International Journal Of Advanced Engineering & Innovative Technology (IJAEIT)
ISSN: 2348 7208
ISRA IMPACT FACTOR: 0.42

Stock Prediction Using Genetic Algorithm

Abhimanyu R. Barge , Akshay A. Gawade, Dhiraj B. Jagtap, Prashant B. Barate

Abstract— The prediction of the stock market is a fascinating subject. It changes the lives of investors on a daily basis based on the decisions they make whether to buy or sell shares. The stock market is most important resource for companies to raise money. In this work we will present an Artificial Neural Network approach to predict stock market indices. We'll outline the design of the Neural Network model with its salient features and customizable parameters. We will implement a number of activation functions along with options for cross validation sets. Stock market Price prediction using data mining is one of the most fascinating issues to be investigated and it is one of the important issues of stock market research over the past decade. However, determining the best time to buy, sell or hold a stock remains very difficult because there are lots of factors that may influence the stock market prices like fundamental, Technical indexes, unknown factors. Data mining technique like Association rule mining (ARM) focuses on finding most frequent item sets and corresponding association rules. Fragment based rule mining method by taking just historical datasets as input is proposed. We will use this algorithm which generate an extremely large number of association rules, often in hundreds or even thousands. Further, the association rules are sometimes very large. It is nearly impossible for the end users to understand easily.

Index Terms— Stock Prediction, Genetic Algorithm, Artificial Neural Network approach

I .INTRODUCTION

Data mining additionally famously known as Knowledge Revelation in Databases (KDD) alludes to the nontrivial extraction of implied, at one time obscure and possibly helpful data from information in databases. While Data mining what's more information revelation in databases (or KDD) are as often as possible treated as equivalent words, Data mining is really piece of the information revelation process. The accompanying (Figure 1) shows Data mining as a venture in an iterative learning disclosure process. [1] Data mining comprises of valuable procedures, for example, Bunching and Association leads, these procedures can be utilized to foresee the future patterns focused around the Item-sets [6]. Bunching is utilized to gathering comparative thing sets while affiliation is utilized to get summed up guidelines of ward variables. Helpful thing sets can be acquired from gigantic exchanging information utilizing these guidelines. [2]

Abhimanyu R. Barge , Akshay A. Gawade, Dhiraj B. Jagtap, Prashant B. Barate, Research students of DIFOE College of Engineering Swami Chincholi (Pune))

Affiliation mining, which is generally utilized for finding affiliation controls in single and multidimensional databases, can be grouped into intra and entomb exchange affiliation mining. Intra-exchange affiliation alludes to relationship in the same exchange; between exchanges affiliation shows affiliation among diverse exchanges [3]. Most commitments in affiliation mining concentrate on intra-exchange affiliation moreover alluded to customary affiliation mining. Between exchange affiliation mining was proposed in 2000 [3] and has an expansive scope of uses, however its essential thought stretches out from intra-transaction affiliation mining. [4] Stock Prices are thought to be exceptionally alterable and powerless to fast changes due to the hidden nature of the budgetary area and to some extent as a result of the mixof known parameters (Previous Day's Closing Price, P/E Ratio and so forth) and obscure elements (like Election Results, Rumors and so on).

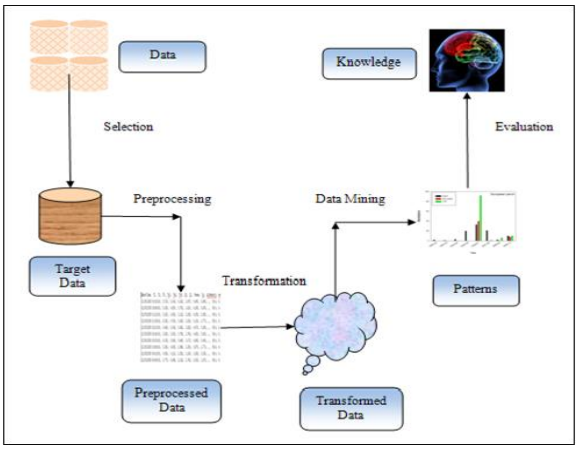
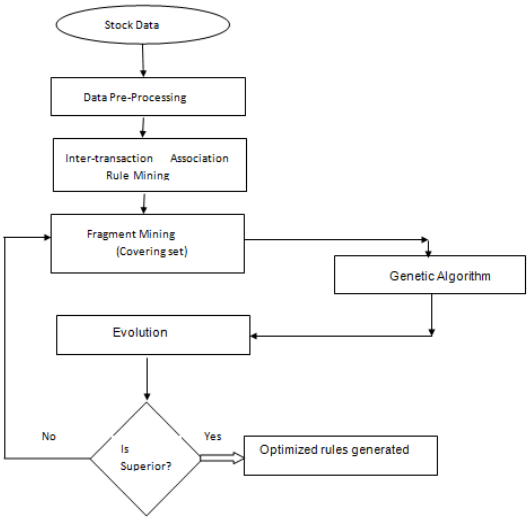


Fig. 1 KDD Process

In this examination we have taken the first information sets of Bombay Stock Exchange (BSE) of diverse organizations, for example, Infosys, TCS, and Oracle and so forth from Yahoo Finance and attempt to find the relationship among the expansive scale IT organizations and Little scale IT organizations. As we realize that there are constantly a few conditions between diverse fields in stock exchange. Our point is to discover whether vast scale organizations influence the little scale organizations' shares. Some test results demonstrates that there is a solid connection in the middle of vast and little scale organizations, we found that major of the times when the offerestimation of vast Rajesh V. et al,/ (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012,3493-3497 3493 organizations go high, little scale organizations imparts likewise goes high and the other way around. Granule mining [4] discovers fascinating relationship between granules in databases, where a granule is a predicate that depicts normal peculiarities of a set of items (e.g., records, or exchanges) for a chose set of traits (or things). For case, a granule alludes to a gathering of exchanges that have the same trait values. Granule mining broadens the thought of choice tables in unpleasant set hypothesis into affiliation mining. The properties in a data table comprise of condition traits furthermore choice traits, with clients' necessities. As in granule mining, part based methodology pieces the information sets into pieces for transforming subsequently diminishing the information size of information sets sustained to the calculation. Rather than granule mining, in piece based mining the condition and choice properties are summed for acquiring summed up affiliation guidelines.

Flow Of Project



II . RELATED WORK

In the past exploration, diverse entomb exchange methods for multidimensional information has been proposed for Data mining; Anthony J.t. Lee, Chun-Sheng Wang, Wan-Yu Weng, Yi-A Chen, Huei-Wen introduced " An Efficient calculation for mining shut between exchange thing sets" an Icminer, for mining shut between exchange thing sets. He performed on the engineered, genuine & "most detrimental possibility" datasets and closed Icminer is more proficient than the EH-Apriori, FITI approaches. Ahmed et al. [9] introduced the information distribution center backboned framework incorporated Data mining and OLAP systems. This framework makes utilization of a switch to embrace the past mining result put away in the information stockroom, appropriately dodging handling a lot of the crude information. [8] Both fundamentalists and specialists have created certain methods to anticipate costs from money related news articles. In one model that tried the exchanging rationalities; Lebaron et. al. set that much can be gained from a mimicked stock exchange with reenacted dealers (Lebaron, Arthur et al. 1999). Wanzhong Yang additionally proposed one creative procedure to process the stock information named Granule mining method, which decreases the width of the exchange information and creates the affiliation standards. [4] Our point is to augment the work in this field and give some essential reflections (Fragments)In this section we describe the structure of

Artificial Neurons and how they are connected to construct Artificial Neural Network.

III .Artificial Neural Networks

A .Artificial neurons

Artificial neurons are motivated from organic neuronal structure. The transmission of a sign starting with one neuron then onto the next through neurotransmitters is a complex concoction prepare in which particular transmitter substances are discharged from the sending side of the intersection. The impact is to raise or bring down the electrical potential inside the collection of the accepting cell. In the event that this reviewed potential achieves a limit, the neuron fires. It is this trademark that the manufactured neuron model endeavor to replicate. The neuron model demonstrated in Figure 1 is the particular case that broadly utilized as a part of manufactured neural systems with some minor alterations on it.

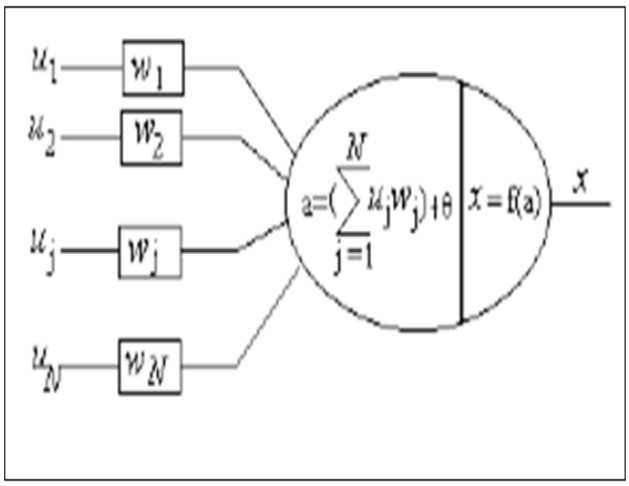


Figure 1: Artificial Neuron Structure

The counterfeit neuron given in this figure has N info, meant as u1, u2, ...un. Each one line interfacing these inputs to the neuron is allocated a weight, which are signified as w1, w2,..,wn separately. Weights in the fake model compare to the synaptic associations in natural neurons. The limit in simulated neuron is typically spoken to by θ and the actuation comparing to the evaluated potential is given by the equation:

$$a = \sum_{j=1}^{N} w_j u_j + \theta$$

The inputs and the weights are genuine qualities. A negative worth for a weight shows an inhibitory association while a positive quality demonstrates an excitatory one. Albeit in natural neurons, has a negative quality, it might be allocated a positive esteem in simulated neuron models. In some cases, the limit is joined for effortlessness into the summation part by accepting a fanciful data u0 =+1 and an association weight w0 = θ. Subsequently the initiation equation gets to be:

$$a = \sum_{j=1}^{N} w_j u_j$$

The yield estimation of the neuron is a capacity of its actuation in a similarity to the terminating recurrence of the natural neurons: There are various capacities that are utilized. Some incorporate paired edge, direct limit, sigmoid, hyperbolic tan and Gaussia.

B .Artificial Neural Networks

While a solitary counterfeit neuron is not ready to execute some boolean capacities, the issue is overcome by uniting the yields of a few neurons as info to the others, so constituting a neural system. Assume that we have associated numerous fake neurons that we acquainted in Section 1.2 with structure a system. In such a case, there are a few neurons in the framework, so we dole out records to the neurons to segregate between them. At that point to express the initiation ith neuron, the recipes are altered as follows:

$$a_i = \sum_{j=1}^{N} w_{ij} x_j + \theta_i$$

where*x*! possibly the yield of an alternate neuron or an outside data. There are various architectures being used for ANNs. In feed-forward neural systems, the neurons are composed as layers. The neurons in a layer get information from the past layer and food their yield to the following layer. In this sort of systems associations with the neurons in the same or past layers are not allowed. The last layer of neurons is known as the yield layer and the layers between the info and yield layers are known as the shrouded layers. The info layer is comprised of extraordinary data neurons, transmitting just the connected outer

information to their yields. In a system if there is the layer of data hubs and a solitary layer of neurons constituting the yield layer then they are called single layer system. On the off chance that there are one or more shroudedlayers, such systems are called multilayer systems. The structures, in which associations with the neurons of the same layer or to the past layers are permitted, are called repetitive system

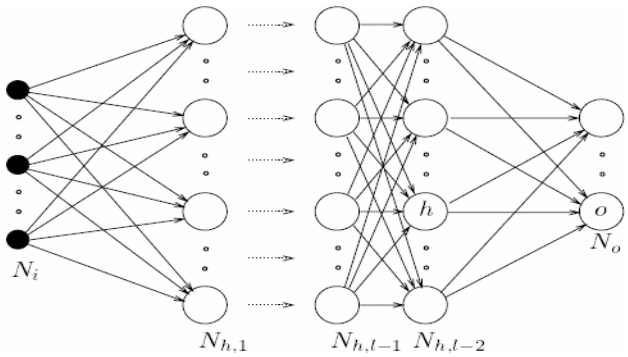


Figure 2: Artificial Neural Network

IV. Background

A .Association Rule Mining

Affiliation standard digging is a method for finding unsuspected information conditions and is one of the best known Data mining methods. The essential thought is to distinguish from a given database, comprising of thing sets (e.g. shopping bushel), whether the event of particular things, infers additionally the event of different things with a generally high likelihood. In standard the response to this inquiry could be effortlessly found by thorough investigation of all conceivable conditions, which is however restrictively costly. Affiliation standard mining in this way takes care of the issue of how to hunt proficiently down those conditions. Created by Agarwal and Srikant1994 Imaginative approach to discover affiliation controls on extensive scale, permitting ramifications conclusions that comprise of more than one thing, Based on least help edge. ssociation guidelines are ramifications of the structure X -> Y where X and Y are two disjoint subsets of all accessible things. X is known as the precursor or LHS (left hand side) and Y is known as the subsequent or RHS (right hand side). Affiliation guidelines have to fulfill stipulations on measures of criticalness and interestingness

B. FITI(First Intra then Inter) Algorithm

The FITI calculation [11] is focused around the accompanying property, a vast between exchange thing set must be comprised of extensive intra-exchange thing sets, which implies that for a thing set to be vast in between exchange affiliation guideline mining, it additionally has to be vast utilizing customary intra-exchange standard mining systems. By utilizing this property, the many-sided quality of the mining procedure can be diminished, and mining between exchange affiliations guidelines can be performed in a sensible measure of time. To start with FITI presents a parameter called maxspan (or sliding window size), meant w. This parameter is utilized as a part of the mining of affiliation standards, and just manages crossing not exactly on the other hand equivalent to w exchanges will be mined. Second, every sliding window in the database structures a mega exchange. A super exchange in a sliding window W is characterized as the set of things W, annexed with the sub window number of everything. The things in the super exchanges are called augmented things. Txy is the situated of super exchanges that contain the set of amplified things X, Y, and Tx is the situated of super exchanges that contain X. The backing of a between exchange affiliation principle X=> Y is then characterized as"

Support = |Txy| /S, Confidence = |Txy|/|Tx

V .Methodology

As FITI calculation takes part of time in handling the information so we concentrate predominantly on decreasing time and deliver more acknowledged affiliation tenets. Section Based methodology takes a shot at conquering the downsides of FITI methodology which bunches the exchanges as opposed to considering all the exchanges from the stock information. Our objective in this exploration is to discover relationship among the Little and Large Scale IT organizations from Indian IT Stock information. In this new approach we consider a solitary exchange as conglomeration of the length of the sliding window, which helps in creating more summed up standards as contrasted with FITI approach. window and Small Scale SUM 2 as the second sliding window et cetera

Table I Indian it stock market transaction (small scale) table iii. Indian it stock market transaction table (large Scale)

ID	Date	A1	A2	A3
1	1/1/2009	315	152	242
2	2/1/2009	320	154	240
3	3/1/2009	320	162	230
4	4/1/2009	310	157	236
5	5/1/2009	310	160	231
6	6/1/2009	315	134	223
7	7/1/2009	320	125	237
8	8/1/2009	300	135	238
.....				
100	6/4/2009	306	140	236
101	7/4/2009	304	140	237
102	8/4/2009	300	145	239
103	9/4/2009	322	158	240

Sliding Window

Small Scale SUM 1

Small Scale SUM 2

Small Scale SUM N

ID	Date	B1	B2	B3
1	1/1/2009	2745	1701	835
2	2/1/2009	2755	1675	815
3	3/1/2009	2760	1590	825
4	4/1/2009	2767	1650	817
5	5/1/2009	2689	1725	835
6	6/1/2009	2735	1698	820
7	7/1/2009	2679	1699	814
8	8/1/2009	2714	1690	810
.....				
100	6/4/2009	2686	1655	825
101	7/4/2009	2695	1724	840
102	8/4/2009	2699	1710	865
103	9/4/2009	2729	1709	849

Sliding Window

Large Scale SUM 1

Large Scale SUM 2

Large Scale SUM N

This huge amount of data is minimized by performing the aggregation based on the sliding window size. Here we defined the size of the sliding window as four and move this sliding window linearly and do the same aggregation. This is expressed in the figure below i.e. Fragmented Transaction table.

Table II. Fragmented data of small scale companies

ID	Small Scale SUM		
	A1	A2	A3
1	1265	625	948
2	1245	554	929
.....			
N	1232	583	952

We do the same operation as that for Small Scale companies; here too we define the Sliding window size asfour, Large Scale Sum 1 as the first sliding window and Large Scale Sum 2 as the second sliding window and so on.

Table III. Fragmented data of large scale companies

ID	Large Scale SUM		
	B1	B2	B3
1	11027	6616	3292
2	10817	6812	3279
.....			
N	10809	6798	3379

Now to generate rules among small and large scale companies data we perform inter transactions among the bothi.e. transaction 1 from small scale companies is related with transaction 4 from large scale companies and so on.

Table V. Inter-transaction among small and large scale Companies

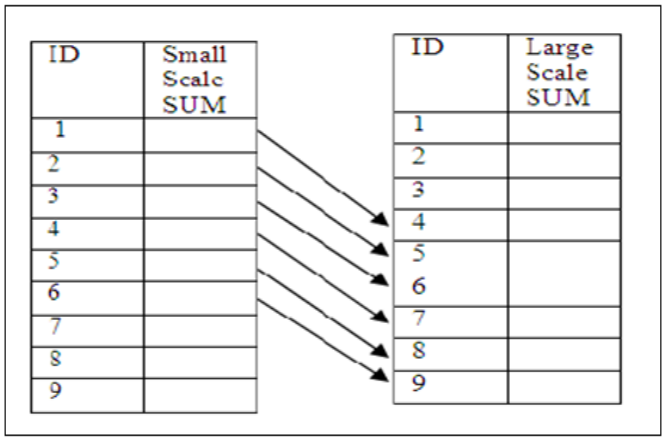


Table VII. Converted table

ID	Small Scale SUM			Large Scale SUM		
	A1	A2	A3	B1	B2	B3
1	0	0	0	1	0	1
2	0	1	1	0	1	1
3	1	1	0	0	0	0
4	0	1	0	1	1	0
5	1	1	1	1	1	1
6	--	--	--	--	--	--

VI . EXPERIMENTS AND RESULTS

A Fragment Based Approach

After applying the fragmentation rule we get the following minimized table. Now we apply the Apriori on this processed data and find the association rules among the attributes. Here KPIT ,Mphasis and Mahi Stym belong to Small Scale Companies where as TCS, Infosys and Wipro belong to Large Scale Companies respectively.

Input Data:

ID	KPIT	Mphasis	MahiStym	TCS	Infosys	Wipro
1	0	1	0	0	0	1
2	0	0	0	0	0	0
3	0	0	1	1	0	0
4	1	0	1	0	0	1
5	0	1	1	0	1	0
		. . .	. . .		. . .	
181	1	1	1	0	0	1
182	0	1	1	0	0	1
183	1	0	0	0	0	0

Output Association Rules after applying Fragment Based Mining

1. Infosys=0 97 ==> TCS=0 75 conf:(0.77)
2. TCS=0 100 ==> Infosys=0 75 conf:(0.75)
3. TCS=1 82 ==> Infosys=1 60 conf:(0.73)
4. Mphasis=0 84 ==> KPIT=0 61 conf:(0.73)
5. KPIT=1 80 ==>Mphasis=1 57 conf:(0.71)
6. Infosys=1 85 ==> TCS=1 60 conf:(0.71)
7. Mphasis=0 84 ==>MahiStym=0 58 conf:(0.69)
8. MahiStym=1 81 ==>Mphasis=1 55 conf:(0.68)
9. MahiStym=0 101 ==> KPIT=0 67 conf:(0.66)
10. KPIT=0 102 ==>MahiStym=0 67 conf:(0.66)

The first association rule shows that Infosys and TCS has .77 confidence, that if Infosys goes high (↑) then TCS will also go high (↑). And the 8th association rule shows that MahiStym and
Mphasis has .68 confidence, that if MahiStym goes low (↓) thenMphasis will also goes low (↓).

VII .CONCLUSIONS

By some experimental analysis we find Fragment Based approach generated more generalized rules as compared to FITI approach. Also time needed to process the data is less as we reduced the size of the input table. The rules generated from Fragment Based approach can be recommended to the customers who invest their money in the stock market.

REFERENCES

[1] OsmarR.Zaiane, "Principles of Knowledge Discovery in Databases",1999.
[2] DattatrayP.Gandhmal, RanjeetsinghParihar,and Rajesh Argiddi "An Optimized approach to analyze stock market using data mining technique", IJCA, ICETT 2011 .
[3] H. Lu, J. Han, and L. Feng (2000). "Beyond intratransaction association analysis: mining multidimensional intertransaction association rules." ACM Transactions on Information Systems 18(4): 423-454.
[4] Wanzhong Yang, "Granule Based Knowledge Representation for Intra and Inter Transaction Association Mining", Queensland University of Technology, July 2009.
[5] J. Dong and M. Han (2007). IFCIA: An Efficient Algorithm for Mining Intertransaction Frequent Closed Item sets. The fourth international conference on fuzzy systems and knowledge discovery, China.