A COMPARATIVE STUDY TO FIND AN OPTIMAL ALGORITHM FOR A STOCK USING ARTIFICIAL INTELLIGENCE

Srikanth Satyala   
*Computer Science engineering*  
*vit-ap university*guntur, India  
srikanthsathyala@gmail.com

Dr. Reeja S R  
*Professor- CSE*  
*vit-ap university*India  
reeja.sr@vitap.ac.in Devika T.S  
*Computer Science engineering Specialisation in Networking security*  
*vit-ap university*  
Chennai,India  
devikasasi2019@gmail.com

Rayudu Vyshnavi Ram  
*Computer Science engineering*  
*vit-ap university*Rajamundry, India  
rayuduvyshnavi@gmail.com

# *ABSTRACT*

Stock markets have different sectors in the market like the industrial sector, technology sector, finance sector, etc. we are using algorithms with analytical methods for efficient, profitable, and optimal trading using machine learning methods. Different mediums like stocks, cryptocurrency, NFT, etc can be traded as an asset. There are many types of trading algorithms like Mean Reversion, Factor-Based Investing ETF Rotation, Smart Beta, etc. and, applying Artificial intelligence models like Random Forest, Support Vector Machine, Naive Bayes, logistic regression, classification regression tree, and DNN models such as recurrent neural network, multilayer perceptron, deep belief network. Algorithmic trading is used for mainly short-term trading, but it can also be used for scalping. Both have some differences, so their respective algorithms are adjusted accordingly for better profits and smooth processing.

Keywords—Artificial Intelligence, HFT, Trading, Multilayer Perceptron, Deep Belief Networks, Logistic Regression, stocks, Random Forest, Short term trading, scalping, intraday trading, machine learning.

# Introduction

Algorithmic trading is a pre-programmed method of running commands and trading instructions that are accounting for variables such as time, price, and volume. Here we are applying artificial intelligence concepts to algorithms to get an OPTIMAL method that can return maximum profits. It can avoid small-scale man-made errors which can lead to a loss on a large scale and there will be no involvement of emotion which is a very risky and highly influencing factor in stocks. All the analytics are done using past data to make sure there are no losses or real money during the testing instead of real money.

There are many classification algorithms in artificial intelligence that can be classified based on trees, distance, probability, and neural networks. Every model has its pros and cons, that’s why different types of classification models are compared to find which models are suitable for which market and we need to find one that is suitable for all the sectors

For training, methods used are done in iterations. Every iteration we train artificial intelligence methods such as

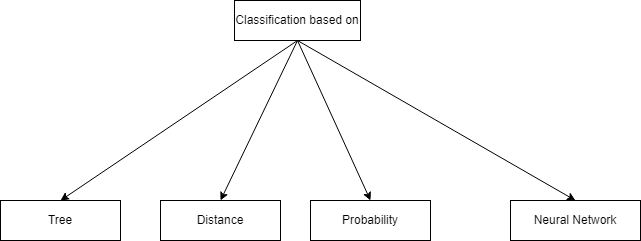
NB(Naive Bayes Classification), RF(Random Forest), LR(Logistic Regression), CART(Classification and Regression Trees), KNN(K Nearest Neighbour ), XGB(XG Boost), RNN(Recurrent Neural Network), MLP(Multilayer Perceptron), SVM(Support Vector Machine), DBN(Deep Belief Networks) are used in different sectors. We chose Annual return of rate (ARR), Winning ratio (WR) as scaling factors to measure which gives the best profits and is optimal for each sector

Fig. 1

Graphical user interface, application

Description automatically generated

Fig. 2

Graphical user interface, application

Description automatically generated

Fig. 3

Graphical user interface, application

Description automatically generated

Fig. 4

Graphical user interface, application

Description automatically generated

Fig. 5

Based on the classification from Fig. 1, we can further add multiple methods that come under the classification where, Fig. 2 is classified based on tree, Fig. 3 is classified based on Distance, Fig. 4 is classified based on probability, Fig. 5 is classified based on Nural Network

# Literature survey

2006 John Duffy [1] experimented on humans about the relation between man and economical decisions made using humans with no prior knowledge of economics using external expert machine agents which are computers explaining the human’s rules and methodologies. Cons are not covered as it is starting to develop.

2012 Marco De Luca [2] Reviewed near zero intelligence and improved algorithms which are not very flexible overall markets and concluded, GDX is better than ZIP which is modifying and applying different algorithms to different markets based on strategies makes it more flexible and can be used for high-frequency trading.

2013 Kirilenko and Lo [3] Upgraded the algorithms so that they covered the cons of previous algorithms which are not so flexible to all types of markets. Not only making them fast, cost-effective, and increasing trade volumes they also gained frequency Strategies. They also predicted future possibilities of threats and opportunities.

2014 MA Goldstein [4] computerized trading as the advancement of technology as time passed added new rules, strategies towards more profit and very less time using real data.

It also covered risks of trading using computers and algorithms at high frequencies and more automation that can be done. Made a milestone leading trading to a new level that is applicable to real-life data and making actual profits that made many researchers realize the potential of artificial agent trading.

2016 Burckhardt & Miller and Shorter [5] Researched on high frequency trading methods like simulating analysis of gains. Selected actions of securing high frequency trading. Also limited issues of research in the past which are mostly time based and some or applying optimization to existing algorithms with new methods.

2019 Te Bao [6] Expanded review on relation between man and computer on including rules, models, strategies, negotiation and dealing. It also covers some unpublished work which had competition between algorithm vs human towards profitability. Also discussed the relation between human psychology on algorithms and how computers can trade as investors in the market and the agent decides that algorithms trade by themselves without humans taking part.

[7] compared 14 different classifications on 7 different sectors to find an optimized method to gain more profit on stocks.

RNN (Fig. 6) is a tree network that mainly works on Memorization and prediction best example can be voice assistants like google voice assistant and Siri which notes the questions that are not in the knowledge base to predict what solution can be given when similar type of question is asked again. It goes in depth concept of patterns in sequential data and memory. We can observe form Fig. 1 we can see it runs like a loop.

A picture containing text, clock, watch

Description automatically generated

Fig. 6 (source: By Kevin Vu, Exxact Corp)

Multilayer Perceptron (MLP Fig. 7) is not ideal for processing sequential data because it needs many parameters to process data. It has three layers: an input layer, an output layer, and a hidden layer. It is mostly applied for supervised learning.

1. IMPLEMENTATION

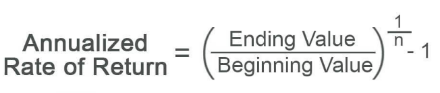
We consider below evaluating factors to train and measure to find our prediction’s accuracy

Win Ratio (WR) is the number of winning trades/total number of trades used in the Kelly Criterion formula. Higher WR indicates that we have a good profit

Diagram

Description automatically generated with medium confidence

The annual return rate (ARR) is the yearly rate calculated by money spent or earned at the end of the year divided by money invested at the start of the year. It also depends on the amount of time the stock period.



ASR is evaluation measure. Suppose the holding period is H, and there are m single periods in a year. In the H period, the ARR of the investment tool is ARRH, the standard deviation of return rate is σH, and Rf is the benchmark such as risk-free return. In this paper, we set Rf = 0. ASR is given as follows.



Drawdown is a measure of historical loss. It is the largest loss compared to the previous highest value of the net value curve. MDD shows the largest decline in the price or value of the investment period H, which is an important risk assessment indicator. In the period of investment τ, we first calculate the Dt at any time r <=H. Then, we can get the MDDH when we go traverse the whole interval.



where pt denotes the value of the net value curve with time t; Dτ represents the drawdown at the time τ, i.e., the difference between the maximum value in [0,τ] and the value of at the time τ. MDDH denotes the maximum drawdown in [0,H].

[ARR, ASR, MDD and WR]

The above four factors are considered to compare optimality.

These sectors are divided into 9 parts as the Fig. 8 below

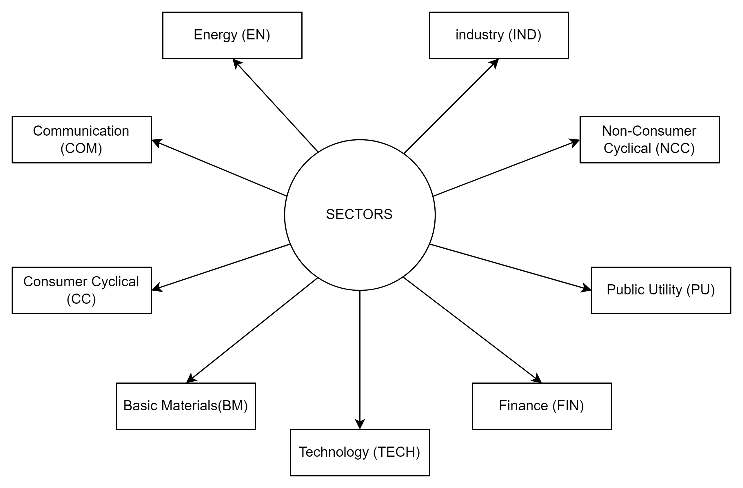


Fig. 8

**Learning algorithm**

Given a training dataset, *D* = {(*x*1,*y*1),(*x*2,*y*2),⋯,(*xP*,*yP*)}, where *xi* = {*xi*1,*xi*2,⋯,*xiP*} is an instance of input; *P*is the number of sample features; *yi* = {0,1} is a class label; *i* = 1,2,3,⋯,*N*, where *N* is the sample size. *D* is a matrix of *N*\*(*P*+1), where the *P*+1-th column of *D* is class label .Main goal of this training is to classify the labels. In this paper, we will use the six traditional machine models, namely LR, SVM, CART, RF, BN, XGB and DNN classifiers to predict the rise and fall of the stock prices. The main model parameters and training parameters of these learning algorithms are shown in the table below.

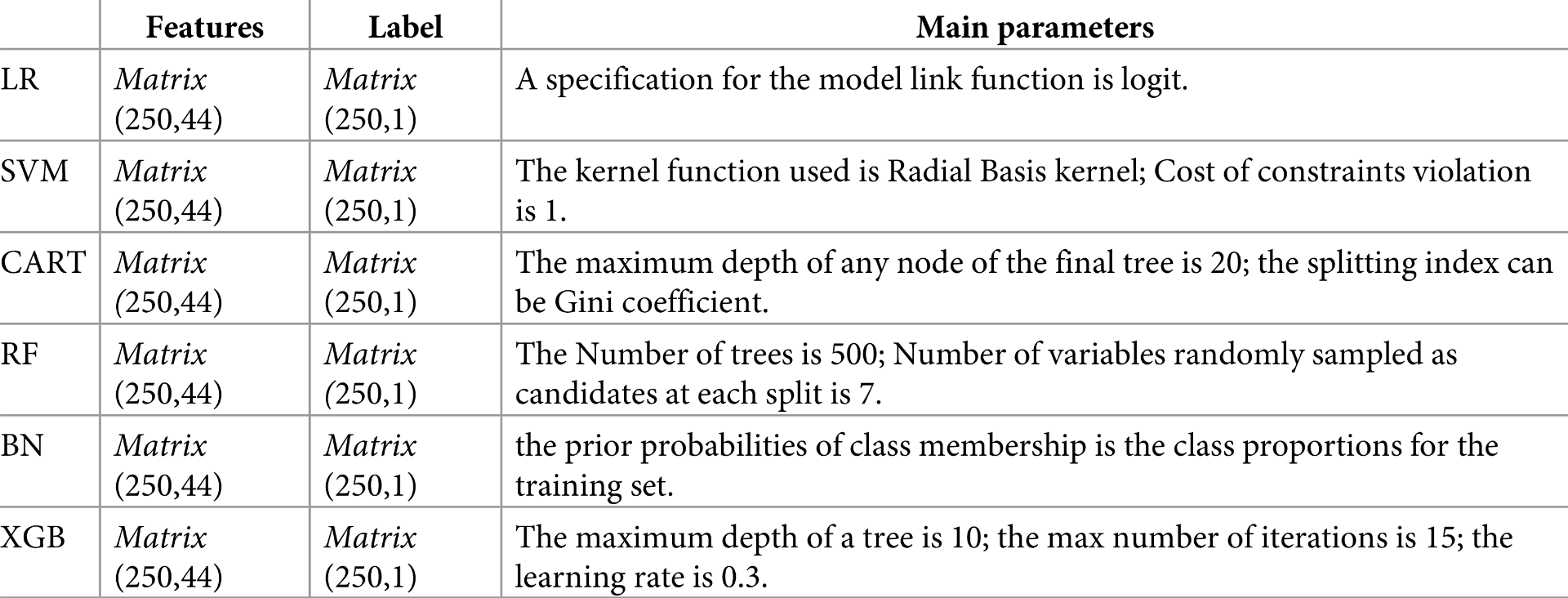
****

Table 1

### **Algorithm for generating trading signals**

we use the data from the past 100 days as the training set and the data for the next 10 days as the test set. Each stock contains data for 1,000 trading days, so it takes (1000–100) / 10 = 90 training sessions to produce a total of 900 predictions that signal the trading. The algorithm for generating trading signals is shown in Algorithm 1

Graphical user interface, application

Description automatically generated

Algorithm 1

Data sets are from SPICS

We analyze difference between parameters like ARR, WR, MDD and ASR to find which algorithm is optimal for different industries

From table 2 Multi-Layer Perceptron (MLP) has the highest Win Ratio (WR) compared to all other methods in all the sectors, but it is close enough to Deep Belief Network (DBN) and SAF

Table

Description automatically generated

Table 2 (WR as parameter)

From Table 3 we can observe that Annualized Return of Rate (ARR) using Classification and Regression Trees is the highest CART is optimal in Basic Materials, Communication, Energy, and Industry. Deep Belief Network is optimal in Consumer Cycle and Public Utility. Multi-Layer Perceptron is the highest in Finance and Non-Consumer Cyclical.Sparse Auto Encoder is the highest in Technology .Frequency of model to the highest number of sectors

Table

Description automatically generated

Table 3

From Table 4 we can observe that ASR of SAE is highest in EN, RF is the highest in the CC, FIN, IND, NCC, PU, and TECH and XGB is highest in COM and BM.

 In the FIN, all algorithms are greater than that of CART; otherwise, there is no significant difference between the ASRs of any two algorithms. In the IND, the ASR of RF is significantly greater than that of DBN; the ASR of CART is significantly lower than that of NB, RF, SVM, XGB; otherwise, there is no significant difference between the ASRs of any two algorithms. In the NCC, the ASR of RF is significantly greater than that of MLP, DBN.

Table

Description automatically generated

Table 4

Table

Description automatically generated

Table 5

From table we can observe that BAH is highest of all sectors

# RESULTS and DISCUSSION

**Selection of the optimal trading model for different industries**

The Optimized Trade Algorithm (TOTA) can be used in all sectors as an indicator. We give a series of rules as follows, where “a>b” represents that the performance of algorithm a is significantly greater than that of algorithm b; “a = b” represents that the performance of algorithm a is not significantly different from that of algorithm b.

For any industry *i*∈ {NB, RF, LR, CART, KNN, XGB, RNN, MLP, SVM, DBN}, for any performance evaluation indicator *j*∈ {*WR*, *ARR*, *ASR*, *MDD*}. Grater the value of j, the better the trading performance of models.

 Benchmark index can be expressed by the relationship among the 3 strategies, which are expressed as *a*, *b*, and *c* respectively.

Text

Description automatically generated

First, we must choose the optimal algorithms for each sector in which performance of the algorithm can be better than that of the benchmark index.

Secondly, the trading performance of the optimal algorithm can be significantly better than the BAH strategy in each sector, which is conducive to take a strategy while reducing risk of loss.

Therefore, if the trading performance of algorithm is not better than that of the index, we hope that it is significantly better than BAH strategy. We select the TOTAs which are significantly better than the rest of the algorithms, as shown in Table 6

Table

Description automatically generated

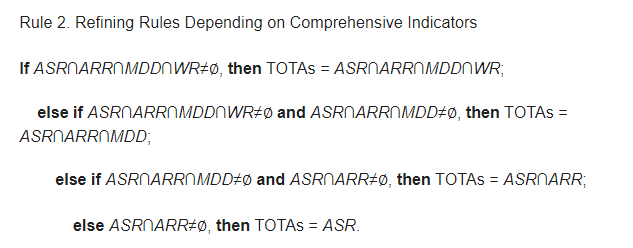
Table 6

From Table6, we find that the optimal model based on the WR is always found in any industry, and MLP is the optimal algorithms in all industries through the analysis of the industries in the SPICS; MLP is the optimal trading model based on ARR in the FIN, and any algorithms can be used in other industries; the optimal trading model based on ASR can be found in the CC, EN, FIN, IND, and NCC, and any algorithms can be used in other industries; the optimal trading model based on MDD can be found in the CC, EN, FIN, IND, and NCC, and any algorithms can be used in other industries; TOTAs based on ARR can be found in the FIN and IND, and any algorithms can be used in other industries; TOTAs based on ASR can be found in the BM, COM, FIN, IND, and NCC, and any algorithms can be used in other industries; TOTAs based on MDD can be found in all industries except NCC. From Table 6, we can observe that there is more than one optimal algorithm. There is no significant difference in performance among the multiple optimal trading algorithms selected. For example, for the BM sector, we obtain the optimal trading algorithms which including MLP, DBN, and SAE based on WR. These three algorithms have no statistically significant difference for WR.

However, we can see from Table 6 that there are too many “ATAUs”, which means that the optimal model for each industry is still not enough, so we need a new set of rules based on Table 6. For each industry, ASR represents risk adjusted returns, it is the most important indicator for evaluating a trading algorithm

It is also an important indicator for evaluating the trading algorithm without considering risk. MDD also describes the potential risks of trading algorithms. WR represents the performance of a trading algorithm in predicting stock price trends, which is not a direct source of stock investment returns.

Therefore, we assume that *ASR*>ARR>MDD>WR according to the importance of the four evaluation indicators, where “m>n” represents the indicator m is more important than the indicator n. The following refining rules are proposed.



For example, we can use the above rules to select TOTAs for NCC in SPICS: *WR* ={*MLP*,*DBN*,*SAE*}, *ARR* = *ATAU* = {*MLP*,*DBN*,*SAE*,*RNN*,*GRU*,*LSTM*,*NB*,*SVM*,*XGB*,*LR*,*RF*,*CART*} *ASR* = {*RF*}, *MDD* = {*RNN*,*GRU*,*CART*,*NB*,*RF*,*XGB*}, now  *ASR*∩*ARR*∩*MDD* = *RF* which means non null value that obeys all the rules(common factor) is RF, so the RF is the TOTA for NCC in SPICS. Like this we can get TOTA for all the other sectors. As shown in Table 7

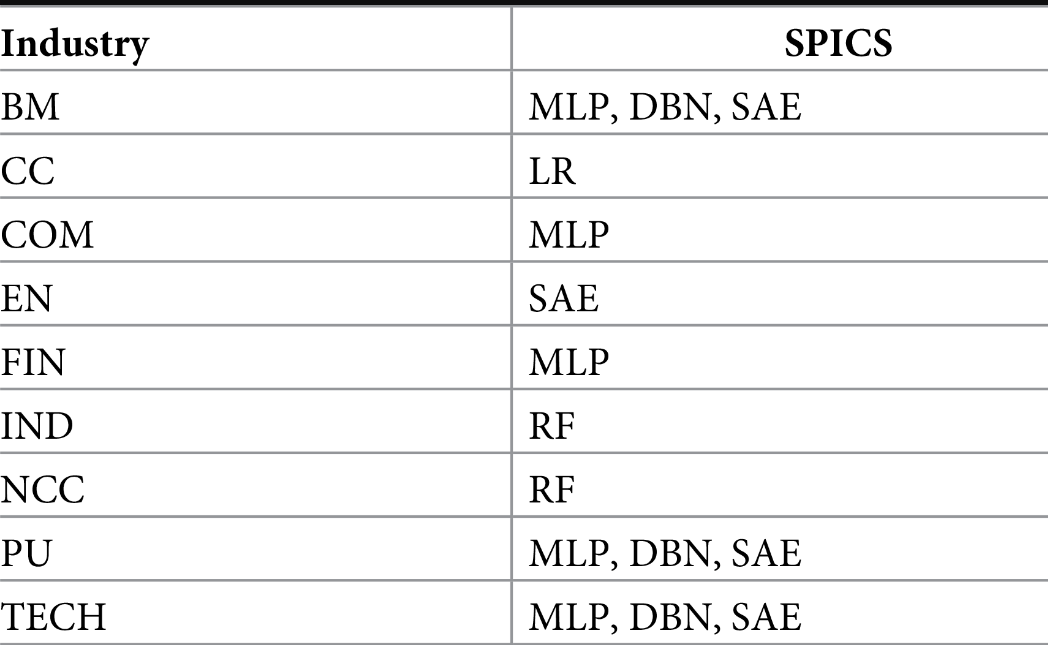


Table 7

As can be seen from Table 7, the number of optimal models selected according to Rule 2 is small because Rule 2 considers the importance between the ASR and the remaining indicators. At the same time, deep neural network algorithms have a good performance in most industries, but LR and RF are very prominent in some industries.

These experimental results show that on SPICS, we can always select TOTAs based on the single indicator and comprehensive indicators in all industries. We can apply TOTAs to implement trading activity in each industry.

# Conclusion

##### Compared to humans, algorithms do a better job at making profits on trading due to experience. Small changes in algorithms based on different strategies for different time segments make the algorithm directed towards respective profitable algorithms. That's why when observing both datasets, we can see that different markets need different methods to have an optimal trade. We give factors like ARR priority to find which a benchmark and compare them to remining algorithms to find the most optimal model. Better Benchmarks for measuring optimality in the future can make finding a method very easy and can be used for High-frequency trading(HFT) using automation and risk management

##### Future Scope

As the world is becoming more data-oriented and new algorithms are being invented as time passes, so much data is being stored and analyzed. as time goes by, we will have more data and thus more chance or scope to recognize patterns.

Machine learning algorithms can help gain more knowledge on past data and predict future prices based on past patterns and factors influencing them just like Elon musk’s tweets influencing tesla stock prices. It can also be used for automation where you can just give money and your personal assistant ai can invest your money and multiply it.

##### References

[1] John Duffy, Chapter 19 Agent-Based Models, and Human Subject Experiments, Editor(s): L. Tesfatsion, K.L. Judd, Handbook of Computational Economics, Elsevier

[2] Cartlidge, J.; Szostek, C.; De Luca, M. and Cliff, D. (2012). TOO FAST TOO FURIOUS - Faster Financial-market Trading Agents Can Give Less Efficient Markets

[3] Kirilenko, Andrei A., and Andrew W. Lo. 2013. "Moore's Law versus Murphy's Law: Algorithmic Trading and Its Discontents."

[4] Goldstein, M.A., Kumar, P. and Graves, F.C. (2014), Computerized and High-Frequency Trading

Beckhardt & Miller and Shorter

[5] Miller, R.S., & Shorter, G.W. (2016). High-Frequency Trading: Overview of Recent Developments.

[6] Bao, Te, and NEKRASOVA, Elizaveta and Neugebauer, Tibor and Riyanto, Yohanes E., Algorithmic Trading in Experimental Markets with Human Traders

[7]Lv D, Huang Z, Li M, Xiang Y (2019) Selection of the optimal trading model for stock investment in different industries

Headings orders are changed accordingly

Added future scope

Added implementation

Updated output