CAIXABANK TECH HACKATON 2022

The Challenge: Forecasting on a given dataset from the IBEX35 index. Predict a target value that indicates if the closing price in the next three days will be lower or greater than the current value.

The Data:

	Date	Open	High	Low	Close	Adj Close	Volume	Target
0	1994-01-03	3615.199951	3654.699951	3581.000000	3654.500000	3654.496338	0.0	0
1	1994-01-04	3654.500000	3675.500000	3625.100098	3630.300049	3630.296387	0.0	1
2	1994-01-05	3625.199951	3625.199951	3583.399902	3621.199951	3621.196289	0.0	1
3	1994-01-06	NaN	NaN	NaN	NaN	NaN	NaN	0
4	1994-01-07	3621.199951	3644.399902	3598.699951	3636.399902	3636.396240	0.0	1

Clear data with High-Low, Close-Open prices, and volume.

How I organized myself to solve it.

- The Data
- The Model
- Predictions

And after all the steps are done, iterate.

<u>The data:</u>

Even though the data is pretty clean, there are some exceptions that could ruin the model. For example, in row three, the values are nan. To solve this issue was implemented a function to fill the values with a median between the two non-nan nearest values.

Another insight about the data is that as it is stock market data, weekends are non-working days, so there is no data recorded in all the datasets.

The data has to be transformed in a way that it packs the most information per point.

For doing so I used the following indexes:

- Stock High minus Low
- Stock Open price minus Close
- Moving Averages of 7, 14, and 21 days
- The standard deviation for the past seven days

The Model:

After some research, I decided to implement a Random Forest Regression model. It is easy to implement and has results very close to mode complex algorithms such as ANN.

The Prediction:

Being a Regression model I used it to predict the Close price, then to get the target I created a function that based on the predicted close price generates the target label

References:

https://www.sciencedirect.com/science/article/pii/S1877050920307924