

IB Analytics Case Study Reference:

[Financial Opinion Mining and Question Answering Open Challenge \(2018\).](https://huggingface.co/datasets/pauri32/fiqa-2018)

Task:

Given a text instance in the financial domain (microblog message, news statement or headline) in English, detect the target aspects which are mentioned in the text (from a pre-defined list of aspect classes) and predict the sentiment score for each of the mentioned targets. Sentiment scores will be defined using continuous numeric values ranged from -1(negative) to 1(positive). You can simplify this into a classification problem with three labels of positive (+1), negative (-1) and neutral (0). If you choose to do so, describe the range of continuous values used to transform into these labels.

Dataset:

<https://huggingface.co/datasets/pauri32/fiqa-2018>

Search this dataset						
sentence	snippets	target	sentiment_score	aspects	format	label
string · lengths	string · lengths	string · lengths	float64	string · lengths	string · classes	int64
38-62 12.6%	32-44 27%	13-16 2.5%	0.24-0.43 23.9%	27-34 32.5%	headline 38.2% 0 65.8%	
How Kraft-Heinz Merger Came Together in Speedy 10 Weeks	['Merger Came Together in Speedy 10 Weeks']	Kraft	0.214	['Corporate/M&A/M&A']	headline	0
Slump in Weir leads FTSE down from record high	['down from record high']	Weir	-0.827	['Market/Volatility/Volatility']	headline	2
\$AAPL bounces off support, it seems	['bounces off support']	AAPL	0.443	['Stock/Price Action/Bullish/Bullish Behavior']	post	0

Please use python with Jupyter notebook for this project. Feel free to use any Machine Learning model.

The notebook should contain the following (using markdown):

1. Exploratory Data Analysis
2. Steps to train a machine learning model for this task – you can choose to use an existing base model with finetuning/transfer learning. You can also choose to train your model on other similar datasets if you believe this can help improve performance. Describe why you chose the model.
3. Describe the metric used for measuring out of sample performance of your model and why you chose to use this metric.

Expected Submission:

Text file (.txt) file with the code and cover letter explaining the same and guide to run it

PDF file containing output of the code (convert .ipynb to pdf)

We will judge the project on:

- Coding style standard
- Quality of data analysis
- Quality of the machine learning algorithm and it's performance.
- Overall presentation of the project

Estimated Duration: 2 days

Unrestricted

For further information about Barclays, please visit our website www.home.barclays

We are an equal opportunity employer and we are opposed to discrimination on any grounds. If this position requires the person to be an Approved Person under the Financial Services and Markets Act 2000 (FSMA), disclosure of spent convictions within the meaning set out in the Rehabilitation of Offenders Act 1974 (Exceptions) Order 1975 will be required.

This document is for information purposes only. It does not impose contractual obligations or confer benefits. It is not intended to be part of a contract of employment or contractually binding

Unrestricted

For further information about Barclays, please visit our website www.home.barclays

We are an equal opportunity employer and we are opposed to discrimination on any grounds. If this position requires the person to be an Approved Person under the Financial Services and Markets Act 2000 (FSMA), disclosure of spent convictions within the meaning set out in the Rehabilitation of Offenders Act 1974 (Exceptions) Order 1975 will be required.

This document is for information purposes only. It does not impose contractual obligations or confer benefits. It is not intended to be part of a contract of employment or contractually binding