WHAT MAKES NEWS SUCCESSFUL?

CHARACTERISING AND PREDICTING USER ENGAGEMENT OF NEWS POSTINGS

Engagement

DIGITAL NEWS ORGANISATIONS → ENGAGEMENT ← SOCIAL MEDIA PLATFORMS

Predicting Engagement

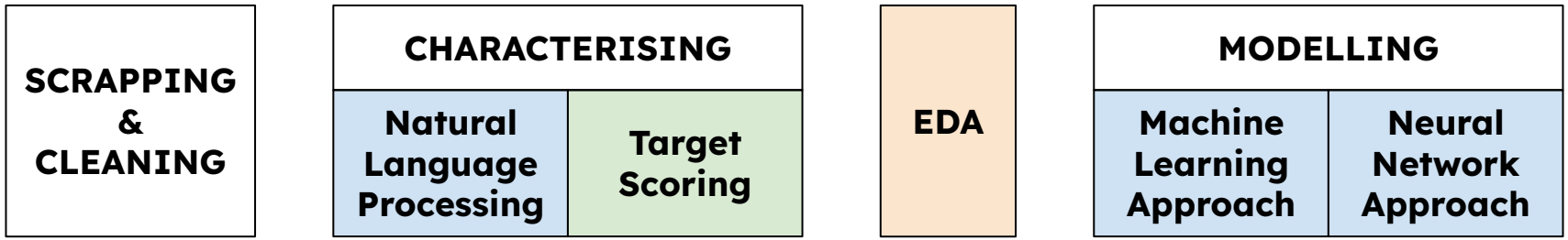
DIGITAL NEWS ORGANISATIONS

PREDICTING ENGAGEMENT

SOCIAL MEDIA PLATFORMS:

- find trends earlier.
- focus on what matters
- content tailoring
- news clustering...

Workflow


SCRAPPING & CLEANING
CHARACTERISING
Natural Language Processing
Target Scoring
EDA
MODELLING
Machine Learning Approach
Neural Network Approach

SCRAPPING & CLEANING

125k news postings

Data:

Title

Body

Views

Likes

Comments

CHARACTERISING

- Natural Language Processing
- Target Scoring

EDA

MODELLING

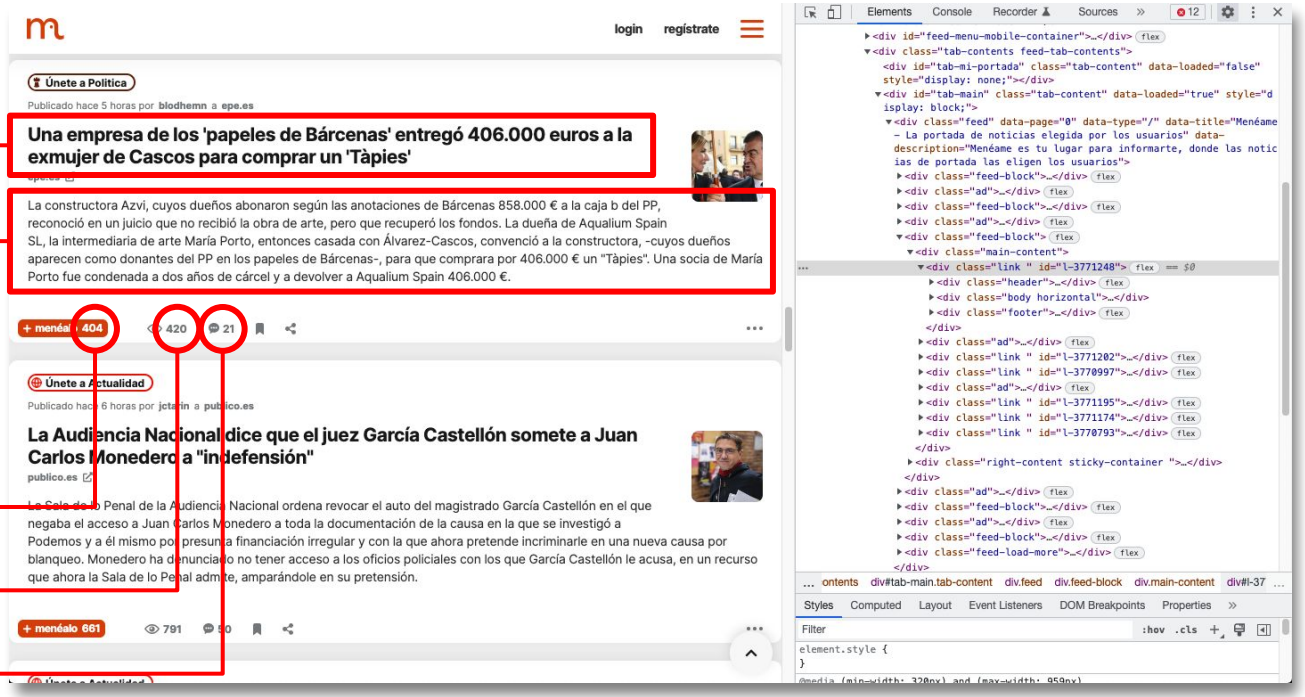
- Machine Learning Approach
- Neural Network Approach

Scrapping & Cleaning

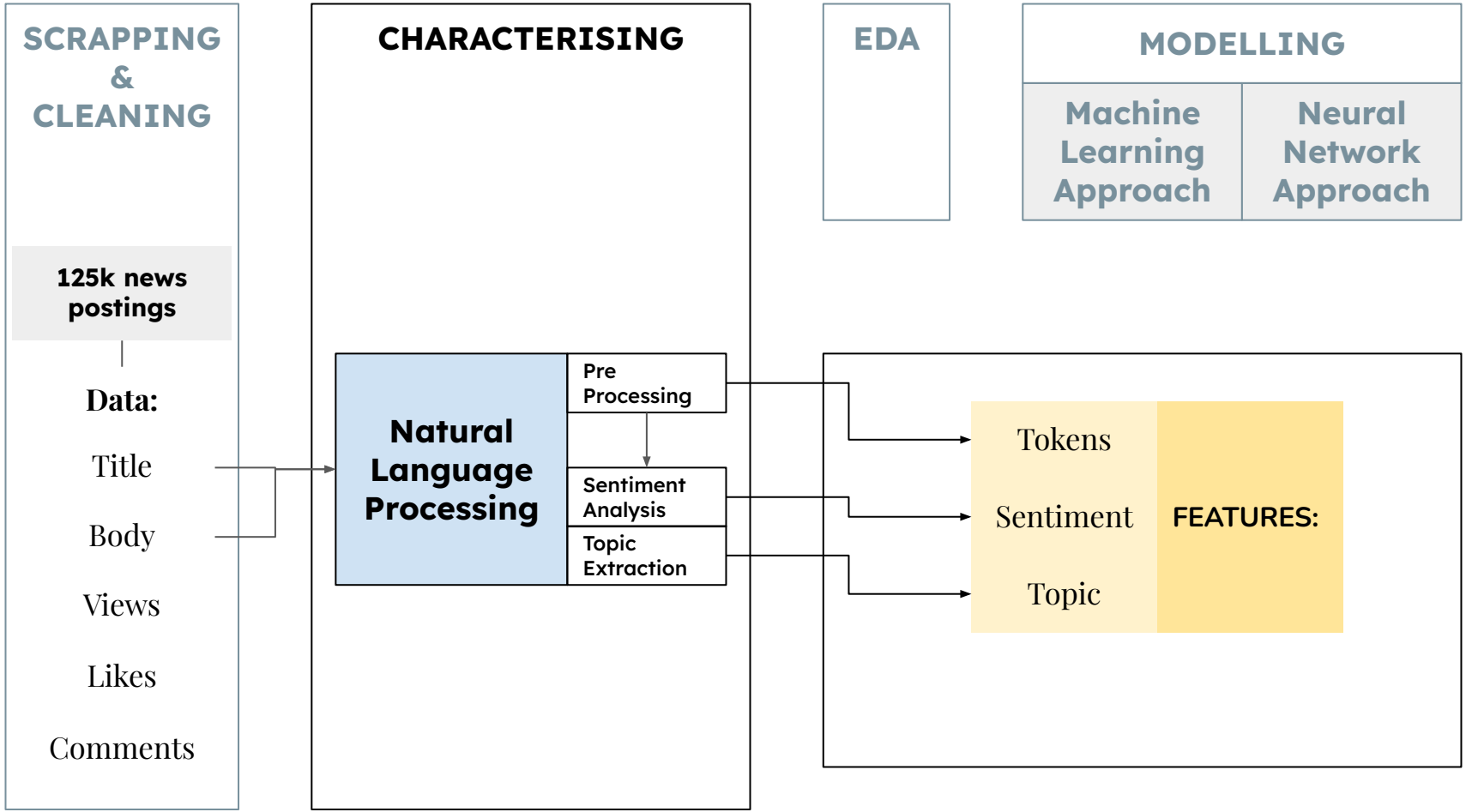
125k news postings

Data:

- Title
- Body
- Views
- Likes
- Comments



- BeautifulSoup


SCRAPPING & CLEANING
125k news postings
Data:
Title
Body
Views
Likes
Comments
CHARACTERISING
Natural Language Processing
Pre Processing
Sentiment Analysis
Topic Extraction
EDA
MODELLING
Machine Learning Approach
Neural Network Approach
Tokens
Sentiment
Topic
FEATURES:

Characterising - Topic + Sentiment

IN: RAW TEXT

PREPROCESSING

TOKEN / LEMMA

N-GRAMS

RELLEVANT

- sciPy
- NLTK

OUT: LIST

TOKENS

- transformers pipeline
- bert-base-multilingual-uncased-sentiment

PRE-TRAINED MODEL

SENTIMENT SCORE

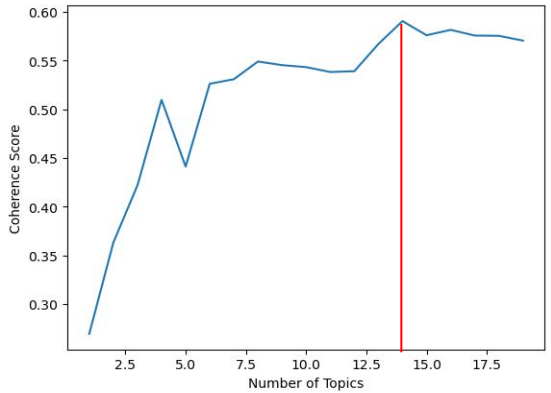
- doc2bow

CREATE DICTIONARY

CREATE CORPUS

LATENT DIRICHLET ALLOCATION

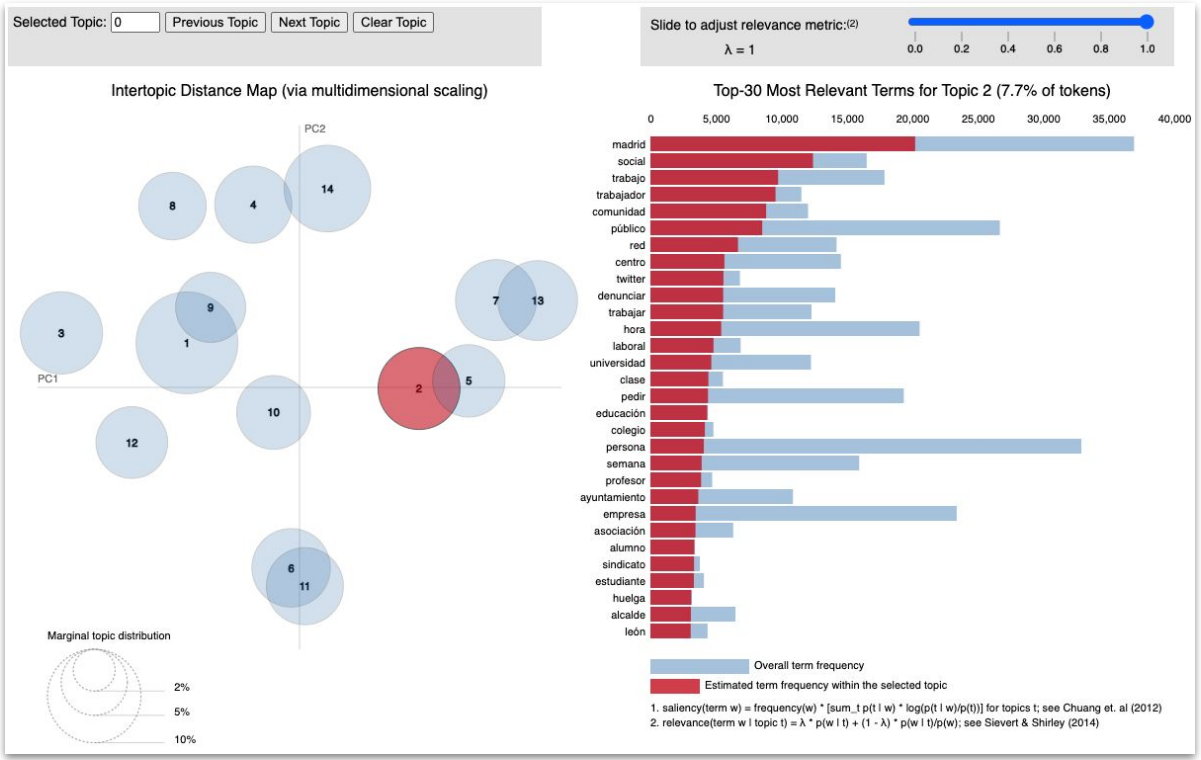
14 TOPICS



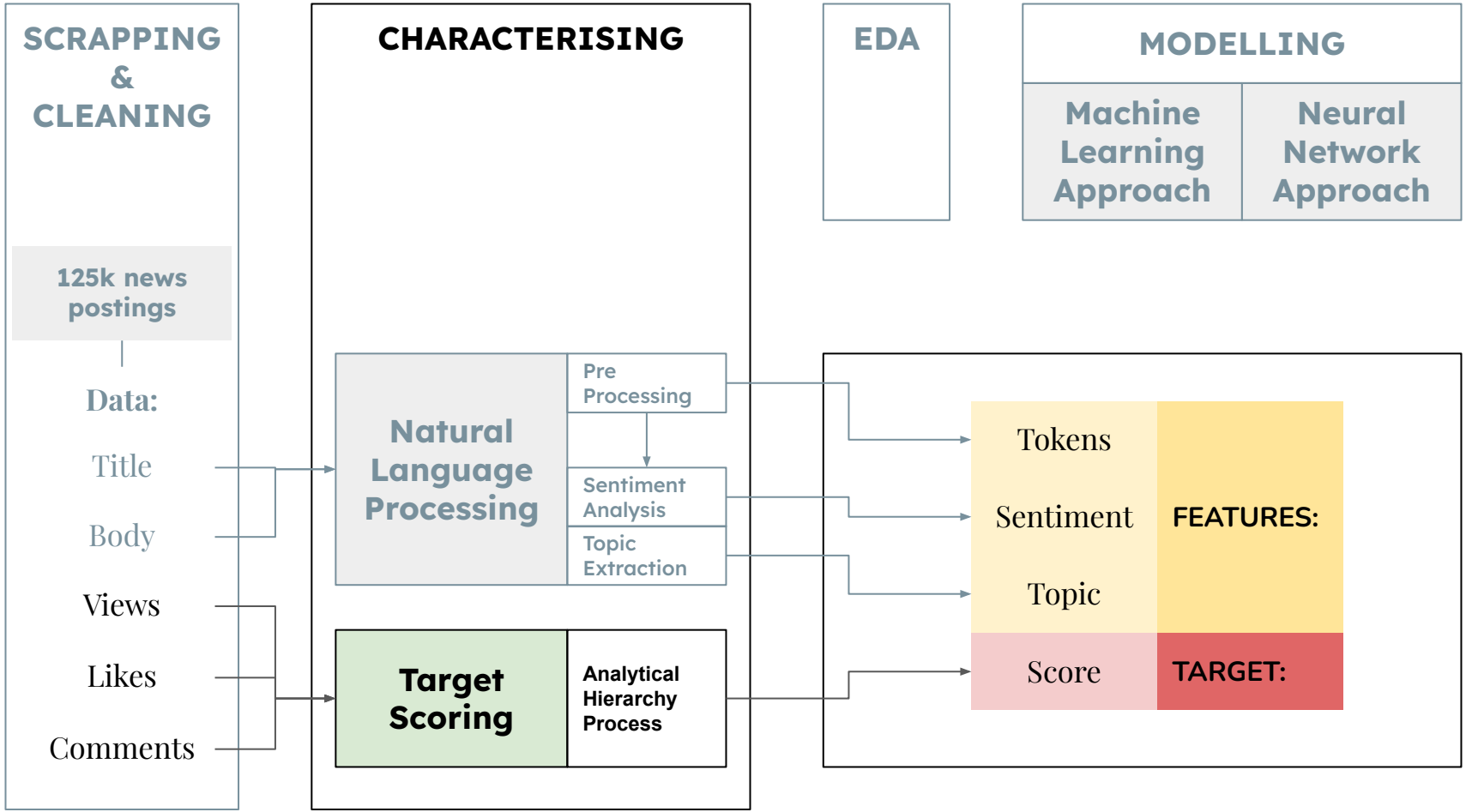
C_V COHERENCE TEST

- gensim

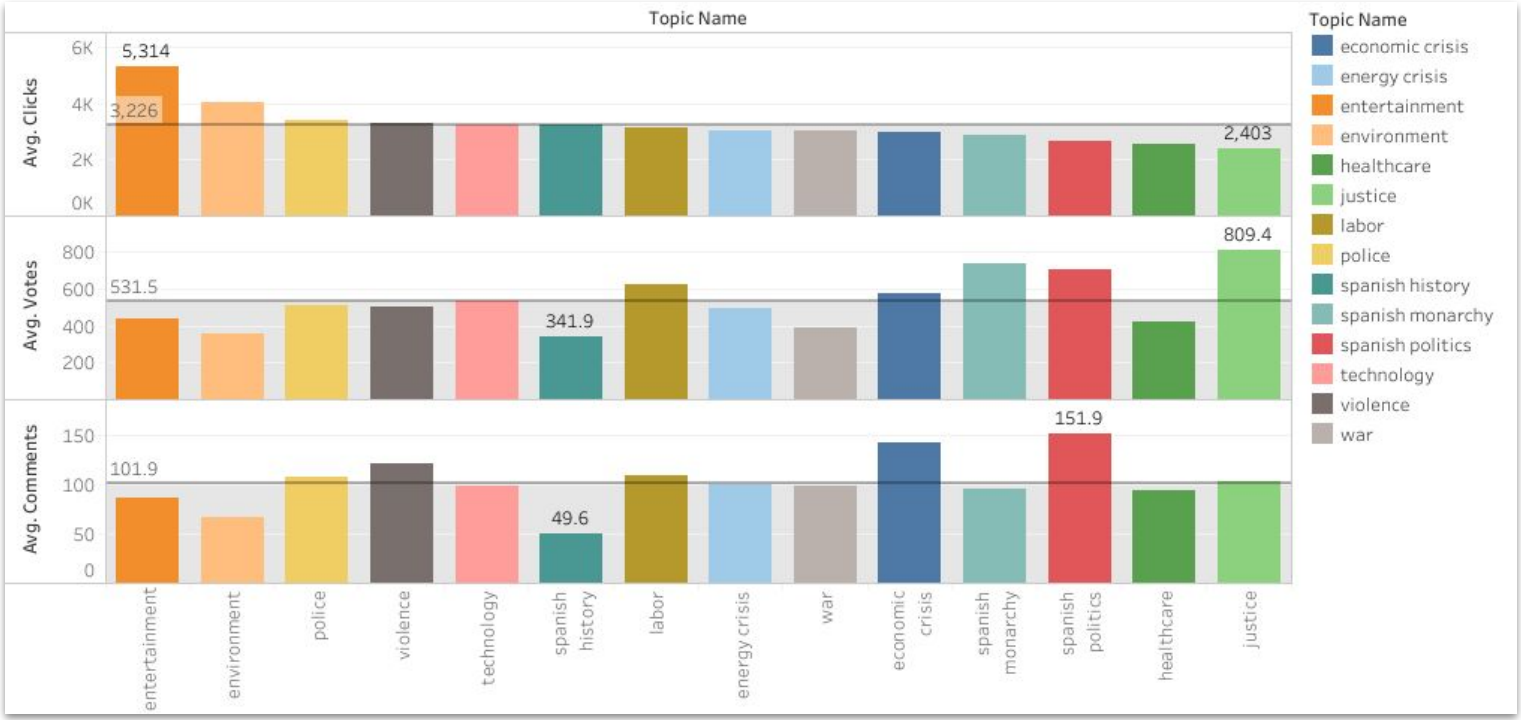
Characterising - Topic + Sentiment



- pyLDAvis


SCRAPPING & CLEANING
125k news postings
Data:
Title
Body
Views
Likes
Comments
CHARACTERISING
Natural Language Processing
Pre Processing
Sentiment Analysis
Topic Extraction
Target Scoring
Analytical Hierarchy Process
EDA
MODELLING
Machine Learning Approach
Neural Network Approach
Tokens
Sentiment
Topic
FEATURES:
Score
TARGET:

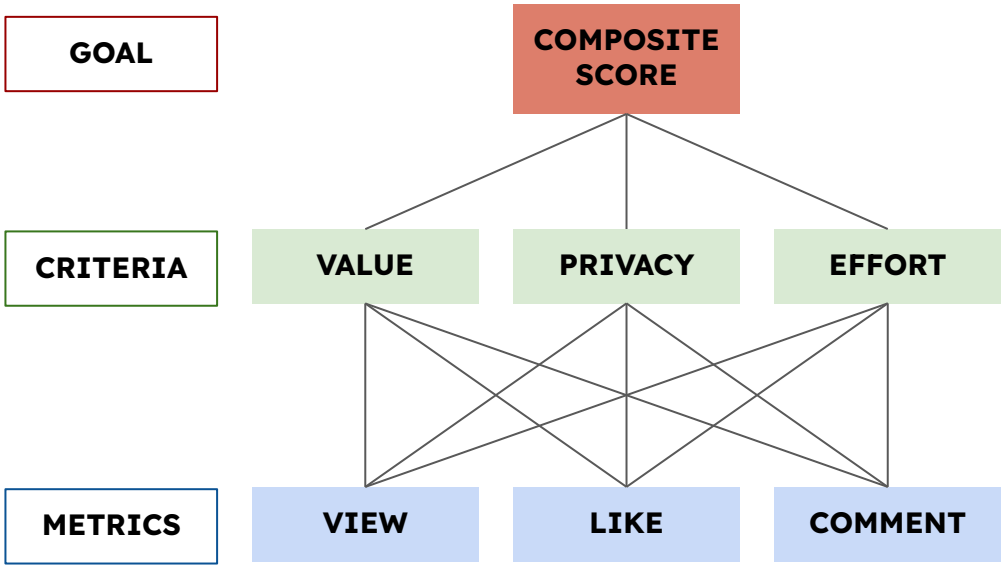
Score targeting: Evaluating scales



Score targeting: A. Hierarchy Process

METRICS:			CRITERIA:	WEIGHTS:
level 4	POSTING	Spreading content by sharing		
level 3	COMMENT	Expressing opinion and feelings by commenting	VALUE	W_1
level 2	LIKE	Exposing use preferences by liking	PRIVACY	W_2
level 1	VIEW	Private engagement by viewing	EFFORT	W_3

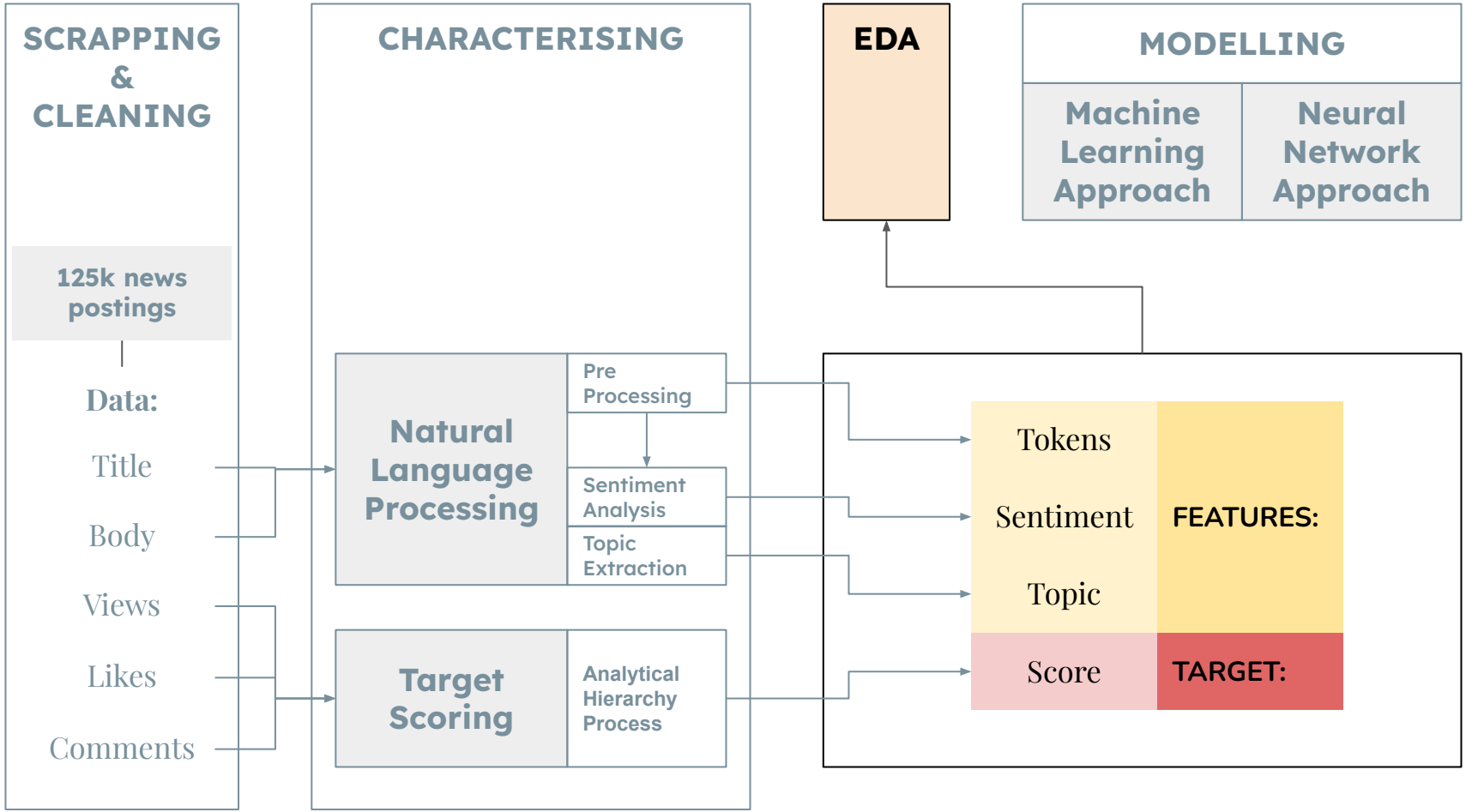
Score targeting: A. Hierarchy Process



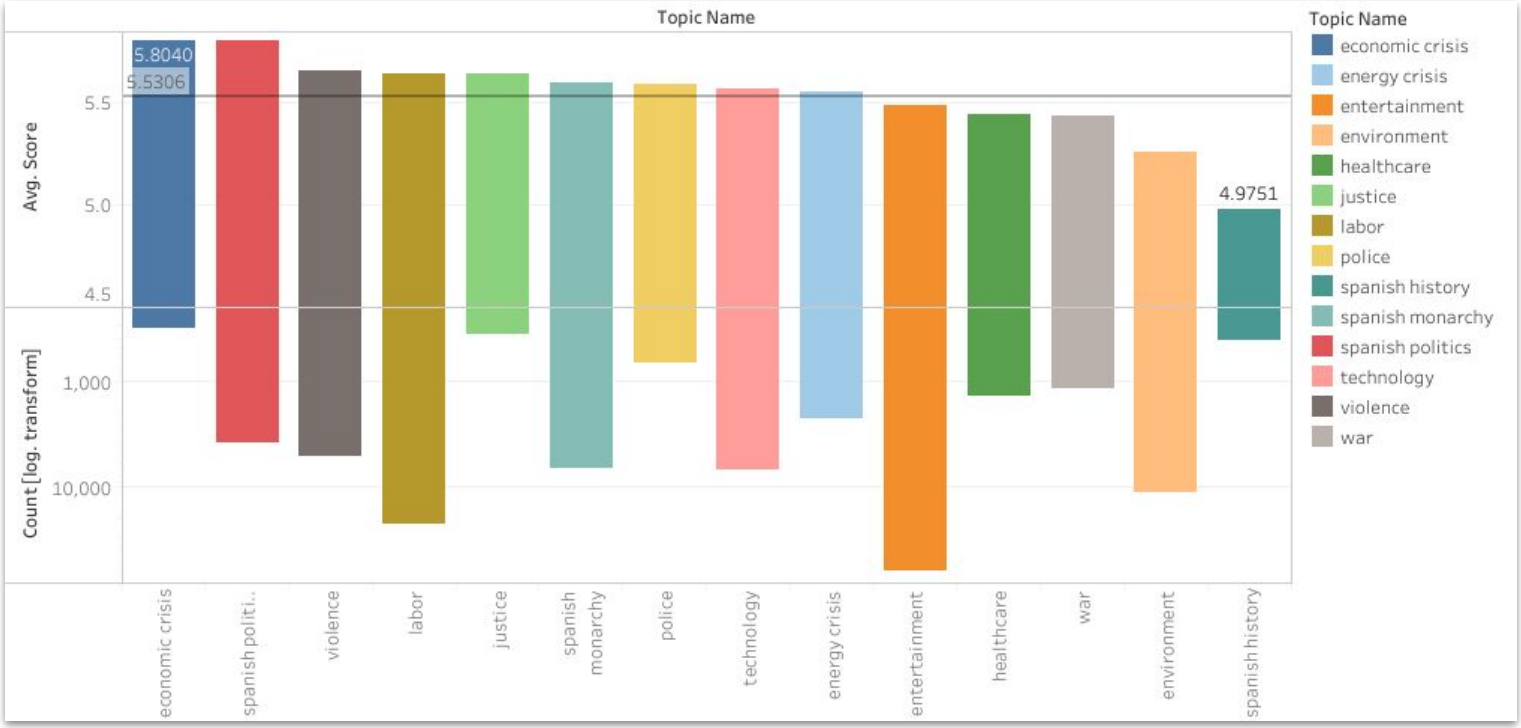
W1 = 0.2125 W2 = 0.2932 W3 = 0.4943

PAIR-WISE COMPARISON

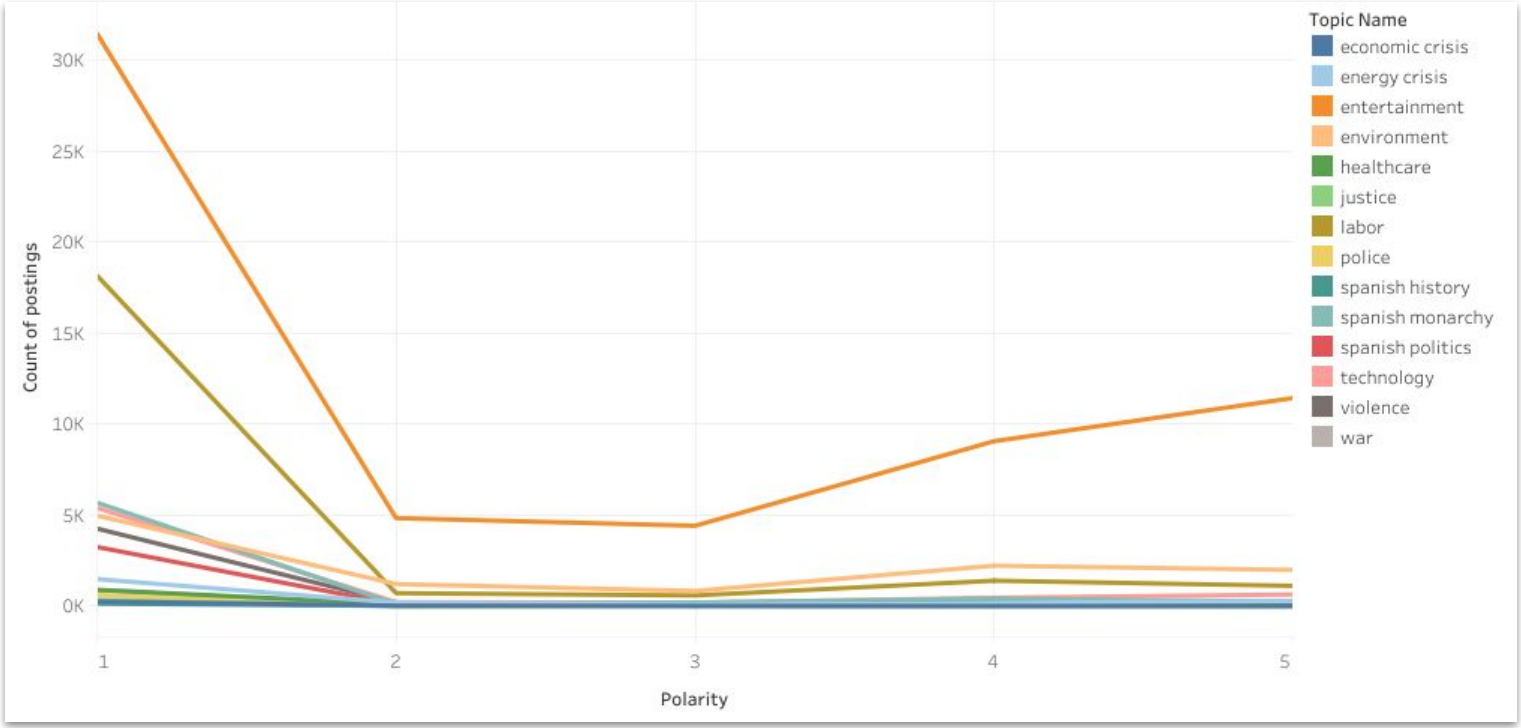
$$A = (w_i/w_j)_{n\times n} = \begin{pmatrix} w_1/w_1 & w_1/w_2 & \dots & w_1/w_n \\ w_2/w_1 & w_2/w_2 & \dots & w_2/w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_n/w_1 & w_n/w_2 & \dots & w_n/w_n \end{pmatrix}$$


SCRAPPING & CLEANING
125k news postings
Data:
Title
Body
Views
Likes
Comments
CHARACTERISING
Natural Language Processing
Pre Processing
Sentiment Analysis
Topic Extraction
Target Scoring
Analytical Hierarchy Process
EDA
MODELLING
Machine Learning Approach
Neural Network Approach
Tokens
Sentiment
Topic
FEATURES:
Score
TARGET:

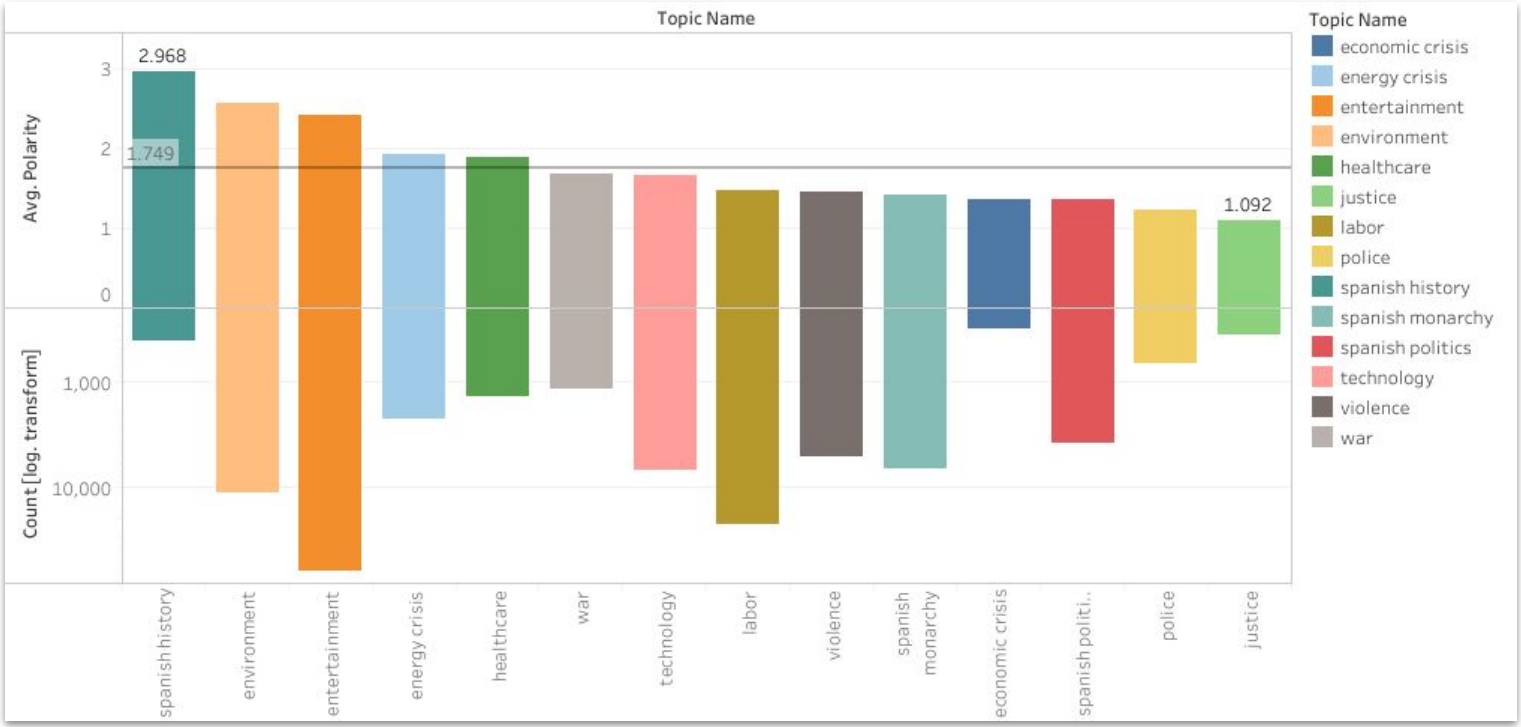
EDA: Score by Topic



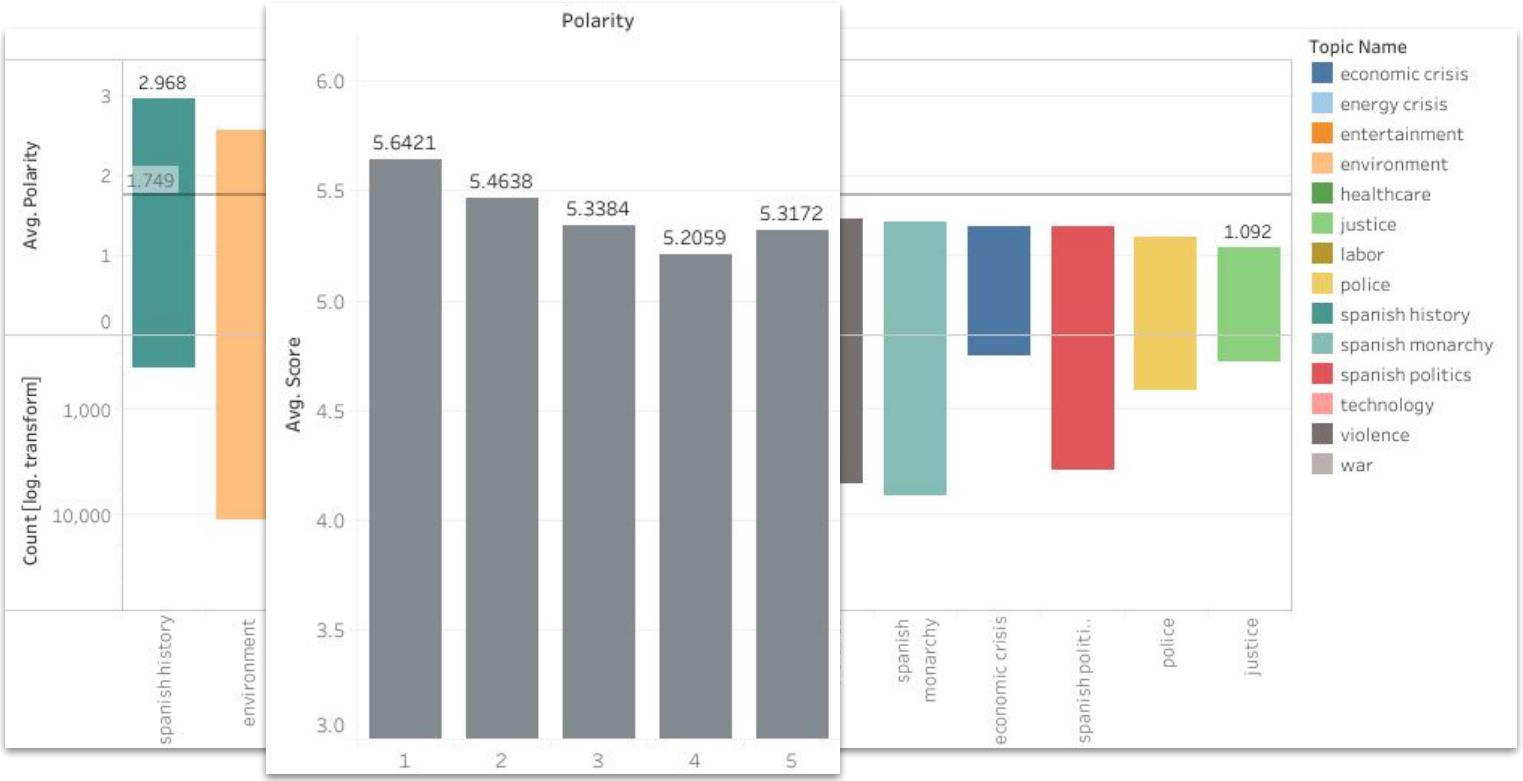
EDA: Score by Sentiment

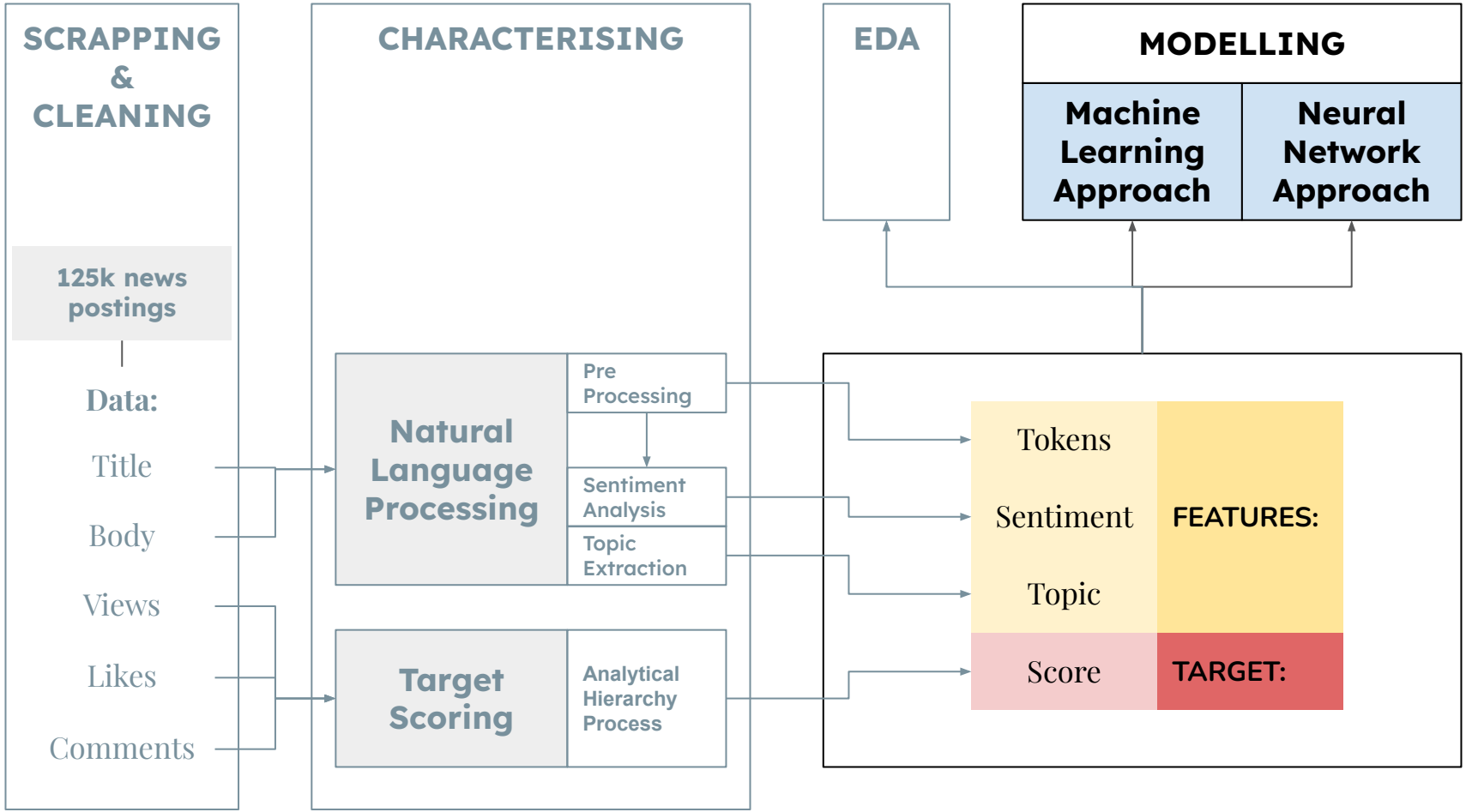


EDA: Topic by Sentiment

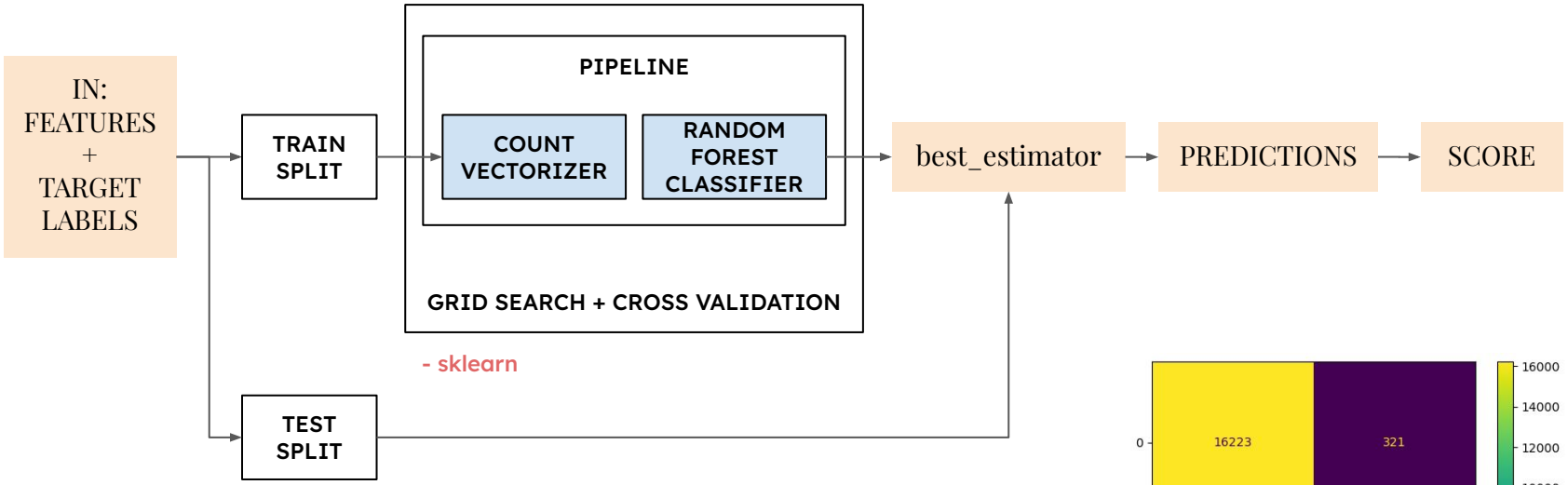


EDA: Score by Sentiment

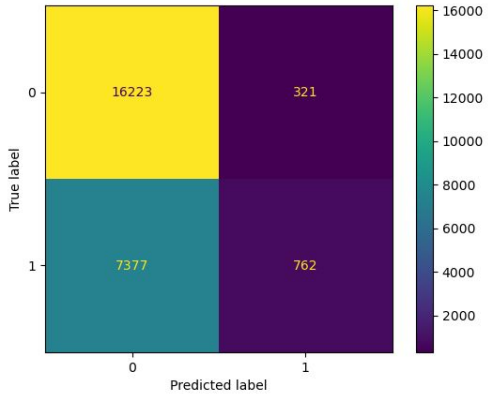



SCRAPPING & CLEANING
125k news postings
Data:
Title
Body
Views
Likes
Comments
CHARACTERISING
Natural Language Processing
Pre Processing
Sentiment Analysis
Topic Extraction
Target Scoring
Analytical Hierarchy Process
EDA
MODELLING
Machine Learning Approach
Neural Network Approach
Tokens
Sentiment
Topic
FEATURES:
Score
TARGET:

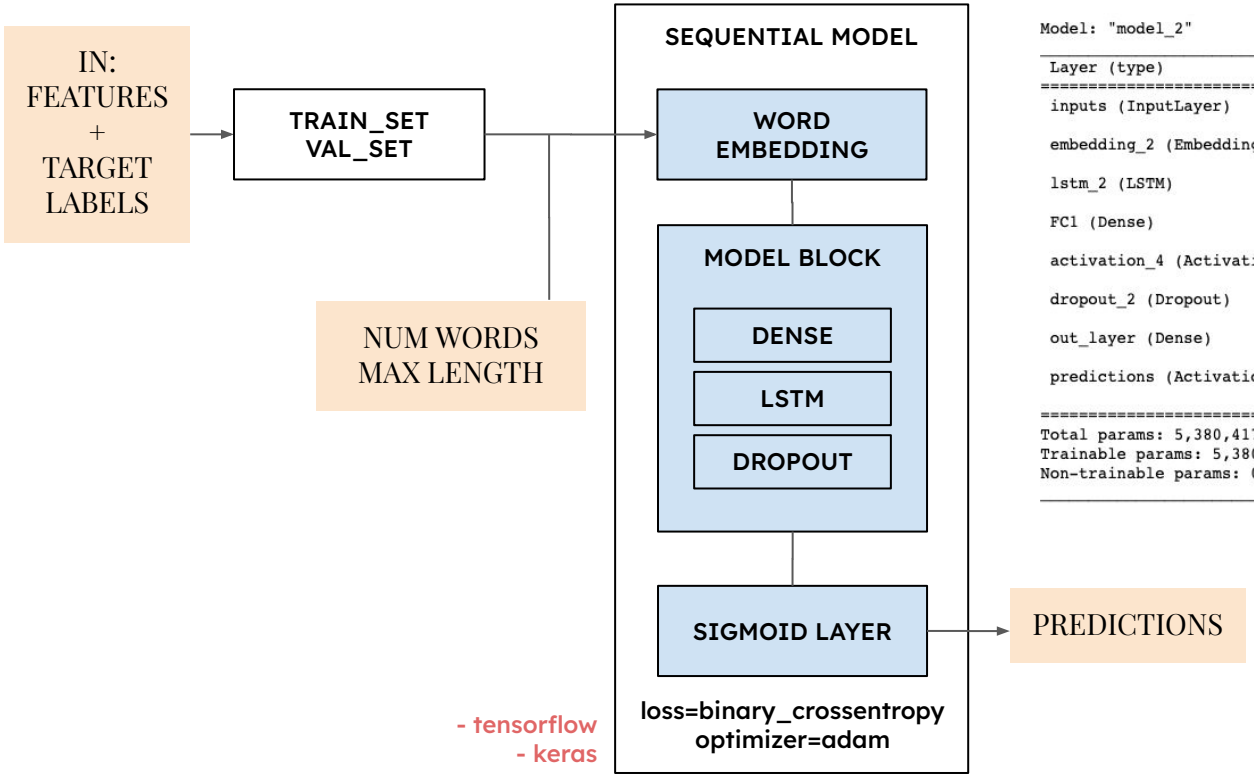
Model: ML Approach



RESULTS:
ACCURACY = 0.69
F1_LABEL(0) = 0.81
F1_LABEL(1) = 0.17



Model: NN Approach

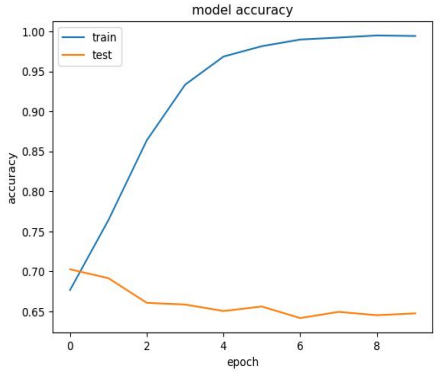


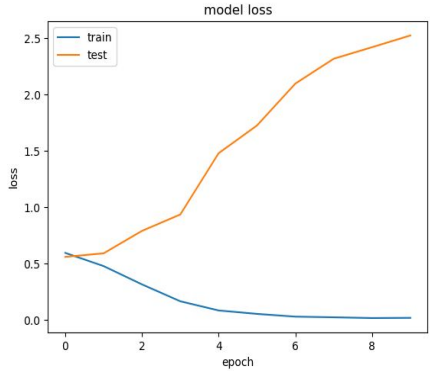
Model: "model_2"

Layer (type)	Output Shape	Param #
inputs (InputLayer)	[(None, 200)]	0
embedding_2 (Embedding)	(None, 200, 50)	5000000
lstm_2 (LSTM)	(None, 256)	314368
FC1 (Dense)	(None, 256)	65792
activation_4 (Activation)	(None, 256)	0
dropout_2 (Dropout)	(None, 256)	0
out_layer (Dense)	(None, 1)	257
predictions (Activation)	(None, 1)	0

Total params: 5,380,417
Trainable params: 5,380,417
Non-trainable params: 0

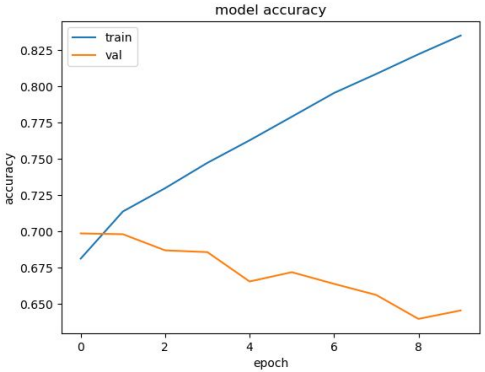
Results

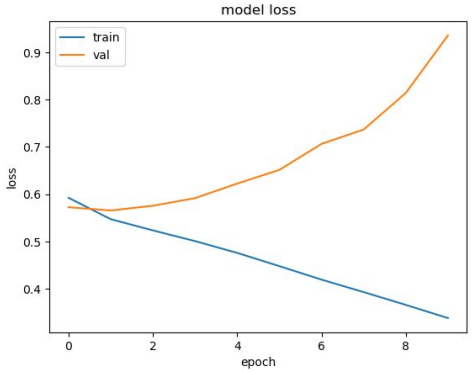




-NUM WORDS = 100000
-MAX LENGTH = 200

Test set
Loss: 0.693
Accuracy: 0.547





-NUM WORDS = 5000
-MAX LENGTH = 100

Test set
Loss: 0.911
Accuracy: 0.651

Further development...

- REVIEW THE ENGAGEMENT SCORE
- RUN REGRESSIONS INSTEAD OF CLASSIFICATIONS
- INCLUDE CREATED FEATURES
- GET BETTER SCORES...