

Multiclass Classifier for Stock Price Prediction

Rebwar M. Nabi^a

^aInformation Technology Department Technical College of Informatics, Sulaimani Polytechnic University, Sulaimani, Kurdistan Region of Iraq, Qaiwan International University, Sulaimani, Kurdistan Region of Iraq

Article History: Received: 10 November 2020; Revised 12 January 2021 Accepted: 27 January 2021; Published online: 5 April 2021

Abstract: The stock market has been a crucial factor of investments in the financial domain. Risk modeling and profit generation heavily rely on the sophisticated and intricate stock movement prediction task. Stock Price forecasting is complex that could have a significant influence on the financial market. The Machine Learning (ML) type of artificial intelligence (AI) provides a more accurate forecast for binary and multiclass classification. Different effective methods have been recommended to resolve the problem in the binary classification case but the multiclass classification case is a more delicate one. This paper discusses the application of multiclass classifier mappings such as One v/s All (OvA) and One v/s One (OvO) for stock movement prediction. The proposed approach comprises four main steps: data collection, assign a multi-label (up, down, or same), discover the best classifier methods, and comparison of classifiers on evaluation metrics of 10k cross-validation for stock price movement. In this study, a stock NASDAQ dataset for about ten years of ten companies from yahoo finance on daily basis is used. The resultant Stock Price prediction uncovers Neural Network classifier has good performance in some case whereas Multiclass (One V/s One) and (One V/s All) have overall better performance among all other classifiers as AdaBoost, Support Vector Machine, OneR, Bagging, Simple Logistic, Hoeffding trees, PART, Decision Tree and Random Forest. The Precision, Recall, F-Measure, and ROC area comparison results show that Multiclass (One V/s All) is better than Multiclass (One V/s one). The proposed method Multiclass classification (One v/s All) yields an accuracy of 97.63% for average prediction performance on all ten stock companies, also the highest accuracy achieved as 98.7% for QCOM. The individual stock-wise evaluation of the Multiclass (One V/s All) classifier is found to achieve the highest accuracy among all other classifiers which is outperforming all the recent proposals.

Keywords: Stock Market Forecasting; Stock Daily Data; Machine Learning; Artificial Intelligence; WEKA; Multiclass Classification; 10k-fold cross-validation; ROC Curve

1. Introduction

Despite the rapid development in the world's economy and technology, the stock market's passion and enthusiasm have never diminished. Although the stock market [1] is one of the most effective financial systems available, it is also unpredictable. The Stock market price forecast has continuously been an area for research and development. It is mainly since the market is non-linear, volatile, and dynamic, and unpredictable. Moreover, the stock market's groups and movements are affected by several economic factors such as political events, general economic conditions, commodity price index, investors' expectations, movements of other stock markets, the psychology of investors, etc. To minimize the high risk, the investor needs information as a reference for decision making of which stock they should buy, sell, and maintain for the future. Consequently, creating an accurate as well as consistent technique is not easy. In recent years to analyze the market trends, the trending technology proposed is machine learning with artificial intelligence [2].

Machine learning is the most powerful tool, including different algorithms to develop their performance in a specific case study effectively. Classification algorithms fall into two kinds: multiclass and binary. Binary classification is classifying instances right into one of two classes, while multiclass classification is identifying instances right into one of three or more unique classes. Multiclass classification is fundamental to a great deal of real-world machine learning applications that need the ability to immediately distinguish between hundreds of different classes. The issue of multiclass classification is come across in different applications such as Medicine, Market, Computer vision, Cybersecurity, etc. Many practical multiclass problems are labeling images, for example, face recognition, or tagging locations in vacation photos.

Most data relating to the stock market and financial fields are binary and multiclass classification. Most research studies have been conducted for stock market movement based on a non-linear SVM model [3] designed and implemented on the real-time stock market data. It used a novel stock technical data transformation technique, text feature extraction method, and non-linear SVM classification algorithm to predict the stock trend on a daily, weekly, monthly basis. According to previous research studies, most of the algorithm for classification is best

suited with binary classification functionality but for multiclass, it leads to a challenging task. That's why multiclass classification attracts researchers for stock movement prediction.

In this stock movement prediction paper, the multiclass problem exists for multi-label classes as UP, DOWN, or SAME differentiating the market movement on daily basis. Although the multiclass problem [4] is more difficult by its very nature, it is possible to achieve significantly better than the random performance by using both the one v/s all and the one v/s one approach. The concept of using the multiclass classification mapping approach is to change the original multiclass problem into smaller-sized binary parts using a binarization approach which is less complicated for classification. It can be recognized that there is a significant absence of studies in the multiclass category for utilizing ONE v/s ONE as well as ONE v/s ALL techniques [5] in the stock market. More importantly, none of the studies have achieved great precision to predict the stock movement to reach maximum accuracy using multiclass classification mapping.

The main research questions to be answered are:

1. Comparison of multiclass classifier (One v/s One and One v/s All) against the most known classifiers like AdaBoost, Support Vector Machine, OneR, Bagging, Simple Logistic, Hoeffding trees, PART, Decision Tree, Random Forest, and Multilayer Perceptron for stock prediction.
2. Which is the best multiclass classifier mapping among One v/s One and One v/s All that achieved the highest accuracy?

The remainder of the research is structured in this manner. In section two, the related work is outlined. In section three, the research methodology is explained in detail. The results and discussion are found in section four. The conclusion and future works of the paper are explained in section five.

2.Related Work

Several recent studies have been conducted for the multiclass problem. The research study [6] modified the multiclass classification technique such as One v/s One, One v/s All, and also Directed Acyclic Graph (DAG) for binary classifiers. They used Iris, wine, glass, and vowel dataset for the research. The highest accuracy on wine data as 100% is using SVM. The One v/s All on the wine dataset is 96.66% and on the Glass dataset is 71.96%. Similarly, The One v/s One on wine dataset is 99.43% and on Glass dataset is 71.49%. The past research [7] provided a study on three sorts of multiclass classification mapping method for categorizing dataset that contains numerous classes on portable based mobile phone gait recognition. The methods were ONE vs. ALL, ONE vs. ONE as well as random correction code (RCC) with 5 sorts of random width variables included. The results showed that using the multiclass classification mapping method partially improved the overall accuracy especially on One v/s One they got 87.78% and One v/s All 70%.

Numerous studies on stock price movements have been completed using feature engineering, feature selection, and various classification techniques. Researchers [8] intended a unique feature engineering approach for multiclass classification for the stock forecast as well as this was taken into consideration the very first study to make use for multiclass classification making use of set techniques. That research study recommended a novel multiclass classification technique called Gradient Boosting Machine with Feature Engineering (GBM-wFE) and Principal Component Analysis (PCA) as the feature selection. It located that GBM-wFE outmatched as well as the overall forecast results of Mean Absolute Percentage Error (MAPE) 0.0406% was achieved. The authors [9] suggested a novel feature engineering technique to predict the stock prices based upon historic data utilizing both binary and also multiclass classification. The monthly stock movement was predicted for each month. The overall prediction result was improved by 25.64% compared to applying the same procedure without feature engineering. The research [10] was presented on the design and implementation of a novel binary classification framework that predicted stock market trends. Also explored and discovered the very best feature selection algorithm. PCA was found to be the very best contrasted to others.

There is rare research found on multiclass classification problems (One v/s One and One v/s All) on financial data such as stock price movement. Effective machine learning approaches offered a rapid evaluation of credit reports while updating older ones on a day-to-day time scale. Relevant studies [11] had based on credit history analyses utilizing the expert system. The study has been conducted on stock market data that shown neural networks and SVM exceed various other strategies by providing much better prediction accuracy. The ONE v/s ONE SVM (77.77 %) and ONE v/s ALL SVM (33.33 %) are made use of yet its performance was not to the anticipated accuracy to forecast corporate credit scores. The research [12] had been done on the Indian stock market on daily basis to predict the future moment of the stock trends using the Artificial Neural Networks (ANN) and SVM formula with Multiclass classification: One v/s All SVM (OVA-SVM) algorithms. It achieved the highest accuracy of 71.14% in the OVA-SVM model.

Based on the above recent research works, it can be observed that there is a lack of researches for stock movement prediction to classify stock instances into a Multiclassification problem. The paper aims to explore a

novel multiclass classifier to solve multiclass (up, down, same) issues for stock movement forecast. In this research, it discovers One v/s ALL multiclass classification mapping achieves the highest accuracy against all latest classifier technique on evaluation metrics of 10k cross-validation basis too. The proposed Multiclass classifiers modes and evaluation metrics are studied and applied in detail in this paper.

3. Research Methodology

In the proposed paper, research is based on the historical stock data that applies supervised machine learning methods. Our study framework is made up of four major phases such as dataset collection, pre-processing, applied Multiclass classification (One v/s One and One v/s All) classifiers, and evaluation analysis. Figure-1 highlights the general study structure in this study.

a. Tools and Software

This research was conducted using the Java application based on WEKA (Waikato Environment for Information Analysis) which was developed by Holmes, Donkin, & Witten in 1994 [13]. Twelve latest and known machine learning classifiers out of the available algorithms in WEKA are used in this research.

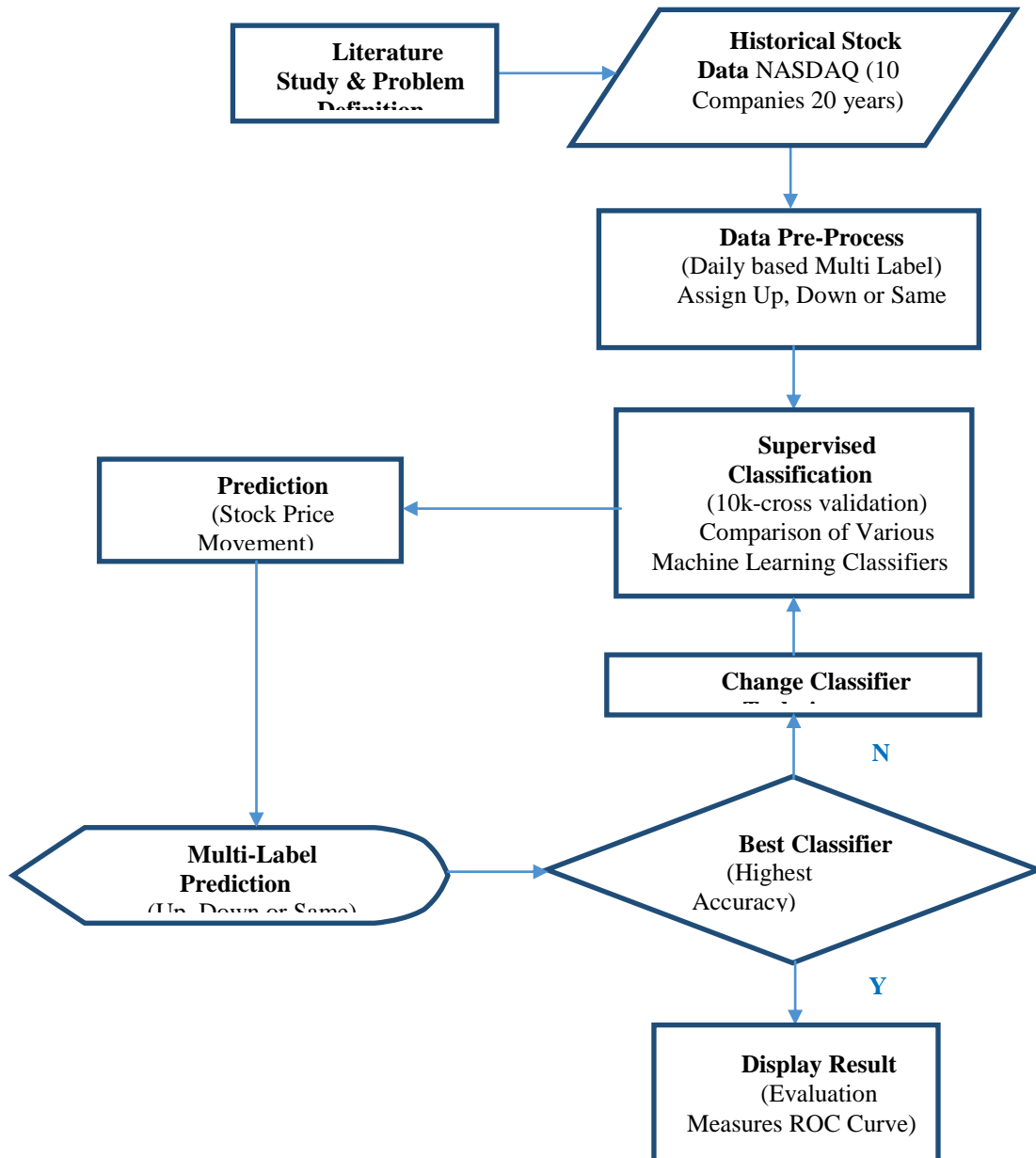


Figure-1 highlights the general study structure

b. Collection of Dataset

The current research work dataset is based on stock data acquired from Yahoo Finance [14]. The historical stock data of NASDAQ is collected from Yahoo finance. Overall, 20 years of historic data are downloaded for 10 companies as Advanced Micro Devices (AMD), Comcast Corporation (CMCSA), Cisco Systems (CSCO), Fastenal Company (FAST), Lam Research Corporation (LRCX), Microchip Technology Incorporated (MCHP), Microsoft Corporation (MSFT), NetApp (NTAP), QUALCOMM Incorporated (QCOM) and Skyworks Solutions (SWKS). The downloaded data is everyday bases each dataset has around 5000+ records and the duration of the information is Jan 2001 to Jan 2021.

Usually, each dataset has six qualities. Sample Dataset of Cisco (CSCO) is shown in Table 1.

- 1. Date The present date of the stock movement
- 2. A close price The closing price of the stock
- 3. Volume Total number of share transaction (buy and sell) of the day
- 4. Open price Open price of a stock
- 5. High price The highest price during a given day
- 6. Low price The lowest price during a given day

Date	Open	High	Low	Close	Adj. Close	Volume
1/18/2001	39.0625	42.75	38.375	41.875	31.315449	79187500
1/19/2001	43.9375	44	40.1875	40.375	30.193714	78838600
1/22/2001	40.6875	42.4375	39.9375	41.4375	30.988285	59726300
1/23/2001	41.6875	42.875	40.4375	42.625	31.876341	55443000
1/24/2001	43.125	43.375	41.5	42.5625	31.829582	70171600

Table 1: Sample Dataset of CSCO

c. Pre-Processing

The study proposed a method to pre-process the data to produce a Multiclass class for day-to-day prediction and assign a label as Up, Down, or Same that shows in Table 2. The stock movement to contrast the forecasted as well as actual percentage change every day designates the class to day-to-day information.

To discover the day-to-day activity right here, stock activity is the difference between the daily close and open price:

- Difference = close price - open rate.

The stock cost motion in regards to the percentage (%) is computed as follows:

- Percentage-Difference = Difference/ open price (Every day)

For appointing the category class in a multi-label classification situation:

- The class is assigned UP label if Percentage-Difference > 1 %.
 - The class is assigned DOWN label if Percentage-Difference <-1 %.
- Or else, the class is assigned SAME.

Day	Close	Volume	Open	High	Low	Result
1	41.875	79187500	39.0625	42.75	38.375	UP
1	40.375	78838600	43.9375	44	40.1875	DOWN
1	41.4375	59726300	40.6875	42.4375	39.9375	UP

1	42.625	55443000	41.6875	42.875	40.4375	UP
1	42.5625	70171600	43.125	43.375	41.5	DOWN

Table 2: Sample Dataset of CSCO Assign the Multi-Label

d. Implementation of Machine learning various Classifier Techniques

Machine Learning [15] is multi-disciplinary and cross-disciplinary. After learning a large amount of historical data, the model itself can have good self-learning ability, thus achieving Artificial Intelligence. Supervised machine learning [16] technique that identifies the category that a new observation belongs to, based on the training set's observations is classification. In this research work, 12 known classifiers implemented which are AdaBoost [17], Support Vector Machine (SVM) [18], Neural Network (NN) Multilayer-Perceptron [19], Random Forest [20], Decision Tree (J48) [21], PART [22], Hoeffding-Tree [23], Simple Logistic [24], Bagging [25], OneR [26], Multiclass Classifier (One V/s ALL), Multiclass Classifier (One V/s One) as shown in figure 2.

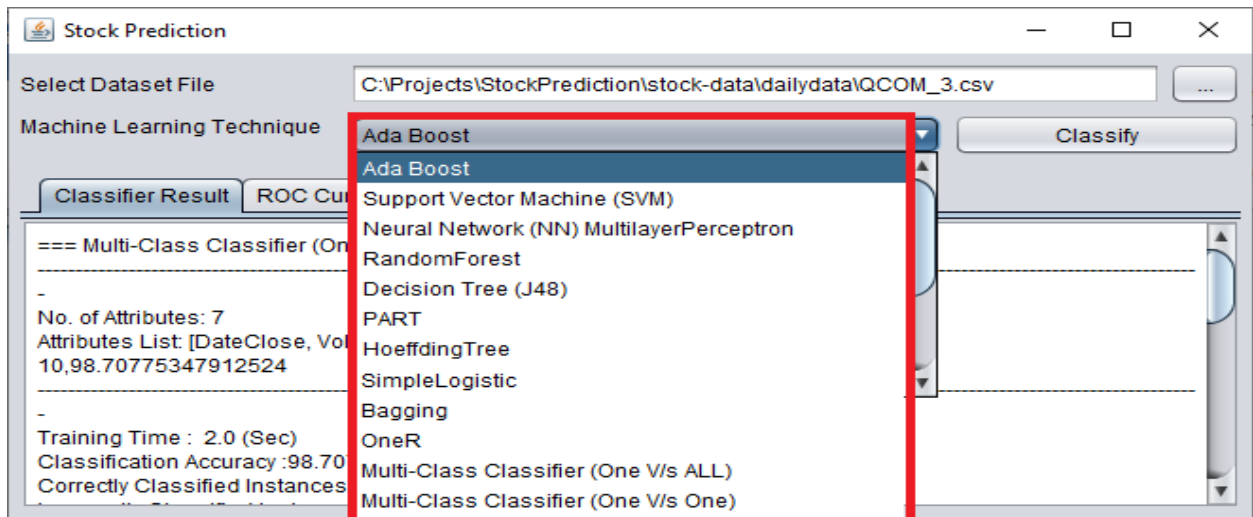


Figure 2: Implementation of various Classifiers

e. Proposed Multiclass Classifier

In multiclass classification, there are mostly 2 sorts of Multiclass classification strategies: One v/s All (OVA) and One v/s One (OVO) that commonly used. OVA and OVA facilitate the application of the data preprocessing techniques to balance the data before the training set goes to the classifier. The OVA approach takes one class as a minority and the remaining classes are combined and transformed into the majority class. This procedure is made for the n classes of the dataset. Multiple class labels are present in the dataset. Here; the variety of classifier models relies on the classification strategy we are putting on. The Confusion matrix is easy to derive but complex to recognize in Multiclass classification [27].

One v/s One: - N-class instances then $N * (N-1) / 2$ binary classifier models. In the One-vs-One classification, for the N-class instances dataset, we have to generate the $N * (N-1) / 2$ binary classifier models. Using this classification approach, we divided the primary dataset right into one dataset for each class opposite to every other class. The One v/s One approach divides a Multiclass classification right into one binary classification issue per each pair of classes.

One v/s All: - N-class instances after that N binary classifier models. In one v/s All classification, for the N-class instances dataset, we have to generate the N-binary classifier models. The variety of class tags present in the dataset and also the variety of generated binary classifiers should coincide. The One v/s All technique splits a Multiclass classification into one binary classification issue per class.

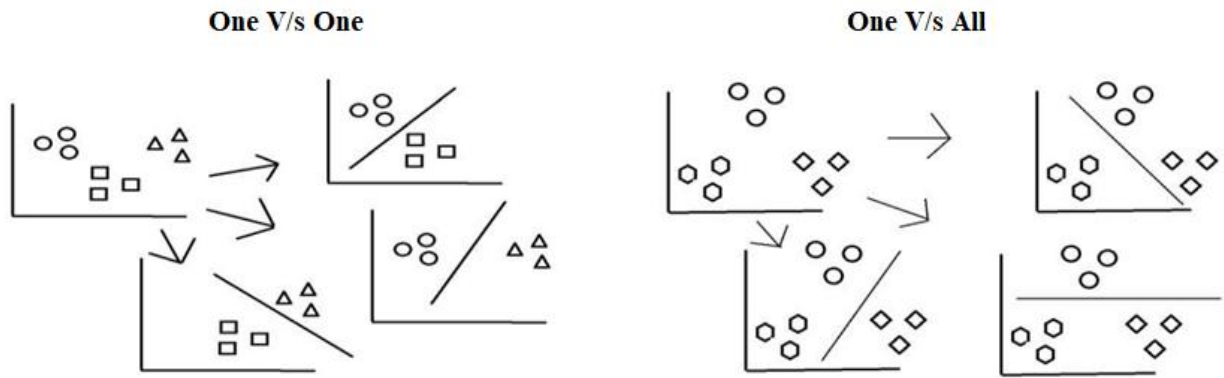


Figure 3: Divide the dataset in Multiclass One v/s one and one v/s all differences

The OVR minimizes the problem of multiclass classification to several binary category problems. It can be categorized right into One v/s All as well as One v/s One. The strategies established based on minimizing the Multiclass issue into numerous binary troubles can likewise be called problem transformation strategies. In a one v/s one strategy it divides the dataset right into the set (like displayed in number) and then does classification. Note that one hypothesis separates just two classes irrespective of the other classes. In a One v/s All strategy divide the datasets such that a hypothesis separates one class label from all of the rest that show in figure 3. The algorithm details and parameters in WEKA as default are shown in the following Table 3.

Parameter	Description	Default Option
-M <num>	sets the method to use	Valid values are 0 for (1_against_all), 1 for (random_codes), 2 for (exhaustive_code) 3 for (1_against_1). (Default 0)
-R <num>	Sets the multiplier when using random codes	default 2.0
-P	Use pairwise coupling	only affects 1-against1
-S <num>	Random number seed	default 1
-D	If a set, classifier is run in debug mode and may output additional info to the console	
-W	Full name of base classifier	default: weka.classifiers.functions.Logistic
-R <ridge>	Set the ridge in the log-likelihood	
-M <num>	Set the maximum number of iterations	default -1, until convergence

Table 3: Multiclass algorithm detail in WEKA

f. 10-k fold cross-validation and Evaluation Metrics

Examining a Machine Learning model can be rather tricky. Generally, we split the data established right into training and testing collections and also use the training readied to educate the model and also testing readied to evaluate the model. K-fold Cross-Validation (CV) gives a solution to this issue by splitting the data right into folds and also making sure that each fold is utilized as testing evaluated some factor. Over-fitting is the most common problem prevalent in most machine learning models. K-fold cross-validation can be conducted to verify if the model is over-fitted at all. The same process takes place for all k folds as shown in figure 4. Cross-validation [28] is a resampling procedure used to evaluate machine learning models on a limited data sample. Cross-

validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data.

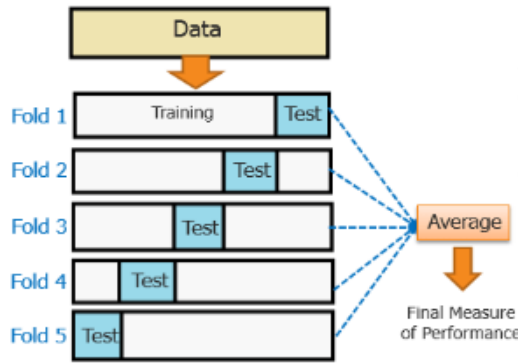


Figure 4: 10-k Cross-Validation

Accuracy is the most significant criterion in the evaluation of the performance success of the prediction methods. Therefore, the commonly used error metrics are used both in evaluating the results of the prediction models and in comparison with each other. Some metrics [29] such as TP rate, FP rate, precision, recall, F-Measures, ROC Curve are used to compare the performance success of the prediction models used in the present paper.

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.957	0.007	0.973	0.957	0.965	0.956	0.996	0.978	UP
0.957	0.007	0.968	0.957	0.963	0.955	0.997	0.971	DOWN
0.982	0.043	0.974	0.982	0.978	0.942	0.991	0.996	SAME
Weighted Avg.		0.973	0.029	0.973	0.973	0.947	0.993	0.988

Figure 5: Evaluation Metrics

Use classification metrics only such as TP rate, FP rate, precision, recall, F-measures as shown in figure 5. Precision quantifies the number of positive class predictions that belong to the positive class. Recall quantifies the number of positive class predictions made out of all positive examples in the dataset.

- F-Measure provides a single score that balances both the concerns of precision and recall in one number.
- Accuracy: Accuracy is a ratio of correctly predicted observation to the total observations
- True Positive (TP): The number of correct predictions that the occurrence is positive.
- True Negative (TN): Number of correct predictions that the occurrence is negative.
- False Positive (FP): The number of incorrect predictions that the occurrence is positive.
- False Negative (FN): Number of incorrect predictions that the occurrence is negative.
- F1- Score: It is the weighted average of precision and recall
- Precision and Recall: Precision is the fraction of relevant instances among the retrieved instances, while recall is the fraction of relevant instances that have been retrieved over the total number of instances. They are used as the measure of relevance.

g. ROC Curve

Making predictions has become an essential part of every business enterprise and all academic fields. One critical aspect of evaluating and comparing these prediction models, algorithms, or technologies is the assessment of their accuracy. Receiver operating characteristic curves address that problem. They show how accurate a model, algorithm, or technology is, depending on its past results. Formally, a receiver operating characteristics (ROC)[30] graph is a technique for visualizing, organizing, and selecting classifiers, set of categories from the dataset, based on their performance.

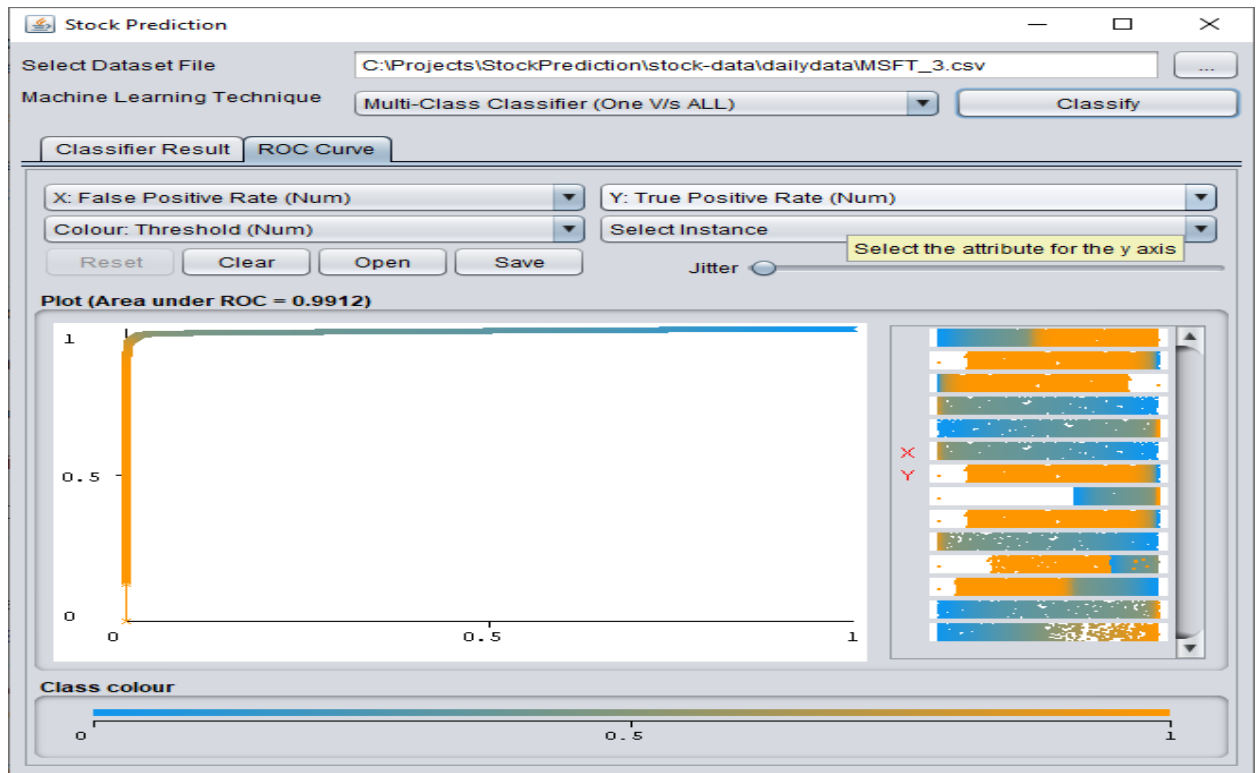


Figure 6: ROC Curve as FP Rate V/s TP Rate

ROC curves provide a comprehensive and visually attractive way to summarize the accuracy of predictions. It offers confidence intervals on forecasted values and gives ROC curves to assist determine the most effective cutoff point for classification. It allows you to verify your outcomes by instantly classifying rows that are not utilized during the analysis as shown in figure 6.

4.Result and Discussion

The experimental conducted on entire 10 company’s dataset on daily basis and tested using 10k fold cross-validation for all the 12 classifiers as AdaBoost, Support Vector Machine (SVM), OneR, Bagging, Simple Logistic, Hoeffding trees, PART, Decision Tree, Random Forest, Multiclass (One V/s One) and (One V/s All). The accuracy of the correction is shown in Table 4.

Table 4: 10k Fold cross-validation Correction Accuracy for each company to all Classifier

Classifier	A MD	C MCS	CS CO	F AST	LR CX	M CHP	M SFT	N TAP	QC OM	S WKS
Ada Boost	3 5.86	56. 46	57. 91	52 .45	42. 80	48. 87	62. 41	47. 32	52.2 5	39. 42
SVM	3 6.36	56. 46	57. 89	52 .45	43. 76	48. 91	62. 41	46. 90	53.3 2	41. 19
Neural Network	7 0.76	86. 98	92. 27	85 .53	67. 67	87. 10	78. 11	90. 04	88.1 3	81. 27
Random Forest	7 1.59	75. 79	79. 09	68 .61	71. 23	75. 07	78. 35	75. 17	74.5 3	71. 83
Decision Tree	4 2.01	65. 83	70. 28	54 .75	57. 42	67. 51	72. 35	70. 68	61.8 9	52. 84
PART	3 9.26	60. 22	64. 04	53 .94	49. 62	52. 54	64. 23	51. 19	55.2 9	41. 57
Hoeffding Tree	3 6.22	56. 46	57. 91	52 .58	43. 70	50. 44	62. 41	47. 26	53.1 2	40. 22

Simple Logistic	5 8.93	57. 06	67. 53	57 .06	51. 11	60. 64	62. 90	70. 06	65.6 9	55. 03
Bagging	6 8.71	70. 42	73. 82	62 .33	66. 38	70. 78	74. 91	71. 63	67.7 9	66. 58
OneR	3 5.35	53. 68	55. 35	50 .28	42. 27	47. 16	60. 14	43. 12	51.5 7	37. 97
Multiclass (OVO)	8 9.11	92. 76	94. 25	92 .49	89. 54	94. 16	94. 27	94. 51	96.4 0	91. 69
Multiclass (OVA)	9 5.83	98. 67	98. 51	97 .67	96. 70	98. 59	97. 30	97. 75	98.7 1	96. 60

Figure 7 shows the correct classification average accuracy of all 10 companies for each classifier. Horizontally, the classifier is listed and vertically the average correction accuracy is listed in the graph. The performance graph contains:

- X-Axis [Classifier techniques Symbol-all 12 classifiers]
- Y-Axis [Average Correction Accuracy in percentage]

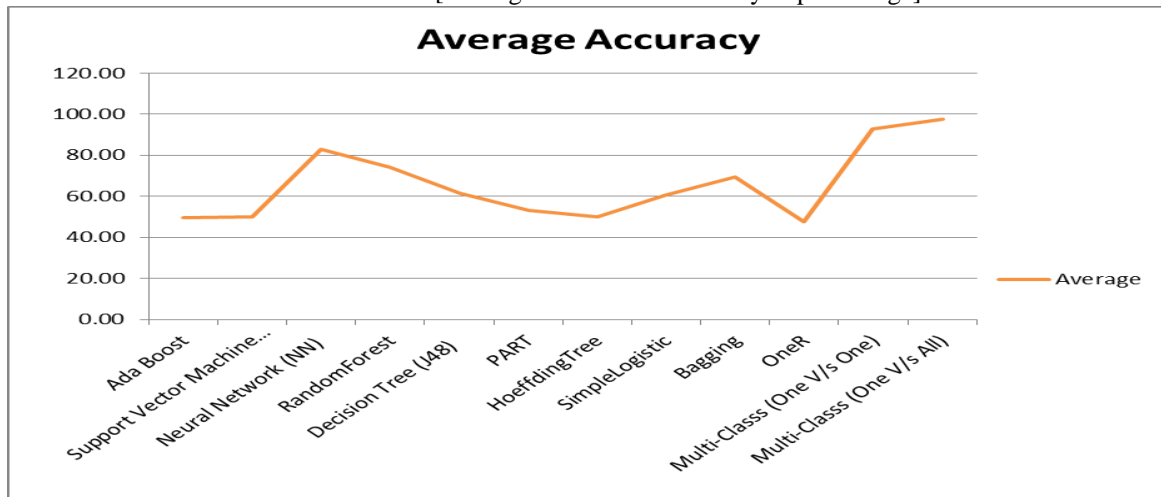


Figure 7: Correct classification average accuracy of all 10 companies for each classifier

Preliminary Investigation and Comparison of Multiclass One V/s One and One V/s All

Table 4 shows the overall accuracy of each company and Figure 7 represents average accuracy for all companies for the various classifier based on that the preliminary prediction is Neural Network has good performance in some case whereas Multiclass (One V/s One) and (One V/s All) have overall better performance among all classifier. The comparison between Multiclass One V/s One (OVO) and One V/s All (OVA) on different heuristics as Precision, Recall, and F-Measure is given below. Table-5 shows Precision, Table-6 shows Recall and Table-7 shows the F-Measures value of both classifiers for UP, DOWN, the SAME label as well as for Average.

Table 5: Precision of Multiclass One v/s one and one v/s all for label UP, DOWN, SAME and Average

Label	Cls.	A MD	C MCS	CS CO	F AST	LR CX	MC HP	M SFT	NT AP	QC OM	SW KS
UP	O	0.	0.9	0.9	0.	0.8	0.8	0.9	0.9	0.96	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9
SAME	O	0.	0.8	0.9	0.	0.9	0.9	0.9	0.9	0.95	0.8
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9
DOWN	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.97	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9

Avg	O	0.	0.9	0.9	0.	0.8	0.9	0.9	0.9	0.96	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9

Table 6: Recall of Multiclass One v/s one and one v/s all for label UP, DOWN, SAME and Average

Label	C	A	C	CS	F	LR	MC	M	NT	QC	SW
el	ls.	MD	MCS	CO	AST	CX	HP	SFT	AP	OM	KS
UP	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.97	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9
SAME	O	0.	0.8	0.9	0.	0.9	0.9	0.8	0.9	0.96	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.99	0.9
DOWN	O	0.	0.9	0.9	0.	0.8	0.9	0.9	0.9	0.96	0.8
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9
Avg	O	0.	0.9	0.9	0.	0.8	0.9	0.9	0.9	0.96	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9

Table 7: F-Measure of Multiclass One v/s one and one v/s all for label UP, DOWN, SAME and Average

Label	C	A	C	CS	F	LR	MC	M	NT	QC	SW
el	ls.	MD	MCS	CO	AST	CX	HP	SFT	AP	OM	KS
UP	O	0.	0.9	0.9	0.	0.8	0.9	0.9	0.9	0.96	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9
SAME	O	0.	0.8	0.9	0.	0.9	0.9	0.8	0.9	0.95	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.99	0.9
DOWN	O	0.	0.9	0.9	0.	0.8	0.9	0.9	0.9	0.96	0.8
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9
Avg	O	0.	0.9	0.9	0.	0.8	0.9	0.9	0.9	0.96	0.9
	O	0.	0.9	0.9	0.	0.9	0.9	0.9	0.9	0.98	0.9

ROC average weighted area presents the whole range of sensitivity and specificity values. Figure 8 shows ROC average weighted area for Multiclass (One v/s One) and Multiclass (One v/s All) classifiers. Here, X-axis represents classifiers and Y-axis represents the average weighted ROC area for all companies

- Y-Axis [ROC average weighted area]
- X-Axis [All companies symbol]
- Red color for Multiclass (OVO) and Blue color for Multiclass (OVA)

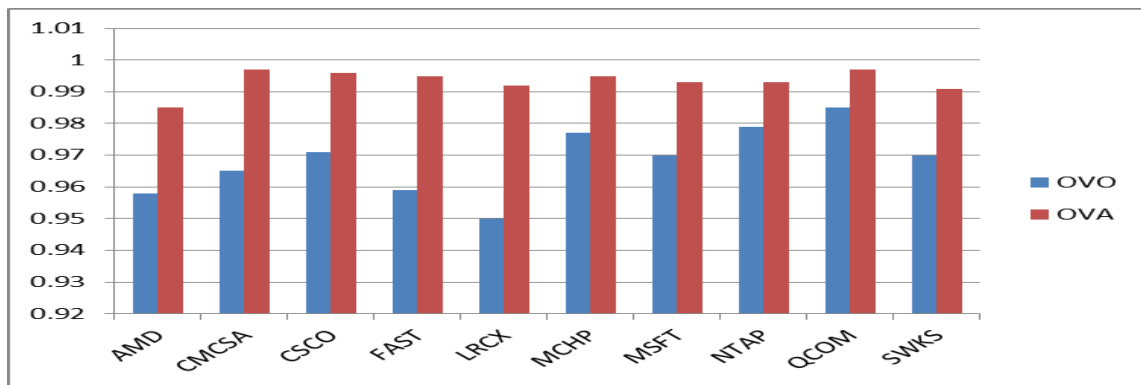


Figure 8: Area under ROC

The evaluation measures of Multiclass One v/s One and One v/s All as Precision, Recall, F-measure for all classes and its average is shown above in tables. Also, the average weighted ROC area is shown for both classifiers. The results of Multiclass One v/s All are better than One v/s One.

5.Achieved the Benchmark

This study results have achieved the highest accuracy on stock market data using Multiclass (One v/s All) among all other classifiers and also complete benchmark with the previous studies that are shown in table 8. The best accuracy achieved by the previous study[12] using Multiclass One v/s ALL (SVM) is 71.14% on the SBIN dataset whereas the proposed the Multiclass One v/s ALL (Logistic) achieved 98.71% for the QCOM company.

Table 8: Benchmark using accuracy with [12]

Benchmark	Model	Dataset	Company	Accuracy %
Their Study	Multiclass One v/s ALL (SVM)	Indian stock market 2 Years daily	SBIN	71.14
Our Study	Multiclass One v/s ALL (Logistic)	NASDAQ stock market 20 Years daily	AMD	95.83
			CMCS	98.67
			CSCO	98.51
			FAST	97.67
			LRCX	96.7
			MCHP	98.59
			MSFT	97.3
			NTAP	97.75
			QCOM	98.71
			SWKS	96.6

6.Conclusion and Future Works

Predicting the direction of movements of the stock market is important for the development of effective market trading strategies. It usually affects a financial trader’s decision to buy or sell a stock. Successful prediction of stock prices may promise attractive benefits for investors. This study attempted to predict the direction of stock price movement in the NASDAQ Stock. This work collects NASDAQ listed stocks from yahoo finance for the last twenty years as a dataset. The dataset contains various companies such as AMD, CMCSA, CSCO, FAST, LRCX, MCHP, MSFT, NTAP, QCOM, and SWKS.

Daily stock movements are predicted for each day evaluations are made with the actual movement to validate the model. The technology uses the interface of Java and WEKA to judge varied styles of machine learning classifiers over the given dataset. Latest and known classifier techniques are applied on dataset such as Ada Boost, Support Vector Machine (SVM), Neural Network (NN) Multilayer-Perceptron, Random Forest, Decision Tree (J48), PART, Hoeffding Tree, Simple Logistic, Bagging, OneR, Multiclass Classifier (One V/s ALL), Multiclass Classifier (One V/s One). The study has tested all the techniques on 10k fold cross-validation on stock market movement as up, down, and same. After the analysis in the WEKA domain, it was found that the performance metric of Neural Network (NN) Multilayer-Perceptron, Multiclass classifier (One V/s One), and (One V/s All) are better.

The individual stock-wise evaluation of the Multiclass (One V/s All) classifier is the highest in accuracy among all classifier techniques. The highest accuracy achieved was 98.7% of QCOM. Also, the average prediction performance of the Multiclass (One V/s All) on all stock companies was found significantly 97.63% best than any other techniques. The Precision, Recall, and F-Measure for all 3 labels and average comparison show that Multiclass (One V/s All) is better than Multiclass (One V/s one). The ROC average weighted area shows Multiclass (One V/s All) as the best classifier among all companies.

Future expansion studies can consider selecting the best features of the dataset through feature selection or optimization for better performance. Other classification techniques under the reinforcement of the project can be

to make the use of Machine Learning algorithms to more efficiently sort out the area where the data related to Sentiment Analysis is being obtained i.e. to more precisely understand the Tweet or Comment made by the user & use that data into the analysis and understanding of the Stock which will eventually produce better prediction results.

References

- Sharma, A., Bhuriya, D., & Singh, U. (2017, April). Survey of stock market prediction using machine learning approach. In 2017 International conference of electronics, communication and aerospace technology (ICECA) (Vol. 2, pp. 506-509). IEEE.
- Ranawat, K., & Giani, S. (2021). Artificial intelligence prediction of stock prices using social media. arXiv preprint arXiv:2101.08986.
- Lin, Y., Guo, H., & Hu, J. (2013, August). An SVM-based approach for stock market trend prediction. In The 2013 international joint conference on neural networks (IJCNN) (pp. 1-7). IEEE.
- Thrampoulidis, C., Oymak, S., & Soltanolkotabi, M. (2020). Theoretical Insights Into Multiclass Classification: A High-dimensional Asymptotic View. arXiv preprint arXiv:2011.07729.
- Scheurer, S., Tedesco, S., Brown, K. N., & O'Flynn, B. (2020). Using domain knowledge for interpretable and competitive Multiclass human activity recognition. *Sensors*, 20(4), 1208.
- Mehra, N., & Gupta, S. (2013). Survey on multiclass classification methods.
- Raziff, A. R. A., Sulaiman, M. N., Mustapha, N., & Perumal, T. (2017, October). Single classifier, OvO, OvA and RCC multiclass classification method in handheld based smartphone gait identification. In AIP Conference Proceedings (Vol. 1891, No. 1, p. 020009). AIP Publishing LLC.
- R. Nabi, S. Ab. M. Saeed, and H. Harron, "A Novel Approach for Stock Price Prediction Using Gradient Boosting Machine with Feature Engineering (GBM-wFE)", *Kurdistan Journal of Applied Research*, vol. 5, no. 1, pp. 28-48, Apr. 2020.
- Rebwar M. Nabi, Soran AB. Saeed, Abdulrahman M. W. Abdi. (2020). Feature Engineering for Stock Price Prediction. *International Journal of Advanced Science and Technology*, 29(12s), 2486-2496. Retrieved from <http://sersc.org/journals/index.php/IJAST/article/view/24722>
- Nabi, R. M., Saeed, S. A. M., Harron, H. B. & Fujita, H. (2019). Ultimate Prediction of Stock Market Price Movement. *Journal of Computer Science*, 15(12), 1795-1808. <https://doi.org/10.3844/jcssp.2019.1795.1808>
- Golbayani, P., Florescu, I., & Chatterjee, R. (2020). A comparative study of forecasting corporate credit ratings using neural networks, support vector machines, and decision trees. *The North American Journal of Economics and Finance*, 54, 101251.
- Shrimalve, H. H., & Talekar, S. A. Comparative Analysis of Stock Market Prediction System using SVM and ANN. *International Journal of Computer Applications*, 975, 8887.
- G. Holmes, A. Donkin, I.H. Witten, "Weka: A machine learning workbench," 357-361, 1994
- Yahoo Finance NASDAQ Companies Historical Data (2021) , "Stock Market Historical Data", <https://finance.yahoo.com/>
- S. Vijayarani, E. Suganya, T. Jeevitha," PREDICTING STOCK MARKET USING MACHINE LEARNING ALGORITHMS", *IRJMETS*, December -2020
- Gomathi, R. M., Rajeshwari, M., Meenakshi, K., Priya, P. A., & Nithya, M. (2020). A FUSION APPROACH AIMED AT FORECASTING OF STOCK PRICE EXCHANGE USING SUPERVISED MACHINE LEARNING ALGORITHM. *European Journal of Molecular & Clinical Medicine*, 7(11), 2391-2398.
- Chang, V., Li, T., & Zeng, Z. An Adaboost Algorithm Based Analysis Method of Nonstationary Financial Data.
- Hao, P. Y., Kung, C. F., Chang, C. Y., & Ou, J. B. (2021). Predicting stock price trends based on financial news articles and using a novel twin support vector machine with fuzzy hyperplane. *Applied Soft Computing*, 98, 106806.
- Kolasani, S. V., & Assaf, R. (2020). Predicting Stock Movement Using Sentiment Analysis of Twitter Feed with Neural Networks. *Journal of Data Analysis and Information Processing*, 8(4), 309-319.
- Ronaghi, F., Salimibeni, M., Naderkhani, F., & Mohammadi, A. (2020, July). ND-SMPF: A Noisy Deep Neural Network Fusion Framework for Stock Price Movement Prediction. In 2020 IEEE 23rd International Conference on Information Fusion (FUSION) (pp. 1-7). IEEE.
- Kia, A. N., Haratizadeh, S., & Shouraki, S. B. (2020). Network-based direction of movement prediction in financial markets. *Engineering Applications of Artificial Intelligence*, 88, 103340.
- Dey, S. K., & Rahman, M. (2020). Effects of machine learning approach in flow-based anomaly detection on software-defined networking. *Symmetry*, 12(1), 7.
- Alhayali, R. A. I., Ahmed, M. A., Mohialden, Y. M., & Ali, A. H. (2020). Efficient method for breast cancer classification based on ensemble hoeffding tree and naïve Bayes. *Indonesian Journal of Electrical Engineering and Computer Science*, 18(2), 1074-1080.

- Nusinovici, S., Tham, Y. C., Yan, M. Y. C., Ting, D. S. W., Li, J., Sabanayagam, C., ...& Cheng, C. Y. (2020). Logistic regression was as good as machine learning for predicting major chronic diseases. *Journal of clinical epidemiology*, 122, 56-69.
- González, S., García, S., Del Ser, J., Rokach, L., & Herrera, F. (2020). A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. *Information Fusion*, 64, 205-237.
- Sandhu, A. K., & Batth, R. S. (2020). Software reuse analytics using integrated random forest and gradient boosting machine learning algorithm. *Software: Practice and Experience*.
- Hameed, N., Shabut, A. M., Ghosh, M. K., & Hossain, M. A. (2020). Multiclass multi-level classification algorithm for skin lesions classification using machine learning techniques. *Expert Systems with Applications*, 141, 112961.
- Soni, A. N. (2020). Diabetes Mellitus Prediction Using Ensemble Machine Learning Techniques. Available at SSRN 3642877.
- Li, Y., & Pan, Y. (2020). A novel ensemble deep learning model for stock prediction based on stock prices and news. *arXiv preprint arXiv:2007.12620*.
- Janssens, A. C. J., & Martens, F. K. (2020). Reflection on modern methods: revisiting the area under the ROC curve. *International journal of epidemiology*, 49(4), 1397-1403.