📄 FAKE NEWS DETECTION USING NLP – PROJECT REPORT

◈ INTRODUCTION:

Fake news poses a significant threat to public trust and accurate information dissemination, especially in the digital age. With the rapid spread of misinformation, it is crucial to build automated systems that can help in identifying fake news articles. This project uses Natural Language Processing (NLP) and Machine Learning (ML) techniques to classify Turkish news articles as either REAL or FAKE.

◈ ABSTRACT:

This project implements a fake news classifier using a dataset of Turkish news articles. Due to the absence of actual labels, simulated labels were used to build a demo model. Text cleaning was performed using a custom Turkish stopword list. TF-IDF was used for feature extraction, and a Logistic Regression classifier was trained. The model was evaluated using accuracy and classification metrics. This project demonstrates the feasibility of using NLP for fake news detection in Turkish language content.

◈ TOOLS USED:

- Programming Language: Python

- Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn

- Text Processing: Custom stopword filtering, TF-IDF

- Model: Logistic Regression

◈ STEPS INVOLVED IN BUILDING THE PROJECT:

1. Data loading and exploration

2. Simulated binary labeling (FAKE = 0, REAL = 1)

3. Text cleaning: Lowercasing, punctuation removal, Turkish stopword filtering

4. TF-IDF vectorization of cleaned text

5. Model training using Logistic Regression

6. Model evaluation using accuracy score and confusion matrix

7. Custom prediction function to classify user-input text

◈ CONCLUSION:

The fake news detection system was successfully built using a simulated dataset.

While the labels were artificial, the process illustrates a realistic pipeline that can be applied to actual labeled data. With access to real labels, this system can be improved and deployed as a tool for verifying the authenticity of news content in the Turkish language.