

Clickbait Detection on Turkish News Articles

Outline

- Problem Definition
- Dataset Characterization
- Research Questions
- Methodology
- Expected Challenges

Problem Definition

- Clickbait titles are misleading and they are not informative.
- Clickbait titles are used to attract users to click on the news article.
- By predicting clickbait titles, we can prevent users from being misled.

Dataset Characterization

- 20,036 news article titles collected from different news websites
- 10,030 clickbait titles and 10,006 non-clickbait titles
- Source: [Kaggle](#)
- I can also try to collect my own data from the APIs of news websites.

Research Questions

- How can we convert the clickbait titles into a meaningful representation for the machine learning models?
- How can we extract some features from the clickbait titles?
- Which machine learning model is the best for clickbait detection?
Do we also need to use deep learning models?

Methodology

- Data Preprocessing
- Feature Extraction
- Tokenization
- Machine Learning Models

Expected Challenges

- NLP is not the topic of the course
- Implementing it for Turkish language could be a challenge