

NLP-based 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.

Dataset Characterization

- 20,036 news article titles collected from different news websites
- 10,030 clickbait titles and 10,006 non-clickbait titles
- [Source](#)

Research Questions

- How can we convert the clickbait titles into a meaningful representation for the machine learning models? (tokenization?)
- Which machine learning model is the best for clickbait detection?

Methodology

- NLP + Tokenization
- Machine Learning Models

Expected Challenges

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