Research Article

Building an Expert System based on Data Mining

Sagar Bhushan Gawde^{*} and Umesh Kulkarni

Department of Computer, Vidyalankar Institute of Technology, Wadala East- 400037, Mumbai, India

Received 01 March 2019, Accepted 02 May 2019, Available online 04 May 2019, Vol.7 (May/June 2019 issue)

Abstract

A novel framework for predicting stock trends and making financial trading, decisions based on a combination of Data and Text Mining techniques. The prediction models of the proposed system extract data in text content of time-stamped web documents in addition to traditional numerical time series data, which is also available from the Web. The financial trading system based on model predictions uses three different trading strategies. In this work, our system is simulated and evaluated on real-world series of news stories and stocks data using Decision Tree Induction Algorithm. Performance is the predictive accuracy of the induced models and, more importantly, the profitability of each trading strategy using these predictions.

Keywords: Expert system, Data mining etc.

1. Introduction

The Efficient Market Hypothesis (EMH), [7] assumes that Stock rates are adequately reflected at every information at any given point in time. As the basis for growth and development of a modern economy, this means information or analysis can be expected to perform out. The market and that stock prices follow 'Random Walks' [9] where a change in stock price over time is purely random and statistically independent of the stock price in the past. However, to this day no one can explain the anomalies in the market, which can be utilized to assure some short term predictive power [6]. In making their forecasts, most financial specialists try to exploit the time gap of the market's adjustment to new information. They reduce their risk by combining both scientific (base future price predictions on past prices) and theoretical base predictions on real economic facts. Such as inflation, trading volume, organizational changes in the company, etc.) analysis strategies, which are mentioned by Gidofalvi [4]. To obtain the data required by both procedures, here refer to various publicly available resources like the stock market itself, the companies, news articles, etc. A somewhat new source for information in the late 20th and the 21st centuries is, of course, the Internet. To exploit this relatively new and additional tool supporting the forecasting task, we need to combine techniques from both time series data mining and web content mining.

In this work, we present a new system for analysis stock trends based on the combination of Data Mining and Web Content Mining techniques. New Financial Trading System which:

1) Creates a "melting pot" of numeric and textual data before running an induction algorithm,

2) Extracts automatically crucial phrases instead of using a prior expert list of phrases,

3) Eliminates the need for word independence assumption by using Decision Trees rather than Naive Bayes,

4) A new method Influence of news articles in the prediction task to dates

Equations

$$S = 1/3 * TF/N + 1/3(P/L * B/L) + 1/3 * AV$$
 (Eq.1)

Where:

L: The time frame, in days, for the word dictionary.

B: The time window between the first and last occurrence of a word.

P: the number of days to the last occurrence of a word.

TF: the number of occurrences of a word during L (known as Term Frequency).

N: The number of words in the dictionary.

AV: The annualized volatility of the stock as calculated by

 $\sigma = \sqrt{\frac{1}{n-1} \sum_{i=1}^{n} (u_i - \bar{u})^2}$ (Eq.2)

280 | Int. J. of Multidisciplinary and Current research, Vol.7 (May/June 2019)

^{*}Corresponding author's ORCID ID: 0000-0002-3906-7080 DOI: https://doi.org/10.14741/ijmcr/v.7.3.6

Sagar Bhushan Gawde and Umesh Kulkarni

Where

 u_i is the $i^{\rm th}$ observation of the stock price, and u bar is the mean of a stock price.

Whereas the proposed system uses

$$S = 1/3 * TF/N + 1/3(P/L * B/L) + 1/3 * AV$$
 (Eq.3)

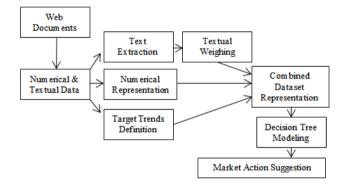


Fig. 1 Proposal with Mixed Numerical and Textual Data

Our system is such a way designed as a full cycle prediction system for stock trends according to past numeric values of the stocks as well as their related textual web articles. It goes through six steps, as shown in Fig. 1, which are:

- Step 1: Data Collection from the Web.
- Step 2: Feature Extraction.
- Step 3: Textual Weighting.
- Step 4: Combined Data-Set Construction.
- Step 5: Classification Model (Decision Tree) Induction.
- Step 6: Market Action Recommendation.

Data Collection from the web itself is a challenging task in itself. In all the standard web scrapping methods we worked through; we found that some programmatic code like HTML or script tag gets included into the scrap and adds to the text noise. Ultimately we found Selenium based automation with chrome to be useful to get exact text content from a website.

Next, we computed features like

L: The time frame, in days, for the word dictionary,

B: The time window between the first and last occurrence of a word,

P: Previous occurred nice of a word number of times, TF: the number of events of a word during L (known as Term Frequency), N: the number of words in the dictionary. These values are used to compute the score which will be used along with other technical benefits of the stock by the Classification Model to get the appropriate action recommendation.

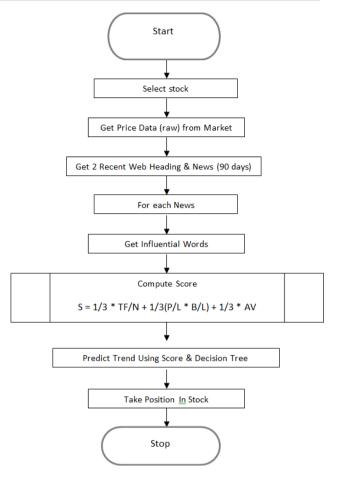


Fig.2 Flowchart

The methods to improve profit include

1. Combination of both numeric and textual data.

2. The use of an automatic text extraction mechanism instead of a predefined expert list.

3. The method of decision tree prediction model.

4. The purpose of smart trading strategies and techniques.

In the flowchart, we compute the score from the features calculated to take into account news articles related to the company. This score is in turn used by the decision tree model to extract the strategy to be followed for making a profit.

Experimental Results

Honest, Businessman, INDUSTRY, India, Wine, Cooler, and, Chest, Freezer, Market, Changing, Lifestyle, Consumption, Pattern, of, Consumers, Drives, Growth, finds, TMR, March, Views, Min, Read, ajinkya, tmrresearch, com, Share, This, Some, the, main, players, in, wine, cooler, chest, freezer, market, are, AB, Electrolux, Haier, Inc, Kieis, Ltd, Elan, Professional, Appliances, Pvt, Westinghouse, Electric, Corporation, Rockwell, Industries, The, Middleby, Western, Refrigeration, Private, Limited, Whirlpool, Williams, Such, Leading, Players, FedEx, UTi, Worldwide, Ryder, System, CEVA, Holdings, Deutsche, Bahn, Agility, Schneider, UPS, Expeditors, APL, SCIENCE, Autonomous, Underw

ater, Vehicle, projected, reach, USD, million, Bluefin, Robotics ,Saab, ECA, Group, About, author, VIEW, ALL, POSTS, Latest, Ne ws, Urban, Gas, Helium, Sales, Supply, Forecasts, COMMERCI AL, AND, SERVICE, MACHINERY, MANUFACTURING, MARKET, GLOBAL, TREND, SEGMENTATION, OPPORTUNITIES, FORECA ST, TO, HEALTH, Industrial, Institutional, Cleaning, Chemicals, register, forecast, Dunnage, Air, Bags, Expected, Witness, Sust ainable, over, Key, Cordstrap, Bates, Cargo, Pak, Stopak, Inc, Co pyright, Created, Meks, Powered, WordPress

Fig.3 Word collection GUI

The above screen shows the collection of keywords from the text content consisting of dated news articles from various websites related to the industry. Subsequently, features computed from these keywords. The news is collected from official sites only which listed in the list box.

From the experiments carried out on the web. Word selected for a selected stock for a specified period; this step includes Textual extraction from authentic sites

The chart of the selected stock can also analyse in this step.

TCS				construction and the				
165			Tata Consultancy Se	rvices Limit	ted (Listed	On: 25-/	Aug-2004)	
News Date:								
Fri, 01-Mar-2019 ~	-	1	1					
		Word	Occurrence., Textual We.,	Time Frame Tim 90	e Wind., Last 0 74	Accurrence Dic	tionary Count; Score (S) 2000027195	512811
	-		10.29					
Get News	2	Records	10.29	90	82	22	20000 11049.5	51.2811
	3	The	4114	90	12	32	20000147	51.2811
Find Frequency	- 4	Indian	10.29	90	83	74	20000 37619	51.2811
No. Of Past Days:	5	Security	10.29	90	57	2	20000 698.25	51.2811
	6	System	10.29	90	9	4	20000 220.5	51.2811
90	7	Integrators	10.29	90	26	16	20000 2548	51.2811
50	8	Market	144	90	50	59	20000 92.19	51.2811
Chart	9	to	61.71	90	87	66	20000 976.94	51.2811
	10	Set	10.29	90	87	59	20000 31439	51.2811
Data Table	11	Exceed	10.29	90	15	41	20000 3766	51.2811
Get Data	12	Business	10.29	90	88	50	20000 26950	51.2811
ouround	13	Wire	10.29	90	87	27	2000014387	51.2811
Authentic Site Filter:	14	press	30.86	90	44	68	20000 2036	51.2811
https://economictimes \land	15	release	30.86	90	73	18	20000 894 25	51,2811
https://timesofindia.inc	16	Mar	11.3.14	90	13	62	20000408	51 2811
https://www.moneycon	17	Kev	30.86	90	38	17	20000439.64	51,2811
in tps.//www.nsemula.ci		players	30.86	90	83	34	20000 1920	51.2811
https://www.finmarket	18	in	5143	90	72	57	200001005	51,2811
https://conitalmind in/ ×					87			
c 3	20	the	41.14	90	87	71	20000 2364	51.2811

Fig.4 Frequency GUI

In this screen, we compute the features from the keywords collected from the news articles. We also calculate the score from the elements to be subsequently used by the decision tree model.

In this step, the occurrence of the words for a particular period has calculated. Helps us to find Textual weight and score for the same.

Select Symbol:	Use Proposed Score													
RELIANCE		Reliance Industries Limited (Listed On: 29-Nov-1995)												
News Date:	Get Data			richardo 2	1000111001	control (c)	5100 0111 2		•)					
Thu, 04-Apr-2019														
	Volatility		SYMBOL	TIMESTAMP	OPEN	HIGH	LOW	CLOSE	TOTTRDQTY	VOLATILITY				
	Score	1	RELIANCE	15-Feb-2019	1229.75	1249.9	1214	1244.45	9597961	27.4241	7			
Get News	Octive -	2	RELIANCE	18-Feb-2019	1250	1252.5	1215	1220.1	9649017	30.0166				
	Proposed Score	3	RELIANCE	19-Feb-2019	1218	1239.7	1211.2	1216.1	6244189	31.0672				
Find Frequency		4	RELIANCE	20-Feb-2019	1223.85	1240	1219	1234.35	6298179	24.2998				
Textual Weight	Decision Tree	5	RELIANCE	21-Feb-2019	1236	1257.8	1229.35	1246.9	10580178	18.7965				
Textual weight	Action	6	RELIANCE	22-Feb-2019	1244.6	1245.3	1226	1232.35	8755865	14.0922				
Time Frame	Action	7	RELIANCE	25-Feb-2019	1236	1243	1220.65	1232.3	7852528	12.9889				
	Actual Change (%)		RELIANCE	26-Feb-2019	1209.5	1234.8	1206	1220.25	10131050	11.4005				
Time Window		9	RELIANCE	27-Feb-2019	1228.05	1244.9	1209	1223.5	11113182	10.4525				
Last Occurrence	TP TN FP FN	10	RELIANCE	28-Feb-2019	1233.75	1239.85	1226.55	1231.05	11286916	10.2935				
Last Occurrence	Authentic Site Filter:	11	RELIANCE	01-Mar-2019	1237	1242.35	1222.25	1226.05	7922513	9.0155				
Dictionary Count	of mancenews.com	12	RELIANCE	05-Mar-2019	1223.4	1239.8	1218.6	1237.65	7121509	8.9509				
	amarketresearchgazet	13	RELIANCE	06-Mar-2019	1239.8	1273.1	1235.1	1264.8	12038231	12.8964				
Score (S)	azdailysun.com	14	RELIANCE	07-Mar-2019	1264	1279.8	1258.15	1270.25	8109259	17.0485				
Annualized Volat	b2breportsnews.com beaumontdaily.com	15	RELIANCE	08-Mar-2019	1266.05	1274.45	1262	1267.1	6040052	19,2124				
Annualized Volat.	capitalmind in	16	RELIANCE	11-Mar-2019	1270.05	1312	1268	13041	9718840	27.4499				
Prop. Score (SD)	chiefobserver.com	17	RELIANCE	12-Mar-2019	1316.9	1334	1314.25	1331.35	11228736	37.3565				
	connectivitymarket.com	18	RELIANCE	13-Mar-2019	1337	1360	13281	1347.3	11236048	44.214				
No. Of Past Days:	c sonewisheroid.com	19	RELIANCE	14-Mar-2019	1349.75	1362	1336.1	1341.55	10402048	46.0472				
90	Get Data		<								, î			

Fig.5 GUI - Data Assembling

Here we consider the technical data of 90 days period of the stock along with annualized volatility. This data s acquired from the above copy file provided by NSE India on its website for each share.

RELIANCE				Reli	ance Indust	tries Limite	d (Listed C	n: 29-Nov	-1995)		
News Date:	Get Data										
Thu, 04-Apr-2019	Volatility										
	voidniny		OPEN	HIGH	LOW	CLOSE	TOTTROQTY	VOLATILITY	SCORET	SCOREZ	Act
Get News	Score	1	1229.75	1249.9	1214	1244.45	9597961	27.4241	26718.77		
		2	1250	1252.5	1215	1220.1	9649017	30.0166			
Find Frequency	Proposed Score	3	1218	1239.7	1211.2	1216.1	6244189	31.0672	28171.77		
rind Prequency	Decision Tree	4	1223.85	1240	1219	1234.35	6298179	24.2998	44079.7		
Textual Weight	Decision Thee	5	1236	1257.8	1229.35	1246.9	10580178		52785.7		
	Action	6	1244.6	1245.3	1226	1232.35	8755865		45813.77		
Time Frame		7	€ 1236	1243	1220.65	1232.3	7852528		87608.7		
-	Actual Change (%)	8	1209.5	1234.8	1206	1220.25	10131050	11,4005	51876.77		
Time Window		9	€ 1228.05	1244.9	1209	1223.5	11113182	10.4525	94433.7		
ast Occurrence	TP TN FP FN	10	1233.75	1239.85	1226.55	1231.05	11286916	10.2935	63841.77		
Cash Orecan ence	Authentic Site Filter:	11	2 1237	1242.35	1222.25	1226.05	7922513	9.0155	95840.7		
Dictionary Count	ofinancenews.com	12	9 1223.4	1239.8	1218.6	1237.65	7121509	8.9509	24141.77		
	amarketresearchgazet	13	9 1239.8	1273.1	1235.1	1264.8	12038231	12.8964	67683.7		
Score (S)	azdailysun.com b2breportenews.com	14	9 1264	1279.8	1258.15	1270.25	8109259	17.0485	28977.7		
Annualized Volat.	beaumontidaily.com	15	9 1266.05	1274.45	1262	1267.1	6040052	19.2124	67251.77		
white void.	capitalmind.in chiefobserver.com consectivitymarket.com ddnewsberald.com	16	1270.05	1312	1268	1304.1	9718840	27.4499	69585.7		
Prop. Score (SD)		17	1316.9	1334	1314.25	1331.35	11228736	37.3565	6870.77		
		18	> 1337	1360	1328.1	1347.3	11236048	44.214	25550.7		
lo. Of Past Days:	< >	19	1349.75	1362	1336.1	1341.55	10402048	46.0472	22475.7		
90	Get Data		¢								3
Chart	Results										

Fig.6 GUI - Score Computation

In this screen compute the score to be passed on to the decision tree model for each of the 90 days. The core data is based on the dated news articles related to the company for each of the 90 days.

RELIANCE				Reli	ance Indust	tries Limite	d (Listed C	m: 29-No	-1995)	
News Date:	Get Data			Pagin	ance arrows	in fee citiers	in (clased c		-1995)	
Thu, 04-Apr-2019	Volatility									
		1)	0PEN 1229.75	HIGH 1249.9	LOW 1214	CLOSE 1244.45	TOTTRDQTY 9597961	VOLATEUTY 27.4241	SCORE1 SCORE2 26718.77 26727.91	Act
Get News	Score	2)	1250	1252.5	1215	1220.1	9649017	30.0166	15436 77 15446 77	
	Proposed Score		1218	1239.7	1211.2	1216.1	6244189	31.0672	28171.77 28182.12	
Find Frequency	Troposed over e	4.2	1223.85	1240	1219	1234.35		24 2998	44079.7_ 44087.8_	
	Decision Tree	5)	1236	1257.8	1229.35	1246.9	10580178	18.7965	52785.7_ 52792.0_	
Textual Weight	Action	6)	1244.6	1245.3	1226	1232.35	8755865	14.0922	45813.77_45818.46_	
Time Frame		77	1236	1243	1220.65	1232.3	7852528	12.9889	87608.7_ 87613.09_	
	Actual Change (%)	8 7	1209.5	1234.8	1206	1220.25	10131050	11.4005	51876.77_51880.57_	
Time Window		0 3	1228.05	1244.9	1209	1223.5	11113182	10.4525	94433.7_ 94437.2_	
Last Occurrence	TP TN FP FN	10 9	1233.75	1239.85	1226.55	1231.05	11286916	10.2935	63841.77_63845.2_	
cust occurrence	Authentic Site Filter:	11 7	1237	1242.35	1222.25	1226.05	7922513	9.0155	95840.7_ 95843.7_	
Dictionary Count	afinancenews.com	12 9	1223.4	1239.8	1218.6	1237.65	7121509	8.9509	24141.77_ 24144.75_	
	amarketresearchgazet	13 9	1239.8	1273.1	1235.1	1264.8	12038231	12.8964	67683.7_ 67688.0_	
Score (5)	ozdailysun.com b2breportsnews.com	14 9		1279.8	1258.15	1270.25		17.0485	28977.7_ 28983.4_	
Annualized Volat	beaumontdaily.com	15 9	1266.05	1274.45	1262	1267.1	6040052	19.2124	67251.77_67258.17_	
reneration rener.	capitalmind in	16)	1270.05	1312	1268	1304.1		27.4499	69585.7 69594.9	
Prop. Score (SD)	chiefobserver.com connectivitymarket.com	17)	1316.9	1334	1314.25	1331.35	11228736		6870.77_ 6883.22	
	ddnewsherold.com	38)	1337	1360	1328.1	1347.3	11236048		25550.7_ 25565.5_	
No. Of Past Days:	¢ >	19 9	1349.75	1362	1336.1	1341.55	10402048	46.0472	22475.7 22491.11	
90	Get Data		<	_						>
Chart	Results									

Fig.7 GUI - Proposed Score Computation

Here we compute the proposed score which takes into account an essential factor called as annualized volatility. Which is an indicator of motion in the stock price? If the volatility is high, there is a more upper movement in stock price which needs to be tapped in for profit.

Select Symbol:	Use Proposed Score											
RELIANCE	*			D	elionce Ted	untrine I im	ited (Listed	On: 20-5	lov-1995)			
News Date:	Get Data					an les cim	neo (cisiec	010 27 1				
Thu, 04-Apr-2019	·											
	Volatility		OPEN	HIGH	LOW	CLOSE	TOTTRDQTY	VOLATILITY	SCORE1	SCORE2	Acti	on
Get News	Score	1	1229.75	1249.9	1214	1244.45	9597961	27.4241		.26727.91.		
Get News		2	1250	1252.5	1215	1220.1	9649017	30.0166		.15446.77.		
-	Proposed Score	3	1218	1239.7	1211.2	1216.1		31.0672		28182.12		
Find Frequency		4	1223.85	1240	1219	1234.35	6298179	24.2998	44079.7	44087.8_	Hold	
Textual Weight	Decision Tree	5	1236	1257.8	1229.35	1246.9	10580178	18.7965	52785.7	52792.0	Hold	
rexitial weight	Action	6	1244.6	1245.3	1226	1232.35	8755865	14.0922	45813.77.	45818.46.	Buy	
Time Frame	Henon	7	1236	1243	1220.65	1232.3	7852528	12.9889	87608.7	87613.09.	Hold	
	Actual Change (%)	8	1209.5	1234.8	1206	1220.25	10131050	11.4005	51876.77.	51880.57.	Hold	
Time Window		9	1228.05	1244.9	1209	1223.5	11113182	10.4525	94433.7_	94437.2	Sell	
Last Occurrence	TP TN FP FN	10	1233.75	1239.85	1226.55	1231.05	11286916	10.2935	63841.77.	63845.2	Hold	
Lost Occurrence	Authentic Site Filter:	11	1237	1242.35	1222.25	1226.05	7922513	9.0155	95840.7	95843.7	Buy	
Dictionary Count	of increases com	12	1223.4	1239.8	1218.6	1237.65	7121509	8.9509	24141.77_	24144.75.	Hold	
	amarketnesearchgazet	13	1239.8	1273.1	1235.1	1264.8	12038231	12.8964	67683.7_	67688.0_	Sell	
Score (S)	azdailysun.com b2breportsnews.com	14	1264	1279.8	1258.15	1270.25	8109259	17.0485	28977.7	28983.4_	Buy	
Annualized Volat	beaumontdaily.com	15	1266.05	1274.45	1262	1267.1	6040052	19.2124	67251.77.	67258.17.	Sell	
Annualized voidt.	capitalmind.in	16	1270.05	1312	1268	1304.1	9718840	27.4499	69585.7_	69594.9_	Buy	
Prop. Scone (SD)	chiefobserver.com	17	1316.9	1334	1314.25	1331.35	11228736	37.3565	6870.77_	6883.22	Hold	
	connectivitymarket.com	18	1337	1360	1328.1	1347.3	11236048	44.214	25550.7	25565.5	Sell	
No. Of Past Days:	< >	19	1349.75	1362	1336.1	1341.55	10402048	46.0472	22475.7_	2249111	Hold	
90	Get Data		6	1	and the second states							5

Fig.8 GUI – Action Strategy Computation

Here the computation of action, i.e. the strategy of trading to be followed is shown on each day for the stock

under consideration. Hold means we carry on with the approach adopted on the previous day.

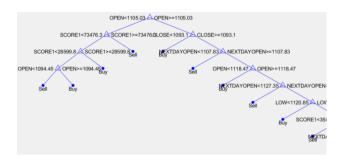


Fig.9 GUI – Decision Tree Model

This screen shows the decision tree constructed based on the technical data values and the computed score. It shows how an action decision arrives.

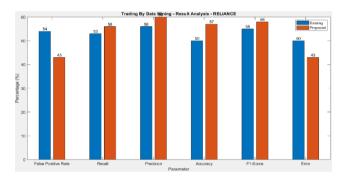


Fig.10 Result Analysis

Here we show the results of the proposed method based on the parameters of False Positive Rate, Recall, Precision, Accuracy, F1-Score and Error Per cent.

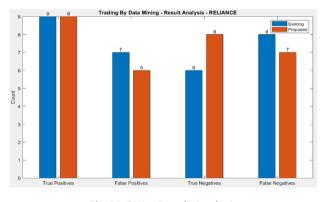


Fig.11 GUI – Result Analysis

Here we show the count of each of the parameters like True Positives, False Positives, True Negatives and False Negatives. Here Positive means Buy strategy suggested by the system and Negative means Sell.

Conclusion

Proposed method demonstrates a new method for the advance reading of stock movements and helping

financial decisions in trading resting on a mix of financial information and news mining methods. The trend deciding factors of the proposed system make use of text material that dated present in web pages along with customary numeric price time series information, present on the Internet. The proposed trend deciding system makes use of more than one system thereby increasing chances of accurate decisions. The method is simulated and assessed based on actual time series data and news information.

References

- J. Clerk Maxwl. S. Jacobs and C. P. Bean, "Fine particles, thin films, and exchange and T. Bollerslev, Generalized Autoregressive Conditional Heteroscedasticity," Journal of Econometrics, 31, 307-327, 1986.
- [2]. T. Bollerslev," A Conditionally Heteroskedastic Time Series Model For Speculative Prices and Rates of Return," Review of Economics and Statistics, 69(3), 542-546, 1987.
- [3]. E.F. Fama, Random Walks in Stock Market Prices, Financial Analysts Journal, September/ October 1965 (reprinted in January-February 1995).
- [4]. G. Gid6falvi, 2001. Using News Articles to Predict Stock Price Movement, Online at, [http://citeseer.nj.nec.com/ 517027.html].
- [5]. V. Lavrenko, M. Schmill, D. Lawrie, P. Ogilvie, D. Jensen, and J. Allan, Language Models for Financial News Recommendation, CIKM 2000, McLean, VA USA, ACM 2000.
- [6]. M.A. Kaboudan, Genetic Programming Prediction of Stock Prices, Computational Economics 16: 207-236, 2000.
- [7]. E.F. Fama, Long Term Returns and Behavioral Finance, Social Science Research Network.
- [8]. M. Last, Y. Klein and A. Kandel, Knowledge Discovery in Time Series Databases, IEEE Transactions on Systems, Man and Cybernetics -Part B: Cybernetics, Vol. 31 No. 1, February 2001.
- [9]. Z. Bodie, A. Kane, A.J. Marcus, Investments, 4th Edition, McGraw Hill, 2001.
- [10]. E. F. Fama, Efficient Capital Markets: A Review of Theory and Empirical Work, Journal of Finance, 25 (May 1970): 383-417.
- [11]. R.A. Huagen, The New Finance: The Case Against Efficient Markets. Prentice-Hall, 1995.
- [12]. R. Engle and T. Bollerslev, "Modelling the Persistence in Conditional Variances," Econometric Reviews, 5, 81 -87, 1986.
- [13]. A.K. Jain, M.N. Murty, P.J. Flynn, Data Clustering: A Review, ACM Computing Surveys, Volume 31, No. 3, September 1999.
- [14]. R. Cooley, B. Mobasher, and J. Srivastava, "Web Mining: Information and Pattern Discovery on the World Wide Web," In Proceedings of the IEEE International Conference on Tools with Artificial Intelligence (ICTAI'97), Newport Beach, CA, November 1997.
- [15]. R. Kosala and H. Blockeel, "Web Mining Research: A Survey, SIGKDD Explorations," Volume 2, Issue 1.
- [16]. O. Maimon, A. Kandel and M. Last, Knowledge Discovery and Data Mining, The Info-Fuzzy Network (IFN) Methodology, Norwell, MA: Kluwer, 2000.
- [17]. D. Peramunetilleke, R.K. Wong, "Currency Exchange Rate Forecasting from News Headlines," Thirteenth Australasian Database Conference (ADC2002), Melbourne, Australia, Conferences in Information Technology, Vol. 5.
- [18]. B. Wuthrich, V. Cho, S. Leung, D. Permunetilleke, K. Sankaran, J. Zhang, W. Lam, "Daily Stock Market Forecast from Textual Web Data," In IEEE International Conference on Systems, Man. and Cybernetics, Volume: 3, Page(s): 2720 -2725, 1998.
- [19]. L. Torgo, The TNT Financial Trading System: a midterm report, ECML-PKDD Workshop on Data Mining for Business, 2005.
- [20]. J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufman Publishers Inc., San Francisco, CA, 1993.
- [21]. R. Landry Jr., R. Debreceny, G.L. Grey, "Grab Your Picks and Shovels! There's Gold in Your Data, Strategic Finance", January 2004, (85, 7).
- [22]. Extractor DBI technologies (2003) [http://www.dbitech.com]