

# **Cryptonews articles with price momentum labels**

Second Phase of Porject

Sahand nazarzadeh  
97522265

## First part

The following are the flags that can be used as arguments in a Python script which can be found in the script directory:

--textual-source: This flag is used to specify the textual source for the data that will be used in the script. It takes a string argument, with a default value of 'cryptonews'. This argument is not required.

--labeling-method: This flag is used to specify the labeling method for the data. It takes a string argument, with a default value of 'color'. This argument is not required.

--label: This flag is used to specify the label for the data. It takes an integer argument, with a default value of None. This argument is not required.

--load-model: This flag is used to specify whether a pre-trained model should be loaded. It takes a boolean argument, with a default value of False. This argument is not required.

--save-model: This flag is used to specify whether the trained model should be saved. It takes a boolean argument, with a default value of False. This argument is not required.

--save-image: This flag is used to specify whether the output image should be saved. It takes a boolean argument, with a default value of False. This argument is not required.

--embedding-size: This flag is used to specify the dimensionality of the embedding space. It takes an integer argument, with a default value of 32. This argument is not required.

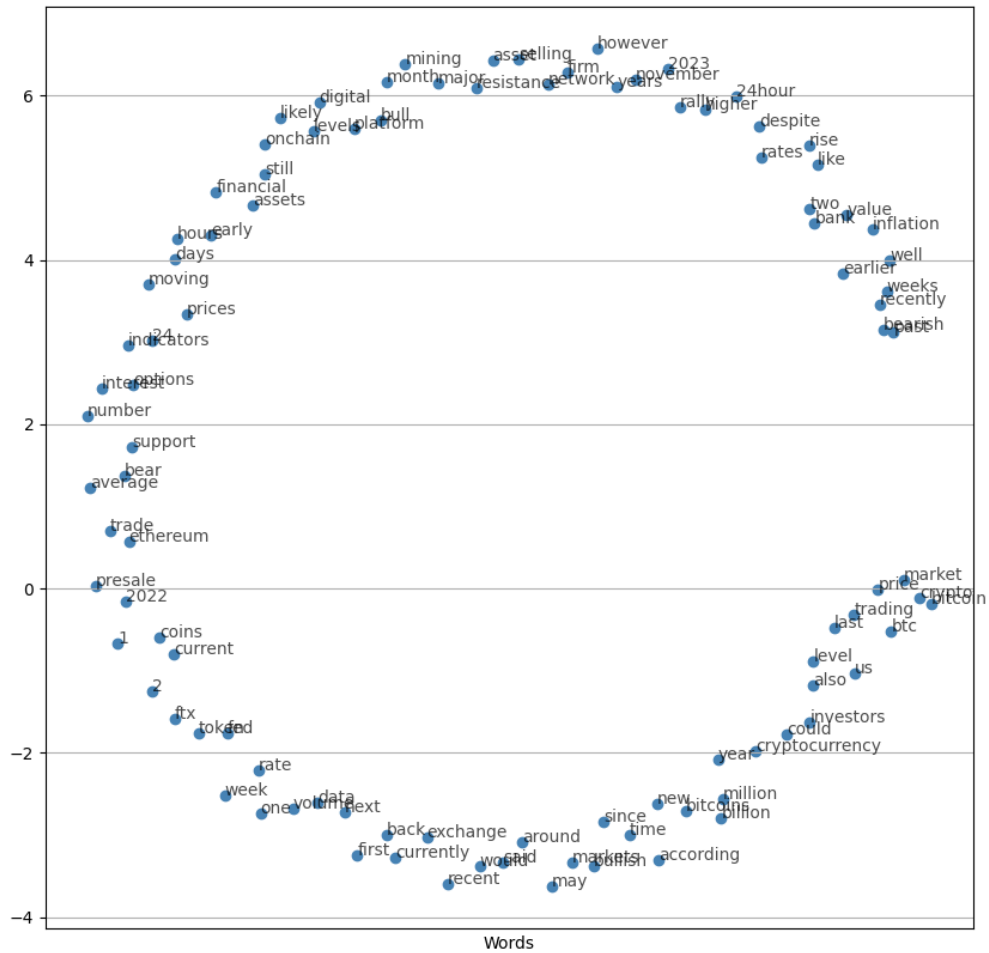
--negative-sampling: This flag is used to specify the number of negative samples to use during training. It takes an integer argument, with a default value of 10. This argument is not required.

--epochs: This flag is used to specify the number of epochs to train the model for. It takes an integer argument, with a default value of 1000. This argument is not required.

--batch-size: This flag is used to specify the size of each batch during training. It takes an integer argument, with a default value of 512. This argument is not required.

--visualizing-images: This flag is used to specify the number of words in image to use for visualization. It takes an integer argument, with a default value of 100. This argument is not required.

Generated image of word2vec which can be found in the stats directory.



## **Second Part**

Codes for this part can be found in the src/tokenizer directory.

## **Third, Fifth and 6<sup>th</sup> Part**

The output of my model is not in the form of a sentence or text; rather, it is a numerical value that enables me to make predictions regarding the future momentum of the market. Due to the nature of the output, it is not possible to use large language models to create additional sentences for each class or generate new samples.

## **Fourth Part**

The codes for this part can be found in the src/feature directory. All feature sets have been implemented, but due to time constraints, only some of them may have been run to generate results for this part.