# Introduction:

Content analysis has always been one of the key methods in communication research. However, with the advent of web 3.0, the data sources for communication and political content analysis has exponentially grew in volume and complexity. The advances in computational methods, such as dictionary analysis or supervised machine learning, allow researcher to process vast quantities of text. Such methods have become more convinient to use with the increased availability and ease-of-use of softwares such as Quanteda (Benoit et al., 2018), NLTK (Loper & Bird, 2002), caRet(Kuhn, 2008) and SciKitLearn(Pedregosa et al., 2011).

Yet, communication rarely happens via single modality in legacy and contemporary communication channels. For example, one of the key political actors in Europe, the European Union has increased its Twitter communication almost 4 folds in the last decade where more than 40% of the messages contain at lease one embedded image (Özdemir & Rauh, 2022). Dictionary based and SL methods, unfortunately, have harder time incorporating multi-modality into the analysis.

Deep learning (DL) carries possiblity to extend content analysis inquiries to multi-modal materials. Previous studies have demonstrated the flexibility of embeddings to analyze multimodel data(Li et al., 2022; Niu et al., 2019; Tseng et al., 2021). With the advent of DL, it is now possible to include more than one modality into the analysiss.

In this paper, we evaluate the feasibility of using multi-modal DL embeddings to classify political messages where the message is delivered with a combination of visual and textual modalities in a computational experiment. We utilize series of SL models and multi-modal DL embedding to classify manually annotated tweets from EU executives. We then compare the classification performance of these models. Our results indicate [...]. Finally, we conclude with some recommendations for researchers who would like to use multi-modal data in automated content analysis.

# Research design and data:

We use 900 tweets from the EU executive accounts sent out between December 1st 2019 and July 31st 2020. We focus on these tweets in our experiment specifically because they deliver their message by combining text and images or supplemented the textual message with imagery. Examples of them is presented in Figure 1. The tweets are manually annotated as a part of the project Trondheim Analytica (Özdemir, Graneng, de Wilde, forthcoming). The orginal dataset is composed of all tweets from the 117 verified accounts of EU executives which includes commissioners, director generals, institutions and agencies responsible for policy making and implementation at the EU level. Tweets were manually coded as a whole based on whether the content of the tweet provide an update and information on political operations, policies, programs, and reports published by the EU, its institutions, or its bureaucrats in binary format. Overall there are 447 tweets that contain such message (code:1) and 379 tweets that do not (code:0). To ensure the data quality, three rounds of intercoder reliability tests were conducted between coders before coding the full sample. The first two rounds showed insufficient reliability scores, but after intensive training and discussion among coders, we reached sufficient reliability scores across all coding categories (Krippendorf α >.8).

Graphical user interface, text

Description automatically generated

Figure 1: Example tweets

In our experiment, we take each the binary indicator as the outcome and learn individual SL and DL predictive models for them. From SL classification family we use Naïve Bayes (NB), Support Vector Machine (SVM), Random Forest (RF), logistic regression and XGBoost (XGB) while we use XXXX from DL classification family. We extensively use Quanteda (Benoit et al., 2018) and caRet (Kuhn, 2008) packages in R for text preprocessing and SL models. For DL models, we utilize XXXX in python.

In our experimental setting, we test the predictive capacity of these models on three different datasets. First dataset is document feature matrix of the tweet texts with no weighting. The second dataset is term-frequency inverse-document-frequency representation of tweet texts. The third dataset combines textual and visual elements in [] embedding format as input data. We train predictive using these tree datasets and compare the predictive capacity of all models across three different datasets using F1 scores.

# Results:

Table 2: F1 scores of predictive models

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | DFM | | | TFIDF | | | Multi-modal embeddings | | |
|  | F1 | Precision | Recall | F1 | Precision | Recall | F1 | Precision | Recall |
| Naïve Bayes | 0.57 | 0.48 | 0.70 | 0.37 | 0.51 | 0.29 | - | - | - |
| **Logistic regression** | 0.66 | 0.68 | 0.64 | 0.70 | 0.70 | 0.71 | - | - | - |
| Support Vector machine | 0.70 | 0.71 | 0.69 | 0.77 | 0.73 | 0.82 | - | - | - |
| **Random forest** | 0.72 | 0.75 | 0.69 | 0.78 | 0.78 | 0.77 | - | - | - |
| XGBoost | 0.67 | 0.64 | 0.70 | 0.71 | 0.72 | 0.70 | - | - | - |
| DeepL | - | - | - | - | - | - | - | - | - |