

READ.me for the project:

The project contains 3 .csv data sets:

1. News_headlines_tagged_uncleaned.csv – original data contain true labels of news headlines.
Source: <https://www.kaggle.com/datasets/ankurzing/aspect-based-sentiment-analysis-for-financial-news>
2. News_headlines_tagged.csv – cleaned data with original labels.
3. News_headlines_tagged_cryptobert.csv – data containing predictions of the model.

The project contains 3 .ipynb notebooks to be run in the following order:

1. CleanData.ipynb – we run it first to preprocess the testing data. It outputs News_headlines_tagged.csv
2. CryptoBert.ipynb – we run it second to obtain forecast from CryptoBERT.
Input: News_headlines_tagged.csv file.
It outputs: News_headlines_tagged_cryptobert.csv
Source: #<https://huggingface.co/ElKulako/cryptobert>
3. NewsHeadlinesAnalysis.ipynb – finally we run this code where we analyze the performance of the model.
Input 1: News_headlines_tagged.csv – cleaned data.
Input 2: News_headlines_tagged_cryptobert.csv