Stocks Selection Using Neural Network Based on Financial Ratios

Thanasit Chiaravanont

Advanced Machine Learning

Objective

This project explores the application of Neural Networks to stock selection using financial ratios, leveraging data from the S&P 500 index over a 13-year period (2010–2023). The dataset, sourced from Bloomberg, includes 15 quarterly financial ratios and stock prices.

Prediction

- BUY recommendation (classification)
- SELL recommendation (classification)
- Stock return prediction (regression)

Data

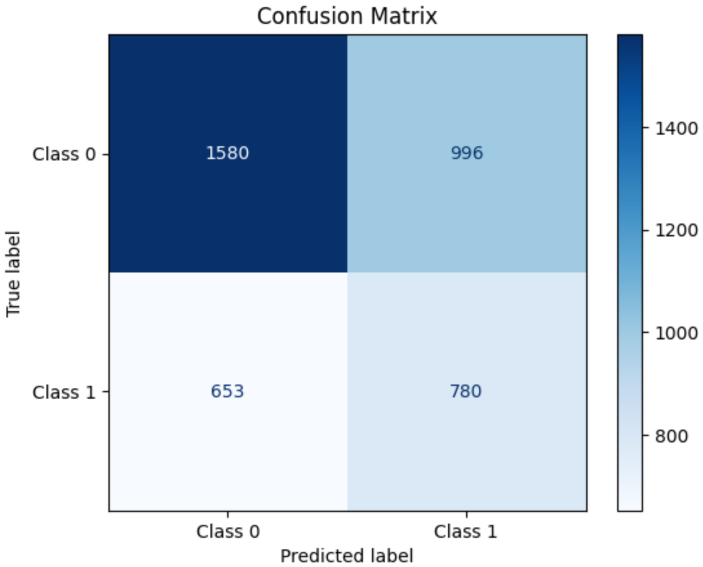
- Original data; 26,730 rows
- After drop all missing value; 13,361 rows
- Stock return; $r_{i,t} = \ln(\frac{P_{i,t}}{P_{i,t-1}})$
- I classified the target variable into stocks to BUY and stocks to SELL, using a cut-off threshold of 5% (or -5% for a SELL action) in price return over the next 3-month period, where a value of 1 indicates a stock return exceeding 5% (or below -5%), and a value of 0 otherwise.
- I observed that the data is slightly imbalanced, with 30% of instances classified as 1 and 70% as 0.
- A false positive is more costly than a false negative because a false positive prediction could result in a direct loss of investment.

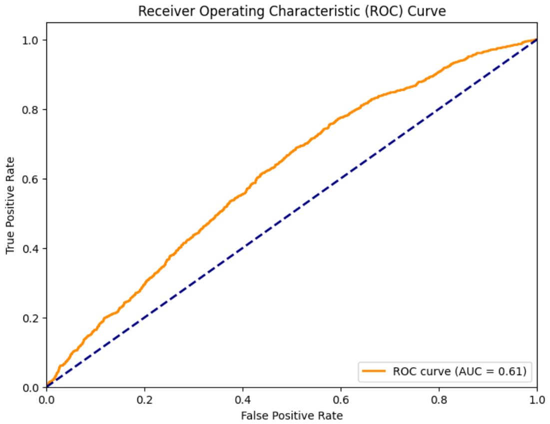
Model building

Model	1	2	3	4	5	6	7	8	9
Class	Binary	Binary	Binary	Binary	Binary	Binary	Cont.	Cont.	Cont.
Type	Buy	Buy	Buy	Sell	Sell	Sell	Return	Return	Return
Hidden Layers	1	4	4	1	4	4	1	4	4
Nodes	32	32 64 128 64	64 128 256 128	32	32 64 128 64	64 128 256 128	32	32 64 128 64	64 128 256 128
Matrices	Accuracy Precision Recall	Accuracy Precision Recall	Accuracy Precision Recall	Accuracy Precision Recall	Accuracy Precision Recall	Accuracy Precision Recall	Mae RMSE R-sq	Mae RMSE R-sq	Mae RMSE R-sq
Activation function	Sigmoid	Sigmoid	Sigmoid	Sigmoid	Sigmoid	Sigmoid	-	-	-
Optimizer	adam	adam	adam	adam	adam	adam	adam	adam	adam
Learning Rate			Adj			Adj			Adj
Data Augmentation	L2 Dropout	L2 Dropout	L2 Dropout Batch normalize	L2 Dropout	L2 Dropout	L2 Dropout Batch normalize	L2 Dropout	L2 Dropout	L2 Dropout Batch normalize

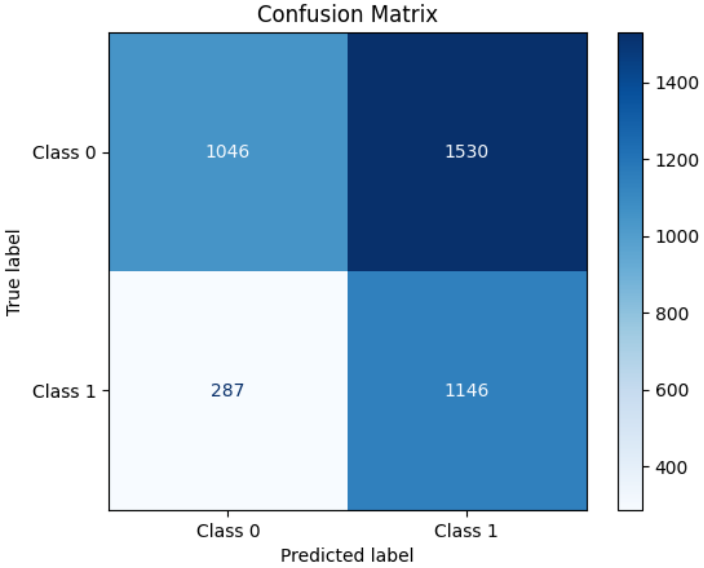
Result
BUY-Recommendation model

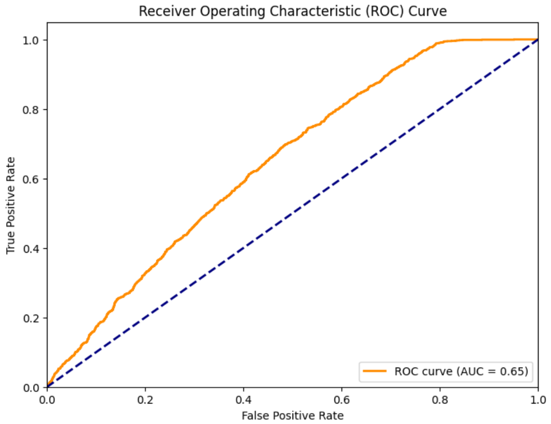
Baseline model



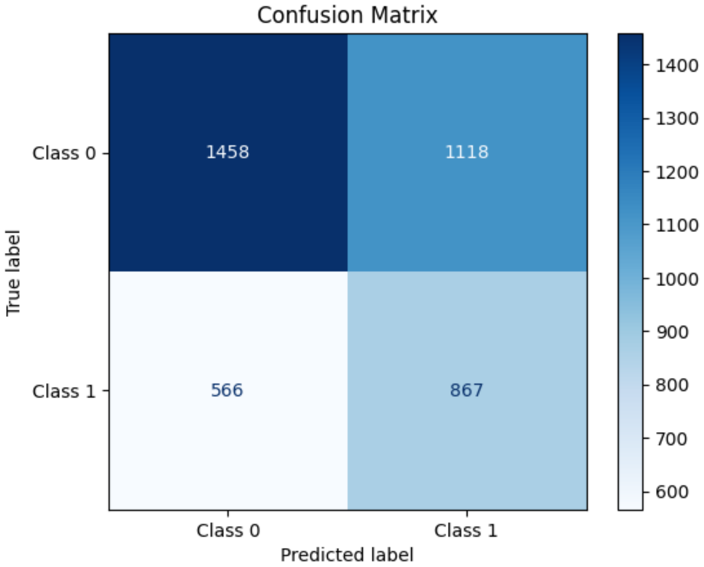


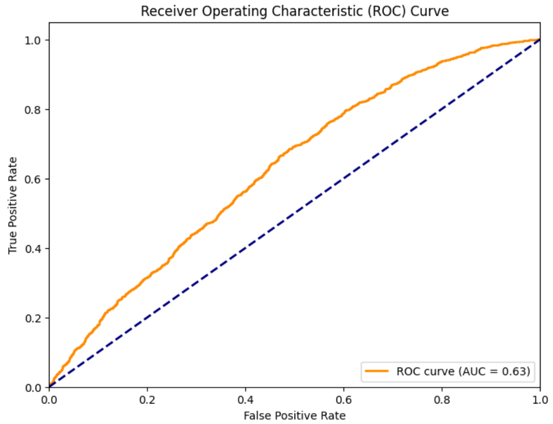
Model-2





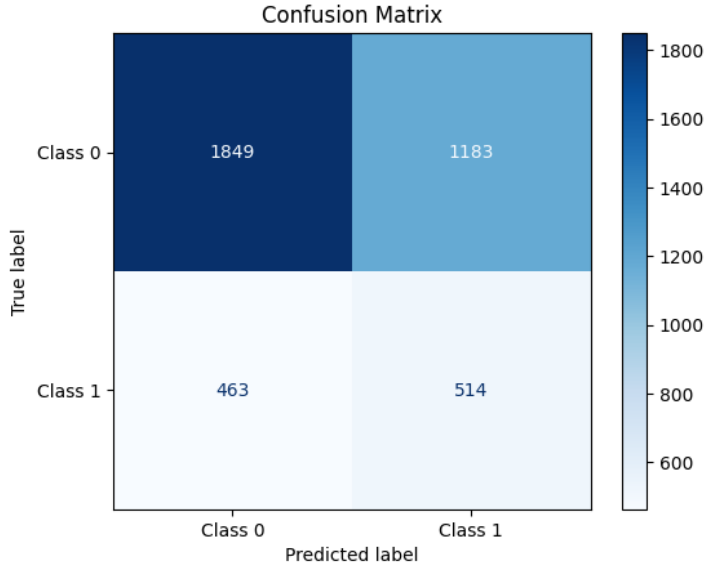
Model-3

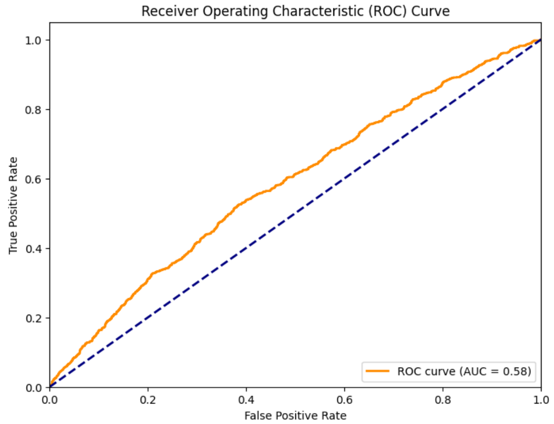




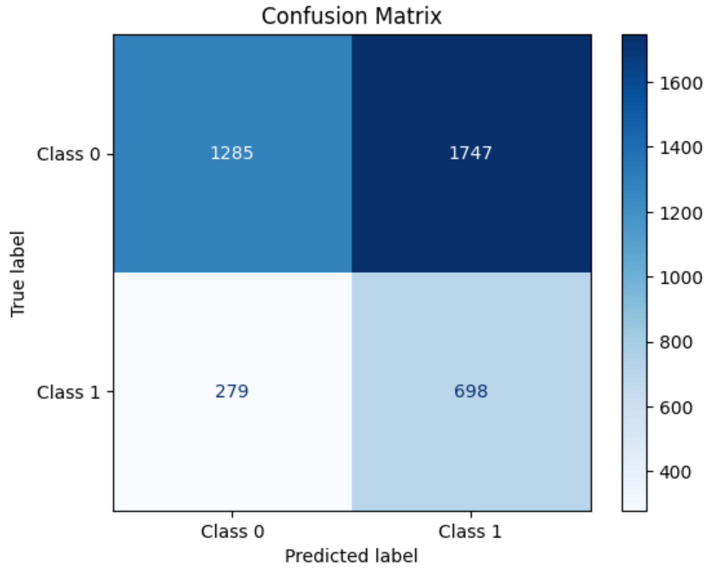
Result
SELL-Recommendation model

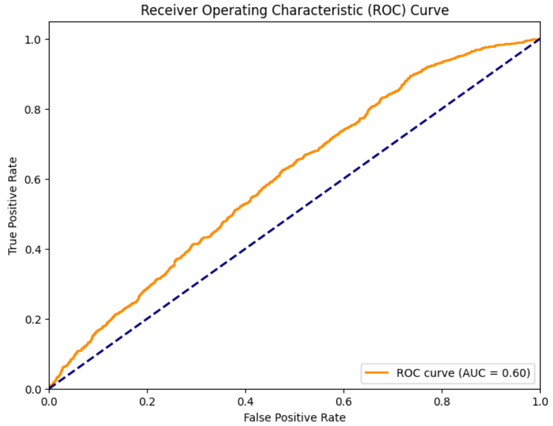
Baseline model



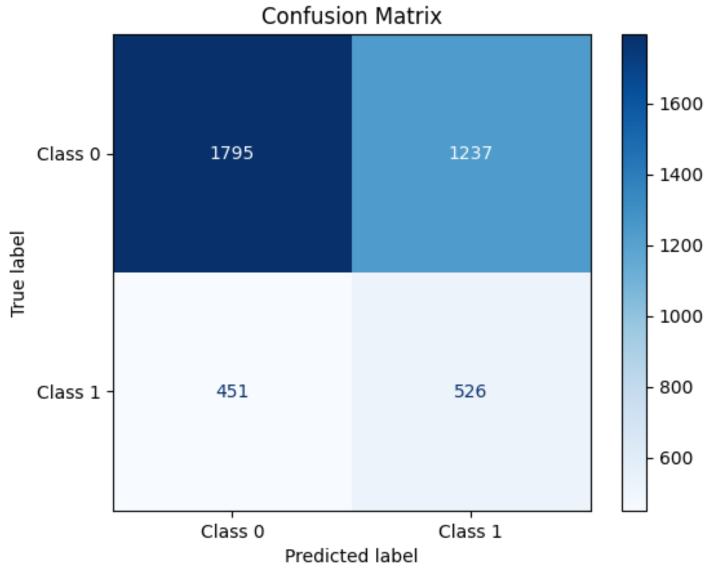


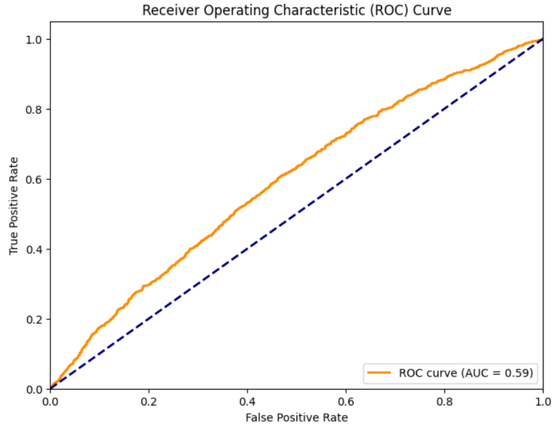
Model-2





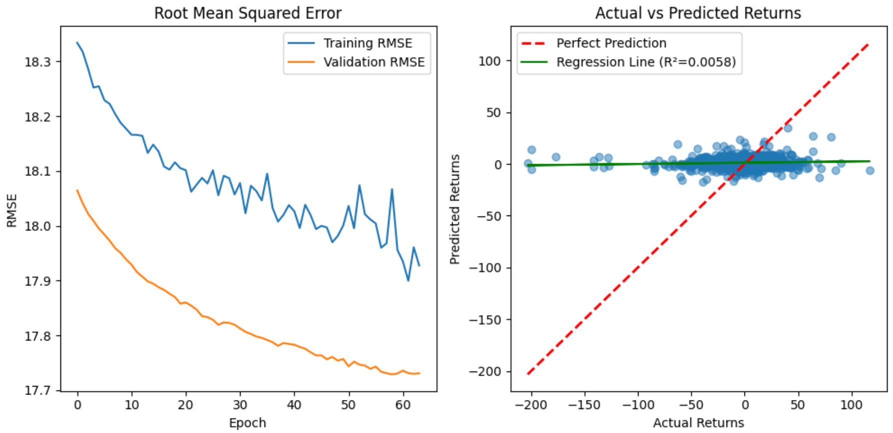
Model-3



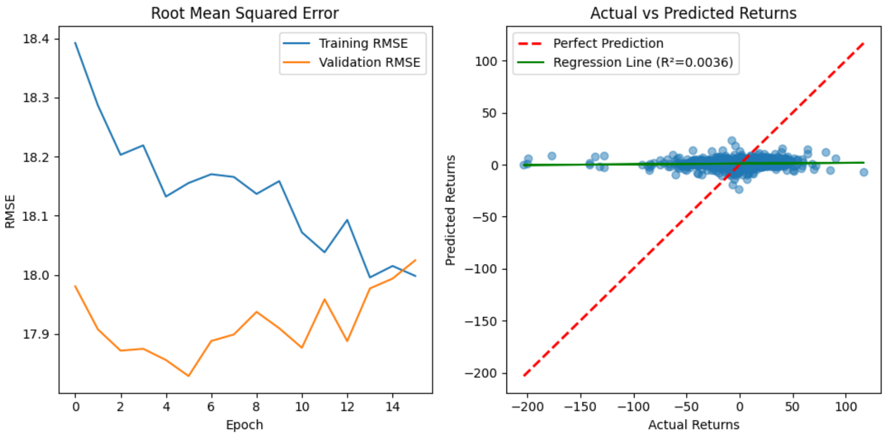


Result
Return-Recommendation model

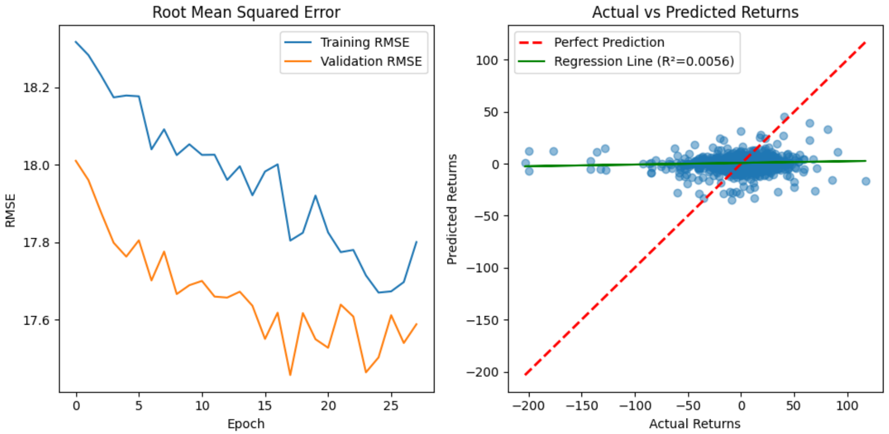
Baseline model



Model-3



Model-2



Summary

Model	1	2	3	4	5	6	7	8	9
Tr. Acc	0.635	0.645	0.635	0.755	0.76	0.76			
ROC	0.61	0.65	0.635	0.58	0.6	0.59			
Te. Acc	0.589	0.54	0.58	0.6	0.49	0.58			
Te. Pre	0.544	0.8	0.6	0.526	0.7	0.53			
Te. Rec	0.439	0.428	0.437	0.306	0.285	0.425			
F1	0.239	0.342	.262	0.16	.20	0.225			
Epochs	35	37	50	41	29	49			
R-sq							0.0013	0.0028	0.0036
RMSE							16.98	16.99	16.97

- The classification models demonstrated moderate performance, with precision values as high as 0.8 in some models. However, they suffered from low recall, indicating challenges with false positive predictions, which carry higher financial risks.
- Despite incorporating advanced techniques such as adjustable learning rates, batch normalization, and regularization, the return prediction models exhibited minimal predictive power.