Name: Teng Lung Yun

Student Number: WQD180024

OBJECTIVE

Provide a user with recommendations around optimal actions to achieve investment objectives such as profits and return on investment (Individual)

BACKGROUND

Technical Analysis (TA) and Fundamental Analysis (FA) are the two main methods used by market participants when buying and selling shares. Technical Analysis studies the behavior of the price-volume movement of shares and uses it to predict its future price movements. It often uses complex charts and trend lines. Fundamental Analysis revolves around analyzing the performance and health of a business by carefully dissecting its financial statements and broad economic factors.

METHODS

Machine Learning

In this project, we will focus on applying machine mining algorithms on the technical indicators and sentiment score to predict the price up and down for the next day.

We will identify mainly 4 industries to tackle the problem. Our hypothesis is each industry/stock may have its own machine learning algorithm with is the best fit for stock price prediction.

The four industries with respective stocks are as follows:

Steel Industries	Banking	Energy	Technology
Lionind, Ssteel, Annjoo, Masteel	CIMB Bank	Sapura Kencana	Dufu

Table 1: Four Industries with Respective Stocks

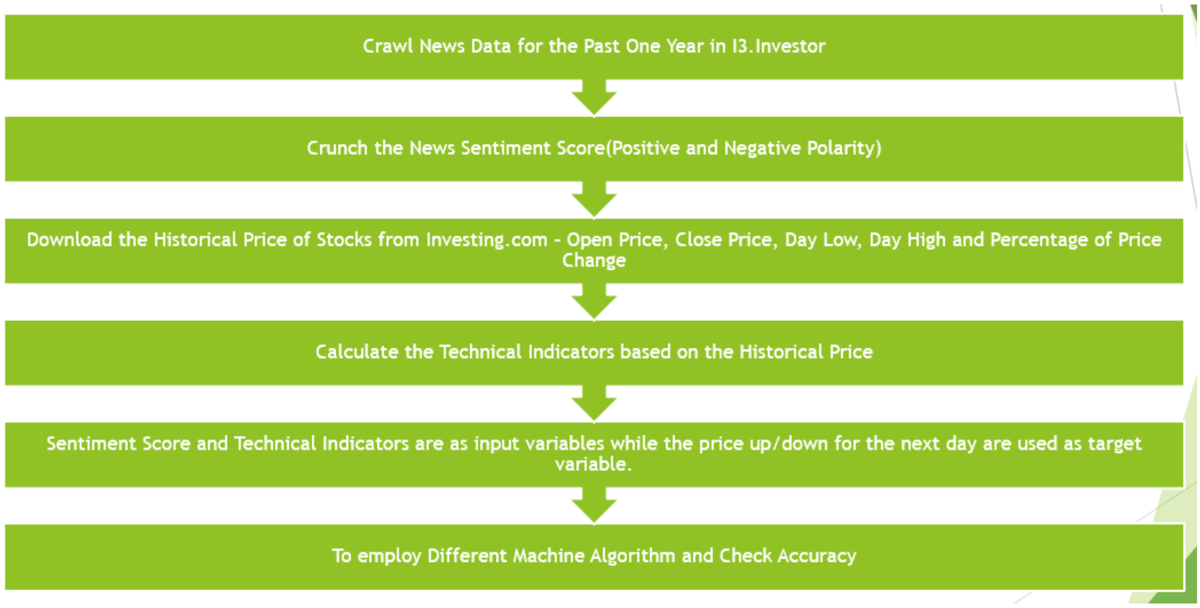


Figure 1: Data Flow Plan

Clustering

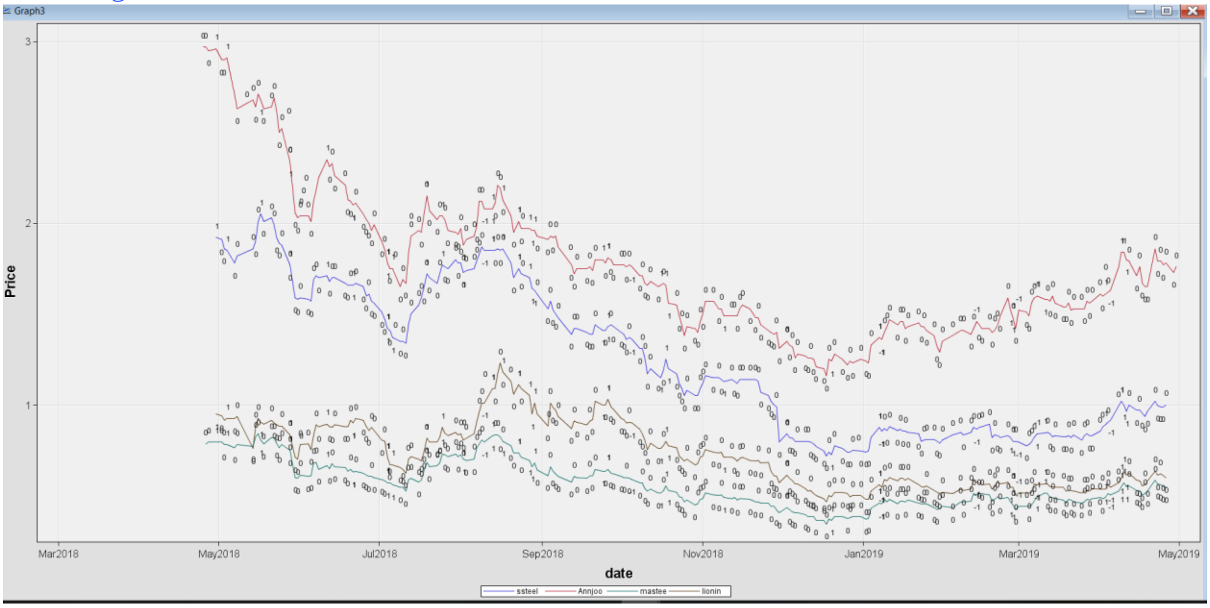


Figure 2: Price Trend of Lionind, Ssteel, Annjoo and Masteel

Lionind, Ssteel, Annjoo and Masteel are belonged to long steel subsector. Hence, from figure 2, the price trend of the stocks is in the same trend. Since investors seek the best opportunities for investment, the results of clustering can help them to make better decision.

First step is to determine appropriate criteria for clustering. I selected the following profitability ratios for easier analysis.

Variables	Definition
Return on assets(ROA)	Net Profit/ Total Assets
Return on equity(ROE)	Net Income/Shareholder Equity
Gross Margin	The Difference between Revenue and Cost of Goods Sold(COGS) divided by Revenue
Earning Per Share(EPS)	Net Income – Dividends on preferred Stock/ Average Outstanding Shares
Operating Profit Margins	Operating Profit/ Sales

Table 2: Variables Definition

COLLECTING/PRE-PROCESSING DATA

Historical News are crawled from I3.Investor.com using Scrapy. The historical news can stretch back to last year June 2018. Below snapshot is the historical news for Sapura Kencana in I3.

02-Jul-2018	SAPE Reports 13.8% QoQ Earnings Improvement
02-Jul-2018	Sapura Energy - Gas & equity raising issues cloud brighter prospects
25-Jun-2018	现处亏损阶段．沙布拉能源仍看好？
25-Jun-2018	RM1.8bn Contract Wins a Catalyst for SAPE
24-Jun-2018	沙布拉能源 新增合约预测上修
22-Jun-2018	Sapura Energy - Positive Momentum On Contract Wins And Rig Recovery
22-Jun-2018	Sapura Energy - Multiple Jobs Secured- HLIB
22-Jun-2018	Sapura Energy Behad - Building Up Orderbook
22-Jun-2018	Mplus Market Pulse - 22 Jun 2018
22-Jun-2018	Sapura Energy Berhad - RM1.8b Haul
18-Jun-2018	Oil & Gas Sector - Huge Kasawari project back on the radar
18-Jun-2018	1Q18 Results Wrap - Latest Stock Picks!
14-Jun-2018	沙布拉能源 股价被低估不合理
12-Jun-2018	Sapura Energy - Pepulut discovery enhances SK408 viability
12-Jun-2018	PublicInvest Research Headlines - 12 Jun 2018
12-Jun-2018	Mplus Market Pulse - 12 Jun 2018
12-Jun-2018	Sapura Energy – Another Exploration Success
11-Jun-2018	黄金10年-391-沙布拉能源
08-Jun-2018	MQ Research: Sapura Energy Upgraded to Outperform
04-Jun-2018	Oil & Gas Sector - Fair 1Q results albeit MISC's low tanker rates

Showing 251 to 295 of 295 entries **Previous** 1 2 3 4 5 6 Next

Figure 3: Historical News from I3

Sapura Energy - 1QFY19: Short-term Rig Setback But Long-Term Revenues Improving

Author: UOBKayHian | **Publish date:** Mon, 2 Jul 2018, 10:00 AM

SAPE's 1QFY19 core loss was due to poor rig utilisation. However, the all-time low revenue in 1QFY19 will improve in the coming quarters, based on orderbook visibility and rig recovery. Despite a wider near-term loss forecast, we note that the orderbook guidance now closely matches our assumption. SAPE's new guidance for capital raising signals more mega contract wins, which is positive for future earnings. Its energy reserves will also have long-term upside. Maintain BUY. Target price: RM0.70.

RESULTS

1QFY19 core loss was a slight negative surprise vs our and consensus loss forecasts of RM119m/RM82m respectively, even though 2H is expected to be stronger. Sapura Energy (SAPE) recorded an all-time low quarterly revenue, a negative surprise due to a lower-thanexpected rig utilisation. There were only close to four rigs working, which is below the guidance of 5 out of 15 rigs working. The energy division realised a 1Q oil price of US$70/bbl at unchanged volumes of 1.1mmboe (1QFY18: US$52/bbl), and also average gas prices of >US$3.5/mcf. However, energy's PBT declined despite the higher O&G prices due to the offsetting factors of higher depreciation and interest cost base. E&C activity was stable from the ramp up of new contracts secured, despite being a typically low season. Brazilian pipe-laying support vessels (PLSV) continued to be fully utilised, contributing to the JV line. Its net gearing remains at 1.6x, while cash balance declined significantly from RM1.7b to RM1.3b due to a lower EBITDA of RM0.2b and capex incurred for energy and drilling.

STOCK IMPACT

Rig updates. The wider rig loss was a negative surprise to us. Although we were aware that SKD Esperenza's contract was completed and in transition for its new contract, the idle time was about 2 months and was longer than expected. Hence in 1QFY19, the rig utilization was closer to 4/15 tender rigs. We believe 2QFY19 will likely see 5 rigs working on SKD Alliance contract commencement in Apr 18. 2HFY19 will see six rigs working, as the incoming commencements of SKD Esperenza and SKD Berani will be offset by the expiry of SKD T- 17 by Jun/Jul 18. We believe utilisation recovery to 6-7 working rigs could enable the rig division to achieve profit breakeven. We understand rig charter rates have stabilised, after a decline of 30-40% from three years ago.

Figure 4: Content of the Financial News

stock code	stock name	date	title	subhead	para
5218	SAPURA ENERGY BHD	Mon, 20 May 2019, 12:08 A	Sapura Energy Berhad (KLSE) #RTable - Y		
5218	SAPURA ENERGY BHD	Fri, 19 Apr 2019, 5:52 PM	Evening Market Summary - 19 Apr 2019		The FBM KLCI eked out gains on Fr
5218	SAPURA ENERGY BHD	Wed, 17 Apr 2019, 1:41 PM	æ²™å¸æ‹‰èƒ½æº çŸæœŸç›ˆåˆ©ç»ç		
5218	SAPURA ENERGY BHD	Mon, 29 Apr 2019, 6:40 PM	Beneficiaries, losers of rallying oil prices		
5218	SAPURA ENERGY BHD	Tue, 16 Apr 2019, 6:02 PM	Evening Market Summary - 16 Apr 2019		The FBM KLCI succumbed to sellin
5218	SAPURA ENERGY BHD	Mon, 15 Apr 2019, 5:35 PM	Sapura Energy	Secured 2 Drilling Contra	Sapura Energy (SAPE) announced t

Figure 5: Crawled Timestamp Financial News

A	B	C	D	E	F	G
	V1	V2	V3	mean	date	convertvalue
1	0.1	0	0.072464	0.057488	6/4/2018	positive
2	0.4	-0.11111	0.048889	0.112593	6/8/2018	positive
3	0	NA	NA	0	6/11/2018	neutral
4	0.166667	0	0	0.055556	6/12/2018	positive
5	0	0	0.039422	0.013141	6/12/2018	positive
6	0	0	0	0	6/12/2018	neutral
7	0.166667	0.090909	0.073171	0.110249	6/12/2018	positive
8	0	NA	0	0	6/14/2018	neutral
9	0	0	0	0	6/18/2018	neutral
10	0	0.333333	0.028571	0.120635	6/18/2018	positive
11	0.375	0	0.226415	0.200472	6/22/2018	positive
12	0.166667	NA	0	0.083333	6/22/2018	positive
13	0	NA	0.113636	0.056818	6/22/2018	positive
14	0	0	0	0	6/22/2018	neutral
15	0	NA	0.103448	0.051724	6/22/2018	positive
16	0	NA	0	0	6/24/2018	neutral

Figure 6: Sentiment Analysis on Timestamp Financial News

The crawled financial news is then converted into time-stamp sentiment polarity.

Crawled stock price data only available since early this year. Full correlation cannot be done due to incomplete crawled stock price data. Hence, to correlate the sentiment polarity with KLCI stock price, historical stock data is downloaded from investing.com.

Date	Price	Open	High	Low	Vol.	Change %
17-May-19	5.16	5.15	5.18	5.13	3.39M	0.58%
16-May-19	5.13	5.14	5.16	5.13	4.66M	0.00%
15-May-19	5.13	5.18	5.21	5.1	7.88M	-0.58%
14-May-19	5.16	5.07	5.19	5.04	14.85M	0.78%
13-May-19	5.12	5.14	5.17	5.07	9.40M	0.00%
10-May-19	5.12	5.14	5.16	5.11	8.45M	-0.19%
9-May-19	5.13	5.2	5.2	5.12	11.43M	-1.35%
8-May-19	5.2	5.18	5.22	5.17	18.91M	0.00%
7-May-19	5.2	5.21	5.24	5.19	13.70M	0.00%
6-May-19	5.2	5.21	5.24	5.18	9.07M	-0.76%
3-May-19	5.24	5.25	5.26	5.22	14.37M	0.19%

Figure 7: Downloaded Stock Data from Investing.com

Using table in Figure 7, below technical indicators are calculated using available package in R.

- Relative Strength Index – Momentum Indicator that measures the Magnitude of Recent Price Changes to evaluate overbought or oversold conditions in the price of stock or other asset.
- Exponential Moving Average – A type of Moving Average(MA) that places a greater weight and significance on the most recent data points.
- Moving Average Convergence Divergence(MACD) – MACD is calculated by subtracting 26-period Exponential Moving Average(EMA) from the 12-period EMA.
- Commodity Channel Index(CCI) – Momentum-based oscillator used to help determine when an investment vehicle is reaching condition of being overbought or oversold.
- Change Momentum Indicator(CMO) – Used to gauge price momentum just like the Relative Strength Index(RSI).
- Williams' Percent Change(WPR) – A Dynamic Indicator which determines whether the market is overbought/oversold.
- Average Strength Index(ADX) – ADX Indicator is used to find whether Stock is in trend and also finds the strength of the trend.
- Rate of Change(ROC) – The ROC indicator is a momentum-based technical indicator that measures the percentage change in price between the current price and the price a certain number of periods ago. It can be used to spot overbought and oversold conditions.

	date	Price	Open	High	Low	Vol.	Change..	convertval	Sentiment	Price.up	RSI	EMAcross	MACD	WPR	ADX	CCI	CMO	ROC
1	5/17/2019	5.16	5.15	5.18	5.13	3.39M	0.58%	neutral	0	1	NA	NA	NA	NA	NA	NA	NA	NA
2	5/16/2019	5.13	5.14	5.16	5.13	4.66M	0.00%	neutral	0	0	NA	NA	NA	NA	NA	NA	NA	NA
3	5/15/2019	5.13	5.18	5.21	5.1	7.88M	-0.58%	neutral	0	0	NA	NA	NA	NA	NA	NA	NA	-0.00583
4	5/14/2019	5.16	5.07	5.19	5.04	14.85M	0.78%	neutral	0	1	25	NA	NA	NA	NA	NA	NA	0.005831
5	5/13/2019	5.12	5.14	5.17	5.07	9.40M	0.00%	positive	1	0	54.71698	0.004	NA	NA	NA	NA	NA	-0.00195
6	5/10/2019	5.12	5.14	5.16	5.11	8.45M	-0.19%	neutral	0	0	54.71698	0.002667	NA	NA	NA	NA	NA	-0.00778
7	5/9/2019	5.13	5.2	5.2	5.12	11.43M	-1.35%	neutral	0	0	74.33155	0.041778	NA	NA	NA	NA	NA	0.001951
8	5/8/2019	5.2	5.18	5.22	5.17	18.91M	0.00%	positive	1	0	61.0989	0.014519	NA	NA	NA	NA	NA	0.015504
9	5/8/2019	5.2	5.18	5.22	5.17	18.91M	0.00%	positive	1	0	61.0989	0.009679	NA	NA	NA	NA	NA	0.013553
10	5/8/2019	5.2	5.18	5.22	5.17	18.91M	0.00%	positive	1	0	61.0989	0.006453	NA	NA	NA	NA	NA	0
11	5/7/2019	5.2	5.21	5.24	5.19	13.70M	0.00%	neutral	0	0	79.53905	0.024302	NA	NA	NA	NA	NA	0
12	5/6/2019	5.2	5.21	5.24	5.18	9.07M	-0.76%	neutral	0	0	79.53905	0.016201	NA	NA	NA	NA	NA	0
13	5/3/2019	5.24	5.25	5.26	5.22	14.37M	0.19%	neutral	0	1	91.55232	0.037467	NA	NA	NA	NA	NA	0.007663
14	5/2/2019	5.23	5.27	5.29	5.21	10.41M	-0.76%	positive	1	0	94.13498	0.038312	NA	0.083333	NA	102.5641	NA	0.005753
15	4/30/2019	5.27	5.21	5.28	5.2	14.75M	1.35%	neutral	0	1	39.62316	-0.01446	NA	0	17.20588	141.5858	40.74074	0.005709
16	4/29/2019	5.2	5.19	5.24	5.17	7.58M	0.19%	positive	1	1	30.72657	-0.02297	NA	0.466667	15.4888	24.8227	22.58065	-0.00575
17	4/29/2019	5.2	5.19	5.24	5.17	7.58M	0.19%	positive	1	1	30.72657	-0.01532	NA	0.466667	14.40509	18.60941	22.58065	-0.01337
18	4/29/2019	5.2	5.19	5.24	5.17	7.58M	0.19%	neutral	0	1	30.72657	-0.01021	NA	0.466667	13.39573	14.23729	14.28571	0
19	4/29/2019	5.2	5.19	5.24	5.17	7.58M	0.19%	positive	1	1	30.72657	-0.00681	NA	0.466667	12.45581	2.24359	33.33333	0
20	4/26/2019	5.19	5.2	5.24	5.18	6.53M	0.00%	negative	-1	0	55.83014	0.002129	NA	0.571429	11.69815	-52.4345	28	-0.00192
21	4/26/2019	5.19	5.2	5.24	5.18	6.53M	0.00%	positive	1	0	55.83014	0.001419	NA	1	10.97895	-75.8333	25	-0.00192
22	4/26/2019	5.19	5.2	5.24	5.18	6.53M	0.00%	positive	1	0	55.83014	0.000946	NA	1	10.29719	-71.5746	-5.88235	0
23	4/26/2019	5.19	5.2	5.24	5.18	6.53M	0.00%	neutral	0	0	55.83014	0.000631	NA	1	9.651732	-67.4699	-5.88235	0
24	4/26/2019	5.19	5.2	5.24	5.18	6.53M	0.00%	positive	1	0	55.83014	0.000421	NA	1	9.0414	-63.5108	-5.88235	0
25	4/26/2019	5.19	5.2	5.24	5.18	6.53M	0.00%	positive	1	0	55.83014	0.00028	NA	1	8.464939	-59.6899	-5.88235	0
26	4/25/2019	5.19	5.25	5.3	5.17	14.54M	-1.14%	neutral	0	0	97.95877	0.03352	NA	1	13.34575	-56	-5.88235	0

Figure 8: Calculated Technical Indicators Merged with Sentiment Polarity

The calculated technical indicators are then joined with the sentiment polarity, using the date as the key value. The result of the joined data is in Figure 8.

A	B	C	D	E	F	G	H	I	J
Sentiment	RSI	EMAcross	MACD	WPR	ADX	CCI	CMO	ROC	Price.up
1	7.217656	-0.0158358	-0.10276	1	7.137313	-116.667	-50	0	1
-1	7.217656	-0.0105572	-0.20259	1	6.85279	-93.3333	-50	0	1
-1	7.217656	-0.0070381	-0.28853	1	6.570706	-77.7778	-50	0	1
0	1.019636	-0.0380254	-0.37629	1	6.035593	-85.2713	-53.8462	-0.00395	0
0	56.74116	0.00798305	-0.44412	1	5.663088	-74.6667	-53.8462	-0.00395	0
0	67.66355	0.01865537	-0.48863	1	8.514756	-67.1345	-53.8462	0	0
0	56.89039	0.00577024	-0.51821	0.95	10.9185	-50.5249	-48.1481	0.001978	0
0	56.89039	0.00384683	-0.53571	0.916667	10.68004	-47.4576	-90.4762	0.001978	0
0	68.26062	0.00923122	-0.54037	0.916667	14.37004	-45.2196	-84.6154	0	0
0	77.25798	0.01282081	-0.53235	0.75	13.72078	37.33333	-60	0.003945	0
0	84.04309	0.01521388	-0.51247	0	14.24413	176.2963	-50	0.005911	0
0	84.04309	0.01014258	-0.48502	0	14.756	146.4052	33.33333	0.001967	0
0	84.04309	0.00676172	-0.45315	0	15.25542	207.4074	50	0.003922	1
0	94.70551	0.01784115	-0.4129	0.166667	14.14171	118.9542	33.33333	0.001963	0
0	94.70551	0.0118941	-0.36855	0	13.27216	152.2807	45.45455	0.001955	0
0	37.82997	-0.0054039	-0.32939	0	12.60445	122.5806	45.45455	0.003914	1
0	37.82997	-0.0036026	-0.29478	0	11.95665	101.9608	45.45455	0	1
-1	37.82997	-0.0024017	-0.26416	0	11.32958	89.74359	77.77778	0	1
0	89.75302	0.03173217	-0.22142	0.571429	15.06331	-31.1111	23.07692	-0.00784	0
0	59.79122	0.00782145	-0.17907	0	15.46631	154.1667	50	0.005842	1
1	23.89535	-0.0347857	-0.15752	0.444444	14.04478	29.88506	21.73913	0.005888	1
-1	59.9561	0.01680953	-0.13104	0.888889	12.36228	-104.762	3.703704	-0.01566	0
0	64.19722	0.01787302	-0.09943	0.375	15.81511	53.63985	18.75	0.001955	0
0	75.75327	0.03191535	-0.05671	0.375	15.91408	47.25738	13.33333	0.009814	0
0	94.47567	0.1612769	0.055991	0	25.64762	417.0833	48	0.040197	0
-1	95.02451	0.12085127	0.21215	0	28.88784	228.9638	48	0.040197	0
-1	96.73007	0.12723418	0.407771	0	27.14681	188.8041	51.85185	0.011194	0

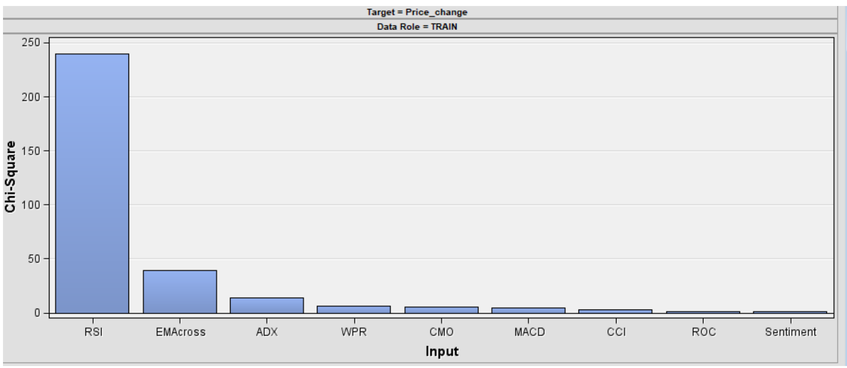
Figure 9: Pre-processed Data

Figure 9 shows the required data for data analysis. All the unwanted columns such as the date, day high, day low etc are removed.

These pre-processed data are used as input data for Enterprise Miner to perform data mining.

<u>Descriptive Statistics</u>

Before model building, preliminary investigation is done to understand the data. Association Chi Square test was performed to confirm the significance of the predictors with the price change as target variable.



Input	Chi-Square	Df	Prob
RSI	239.6085	4	<.0001
EMAcross	39.6085	4	<.0001
ADX	14.2829	4	0.0064
WPR	6.3814	4	0.1724
CMO	5.8673	4	0.2093
MACD	5.0263	4	0.2846
CCI	2.8684	4	0.5801
ROC	1.1744	2	0.5559
Sentiment	1.0684	2	0.5861

Figure 10: Results of Chi-Square Test from Enterprise Miner

From the results of Chi-Square Test, we can conclude that the predictors: RSI and EMA cross seem to have stronger association based on the Chi Square. The news sentiment polarity is the worst predictor amongst all.

DATA PREPARATION

The variable summary is given in Table 3. There is no missing values and outliers, with exception of few continuous predictors showing skewness and kurtosis.

Variable Summary

Role	Measurement Level	Frequency Count
INPUT	INTERVAL	8
INPUT	NOMINAL	1
TARGET	BINARY	1

Table 3: Variable Summary Result from SAS Enterprise Miner

Variable	Role	Mean	Standard Deviation	Non Missing	Missing	Minimum	Median	Maximum	Skewness	Kurtosis
ADX	INPUT	20.79033	6.519028	1018	0	5.857793	20.62401	37.22334	0.175192	-0.71716
CCI	INPUT	15.25296	112.5506	1018	0	-458.437	38.67403	347.7987	-0.46927	-0.27922
CMO	INPUT	8.554526	42.55501	1018	0	-91.6667	7.692308	100	-0.03895	-0.84768
EMAcross	INPUT	0.000206	0.077525	1018	0	-1.60075	0.00553	0.313265	-12.1482	220.2545
MACD	INPUT	-0.50262	7.253724	1018	0	-38.2203	0.66232	8.206001	-2.87205	10.88325
ROC	INPUT	0.000261	0.090509	1018	0	-1.69553	0.007968	0.231512	-12.9051	238.7987
RSI	INPUT	55.25043	29.12344	1018	0	1.168876	58.7125	99.14057	-0.22574	-1.2698
WPR	INPUT	0.427345	0.404522	1018	0	0	0.319149	1	0.254416	-1.63264

Table 4: Interval Variable Summary Statistics from Enterprise Miner

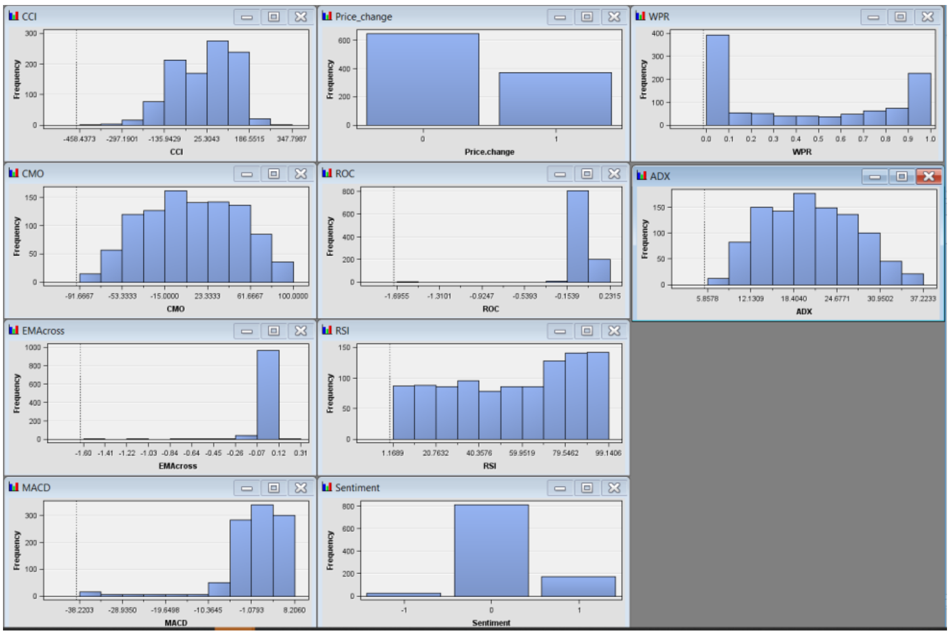


Figure 11: Skewness of the Variables from SAS Enterprise Miner

The skewness could have impact in the accuracy and performance of Stock Price Prediction. These independent variables need to be transformed to be more symmetrical to improve the model performance.

DATA TRANSFORMATION

ROC, EMAcross and MACD were transformed respectively in SAS Enterprise Miner.

Transformations Statistics

Source	Method	Variable Name	Formula	Skewness	Kurtosis
Input	Original	EMAcross		-11.8005	189.6673
Input	Original	MACD		-2.98947	11.9196
Input	Original	ROC		-12.4372	201.9485
Output	Computed	PWR EMAcross	(max(EMAcross--...	-1.24653	25.22826
Output	Computed	PWR MACD	(max(MACD--38.2...	-0.0429	-0.52186
Output	Computed	PWR ROC	(max(ROC--1.695...	-1.04318	11.86279

Table 5: Transformation Statistics from SAS Enterprise Miner

The value of skewness and kurtosis are reduced considerably, and this could significantly improve the performance of the system.

PREDICTIVE MODELING

Data partition is done prior to model building. 75% of data is used for training, 15% is used for validation while 10% is used for test. Decision Tree, Logistic Regression, Random Forest, Support Vector Machine(Linear) and Support Vector Machine(Quadratic) are built for the stock price prediction.

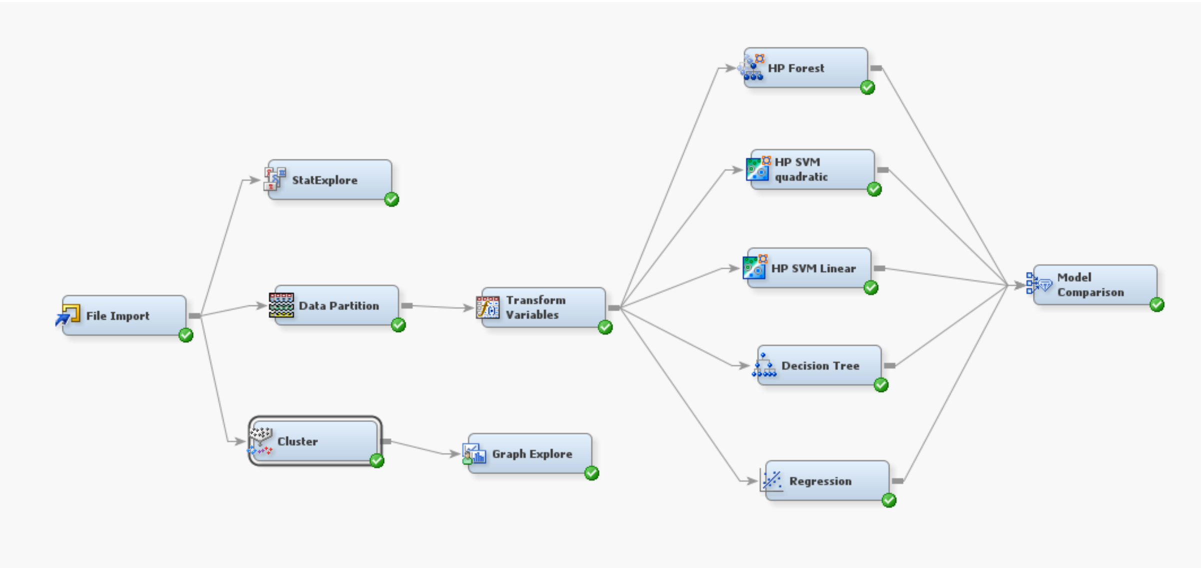
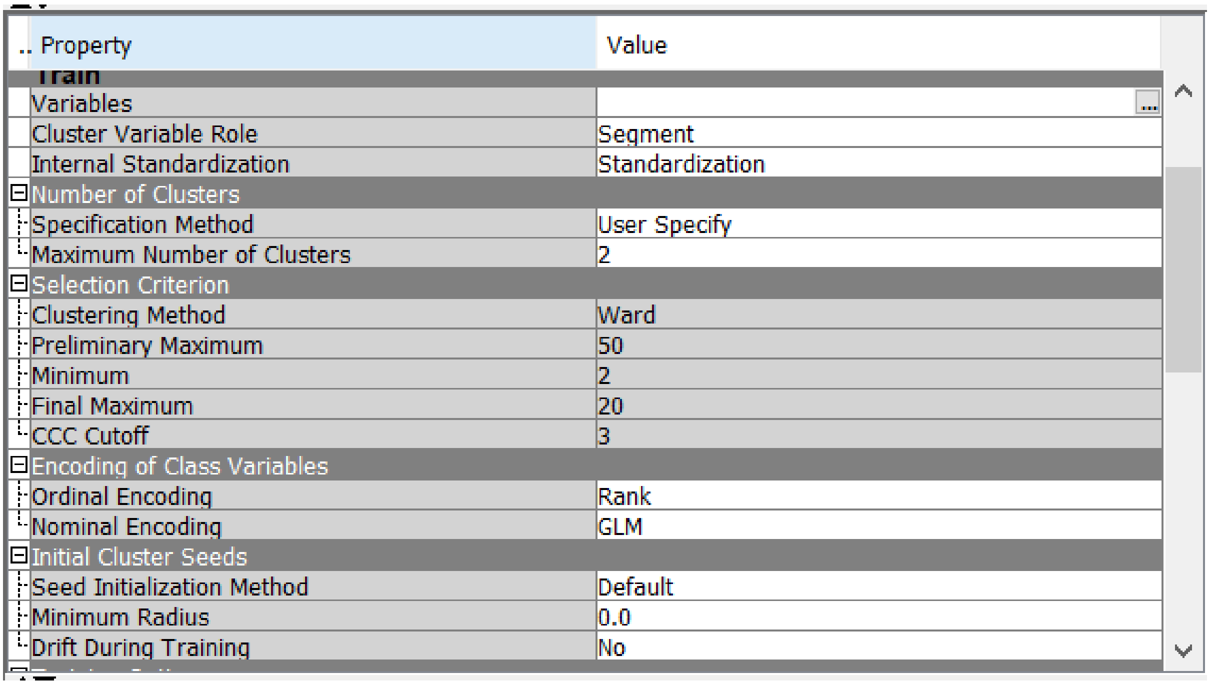


Figure 12: Models Built in SAS Enterprise Miner

CLUSTERING

Since there are only 4 stocks to perform clustering, only 2 cluster is sufficient.



Property	Value
Train	
Variables	...
Cluster Variable Role	Segment
Internal Standardization	Standardization
Number of Clusters	
Specification Method	User Specify
Maximum Number of Clusters	2
Selection Criterion	
Clustering Method	Ward
Preliminary Maximum	50
Minimum	2
Final Maximum	20
CCC Cutoff	3
Encoding of Class Variables	
Ordinal Encoding	Rank
Nominal Encoding	GLM
Initial Cluster Seeds	
Seed Initialization Method	Default
Minimum Radius	0.0
Drift During Training	No

Figure 13: Configuration for Clustering

	ROA	ROE	Gross Margin	Earning Per Share	Operating Profit Margin
Lionind	2.08	3.58	14.9	9	6.8
Ssteel	1.92	4.94	7.8	10.34	2.1
Masteel	0.5	1.03	5.9	1	1.4
Annjoo	3.16	6.23	13.5	14.56	8

Table 6: Data Imported for Clustering

Results

Machine Learning

Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Reg	Regression	0.13245	0.12813	0.17955	0.09961
	HPSVM2	HP SVM quadratic	0.14570	0.15228	0.15858	0.14928
	HPSVM	HP SVM Linear	0.15232	0.15182	0.18349	0.14017
	Tree	Decision Tree	0.16556	0.11761	0.15203	0.12555
	HPDMForest	HP Forest	0.17881	0.01868	0.00000	0.12388

Table 7: Steel Industry Model Comparison from SAS Enterprise Miner

Fit Statistics
Model Selection based on Valid: Misclassification Rate (_VMISC_)

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Tree	Decision Tree	0.10638	0.14254	0.18803	0.09963
	HPDMForest	HP Forest	0.12766	0.02452	0.00000	0.09762
	Reg	Regression	0.12766	0.12823	0.19231	0.11172
	HPSVM2	HP SVM quadratic	0.14894	0.15009	0.17949	0.14846
	HPSVM	HP SVM Linear	0.17021	0.15120	0.17949	0.14355

Table 8: Banking Industry Model Comparison from SAS Enterprise Miner

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Tree	Decision Tree	0.12121	0.11890	0.15758	0.10754
	HPDMForest	HP Forest	0.15152	0.02837	0.00000	0.09572
	HPSVM2	HP SVM quadratic	0.18182	0.16134	0.19394	0.15253
	Reg	Regression	0.21212	0.13673	0.20606	0.15355
	HPSVM	HP SVM Linear	0.24242	0.16482	0.26667	0.15881

Table 9: Technology Industry Model Comparison from SAS Enterprise Miner

Selected Model	Model Node	Model Description	Valid: Misclassification Rate	Train: Average Squared Error	Train: Misclassification Rate	Valid: Average Squared Error
Y	Tree	Decision Tree	0.18519	0.15547	0.19403	0.15107
	HPSVM	HP SVM Linear	0.18519	0.17012	0.22761	0.15585
	HPSVM2	HP SVM quadratic	0.20370	0.16190	0.21269	0.15701
	HPDMForest	HP Forest	0.22222	0.02272	0.00000	0.13947
	Reg	Regression	0.24074	0.14112	0.21642	0.12561

Table 10: Energy Industry Model Comparison from SAS Enterprise Miner

Clustering

Distance to Nearest Cluster	Earning Per Share	Gross Margin	Operating Profit Margin	ROA	ROE
3.602921	11.3	12.06667	5.633333	2.386667	4.916667
3.602921	1	5.9	1.4	0.5	1.03

Table 11: Cluster Centers in Steel Industry

Companies	ROA	ROE	Gross Margin	Earning Per Share	Operating Profit Margin	Segment Id
Lionind	2.08	3.58	14.9	9	6.8	1
Ssteel	1.92	4.94	7.8	10.34	2.1	1
Masteel	0.5	1.03	5.9	1	1.4	2
Annjoo	3.16	6.23	13.5	14.56	8	1

Table 12: Segment ID – Cluster that Steel Companies Belong to

Compare Output with Hypothesis

The hypothesis stays true. Each industry/stock has its own choice of machine learning algorithm which is the best fit for stock price up/down prediction.

Discussion

We selected stocks of companies from each industry to train and test the system. Steel industry has Regression being the best machine learning algorithm for stock price up/down prediction. Banking Industry, Technology Industry and Energy Industry have Decision Tree being the machine algorithm which is the best fit for stock price up/down prediction, albeit different accuracies among them.

Summary and Conclusion

A particular Machine Learning Algorithm may be better suited to a particular stock, say Energy Stock, whereas the same algorithm may give lower accuracies while predicting other type of stock, say Technology Stock.

Recommendation for Future Studies

More stocks to be used as sample test the machine learning algorithm and clustering.

In this project, the machine learning algorithm only able to predict stock price for the next day. In future, a hybrid machine learning model involving both Fundamental and Technical Analysis can be used to predict stock price.

Youtube Video: https://www.youtube.com/watch?v=elXcd3yue5s&feature=youtu.be