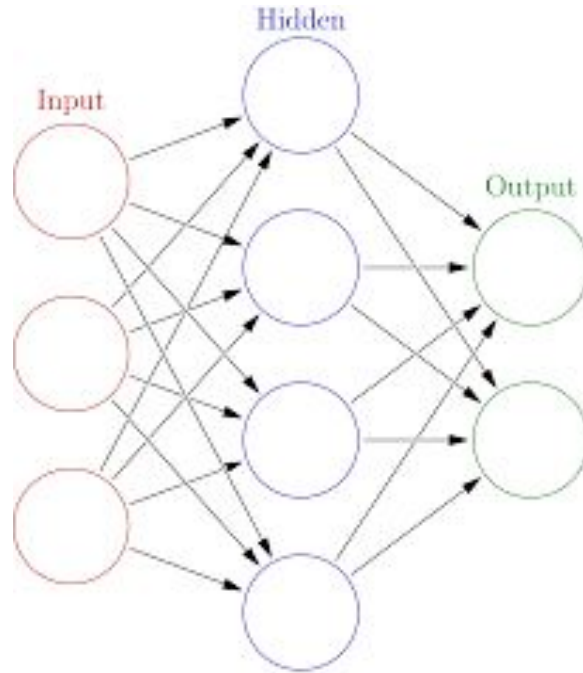

Predicting the direction of prices on the stock market using recurrent neural networks

— David M. Q. Nelson —
davidmn@ufmg.br

Stock price prediction



Artificial Neural Networks

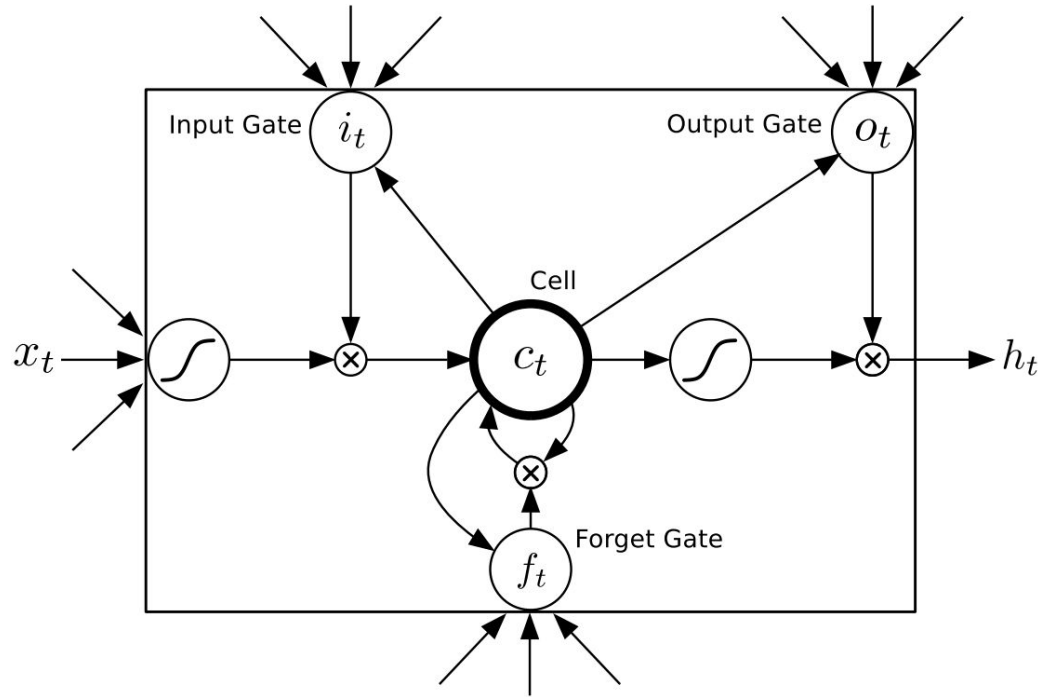


Recurrent Neural Networks

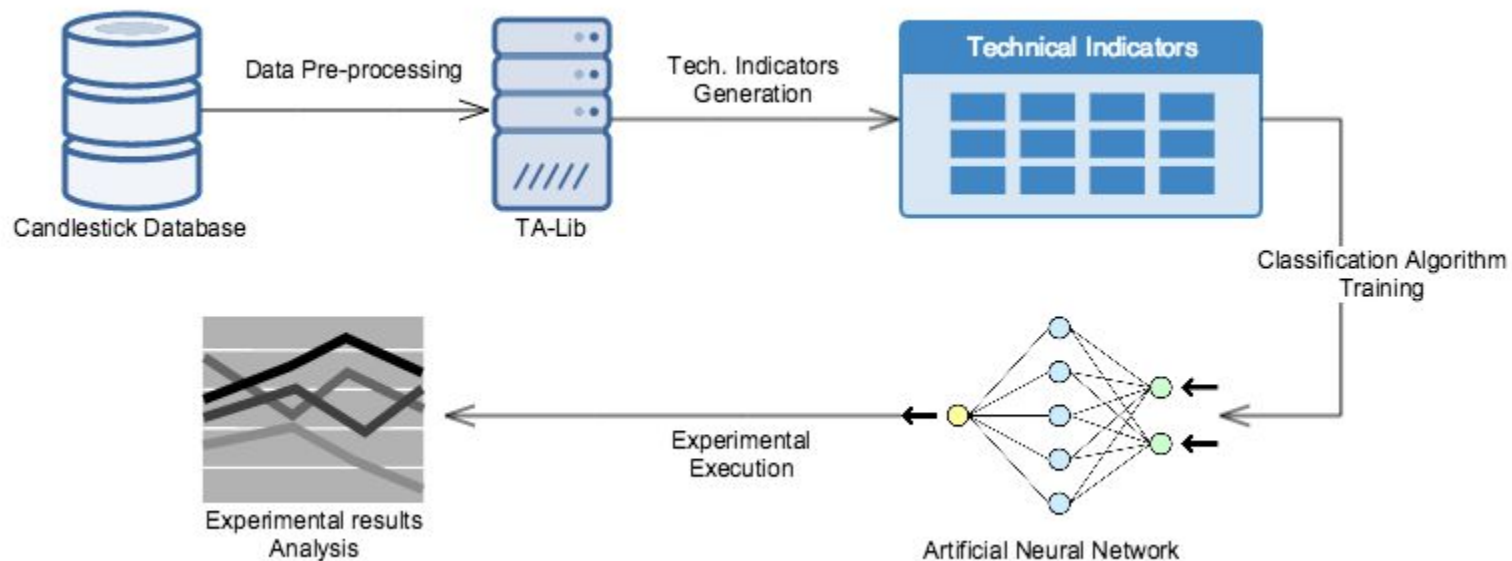
That's very good, not bad at all!

That's very bad, not good at all!

Long short term memory



Methodology



Features

- Features (180):
 - Candlestick data: {open, close, high, low, volume}
 - Technical indicators

Código	Nome do Indicador	Código	Nome do Indicador
AD	Chaikin A/D Line	CDLONNECK	On-Neck Pattern
ADOSC	Chaikin A/D Oscillator	CDLPIERCING	Piercing Pattern
ADX	Average Directional Movement Index	CDLRICKSHAWMAN	Rickshaw Man
ADXRS	Average Directional Movement Index Rating	CDLRRISEFALL3METHODS	Rising/Falling Three Methods
APO	Absolute Price Oscillator	CDLSEPARATINGLINES	Separating Lines
AROON	Aroon	CDLSHOOTINGSTAR	Shooting Star
AROONOSC	Aroon Oscillator	CDLSHORTLINE	Short Line Candle
ATR	Average True Range	CDLSPINNINGTOP	Spinning Top
AVGPRICE	Average Price	CDLSTALLEDPATTERN	Stalled Pattern
BBANDS	Bollinger Bands	CDLSTICKSANDWICH	Stick Sandwich
BETA	Beta	CDLTAKURI	Takuri (Dragonfly Doji with very long lower shadow)
BOP	Balance Of Power	CDLTASUKIGAP	Tasuki Gap
CCI	Commodity Channel Index	CDLTHRUSTING	Thrusting Pattern
CDL2CROWS	Two Crows	CDLTRISTAR	Tristar Pattern
CDL3BLACKCROWS	Three Black Crows		
CDL3WHITESWANS	Three White Swans		

Input and Output

- $X = \{\{x_t, x_{t-1}, x_{t-2}, \dots, x_{tn}\}, \{x_{t-1}, x_{t-2}, x_{t-3}, \dots, x_{tn}\}, \dots\}$
- $Y =$
 - 0 = Price does not go up (no action)
 - 1 = Price goes up (triggers action)

- Sliding window model
 - n previous instances of x to determine y on the next time step.

Data

- Candlesticks for every 15 minutes
- Train: 10 months of data
- Test: 1 month of data

Stock	% triggers	Begin	End	B&H
BOVA11	0.471875	55.53	55.73	20
BBDC4	0.435937	15.09	13.26	-183
ITUB4	0.442188	13.81	14.73	92
CIEL3	0.453125	20.42	21.97	155
PETR4	0.479687	21.71	18.52	-319

Operations

- Every time **y = 1**
 - Buys a set of 100 stocks at the current market value
 - Sells 15 minutes later at current market value
 - Uses close price to determine
 - Disregards costs and taxes

Results

Accuracy = correct predictions / total predictions

Precision = $tp / (tp + fp)$

Recall = $tp / (tp + fn)$

F1 = $2 (Precision \times Recall) / (Precision + Recall)$

Baselines:

- Multi-layer perceptron

- Random Forest

Stock	Accuracy	Precision	Recall	F1	MLP Accuracy	RF Accuracy
BOVA11	0.564063	0.560847	0.350993	0.431772	0.531250	0.526563
BBDC4	0.575000	0.553846	0.129032	0.209302	0.526563	0.475000
ITUB4	0.551562	0.475000	0.134276	0.209366	0.539062	0.481250
CIEL3	0.540625	0.476190	0.137931	0.213904	0.517188	0.521875
PETR4	0.545312	0.563492	0.231270	0.327945	0.539062	0.534375

Results

Baselines:

- Random
- Optimistic
- Buy and Hold

Stock	\$	\$/Op.	\$(Random)	\$/Op.(Random)	\$(Optimistic)	\$/Op.(Optimistic)	B&H
BOVA11	633	3.35	39	0.13	-162	-0.54	20
BBDC4	274	4.22	-19	-0.07	-122	-0.44	-183
ITUB4	199	2.49	-57	-0.20	500	1.77	92
CIEL3	42	0.50	196	0.71	-57	-0.20	155
PETR4	83	0.66	-2	-0.01	-328	-1.07	-319

Conclusions and next steps

- Results are better than baselines
- Look into ways to improve consistency and general results quality
- Analyze and try to identify possible patterns that determine the network performance
- Add non-pricing data into the mix: sentiment analysis on news related to the finance market