Fake News Detection

M. Affi, M. Wójcik, A. Zalas

Faculty of Mathematics and Information Science

Warsaw, 9.11.2022

Agenda

1. Introduction
2. Background
3. Literature Review
4. Datasets
5. Proposed Methodology
6. Risks and Future Research
7. Questions & Answers Session

Introduction

The project will aim to detect news that contain misinformation, particularly fake news, based on online article texts.

We will test and compare different methods for misinformation detection (MID) using multiple datasets. The research will be focused on testing different approaches for data preprocessing and performing **binary classification** task.

Background

Definition (Misinformation)

Misinformation is a false statement or a set of statements that mislead other people by hiding or "twisting" the facts.

Mathematically, misinformation may be defined as a binary function which assigns *true* or *false* to an arbitrary sentence. Nonetheless, in general, the truth may be somewhere in between.

$$M(a) = \begin{cases} 1, & \text{if } a \text{ is true} \\ 0, & \text{if } a \text{ is false} \end{cases}$$

Types of Misinformation

We distinguish five different types of misinformation

1. false information – the broad concept of misinformation. It is intentionally used to be defined as correct information interchangeably
2. rumour – a story of doubtful truth, which is usually spread widely
3. **fake news** – a modified version of original news or piece of information that is spread intentionally and usually very difficult to identify
4. spam – an unwanted message with irrelevant, inappropriate, or even harmful information used to mislead users
5. disinformation – false facts that are conceived to deceive a user

Types of Misinformation

Type	Characteristics	Objectiveness	Severity	Integrity
Rumors	Ambiguous	Not sure	Low	Not sure
False information	Deception	Yes	High	False
Fake news	Misguided	Yes	Medium	False
Spam	Confused	Yes	Low	Not sure
Disinformation	Mislead/deceive	Yes	Medium	False

Figure: Different misinformation types considered in (Islam et al., 2020)

Impact and Our Goals

- Misinformation may influence society, economy, politics, and emergency response during natural disasters, epidemics, crises...
- Nowadays, due to the popularity of social network platforms like Facebook, Twitter, or Reddit, it has become easier to spread misinformation in seconds.

Our goal is to find potential patterns in fake news and analyse their structure so they can be detected more accurately.

Literature Review

In (Islam et al., 2020) three different categories of models were described for fake news detection task:

1. discriminative
2. generative
3. hybrid models

Discriminative Models

- **Convolutional Neural Networks** (CNN)
  Some researchers extended the usage of CNN to include not only text but also visual information such as images
- **Recurrent Neural Networks** (RNN)
  Used due to the sequential characteristics of text data.
- **Recursive Neural Networks** (RvNN)
  Used to capture the hierarchical structure of the input.

Generative Models

The nature of generative models is well suited for a topic such as misinformation, as authors tend to intentionally sound **as factual as possible to deceive the reader**, the former playing the role of a **generator** and the latter indirectly acting as a **discriminator**.

Example types of generative models that are applicable for this type of problem are:

- Restricted Boltzmann Machine,
- Deep Belief Network, Deep Boltzmann Machine,
- Generative Adversarial Network
- Variational Autoencoder

Hybrid Models

Hybrid models were applied to increase the performance of individual models of one type.

- **Convolutional Recurrent Neural Network** (CRNN)
- **Convolutional Restricted Boltzman Machine** (CRBM)
- **Ensemble-Based Fusion** (EBF)
- **LSTM Density Mixture Model**

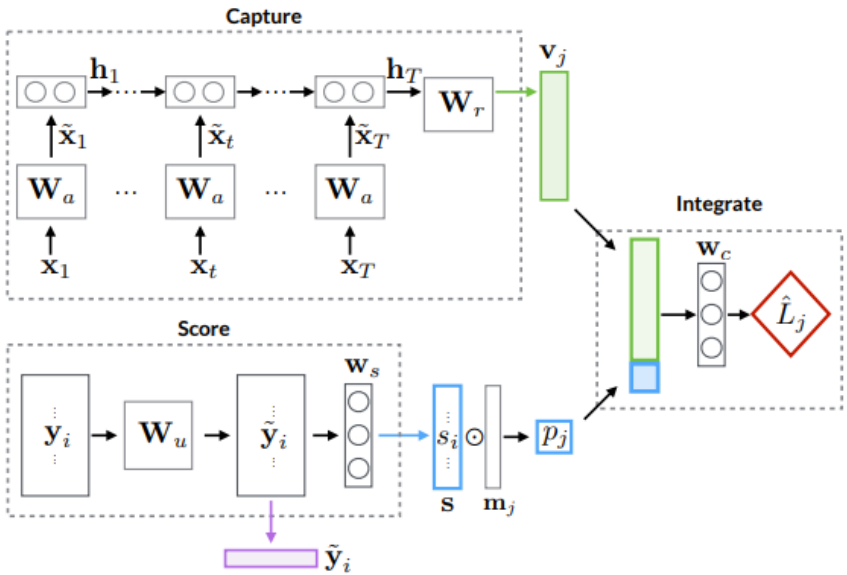

Figure: CSI model specification

Datasets

- **ISOT Fake News Dataset**
  Data collected from Reuters.com source in case of real news, and different unknown sources for the fake news set in 2016 and 2017. There are 21417 articles labeled as real and 23481 labeled as fake.
- **Fake News Kaggle Dataset**
  The dataset consists of news articles data. It was published on Kaggle in the competition Fake News. There are 18285 observations split into training and test set with a total of 43% of fake news.

ISOT Fake News Dataset

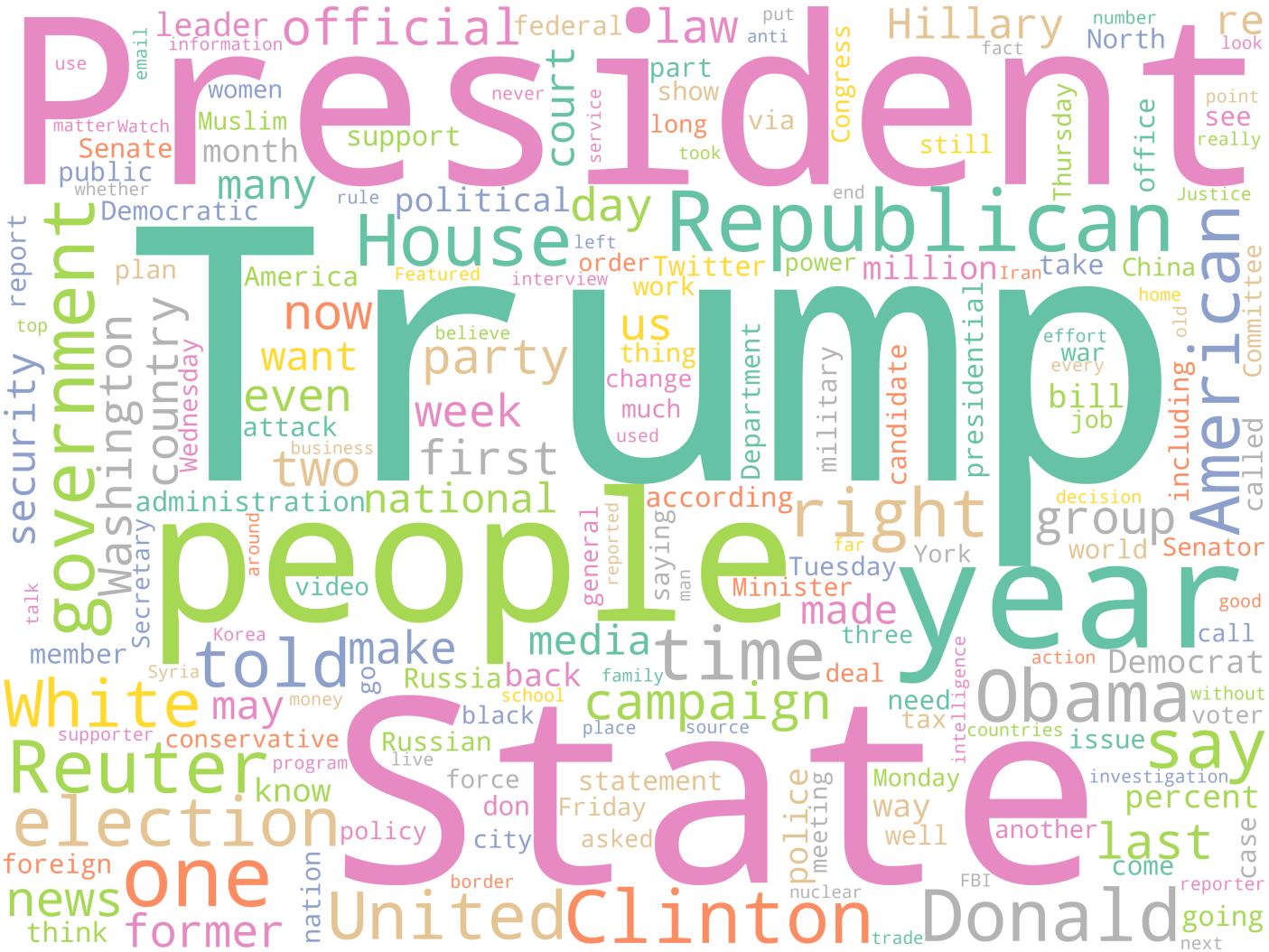

Figure: Word cloud for ISOT Fake News Dataset

ISOT Fake News Dataset

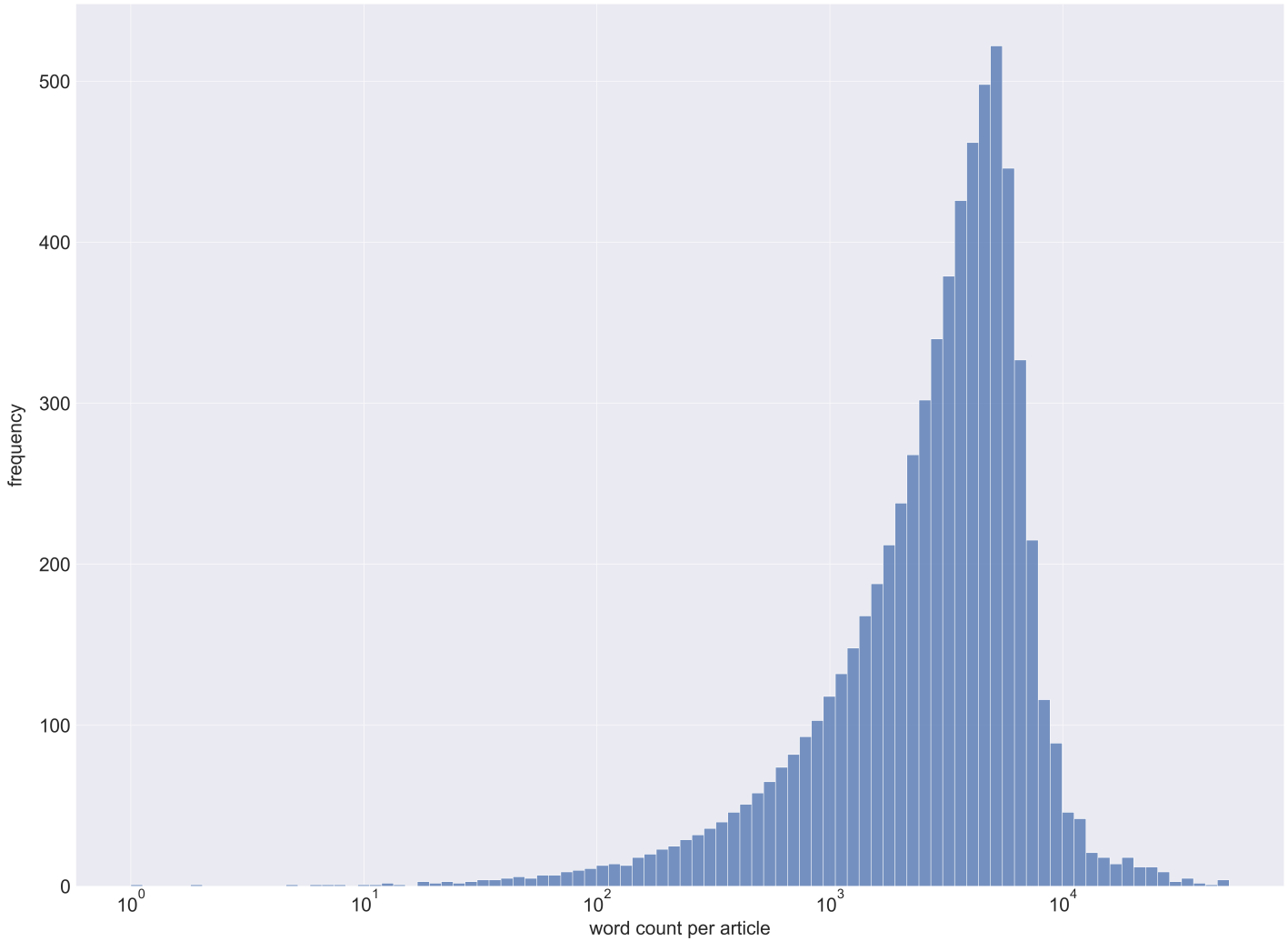

Figure: Word count histogram for ISOT Fake News Dataset

Fake News Kaggle Dataset

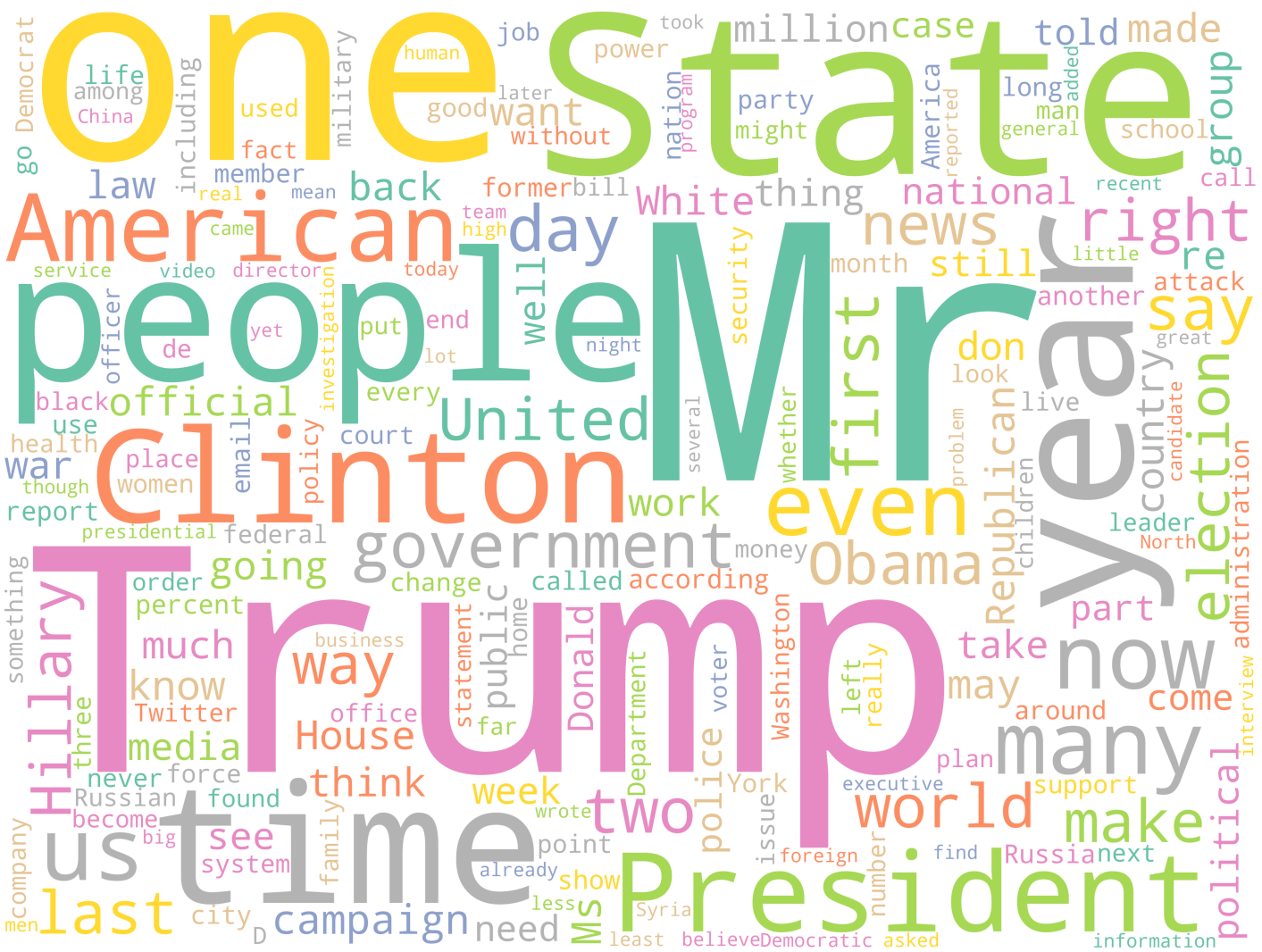

Figure: Word cloud for Fake News Kaggle Dataset

Fake News Kaggle Dataset

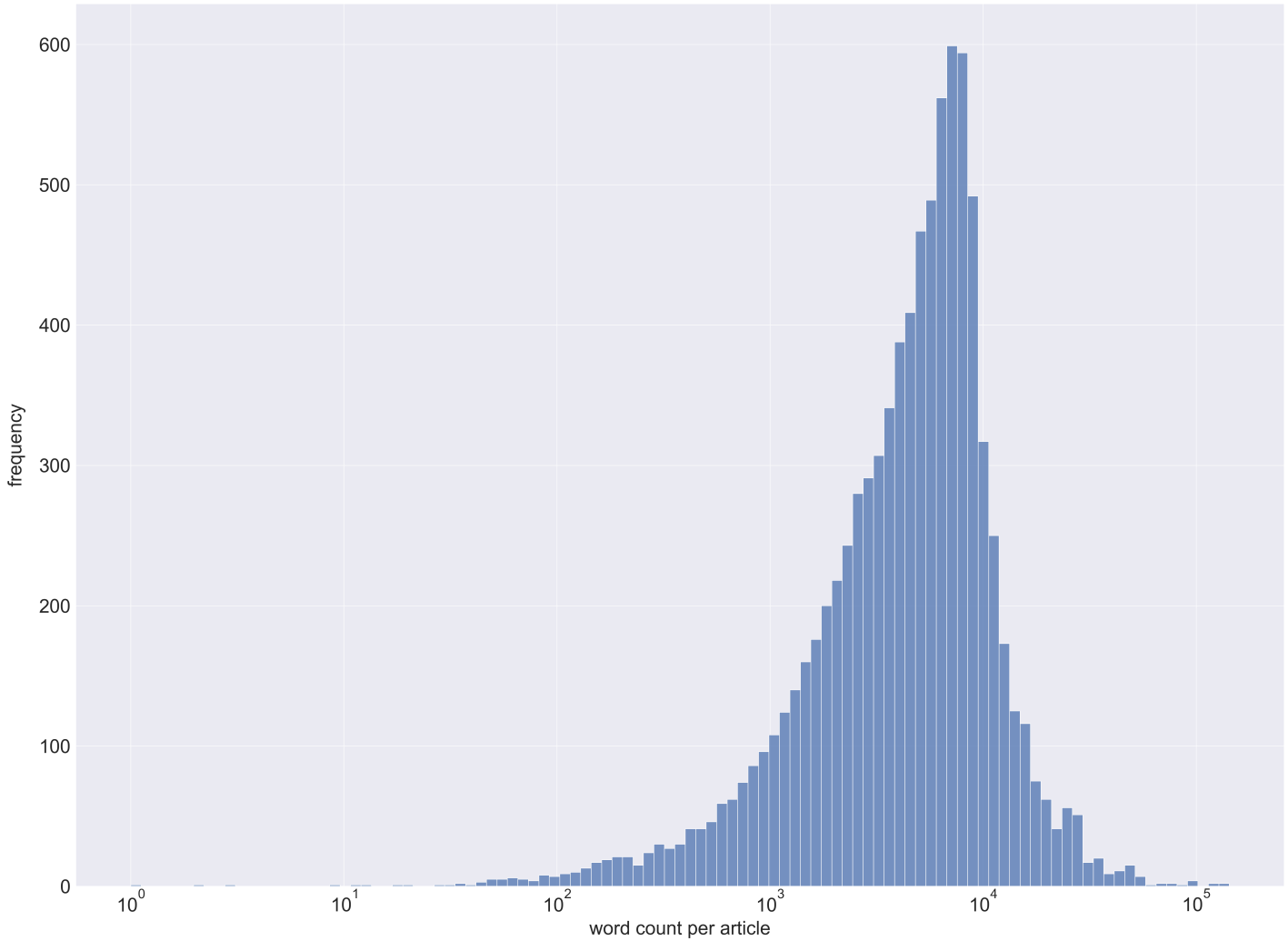

Figure: Word count histogram for Fake News Kaggle Dataset

Methodology

Preprocessing techniques to consider:

- removing punctuations
- removing stop words
- lower casing
- tokenization
- stemming
- lemmatization

We will also test additional preprocessing steps depending on model type and architecture.

Methodology

We are planning to implement and test at least three among the following approaches:

- `word2vec` word embeddings with LSTM
- hybrid model, for example Convolutional Recurrent Neural Network
- ensemble model based on "lighter" architectures and models
- attention-based LSTM
- pre-trained text representation model BERT to extract features with a chosen classifier

Further customizations include different loss functions like cross-entropy loss or log-likelihood and hyperparameters tuning.

Risks and Further Research

Potential risks that we may encounter during research:

- lack of variety in the datasets, which may result in overfitting
- proposed extensions and improvements to the architectures and pipelines will not improve state-of-the-art solutions' results
- hardware limitations while working on more complex architectures

Further research suggestions:

- investigating whether obtained models can be generalised to other benchmarks or public datasets
- multiclass classification – including other misinformation types or degree of certainty
- providing the dataflow with other inputs, for example authors' data, connections between articles, comments, or opinions

Thank you for your **attention**!

$$A(Q, K, V) = \text{softmax}\left(\frac{QK^{\mathsf{T}}}{\sqrt{d_K}}\right) V$$