

GROUP 8

STOCKS' PROFITABILITY PREDICTION BASED ON FINANCIAL RATIOS



GROUP MEMBERS



- Lam Nhut Thinh K194141751
- Thai Tuan Kha K194141725
- Nguyen Tuan Hung K194141723
- Ho Ngoc Quynh Nhu K194141738
- Tran Thanh Truc K194141755

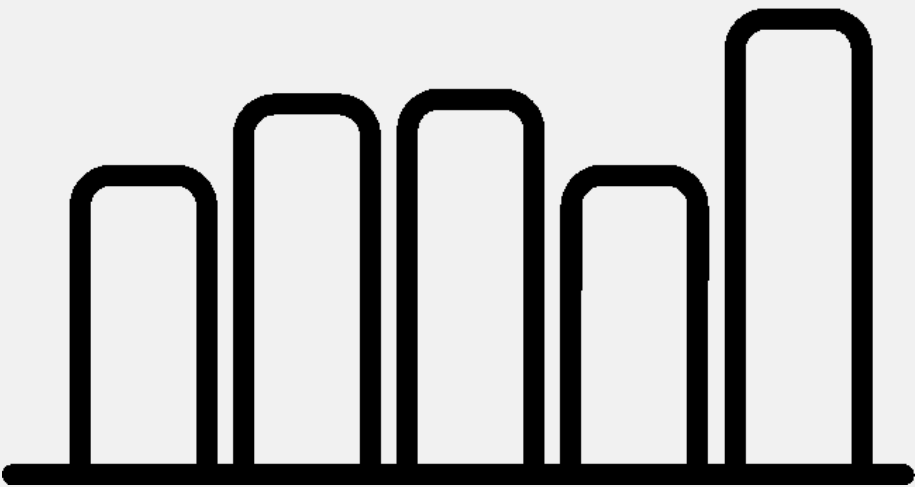
1. BACKGROUND

- Cash flow in the financial market is congested in many different investment channels.
- Stock market 2021 rises and sets many new records.
- Financial statement is an important basis for investors to refer to and evaluate the profitability of a company's stock.



2. DATA PROCESSING

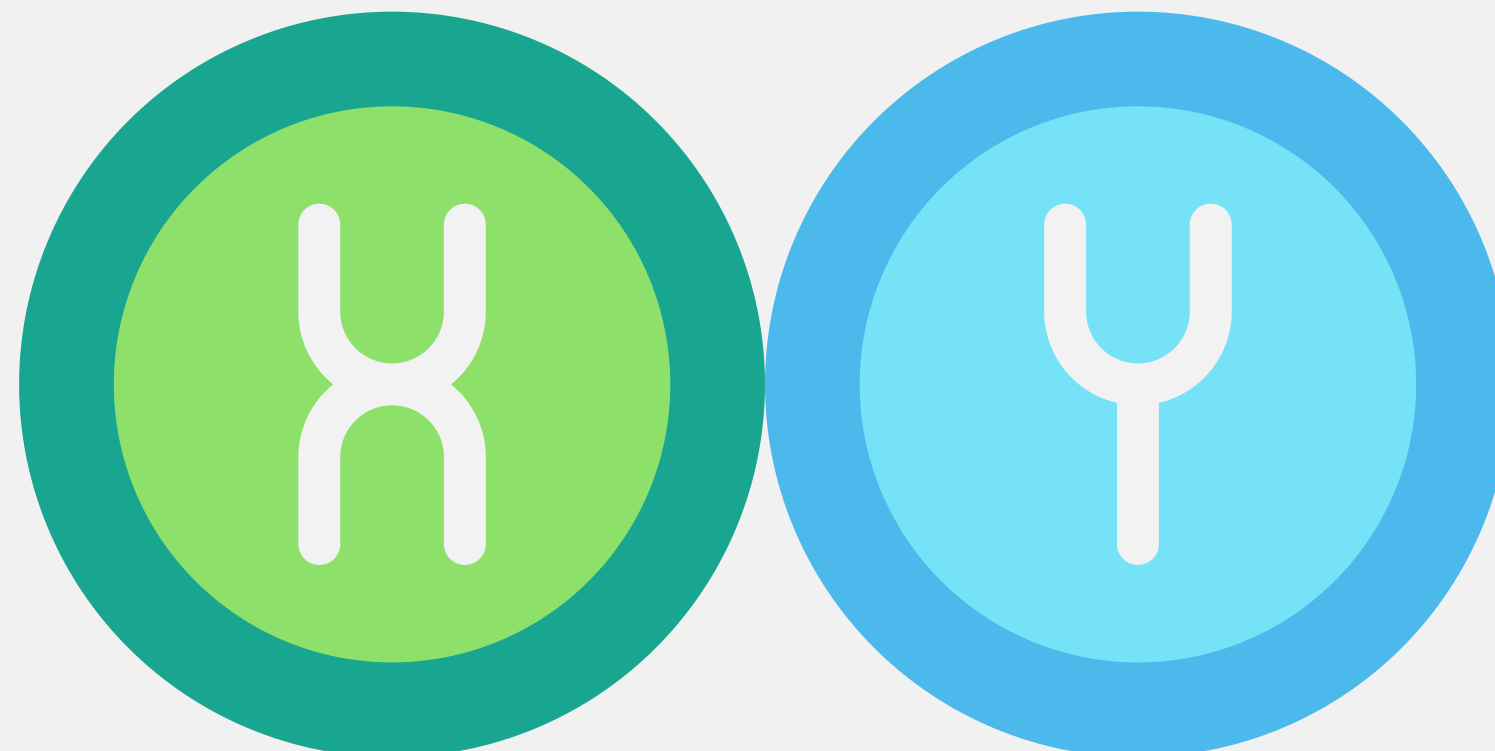
48 financial ratios, 447 companies in Vietnam.



	Research And Development	Total Operating Expense	Operating Income	Interest Expense, Net - Operating	Income Tax - Total	Net Income Incl Extra Before Distributions	Preferred Dividends	Basic Normalized EPS	Diluted Normalized EPS	Basic Weighted Average Shares	...	Cas Ope Act Cum
Name												
CHP.HM	NaN	5.329274e+11	2.545201e+11	1.158986e+11	1.241771e+10	2.420746e+11	NaN	1647.40909	1647.40909	146912670.0	...	3.10915
POM.HM	NaN	1.378265e+13	2.176012e+11	3.947202e+11	3.336702e+10	1.822244e+11	NaN	642.65118	642.65118	278543550.0	...	-4.61051
TDC.HM	NaN	1.484144e+12	1.740635e+11	1.546102e+11	5.008222e+10	1.207732e+11	NaN	1901.05222	1901.05222	100000000.0	...	-4.34649
OGC.HM	NaN	3.493755e+11	6.046720e+10	1.372862e+10	3.915555e+09	1.445611e+11	NaN	482.00000	482.00000	299919350.0	...	-1.34000
NVT.HM	NaN	1.925600e+11	-6.593012e+10	1.649418e+10	-1.715912e+09	-5.311811e+10	NaN	-586.94047	-586.94047	90500000.0	...	2.59306
...
SVI.HM	NaN	1.722265e+12	1.149172e+11	6.010629e+09	2.321289e+10	9.211948e+10	NaN	7441.51804	7441.51804	12832440.0	...	1.20609
BVH.HM	NaN	4.316567e+13	-5.587410e+12	NaN	3.730695e+11	1.896699e+12	NaN	2553.11463	2553.11463	742322760.0	...	1.23216
LIX.HM	NaN	2.402414e+12	2.105892e+11	4.122230e+09	4.365618e+10	1.673662e+11	NaN	4636.13657	4636.13657	32400000.0	...	2.13004
VNG.HM	NaN	1.791683e+11	3.310400e+09	6.585747e+10	3.473361e+09	1.905561e+09	NaN	-1177.22642	-1177.22642	97276610.0	...	-1.75041
BTT.HM	NaN	6.594758e+10	2.363102e+09	NaN	4.515149e+08	2.777326e+09	NaN	176.08342	176.08342	13500000.0	...	4.31694

447 rows × 48 columns

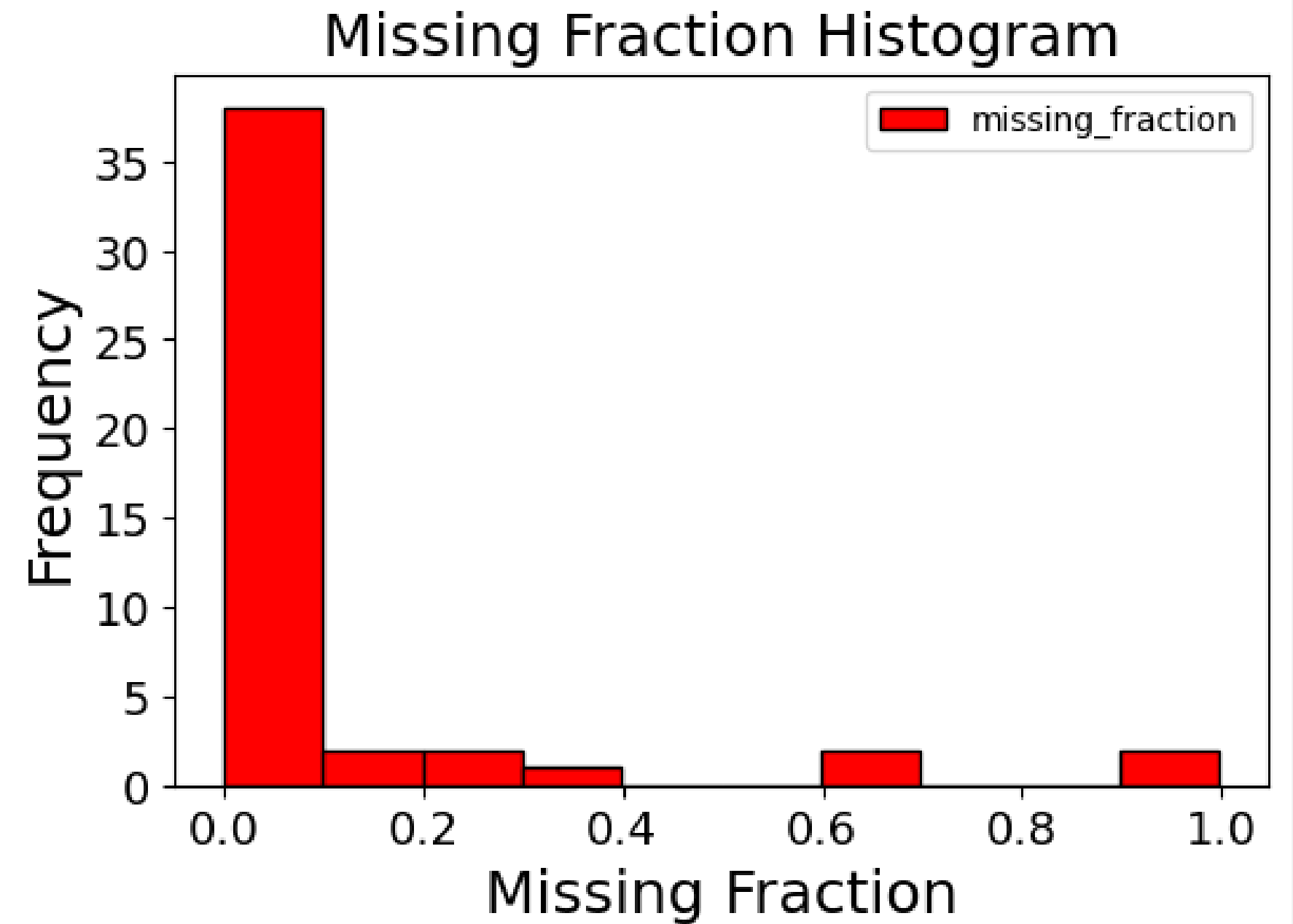
Cash from Operating Activities, Cumulative	Capital Expenditures, Discrete	Cash from Investing Activities, Cumulative	Cash from Financing Activities, Cumulative	FOCF Ex Dividends, Discrete	Net Change in Cash, Cumulative	Pretax Margin, Percent	Operating Profit Margin	Net Profit Margin, (%)	Target
3.109158e+11	8.651382e+09	-4.514561e+09	-4.850702e+11	3.022644e+11	-1.786692e+11	32.31864	NaN	30.74169	1
-4.610515e+11	9.654058e+11	-9.335639e+11	1.725715e+12	-1.426457e+12	3.309851e+11	1.54209	NaN	1.30158	1
-4.346490e+10	3.966927e+09	-1.146927e+09	8.166955e+10	-4.743183e+10	3.705772e+10	10.52129	NaN	6.11606	1
-1.340004e+11	7.905459e+10	-2.804611e+09	1.792958e+10	-2.130550e+11	-1.188754e+11	26.53096	0.0	35.27235	1
2.593067e+10	5.226709e+09	-2.973585e+11	2.878679e+11	2.070396e+10	1.647019e+10	-52.77194	NaN	-41.94755	1
...
1.206092e+11	8.226478e+09	1.720859e+10	-1.056608e+11	1.123828e+11	3.208050e+10	6.27768	NaN	5.01417	0
1.232165e+13	2.377686e+11	-1.428670e+13	5.389117e+11	1.208388e+13	-1.426650e+12	6.32411	0.0	5.04733	0
2.130042e+11	5.415793e+10	-5.364795e+10	-3.240000e+10	1.588462e+11	1.269582e+11	8.07585	NaN	5.75278	0
-1.750412e+11	5.229407e+09	-4.999060e+10	1.350825e+11	-1.802706e+11	-8.994928e+10	3.61356	NaN	1.04427	0
4.316940e+08	8.826258e+09	1.692390e+10	-9.380705e+09	-8.394564e+09	7.973224e+09	4.72670	NaN	4.06573	0



- Dependent variable: The year-to-date percentage change in the stocks' prices calculated from the closing prices at the beginning of the year and year-end closing prices. Declining stocks will be assigned a value of 0, increasing stocks will be assigned a value of 1.
- Independent variables: The indicators in financial statements.

2.1 REMOVE NULL VALUES

- There are some columns where most of the data is null.
- Filter and drop all columns where more than 10% of the data is null.
- Then we can use `dropna()` function to eliminate all null values.



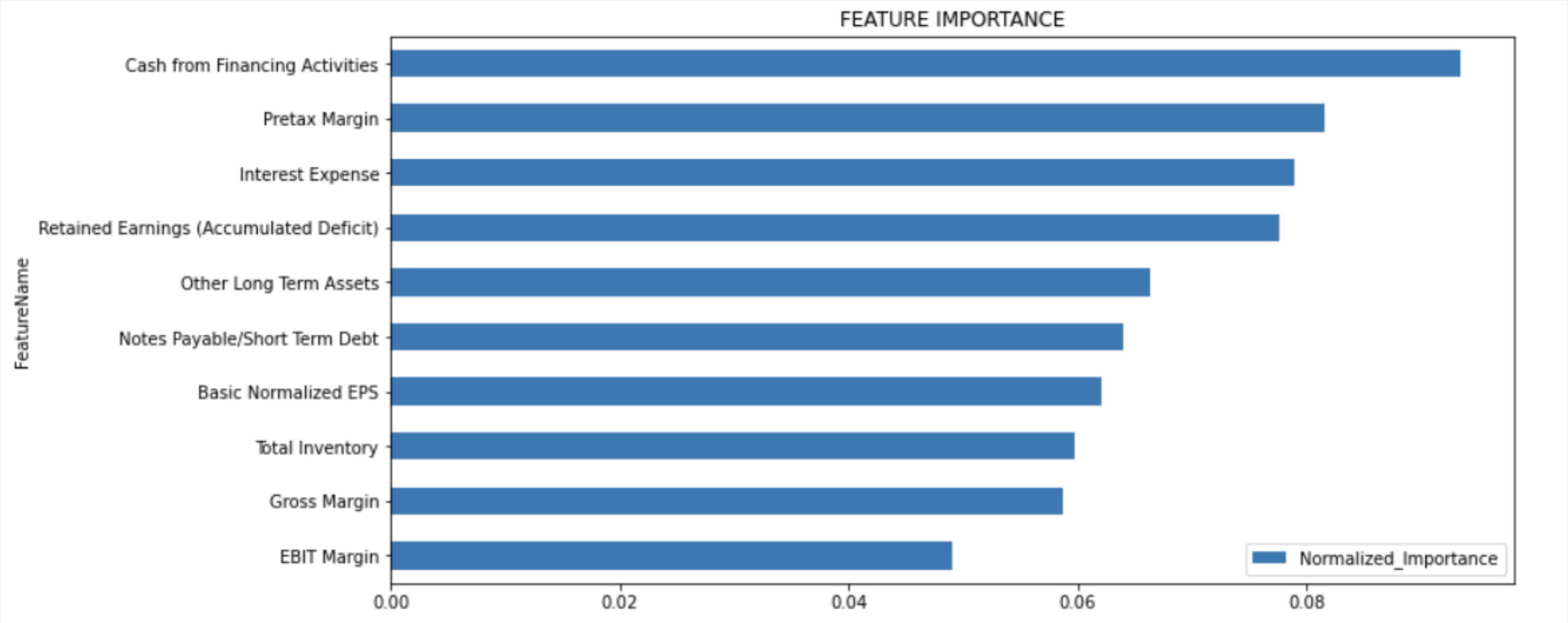
2.2 COLLINEAR FEATURES PROCESSING



Eliminate variables that have the correlation with some other variables greater than 0.9

2.3 FEATURES SELECTION

Calculate the Feature Importance Score of each independent variables and choose the top 10 highest.

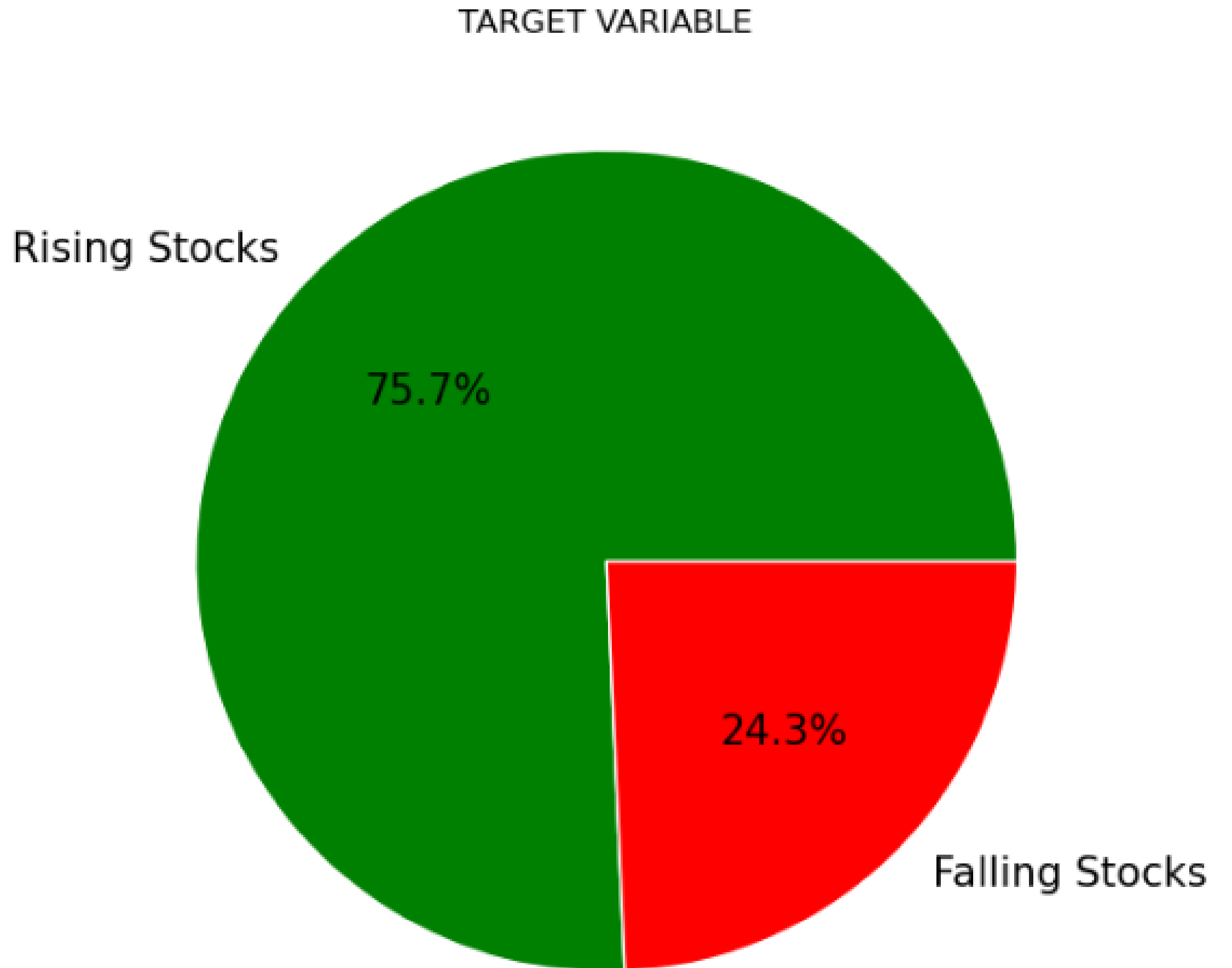


FEATURES IN USE

- Cash from Financing Activities
- Pretax Margin
- Interest Expense
- Retain Earnings
- Other Long Term Assets
- Notes Payable/Short Term Debt
- Basic Normalized EPS
- Total Inventory
- Gross Margin
- EBIT Margin

3. EDA

The amount of stocks with positive return (class = 1) is about 3 times greater than the number of stocks with negative return (class = 0).

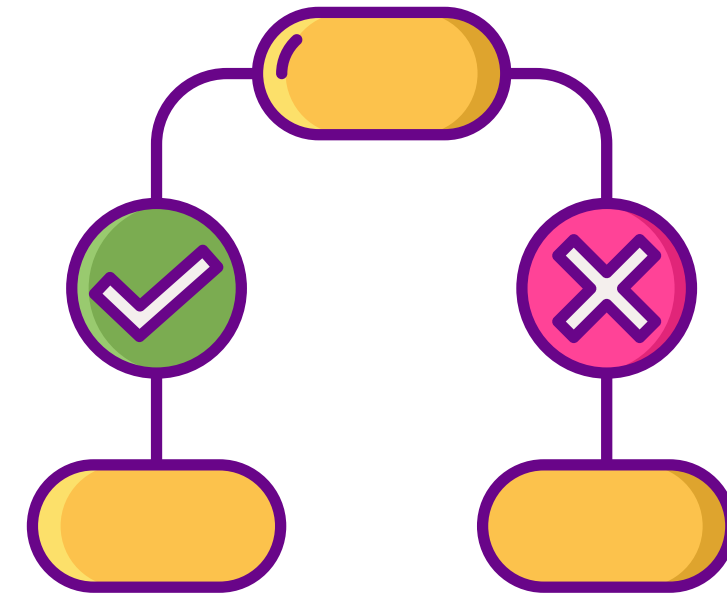


BASIC STATISTICS OF THE INDEPENDENT VARIABLES

	Cash from Financing Activities	Pretax Margin	Interest Expense	Retained Earnings (Accumulated Deficit)	Other Long Term Assets	Notes Payable/Short Term Debt	Basic Normalized EPS	Total Inventory	Gross Margin	EBIT Margin
count	3.460000e+02	346.000000	3.460000e+02	3.460000e+02	3.460000e+02	3.460000e+02	346.000000	3.460000e+02	346.000000	346.000000
mean	4.704137e+11	11.160045	2.110400e+11	9.176588e+11	4.558907e+11	1.054664e+12	2207.943128	2.174935e+12	20.531247	9.054661
std	2.923238e+12	24.972108	1.015217e+12	3.866282e+12	2.591530e+12	3.232403e+12	2994.026831	8.360467e+12	18.938225	19.014343
min	-5.257398e+12	-177.294820	0.000000e+00	-9.328983e+12	8.339020e+06	0.000000e+00	-8647.635000	0.000000e+00	-43.564280	-64.043316
25%	-1.119880e+11	2.808010	5.005780e+09	3.956139e+10	8.714716e+09	2.747669e+10	570.848185	7.952482e+10	8.574710	2.350346
50%	-2.191015e+09	7.998900	2.585647e+10	1.513749e+11	3.308450e+10	2.055295e+11	1540.379765	3.323958e+11	17.120680	7.219609
75%	2.685765e+11	17.753385	9.568722e+10	4.473369e+11	1.529502e+11	6.464484e+11	3066.345740	1.038944e+12	30.123848	14.669197
max	2.567202e+13	162.753580	1.028889e+13	5.625940e+13	4.252726e+13	3.754701e+13	25561.348910	1.101569e+14	96.341580	69.732416

MODELS IN USE

DECISION TREE CLASSIFIER



RANDOM FOREST CLASSIFIER

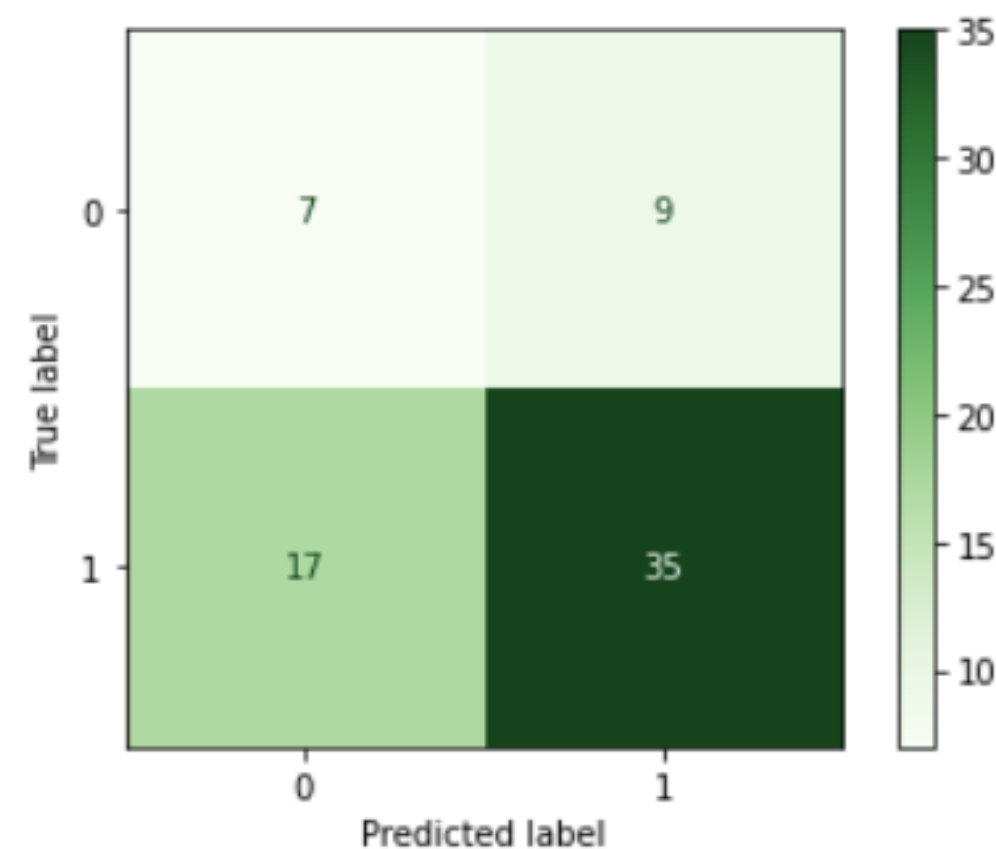


4. RESULTS

DECISION TREE CLASSIFIER REPORT

	precision	recall	f1-score	support
0	0.29	0.44	0.35	16
1	0.80	0.67	0.73	52
accuracy			0.62	68
macro avg	0.54	0.56	0.54	68
weighted avg	0.68	0.62	0.64	68

Decision Tree Accuracy: 62.0%



- Accuracy in the prediction of the decision tree model is quite low at only 62%.
- In addition, most of the stocks that fell (class = 0) were omitted in the prediction results. So recall at this class is very low, only 0.44.

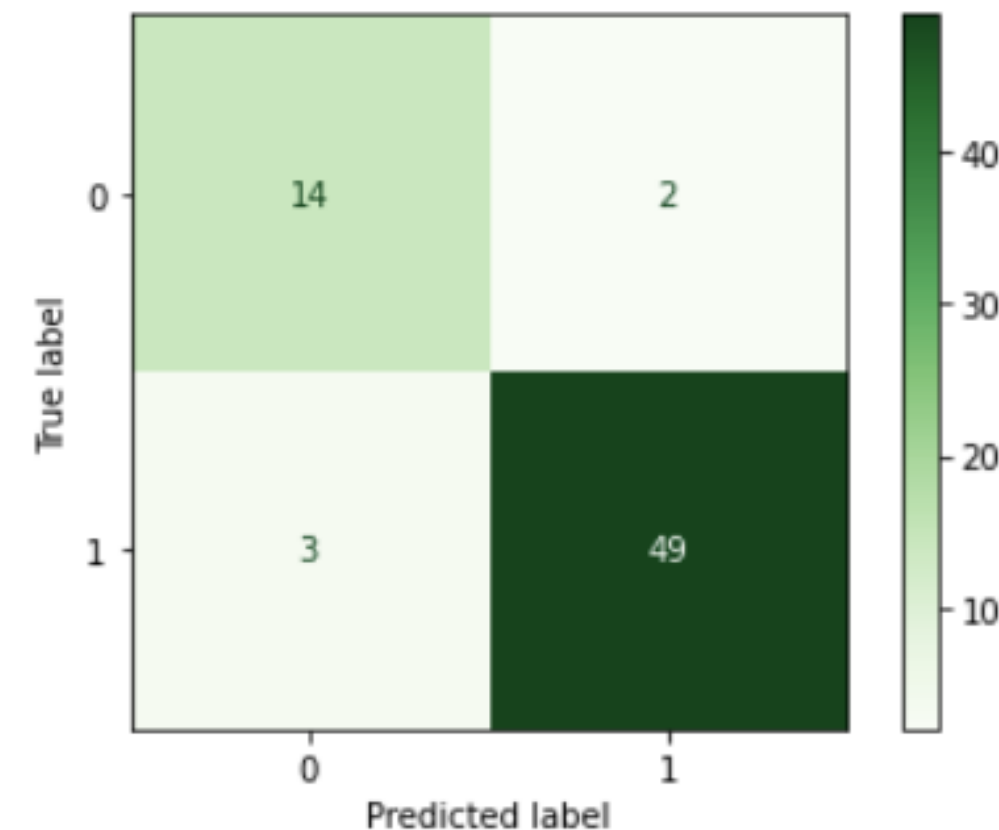
4. RESULTS

RANDOM FOREST CLASSIFIER REPORT

- Accuracy in the prediction of the random forest model is very good at 93%.
- The stocks in each class are forecasted quite accurately, without the "missing" phenomenon as in the decision tree model. So precision and recall are both very high.
- The random forest model shows superior predictive performance compared to the decision tree model.

	precision	recall	f1-score	support
0	0.82	0.88	0.85	16
1	0.96	0.94	0.95	52
accuracy			0.93	68
macro avg	0.89	0.91	0.90	68
weighted avg	0.93	0.93	0.93	68

Random Forest Accuracy: 93.0%





5. CONCLUSION

In finance, financial indicators are not the decisive factor for 100% of investment results, but also depend on many factors such as capital management, money movement, ... Therefore, the error in the model is inevitable.

More importantly, there is no such thing as the holy grail of investing. The model only helps investors to be provided with more perspectives. From having more objective information to making decisions, it is certainly wiser and less emotional to make decisions.

