

# Predict Dow Jones will go up or down using machine learning based on headlines

↳ (e) Date Date OS Bus. Features Headlines (Top 50)

↳ combine the headline w/ Bus.F

↳ ~~make~~ combine all the headlines

↳ make features based on the ML.

pos, neg, bias, unbiased

↳ ML 80% Train

20% Test

↳ Train the model  
input  
all the features  
Bus & Text  
output  
label  
(0, 1)

↳ Test the model

input

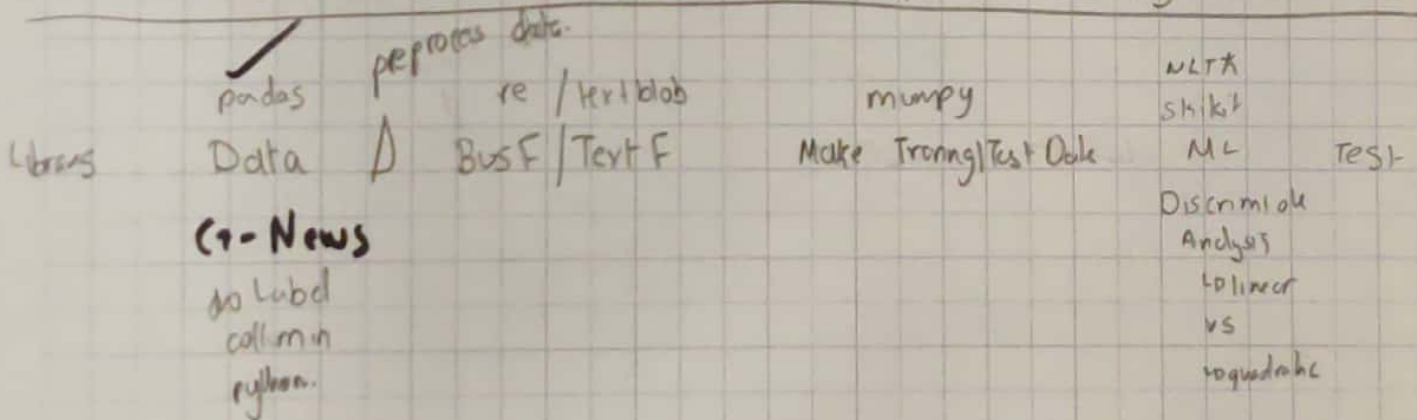
all the features

⇒

compare output to label

already known table.

↳ show what it did the machine predict correctly.



↳ export to excel to see if everything is correct

↳ rename variables

HONG KONG | LONDON | LOS ANGELES | PARIS | PHILADELPHIA

WWW.INSIGNIAM.COM

Sent	Headline
Pos	—
P	—
Neg	—

① convert to 0, 1, 2

① Stop words

① Tokenize

② Normalization

③ noise removal

↳

④ normalization  
is: C → bad

② tokenizer

① lower case

② Stemming

③ Lemmatize etc

**Transform**

**Classifier**

ex: Stop words:  
show freq d. Brian. so  
they know we have  
to take them out  
↳ in each step show  
why you do this

understand  
this.

use sklearn  
book marked  
page.

④ predict

given a tweet that shows that stock should  
classify it and tell that as we can see  
the stock should have shifted down.

Based on the date of the tweet, plot the  
stock and show it actually happened.

Extracting Features from Text.

Tokenizer ⇒ Transform ⇒ Classifier. or Pipeline.