FEATURE SELECTION FOR PRICE CHANGE PREDICTION

Zehra Çataltepe[1*], Savaş Özer[2], Vahide Unutmaz Barın[2]
[1]Istanbul Technical University, Faculty of Computer and Informatics, Maslak, Istanbul, 34469 Turkey.
[2] Rasyonet, İTÜ Arı Teknokent 2, A Blok 4/2, Maslak, Istanbul, 34469 Turkey.
cataltepe@itu.edu.tr, savas@rasyonet.com, vahide@rasyonet.com

Abstract: *We investigate the use of feature selection and classification for better prediction of the change in the next day close price of 13 Istanbul Stock Exchange bank stocks. We use filter type feature selection methods and evaluate feature relevance and redundancies using four different measures, namely, correlation, mutual information and two different kinds of normalized mutual information. We find out that using correlation measure together with scoring using relevance-redundancy performs better than all the other methods in general. We also find out that this fast filter type feature selection is both more accurate and faster than wrapper type forward feature selection. We use radial basis function networks as our classifiers and find out that different stocks perform best using different model parameters, which point out to the fact that model evaluation and selection, as well as feature evaluation and selection need to be performed for each stock separately.*

Keywords: Stock price prediction, Istanbul Stock Exchange (ISE), feature selection, mRMR (minimum Redundancy Maximum Relevance) feature selection, stable feature selection, radial basis function (RBF) networks, mutual information.
Subject headings: 1. Algorithmic trading 5. Computational methods

1. INTRODUCTION

With the advances in computer technology, stock prices and other financial indicators related to major markets are readily and easily available. Machine learning techniques and software for analysis of these data have also become more widely available. Radial Basis Function (RBF) Networks, artificial neural networks (ANNs) or support vector machine (SVMs) are common machine learning/pattern recognition algorithms used for stock price prediction [Atsalakis&Valavanis, 2009].

When there are many different inputs, in order to achieve good generalization (i.e. the training accuracy of a model is a good indicator of it test accuracy), machine learning algorithms usually need a large number of instances. However, for the stock market, the number of instances is usually quite low, necessitating models with as little number of informative inputs (features) as possible. Feature selection methods [Guyon&Elisseeef, 2003] have been used in machine learning to identify a smaller set of informative features, so that the trained classifiers are both fast and accurate on unknown test instances.

In this study, we evaluate different feature selection methods for prediction of the next day stock price direction for 13 bank stocks from the Istanbul Stock Exchange. First of all, we use filter methods which are fast and evaluate the value of a feature using its relevance and redundancy. We evaluate relevance (feature-label similarity) and redundancy (feature-feature similarity) with four different similarity measures, namely, correlation, mutual information and two different kinds of mutual information. As the feature selection algorithm, we use mRMR (minimum Redundancy Maximum Relevance) [Peng et.al., 2005] algorithm, which is a recent and accurate information theoretical feature selection algorithm.

2. METHODS USED

3.1.Notation

* Corresponding author.

We assume that we are given *d* features (inputs) for days *1..T* of a stock in a matrix X^{all}_{Txd}. We use the first *1..N* days for training and days *T+1..N* for testing of our algorithms. We use **x**(t) to denote the d dimensional feature vector for day t.

For each day, we define the *target output* based on whether the stock close price *p(t)* increased or remained the same (+1) or decreased (-1) compared to the previous day:

$$r(t) = \begin{cases} +1, if\ p(t) \geq p(t-1) \\ -1, \qquad otherwise \end{cases} \tag{1}$$

We denote the *training set* using $Z = \{\mathbf{X},\mathbf{r}\}$ where **X,** are the rows 1..N of the matrix X^{all}_{Txd} and **r** are the target outputs for days *1..N.* We use the training set for selection of features as well as training (computing the optimal parameters) of our classifiers. A classifier is a mapping which tries to approximate the target output given the inputs $g(\mathbf{x}(t)) \in \{-1,+1\}$. The accuracy of a classifier is measured on the test set as:

$$acc(g) = \frac{1}{T-N} \sum_{t=N+1}^{T} (g(\mathbf{x}(t) - r(t))^2 \tag{2}$$

3.1. mRMR Feature Selection Algorithm:

mRMR (minimum Redundancy Maximum Relevance) [Peng et.al., 2005] feature selection algorithm is a fast and accurate feature selection algorithm. It is fast, because it is a filter [Guyon&Elisseeef, 2003] type feature selection algorithm, which means that, unlike the wrapper type feature selection algorithms, it does not need to train classifiers in order to determine whether a feature should be included in the selected set or not. It is also accurate, because unlike PCA or ICA, it does not ignore the class labels and uses them to determine how important (i.e. relevant) each feature is. mRMR is a forward type feature selection algorithm, it starts with an empty set of features and adds features one by one, based on their relevance (i.e. how correlated they are with the class label) and redundancy (how correlated they are with the already selected set of features).

For the time being, we will use $\mathbf{f}_i$ to denote the vector of values feature i ($\in$ 1..d) takes on the training set and **r** the vector of target outputs on the training set. $Rel(i) = sim(\mathbf{f}_i,\mathbf{r})$ to denote the *relevance* of feature $\mathbf{f}_i$ with the class label **r** and we will use $Red(i,j) = sim(\mathbf{f}_i, \mathbf{f}_j)$ to denote the redundancy between two features. The *redundancy* between a feature $\mathbf{f}_i$ and a set of features *S, Red(i,S),* is the average redundancy between $\mathbf{f}_i$ and all the features in *S*.

mRMR algorithm is a forward feature selection algorithm. It returns a set of features $S \subseteq \{1,\ldots,d\}$. First of all, it includes the most relevant feature in the selected set of features S. The remaining features are selected to maximize their relevances and minimize their redundancies with S. A score can be computed for a feature $j \in \{1,\ldots,d\}$-S according to either MID(j)=Rel(j)-Red(j,S) or MIQ(j) = Rel(j)/Red(j,S). Both scoring mechanisms have been used in different applications [Peng et.al., 2005; Gulgezen et.al., 2009], we will be using both scoring mechanisms in our experiments. Please see [Peng et.al., 2005] for a detailed analysis of the mRMR algorithm.

3.2. Feature Relevance/Redundacy Measures

In order to compute the feature relevance and redundancy in mRMR, the same similarity measure is used. In the original work of [Peng et.al., 2005], mutual information is used as the similarity measure. Mutual information computation requires binning of the argument vectors, which may affect the results. In this paper, in addition to mutual information, we use feature correlation and two different types of normalized mutual information for feature relevance and redundancy computation. The mutual information has some drawbacks which can be eliminated by the normalized mutual information. We use two different types of normalization. The avg normalization corresponds to the symmetric uncertainty which was used in FCBF feature selection algorithm of [Yu&Liu, 2003]. The max normalized mutual information is used in [Esteves, et. al., 2009]. Both methods divide the mutual information between two variables, by a function of their individual entropies.

- **Correlation:** is the statistical correlation between two N dimensional vectors **x** and **y:**

$$corr(x,y) = \frac{1}{N}\sum_{k=1..N} x(k)y(k) - \frac{1}{N^2}\sum_{k=1...N} x(k)\sum_{k=1...N} y(k) \quad (3)$$

- **Mutual Information:** In order to compute the mutual information two different vectors, first of all they are discretized so that they can take only discrete values and the marginal and joint probabilities can be estimated.

  Entropy of a feature f_i is computed as: $H(f_i) = -\sum_{x \in X} P(x) \log P(x)$ where X shows the values feature x can take and log is taken in base 2 and p(x) is computed based on the occurrances of the features in the training set.

  Mutual information between two features is given as:

$$MI(f_i, f_j) = -\sum_{y \in Y}\sum_{x \in X} p(x,y) log \frac{p(x,y)}{p(x)p(y)} \quad (4)$$

  where features takes values from X and Y and probabilities are computed based on the occurrances of the features in the training dataset.

- **Normalized Mutual Information (avg):**

$$NMI_{avg}(f_i, f_j) = \frac{MI(f_i, f_j)}{H(f_i) + H(f_j)} \quad (5)$$

- **Normalized Mutual Information (max):**

$$NMI_{max}(f_i, f_j) = \frac{MI(f_i, f_j)}{\max(H(f_i), H(f_j))} \quad (6)$$

3. RESULTS

3.1. DataSets

We used stock prices of 13 banks ALNTF, AKBNK, FINBN, DENIZ, VAKBN, HALKB, GARAN, TEKST, TEBNK, YKBNK, ISCTR, TSKB, SKBNK from Istanbul Stock Exchange in between 07.11.2007-28.10.2010. The portion of the data between 07.11.2007 and 30.03.2010 (572 Data points) is used for training, the data between 31.03.2010 and 28.10.2010 (144 Data points) is used for testing.

We used the 49 features given in Table 1. The features were obtained using the Stockground Financial Analysis Software by Rasyonet (www.rasyonet.com). The abbreviations used in the table are: DAX: Deutscher Aktien Index, DJI: Dow Jones Industrial Index, XU100: Istanbul Stock Exchange National 100 Index, ROC: Rate of Change, EMA: Exponential Moving Average indicator, WillR: Williams' %R indicator.

Table 1: *The features used and their explanations.*

Feature No.	Feature Name	Explanation
1	ADI	Accumulation/Distribution indicator
2	BBW	Bollinger Bands Width indicator
3	Close	Close price of security
4	DAX	DAX Index
5	DAX(EMA10)	EMA10 of DAX Index
6	DAX(EMA20)	EMA20 of DAX Index
7	DAX(EMA30)	EMA30 of DAX Index
8	DAX(ROC12-Close)	ROC12 of Close value of DAX Index
9	DAX(ROC12-High)	ROC12 of High value of DAX Index
10	DAX(WillR14)	WillR of DAX Index
11	DAX-2	(Close - EMA14) / (Close + EMA14) of DAX
12	DEMA	Double Exponential Moving Average
13	DJI	DJI index
14	DJI(EMA10)	EMA10 of DJI index
15	DJI(EMA20)	EMA20 of DJI index
16	DJI(EMA30)	EMA30 of DJI index
17	DJI(ROC12-Close)	ROC12 of Close value of DJI index
18	DJI(ROC12-High)	ROC12 of High value of DJI index
19	DJI(WillR14)	WillR of DJI

20	DJI-2	(Close - EMA14) / (Close + EMA14) of DJI
21	EMA	EMA25
22	EOM	Ease of Movement in a period of 14 days
23	F/DD	Price / Book Value ratio
24	F/K	Price / Earnings ratio
25	Inertia	Inertia indicator
26	JPY_DA	TRL/JPY
27	Log(C(t)/C(t-2))	Logrithm of Close(t) / Close(t-2)
28	Log(C(t)/C(t-3))	Logrithm of Close(t) / Close(t-3)
29	Log(C(t)/C(t-4))	Logrithm of Close(t) / Close(t-4)
30	Log(C(t)/C(t-5))	Logrithm of Close(t) / Close(t-5)
31	MACDHist	Moving Average Convergence-Divergence Histogram
32	MF	Market Facility
33	ProjO	Projection Oscillator
34	ROC	Rate of change
35	RSI	Relative Strength Index
36	TEFE	Wholesale Price Index
37	TRIX	Triple exponential
38	TUFE	Consumer Price Index
39	USD_DA	TRL/USD
40	VROC	Volume Rate of Change
41	WAD	Williams' Acc/Dist
42	XU100	XU100 Index
43	XU100(EMA10)	EMA10 of XU100 Index
44	XU100(EMA20)	EMA20 of XU100 Index
45	XU100(EMA30)	EMA30 of XU100 Index
46	XU100(ROC12-Close)	ROC12 of Close value of XU100 Index
47	XU100(ROC12-High)	ROC12 of High value of XU100 Index
48	XU100(WillR14)	WillR of XU100 Index
49	XU100-2	(Close - EMA14) / (Close + EMA14) of XU100

3.2. Machine Learning and Feature Selection Methods

All experiments were performed using Radial Basis Function (RBF) networks as the machine learning method. Different number of clusters (4-9) and the same value of ridge parameter (0.0001) was used.

In order to determine the performance of feature evaluation and scoring methods, all 8 combinations of them were used for each dataset and the number of clusters and the training and test accuracies were recorded for each number of features 1..49.

3.3. Experimental Results

In Figure 1, we show the test accuracies obtained at the best training accuracy while using mRMRM selected features, considering all 8 different similarity and scoring combinations. Since the test data is not available, one would decide on how many features to select based on either the training data or by means of cross validation. Therefore, the accuracies shown with (best acc to Train) are accuracies that would be obtained using a method like this. In the figure, we also show the best possible accuracies if one was able to glance at the test data, which is not possible. The number of RBF clusters used and the feature evaluation and scoring method used for each dataset is different in Figure 1. One interesting and not so surprising conclusion from Figure 1 is the fact that for different stocks, feature selection and classification performs differently. While accuracies above 0.64 are obtained for DENIZ, FINBN, TEKST, ALNTF, accuracies for AKBNK and VAKBN are 0.55.

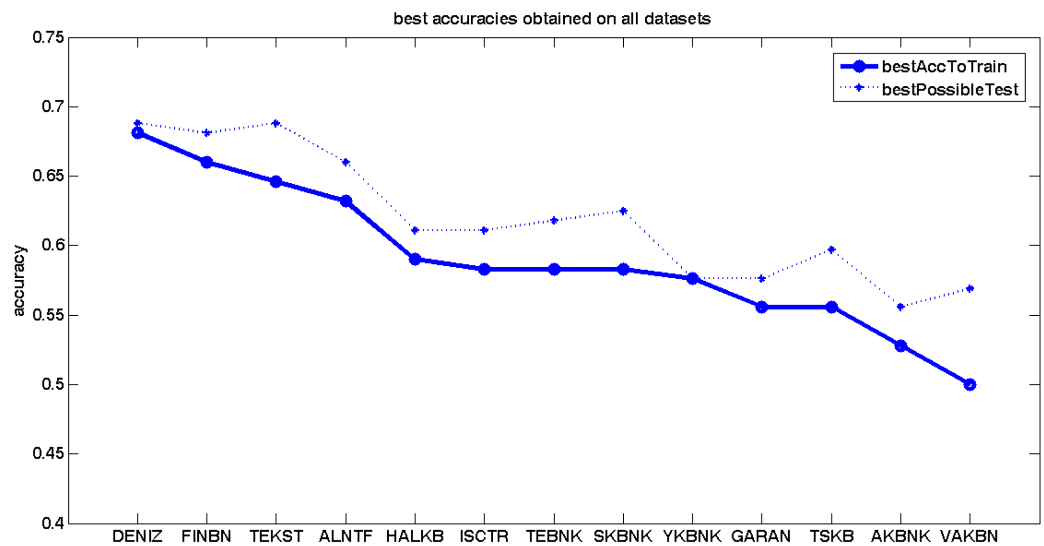

Figure 1: *Best test accuracies obtained using mRMR feature selection on 13 different ISE stocks (shown in bold). For comparison the best possible test accuracies (obtained by choosing the number of mRMR features that gives the best test error) with mRMR(bestPossibleTest) are shown.*

Table 2: *Best test accuracies obtained using different feature relevance/redundancy measures (Corr,MI, NMIAvg,NMIMaxfeature) and their combination for score in mRMR (MID, MIQ) on 13 different ISE stocks. The last line shows the average normalized score of each method over all 13 datasets.*

		Corr		MI		NMIAvg		NMIMax		
Dataset	#RBF cl	MID	MIQ	MID	MIQ	MID	MIQ	MID	MIQ	maxAcc
AKBNK	5	0.514	0.514	**0.528**	0.521	0.521	0.521	0.521	0.521	0.528
ALNTF	7	0.618	0.618	0.611	**0.632**	0.549	**0.632**	0.549	0.549	0.632
DENIZ	5	**0.681**	0.667	0.681	0.667	**0.681**	**0.681**	**0.681**	**0.681**	0.681
FINBN	7	0.639	0.472	**0.66**	**0.66**	**0.66**	**0.66**	**0.66**	**0.66**	0.66
GARAN	9	0.528	0.521	0.514	0.528	**0.556**	0.528	0.514	0.528	0.556
HALKB	9	0.535	0.521	**0.59**	0.521	0.59	0.549	0.549	0.542	0.59
ISCTR	7	**0.583**	0.549	0.514	0.549	0.514	0.528	0.514	0.451	0.583
TEBNK	9	0.563	**0.583**	0.528	0.556	0.583	0.542	0.576	0.542	0.583
SKBNK	9	**0.583**	0.535	0.535	0.507	0.528	0.5	0.472	0.465	0.583
TEKST	5	0.618	0.639	**0.646**	**0.646**	**0.646**	0.556	0.639	0.618	0.646
TSKB	7	**0.556**	0.451	0.528	0.521	0.528	0.507	0.528	0.486	0.556
VAKBN	9	0.493	0.486	**0.5**	0.451	**0.5**	0.451	0.493	0.451	0.5
YKBNK	7	0.514	0.514	0.507	0.465	0.535	**0.576**	0.514	0.486	0.576
Avg Norm. Score		**0.97**	0.92	**0.96**	0.94	**0.96**	0.94	0.94	0.91	
Errorbar		0.01	0.02	0.01	0.02	0.02	0.02	0.02	0.02	

In Table 2, we show the particular number of best RBF clusters for each dataset and the accuracies obtained using different feature evaluation and scoring methods for that particular number of RBF clusters. According to the table, different number of RBF clusters seem to best for each different dataset. Therefore, it is definitely a must that model selection needs to be employed for stock market prediction for each specific dataset. Another observation is that the *MIQ* is in general worse than *MID*. In order to compare each method fairly, the last two lines of Table 2 show the normalized score of each method. The normalized score for each of the 8 feature evaluation and selection methods *(Corr, MI, NMImax, NMIavg)x(MID,MIQ)* on a particular dataset is obtained by dividing the accuracy of the method with the best possible accuracy for that dataset (shown in the next to the last column). According to the normalized scores, the normalized mutual information measures, *Corr-MID* is the best followed by *MI-MID and NMIavg* –MID.

Table 3: *The number of features used for the best test accuracies in Table 2.*

	Corr		MI		NMIAvg		NMIMax	
Dataset	MID	MIQ	MID	MIQ	MID	MIQ	MID	MIQ
AKBNK	3	4	**5**	10	1	1	1	1
ALNTF	12	34	1	**11**	2	**12**	2	2
DENIZ	**5**	6	3	1	**4**	**4**	**4**	**4**
FINBN	6	13	**1**	**1**	**1**	**1**	**1**	**1**
GARAN	8	9	10	25	**15**	34	16	34
HALKB	37	29	**17**	10	**18**	33	19	15
ISCTR	**18**	16	3	10	4	11	4	2
TEBNK	22	**3**	5	16	5	14	6	7
SKBNK	**22**	4	4	8	4	37	17	30
TEKST	1	3	**5**	**6**	**6**	3	3	4
TSKB	**17**	13	28	15	29	17	29	11
VAKBN	28	45	**2**	17	**3**	18	12	18
YKBNK	27	42	4	2	22	5	13	35
Avg	15.85	17	6.77	10.15	8.77	14.62	9.77	12.62
Errbar	3.08	4.23	2.13	1.93	2.53	3.55	2.4	3.56

In Table 3, we show the number of features used in order to obtain the best accuracies shown in Table 2. Considering the last line which shows the average number of features used for each dataset, in general *MID* uses less number of features than *MIQ*. Considering Tables 2 and 3 together, *MID* is better than *MIQ* both in terms of time and accuracy.

We also wanted to see how mRMR feature selection accuracies compare with other methods of feature selection and classification. For comparison we used three different methods:

- **Forward wrapper feature selection (fwd):** We used wrapper feature selection using an RBF with as many clusters as the optimal number of RBF clusters in Table 2 for each data set. We stopped feature selection when the training accuracy did not improve anymore.
- **majority train label (StrategyMajority):** In this trading strategy, the majority class label is computed on the training set and that label is assumed to hold for all test instances.
- **tomorrow's change same as today's (StrategySame):** The label for time t is assumed to be the same as the label at time t-1.

In Table 4, we show the comparison of these four methods with the best accuracy obtained by mRMR feature selection for all 13 datasets. mRMR based methods perform better than all three methods. Forward feature selection is not able to choose good features and is also slow, so mRMR based methods should be considered instead of forward feature selection. It is also worth noting that the simple strategy of assuming tomorrow's price change will be the same as today's is able to perform better than forward feature selection.

Table 4:*Comparison of best accuracies obtained using mRMR with forward feature selection and two simple algorithms.*

Dataset	Max mRMR	Forward Feature Select	Training Data Majority	Same Class As Today
AKBNK	**0.528**	0.69	0.472	0.514
ALNTF	**0.632**	0.528	0.514	0.618
DENIZ	**0.681**	0.542	0.542	**0.681**
FINBN	**0.66**	0.486	0.438	0.639
GARAN	**0.556**	0.521	0.5	0.528
HALKB	**0.59**	0.458	0.472	0.549
ISCTR	**0.583**	**0.583**	0.451	0.514
TEBNK	0.583	**0.653**	0.514	0.549
SKBNK	**0.583**	0.528	0.542	0.556
TEKST	**0.646**	0.528	0.486	0.618
TSKB	**0.556**	0.549	0.465	0.542
VAKBN	0.5	**0.549**	0.542	0.507
YKBNK	**0.576**	0.486	0.528	0.514
Avg Norm. Score		0.94	0.85	0.95
Errorbar		0.05	0.03	0.01

Finally, we found out the highest ranked features for all 13 banks, by summing up their rank in the Corr-MID ranking upto and including 38th feature for each bank. In Table 4 we provide those features. It is interesting (and expected) that for bank stock price prediction, the price from the past as well as the currency exchange rates and Istanbul and other major stock market indices are helpful. TRIX and WAD are bank specific indicators which have been useful for prediction more than the other indicators.

Table 4:*Most commonly selected features by the Corr-MID feature selection for all 13 banks.*

1	XU100(EMA10)	6	DAX(EMA10)	11	DJI(EMA20)	16	DJI(ROC12-Close)
2	USD_DA	7	DJI(EMA10)	12	DJI	17	DAX(EMA20)
3	DJI(ROC12-High)	8	XU100(EMA20)	13	DJI(WillR14)	18	DJI(EMA30)
4	**TRIX**	9	DJI-2	14	DAX-2	19	**WAD**
5	DAX(EMA30)	10	JPY_DA	15	DAX(ROC12-High)	20	DAX(ROC12-Close)

3.3. Related Work

There is some previous related work on feature selection for stock market prediction. Wrapper feature selection was used to select the best features for different classifiers, such as kNN, SVM, neural network in [Huang et.al., 2008]. The outputs of these trained classifiers were then combined using voting. They used 11 features and claim that their method results in the best performance among the filter type feature selection methods that they used. The filter type methods they used did not include mRMR and except CFS, they did not take into consideration the redundancy between the features. Also, they worked on Taiwan and Korea stock price prediction, while we work on price prediction for specific stocks in banking in Istanbul Stock Exchange. [Huang&Tsai, 2009] used filter type feature selection as a step in order to select features for stock price prediction. They clustered instances based on the selected features and then used SVR (support vector regression) to predict the stock price. Genetic algorithms were used by [Kim, 2006] for both instance and feature selection for Korean Stock Price Index. They used neural networks as their classifiers.

There are also a number of studies on Istanbul Stock Exchange prediction, for example, [Boyacioglu&Avci, 2010], [Yumlu, et.al., 2005], [Bildirici&Ersin, 2009]. In these studies different models were used for learning the stock price behavior, however, no feature selection was performed.

4. CONCLUSIONS

In this paper, we have investigated the use of mRMR type filter feature selection algorithms for stock price direction prediction. We have studied 13 different bank stocks from the Istanbul Stock Exchange and used 49 different features as inputs. We have evaluated four different feature similarity measures to determine feature relevance and redundancies, namely, correlation, mutual information, normalized (avg) mutual information and normalized (max) mutual information. We also evaluated two different methods of feature scoring, MID which uses relevance-redundancy and MIQ which uses relevance/redundancy in order to determine whether a feature should be included in the selected feature set or not. We have observed that MID consistently results in less number of features and more accuracy for all four similarity measures. Among the similarity measures, correlation is the most useful measure, followed by mutual information. We have also observed performance of the feature selection and machine learning algorithms are different for each stock.

References

Atsalakis, G.S., Valavanis, K.P., Surveying stock market forecasting techniques – Part II: Soft computing methodsExpert Systems with Applications 36 (2009) 5932–5941

Bildirici, M., Ersin, Ö.Ö., Improving forecasts of GARCH family models with the artificial neural networks: An application to the daily returns in Istanbul Stock Exchange Expert Systems with Applications 36 (2009) 7355–7362.

Boyacioglu, M.A., Avci, D., An Adaptive Network-Based Fuzzy Inference System (ANFIS) for the prediction of stock market return: The case of the Istanbul Stock Exchange, Expert Systems with Applications, Volume 37, Issue 12, December 2010, Pages 7908-7912.

Estévez, P.A., M. Tesmer, M., Perez, C.A., Zurada, J.M., Normalized Mutual Information Feature Selection, IEEE Trans. On Neural Networks, VOL. 20, NO. 2, February 2009 189-201.

Gulgezen, G., Cataltepe, Z., Yu, L., Stable and Accurate Feature Selection, ECML/PKDD 2009, Bled, Slovenia.

Guyon, I., Elisseeef, A., 2003. An introduction to variable and feature selection. Journal of Machine Learning Research 3, 1157-1182.

Huang, C.L., Tsaia, C.Y., A hybrid SOFM-SVR with a filter-based feature selection for stock market forecasting, Expert Systems with Applications, Volume 36, Issue 2, Part 1, March 2009, Pages 1529-1539

Huang , C.J., Yang, D.X., Chuang, Y.T., Application of wrapper approach and composite classifier to the stock trend prediction, Expert Systems with Applications 34 (2008) 2870–2878.

Kim, K., Artificial neural networks with evolutionary instance selection for financial forecasting, Expert Systems with Applications 30 (2006) 519–526.

Lu, C.H., Chen, T.C., A study of applying data mining approach to the information disclosure for Taiwan's stock market investors, Expert Systems with Applications 36 (2009) 3536–3542.

Pan, H., Tilakaratne, C., Yearwood, J., Predicting Australian Stock Market Index Using Neural Networks Exploiting Dynamical Swings and Intermarket Influences, Journal of Research and Practice in Information Technology, Vol. 37, No. 1, February 2005.

Peng, H., Long, F., Ding, C., 2005. Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans. Pattern Anal. Mach. Intell 27 (8), 1226-1238.

Sindhwani, V ., Rakshit, S, Deodhare, D., Erdogmus, D., Principe, J., and Niyogi, P., Feature Selection in MLPs and SVMs based on Maximum Output Information, Neural Networks, IEEE Transactions on, 15:4, 937--948, 2004.

Yu, L., Liu, H., 2003. Feature selection for high-dimensional data: A fast correlation-based _lter solution. In: Fawcett, T., Mishra, N. (Eds.), ICML 2003, August 21-24, 2003, Washington, DC, USA. AAAI Press, pp. 856-863.

Yumlu, S., Gurgen, F. G., & Okay, N. (2005). A Comparison of global, recurrent and smoothed-piecewise neural models for Istanbul Stock Exchange prediction. Pattern Recognition Letters, 26, 2093–2103.