

Price Prediction Based on News

You can download the initial problem file from [this link](#).

Regarding the prediction of the cryptocurrency market trend, it can be said that this trend depends on multiple factors, including factors related to global markets and daily information and news. Technical analyses (which examine price patterns and indicators) and fundamental analyses (which examine economic and political news and events) are two important methods for predicting crypto markets.

Note that due to their dynamism and the influence of unpredictable factors such as news, government decisions, and important global events, cryptocurrency markets usually have more irregularity in their trends compared to traditional markets. Therefore, **market trend predictions**, although they can provide useful information, must be used with caution and by considering the risks of dynamic crypto markets.

In this question, we want to examine the impact of existing news on a cryptocurrency. To examine these cases, we have provided you with the headlines of the news and their effects on a cryptocurrency in a specific time frame. And we ask you to predict the `Label` column for the coming days.

Dataset

The dataset provided to you in this question includes the following columns:

Column	Description
Date	Trading Day Date
Top 1–25	News headlines related to that day
Label	Number 0 or 1

- **Note:** If the adjusted price on the trading day is higher than the previous day, the value of the `Label` column will be `1`, and if it is lower than the previous day, it will be `0`.

Final Requirement

In this question, you must predict the value of the `Label` column for the next 400 trading days using machine learning models.

Evaluation

Your model will be evaluated using the `F1 Score` metric, and the averaging model is `Weighted`.

▼ Attention

During the competition, the score you see is only the result of the `F1 Score` on 30% of the file you upload to Quera. After the competition time ends, your **final score** will be calculated on the remaining 70%.

This is done to prevent `overfitting` and maintain the generality of the model to ensure that models that have been overfitted will have a lower final score.

Question Output

You must save your predictions **in order** in a column named `prediction` as a `csv` file named `submission.csv` that contains one column named `prediction`.

Sample Output

<code>prediction</code>
<code>0</code>
<code>1</code>

prediction
0
1
0

▼ Important Notes About the Submission File

- **Note 1:** Make sure that the mentioned column definitely has a header .
- **Note 2:** Be careful not to save the index in the final file and only have one prediction column.
- **Note 3:** The numbers in the prediction column are merely for example and are not the correct answer!

▼ Warning

Do not forget that **before the end of the competition time**, you **must** send us all the codes for this competition from the **Code Upload** section. Otherwise, you will not get any points from this competition.

Note that if you are using a jupyter notebook , you must get the .py output just like the explanations in the **Code Upload** section and consider it for submission. Submitting jupyter files like .ipynb is not acceptable.