***Dataset***

***Bitcoin info (8)***: Date - Bitcoin price - Open - Close - High - Low - Trading Volume - Market Cap

***Mining info (3)***: Mining difficulty - Mining award - Average Hashrate

***Public attention (4)***: Google Trends - Twitter tweets - Baidu index - Reddit posts

***Macroeconomics Environment (12)***: Brent Crude futures - CBOE volatility index - NYSE Copper Futures - Dow Jones industrial average - New York Stock Exchange Gold futures - HuShen 300 index - Korea composite index - Singapore Straits Times index - Standard & Poor’s 500 index - FTSE 100 index - Tokyo’s Nikkel index - US dollar index - WTI oil future

*So we’re going to have a dataset of 27 columns.*

***Step-by-step planning***

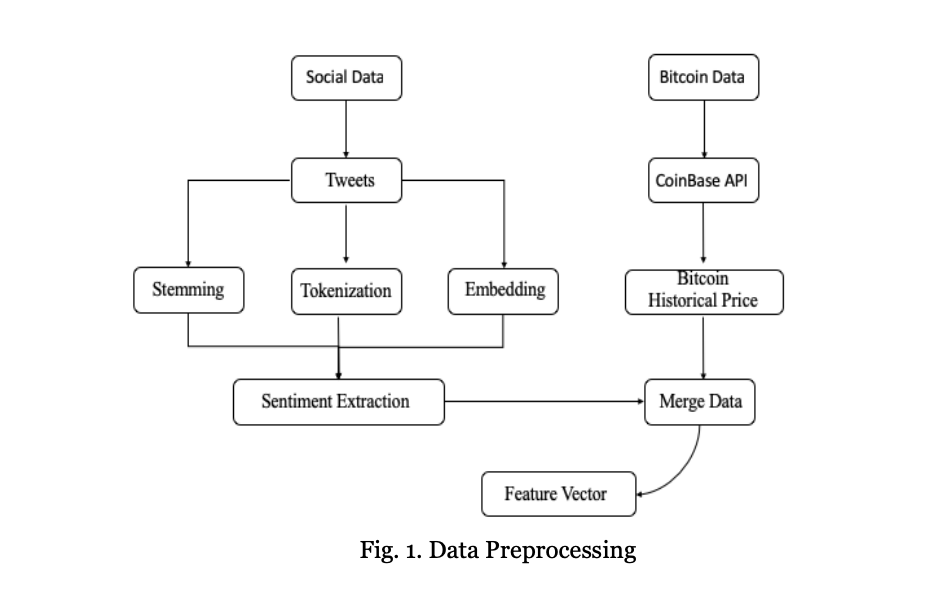
**First: Data Preparation**

After scraping through Google trends, Twitter, Top 10 news headlines (Optional), and Baidu. Each row should be iterated over an algorithm (TextBlob, Haven on demand, Tweepy) that generates the Subjectivity, Objectivity, Positivity, Negativity, and neutral ratios.

→ After Getting these ratios we have to add them into the dataset as a news indicator that may impact the price of BTC.

We have to add all the determinants into one CSV in addition to the BTC info and news ratios.

Then we have to divide our dataset into training, testing, and validation datasets.



**Second: Model Selection**

We have to choose between 4 models: XGBoost, CatBoost, LSTM, and ARIMA.

Then we have to do hyper-parameter tuning to choose the best parameters for our model for better prediction.

Note: There’s no significant difference in prediction between XGboost and LSTM. But XGboost is better in performance, where there’s a significant difference in computation speed between XGboost and LSTM, where *XGboost is faster than LSTM***.**

Why to use LSTM?

**Third: Fitting the model and predicting**

Splitting data: 70:30

After choosing the model and the most optimal parameters, we can start by fitting the model on the training dataset.

**Fourth: Forecasting the BTC price**

After training the model on the training data set and predicting the prices, we have to update the dataset either hourly or every 4 hrs or daily according to the timestamp that we need to predict.

**HOW??**