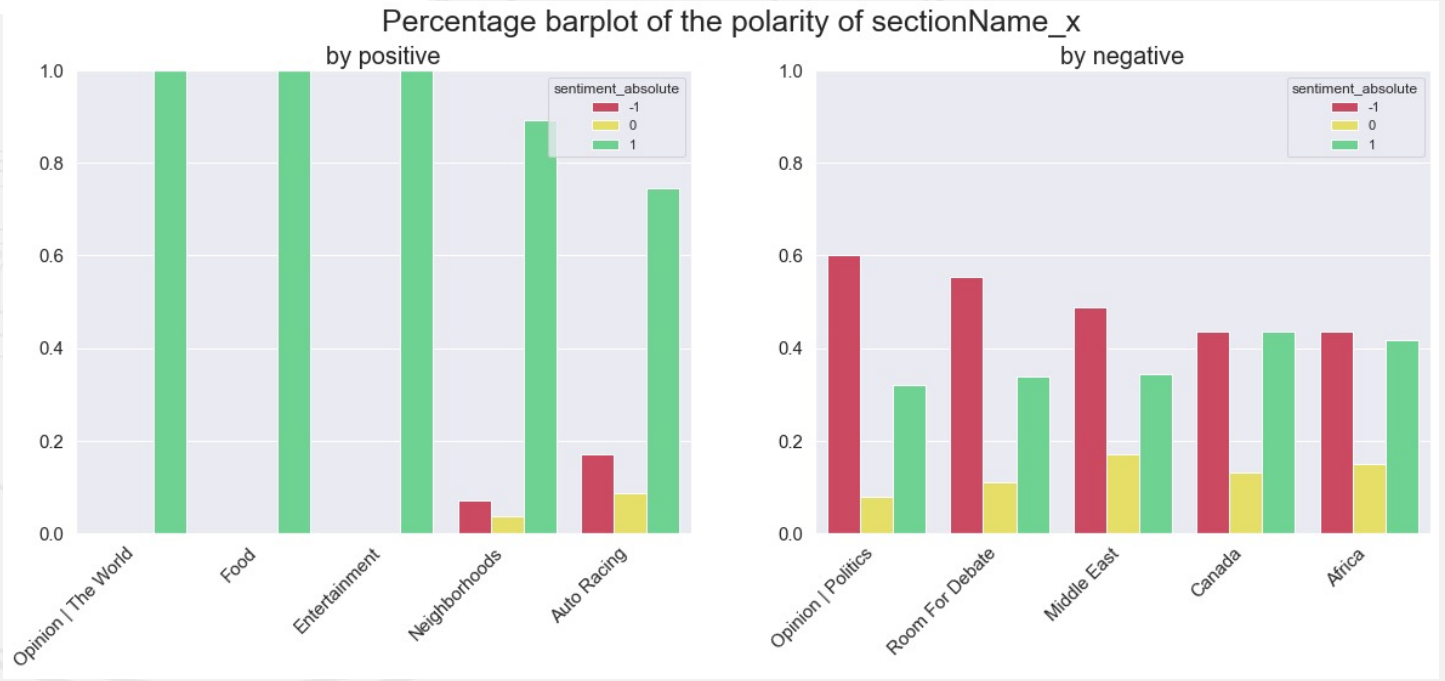




- Same result for top-3 positive topics
- Different result for top-3 negative topics
- Vader finds greater percentage of negative comments

Vader

Variable: `sectionName`



IMPLEMENTATION FOR RETRIEVING INFORMATION ABOUT COMMENTS

Retrieve useful information given the COMMENT ID:

- Text of comment
- Keywords of article
- Predicted labels by the multimodal neural network
- True labels
- Absolute polarity and score of the comment
- Barplots of polarity of topics
- Barplots of controversy of topics


Horst Stern: Reise
DIE
WOCHENZEITUNG FÜR POLITIK · WIRTSCHAFT · HANDEL UND KULT
Tea amo!
21. Juni 1996, 51. Jahrgang
Frankfurter Allgemeine
ZEITUNG FÜR DEUTSCHLAND
Kreml-Intrigen
Russie : un général arbitr
INTERNATIONAL

Q&A


UNIVERSITAS STUDIORUM MEDIOLANENSIS