

First day IPO Stock Price Prediction :

Can we use Machine Learning to outsmart the market ?

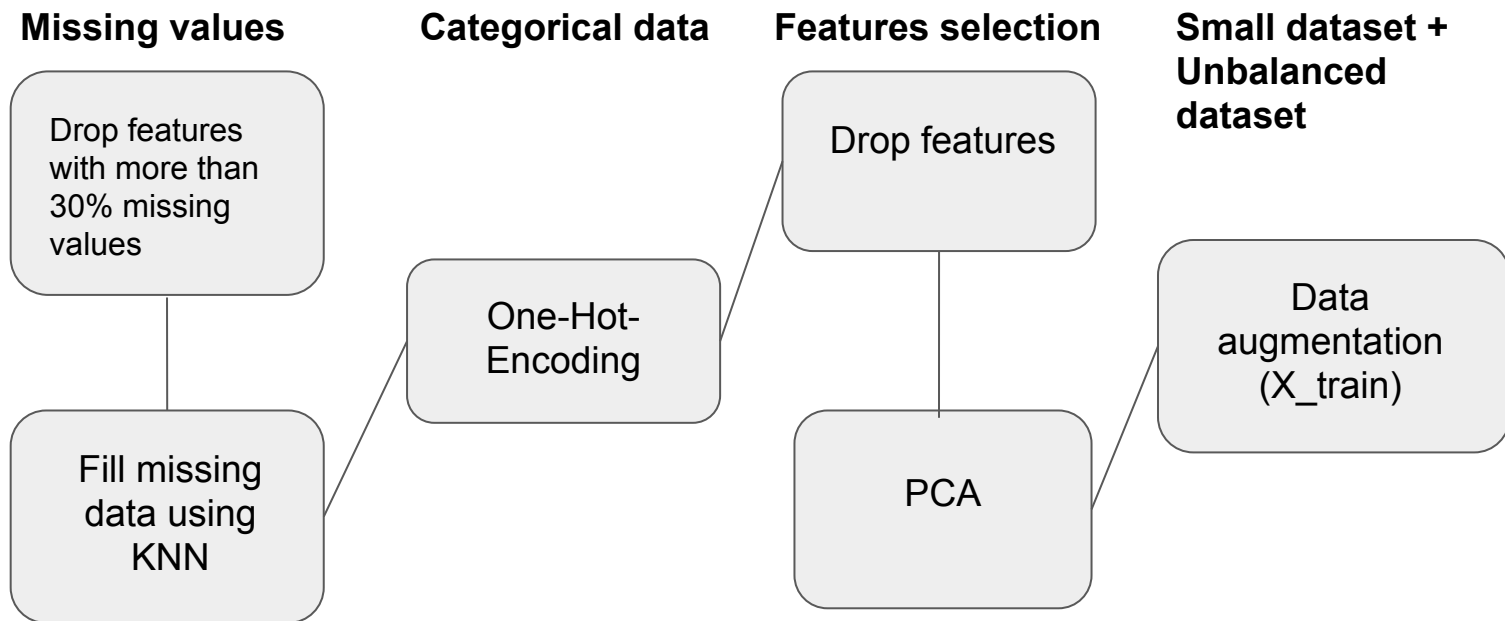
Mortiniera Thevie, Spiess Philipp, Vidal-Revel Elisa

What is the problem and why is it important ?

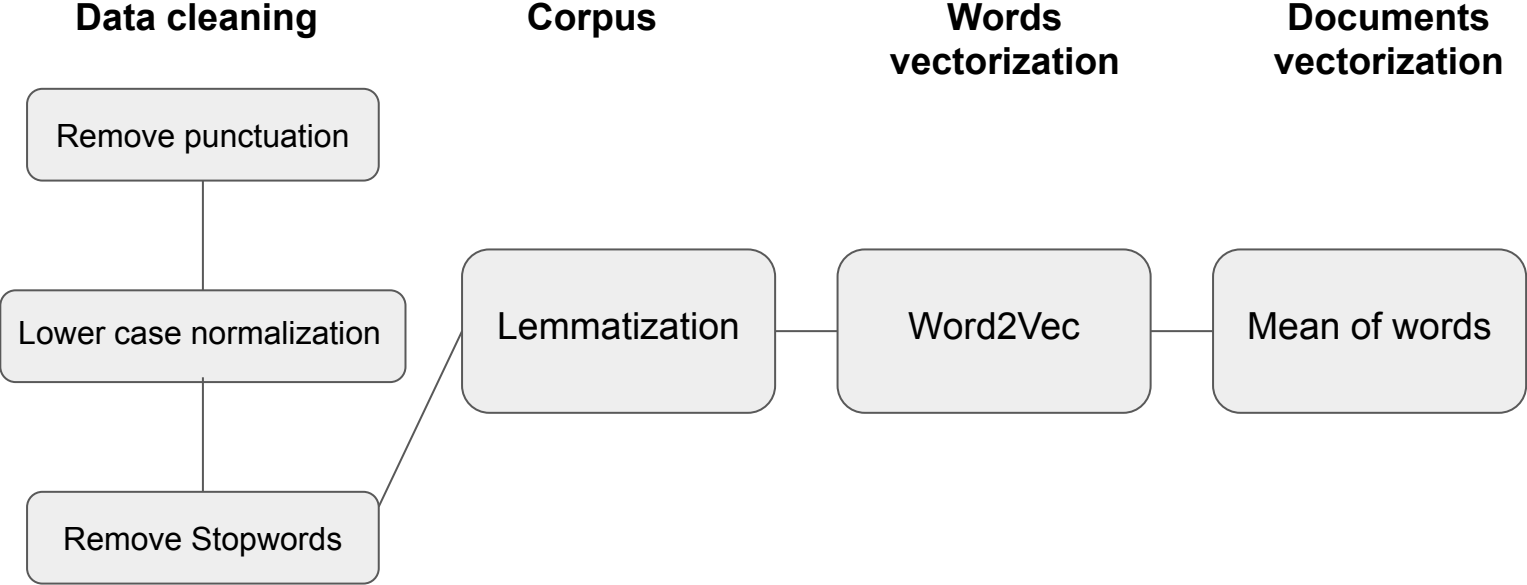
- IPOs pricing is a challenging task which require lot of time and workforce
- We place ourselves as **external consultant** for an investment bank who would like to know if they should use automated tools based on Machine Learning to more accurately price IPOs. In this scenario, we focus on first-day IPO pricing
- The **challenge** is double for the investment bank reputation
 - Setting a price too low in comparison to the price of the stock at the end of the day is not in the interest of the issuing company as it wants an accurate prediction because that helps them maximize the amount of capital that they can raise from the IPO
 - Conversely, overpricing the IPO can lead to a rapid fall in prices, which will hurt the reputation and future prospects of the bank.

P1	Predict whether the closing price will go up or down without the risk factors.
P2	Predict whether the closing price will go up or down using only risk factors, year and industryFF12
P3	Using any features, predict whether the closing price will go up or down
P4	Predict whether the closing price will go up by more than 20% from the original price
P5	Predict whether the closing price will go down by more than 20% from the original price
P6	Predict the share price at the end of the first day
P7	Predict the probability that the closing price will go up by more than 5% from the original price
P8	Predict the probability that the closing price will go up by more than 50% from the original price
P9	Predict the probability that the closing price will go down by more than 5% from the original price

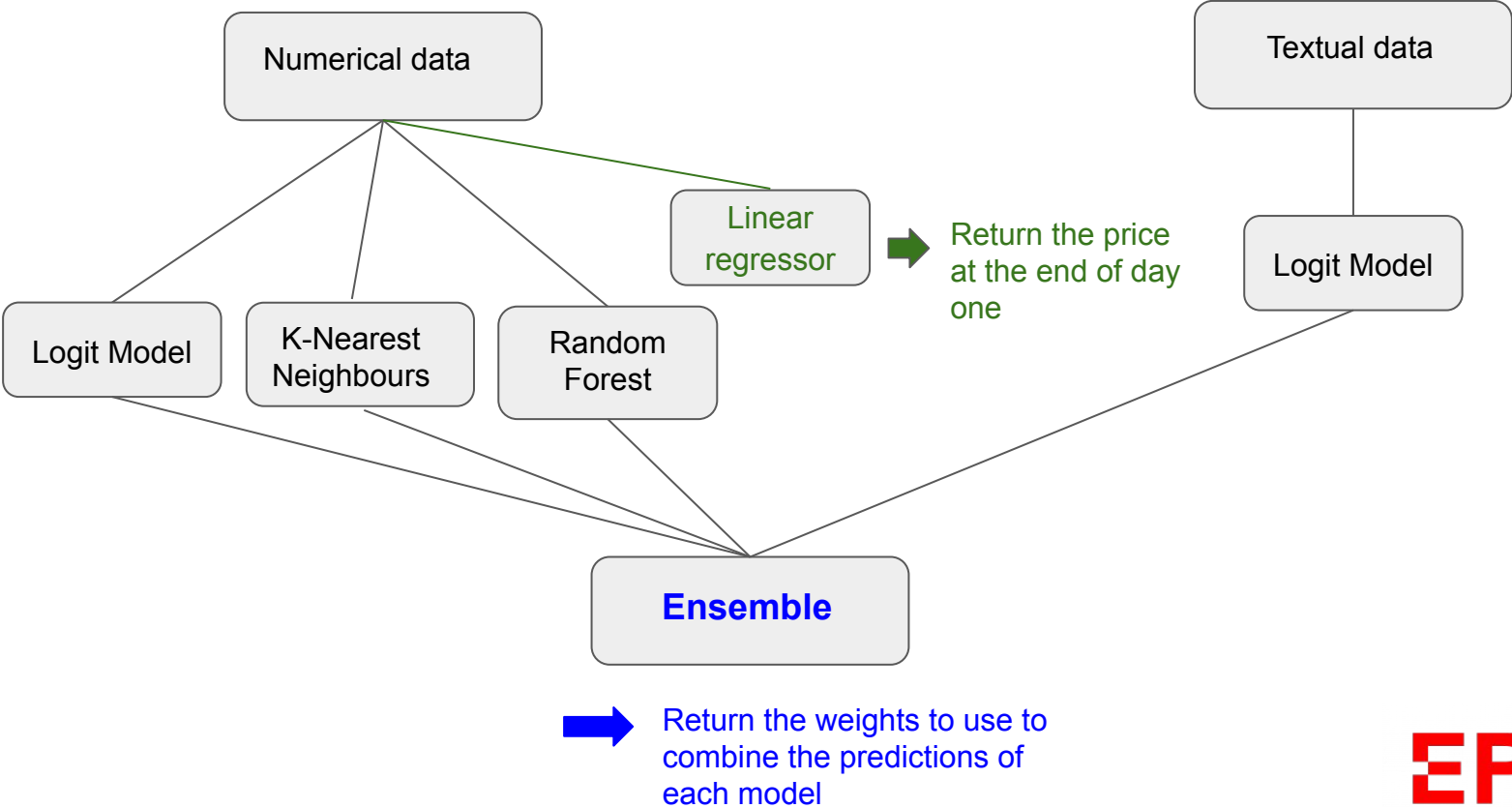
Data processing



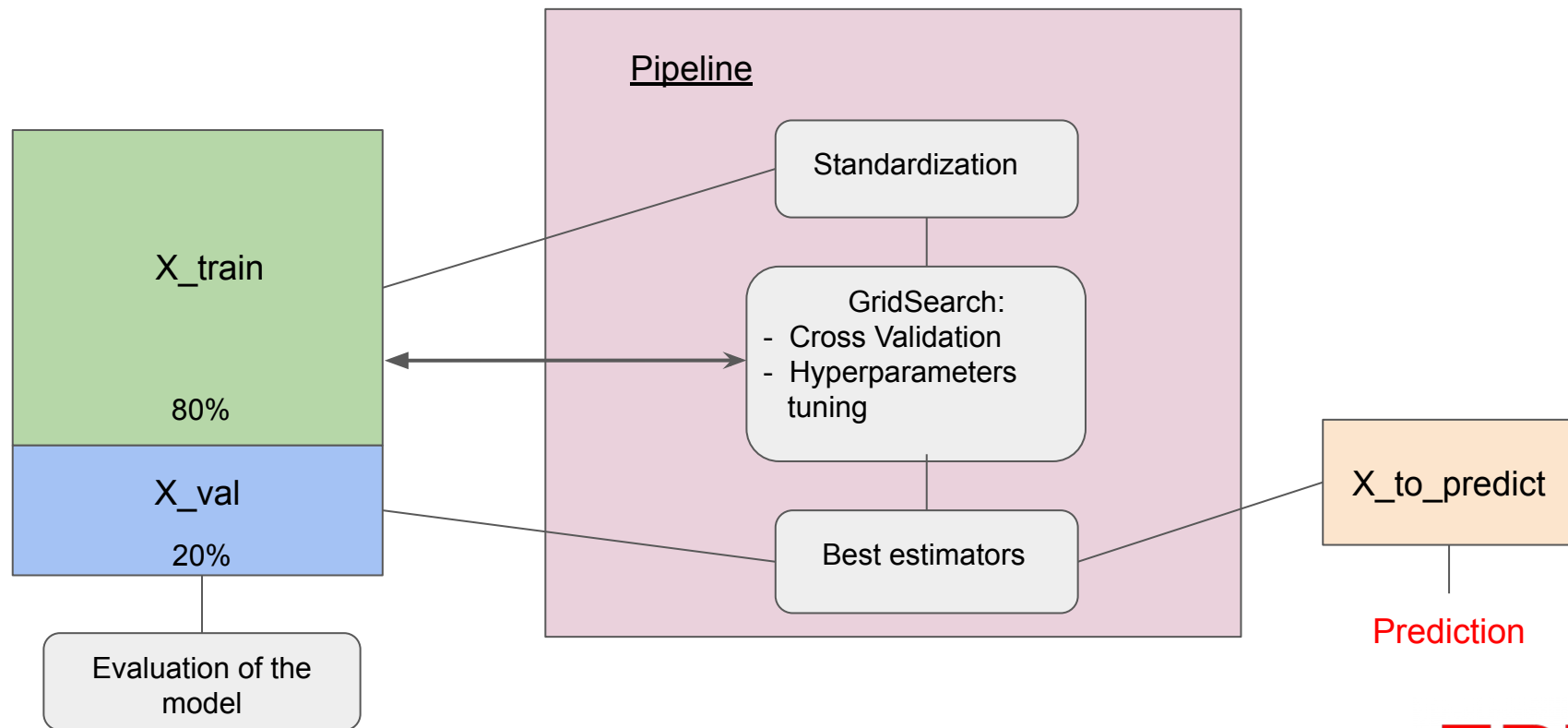
Text processing



Models



Models: exemple



Metrics

P1, P2, P3, P4, P5	F1 score, average=macro
P6	Root mean square error
P7, P8	<ul style="list-style-type: none">- If proba P turns out to be FALSE, the score is $P * P$- If proba P turns out to be TRUE, the score is $(100-P)*(100-P)$
P9	<ul style="list-style-type: none">- If proba P turns out to be FALSE, the error score is P- If proba P turns out to be TRUE, the error score is $2*(100-p)$

Results

Target	Standard	PCA	PCA + augmentation
P1 (f1-score)	0.654	0.655	0.651
P2 (f1-score)	0.618	0.658	0.651
P3 (f1-score)	0.659	0.682	0.687
P4 (f1-score)	0.764	0.759	0.763
P5 (f1-score)	0.832	0.832	0.833
P6 (mse)	5.74	7.79	7.79
P7 (p7-metric)	2548.42	2621.9	2624.4
P8 (p7-metric)	409.25	1233.7	1194.5
P9 (p9-metric)	5.045	22.0	22.0