International Journal of Engineering & Technology, 8 (1.11) (2019) 143-149

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

A Hybrid Model to Forecast Financial Time Series Based on Technical Analysis and Support Vector Machines

Houda Benkerroum[1], Walid Cherif[2], Mohamed Kissi[1]*

[1] *LIM laboratory , Department of Computer Science, University Hassan II Casablanca, Faculty of Sciences and Technology, B.P. 146, 20650, Mohammedia, Morocco*

[2] *Lboratory SI2M, Department of Computer Science, National Institute of Statistics and Applied Economics, B.P. 6217, Rabat, Morocco*

**Corresponding author E-mail:kissim@gmail.com*

Abstract

The aim of this paper is to find functional relations in the behaviour of the USD swaps daily level time series and, in turn, forecast future values of the series through applying a relevant machine learning technique. As our original dataset variables appeared to show strong cross correlation, we decided to use Principal Component Analysis (PCA) to process the data before passing it to our machine learning algorithm. Then, we extract some technical indicators from our historical product price time series and use them as inputs to our model. Finally, Support Vector Machines (SVMs) is applied to our processed data set to realise the forecasting, and the resulting time series can be used to generate signals of when to enter or unwind a trade. Analysis of the results demonstrates that it is advantageous to apply SVMs to forecast financial time series, based on the criteria of Root Mean Square Error (RMSE) and F-measure (F score)

Keywords*: Financial time series; Machine Learning; Principal Component Analysis; Support Vector Machines; USD swaps.*

1. Introduction

An interest rate swap is a financial derivative contract between two parties, who agree to exchange fixed interest rate payments for floating rate payments. Swaps are then useful when a company wants to receive a payment with a variable interest rate, while the other wants to limit future risk by receiving a fixed-rate payment instead.

The fixed interest rate is the quoted rate of a swap (and is then referred to as the ‘price’ of the swap), and depends upon the duration or maturity of the contract. The fixed rate can be seen as the expected value of short term interest rates over the contract duration [1], and is often treated as a proxy for the borrowing costs of large corporations. The swap rate is measured in percent, and daily changes are measured in basis points, where one basis point is equal to 0.01%. – We will adopt this convention in this paper.

Thereafter, modelling and analysing the risk inherent in term structure of interest rates is a crucial task for owners of fixed income securities, such as many banks and investors, who have a large portion of interest rate sensitive structures in their portfolios.

Financial time series prediction is usually considered as one of the most challenging issues among time series predictions, due to its noise and volatile features. Some of the most widely used machine learning techniques to predict financial time series are Artificial Neural Networks (ANNs) [2], [3] and the Support Vector Machine (SVM) [4], [5], [6]. However, we notice recently the emergence of a new trend considering that a deep nonlinear topology would be more accurate for time series prediction [7]. One should note also that feature engineering and extraction are key parts of the machine learning workflow. They are about transforming training data and augmenting it with additional features, in order for our machine learning algorithms to achieve better performance. Hence, considering the complexity of financial time series, combining machine learning techniques with financial market forecasting is definitely a very interesting topic.

2. Methodology and data

2.1. Data description

Our data consist of both USD swaps and USD forward swaps daily closing prices for different tenors and starting dates, from January 2013 through December 2017, imported form Bloomberg. We decided to focus on USD swaps liquid instruments in order to avoid returns being contaminated by illiquidity or stale price issues and to match closely an implementable strategy at a significant trade size.

Also we used a short data history of 5 years because our aim is to capture only short-term changes in yield dynamics, through using both Principal Component Analysis (PCA hereafter) and Support Vector Machines (SVM hereafter) on differenced data. Furthermore, using a too large data history would lead to capturing more rate regimes (a ‘rate regime’ being characterised as ‘easing cycle’, ‘hiking cycle’ or ‘on hold’), and thus to a bias in our model [8].

2.2. Input variables and pre-processing

Usually when raw data is filled with noise, our mining algorithm's accuracy could be affected. In the fixed income market for example, the daily closing price is influenced by multiple factors, resulting in lot of noise and making it harder to observe long-term features of our time series. Therefore, it is better practice to pre-process the original raw data and work on the pre-processed information.

First, we process our dataset by using daily price changes time series instead of daily prices alone as the former are more stationary and approximately i.i.d. (independent and identically distributed), although there is some conditional heteroscedasticity. Under the assumption of i.i.d. observations, our PCA would generalize from sample to population [9].

Second, we scale all our time series (daily price changes of 96 fixed income instruments through a 1229-day period) before implementing PCA.

Finally, we extract 3 of the most popular technical indicators calculated from the historical product prices that we want to forecast (our 'output'): SMA15, MACD26, and RSI14 - these 3 technical indicators will be explained in more details in the paragraph 2.3.2. of our study.

2.3. Methodology

2.3.1. PCA implementation

As discussed in the data description section, we will apply PCA on our processed time series of basis point changes of over 100 fixed income instruments including swaps, forward swaps and butterflies, with different and non-overlapping maturities – "Butterfly" being a trading strategy that consists of buying and selling three different terms of a given swap/forward swap.

Principal component analysis (PCA) is a statistical procedure that uses an orthogonal transformation to convert a set of correlated variables into linearly uncorrelated variables (principal components) [10]. We are using it here as a way to create better features for our SVM, the improved features being the projection, through PCA, of our original dataset and are hence uncorrelated and linear combinations of the original ones.

One should note that by applying PCA, we were able to identify the main factors that have historically driven the shape of the swap curves. Also, in addition to decorrelating our inputs, PCA offers two major advantages: 1) identifying the true sources of variations in term structures, hence risk, in the historical time series, and 2) reducing the number of risk factors down to a more manageable size.

After performing the eigen-decomposition of our dataset's covariance matrix, we choose to keep the two first principal components (i.e. the first two main drivers of our dataset) as they explain more than 98% of the overall dispersion of our data - as shown on the Figure 1 below.

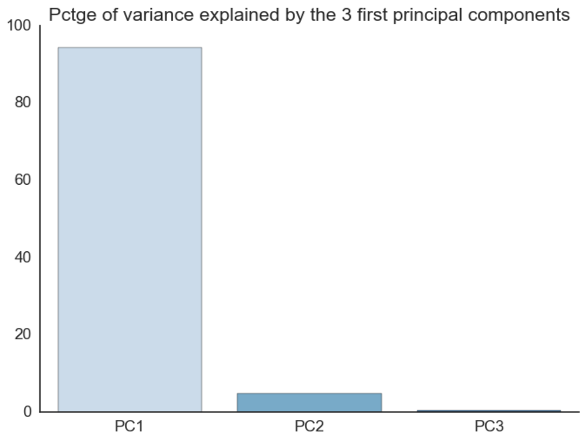

Fig. 1: Percentage of variance explained by the three first principal components

Now that we have implemented the PCA, we are able to reconstruct our dataset, through projecting our data to the new coordinates system, then getting the coordinates of the fitted points in terms of the original coordinate system (Figure 2).

Fig. 2: The 2y5y10y butterfly price levels time series implied by the PCA (in pink) along with our original data.

2.3.2. Technical indicators

In literature, different types of technical indicators are widely used by researchers/investors to provide insights on market trends in financial markets and to generate trading signals [11], [12], [13]. In this study, three of the most popular technical indicators i.e. SMA15, MACD26, and RSI14, are used as input to our model. They are calculated from historical prices as follows:

- Simple Moving Average (SMA):

The SMA is the simple statistical mean of previous n day closing price that normally smooths out the price values. In this study, n is set to 15, and SMA is hence calculated as follows:

$$\mathrm{SMA}_{15} = \frac{1}{25}\sum_{i=1}^{25} cp(i) \tag{1}$$

Where cp(i) is the closing price of our product on the day i.

- Moving Average Convergence and Divergence (MACD):

The MACD is defined as the difference between two exponential moving averages and is calculated as follows:

$$\mathrm{MACD} = EMA_{12}(C) - EMA_{26}(C) \tag{2}$$

Where, EMA stands for Exponential Moving Average and is defined as:

$$EMA_n(C) = \text{n day Exponential Moving Average}$$

- Relative Strength Index (RSI):

RSI is a momentum indicator calculated as follows:

$$RSI\%(i) = 100 - \frac{100}{1 + RS} \tag{3}$$

$$RS = \frac{\text{Average Gain Over past i days}}{\text{Average Loss Over past 14 day}} \tag{4}$$

RSI is a popular momentum indicator that determines whether the financial product is overbought or oversold [13]. A financial product is said to be overbought when the demand unjustifiably pushes the price upwards: this condition is usually interpreted as a sign that the financial product is overvalued and the price is likely to go down – and vice versa when the financial product is said to be oversold. RSI ranges from 0 to 100 and generally, when RSI is above 70, it may indicate that the stock is overbought and when RSI is below 30, it may indicate the stock is oversold.

2.3.3. Time series forecasting: Machine Learning approach through Support Vector Machines:

The principle of Support Vector Machine is as follows [14]:

Let's consider a set of data: (x_1, y_1) … (x_k, y_k) where $x_i \in R^n$ are the input data and $y_i \in \{-1,1\}$ their class labels.

The aim is to construct a hyperplane $\omega^T x + b = 0$, that separates the two classes: $\omega^T x + b = 1$ and $\omega^T x + b = -1$ [15].

The problem consists on minimizing $\frac{1}{2}\omega^T \omega$:

$$\begin{cases} \min \frac{1}{2}\omega^T \omega \\ s.t: \\ y_i((\omega^T x_i) + b) \ge 1 \end{cases} \tag{5}$$

By introducing ξ_i and ξ_i^* (upper and lower constraints on the outputs) and C, a constant that regularizes the equation, the problem becomes:

$$\min \frac{1}{2}\omega^T \omega + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{6}$$

The mapping $x_i \in R^n$ ($1 \le i \le n$) is performed by a kernel function:

$$K(x_i, x_j) = \phi(x_i)\phi(x_j) \tag{7}$$

Different kernel functions are used in the literature [16], the most known are: linear [17], radial [18] and polynomial [19]. The decision function implemented by SVM can be written as:

$$f(x) = sign(\sum_{i=1}^{N} \alpha_i y_i K(x, x_i) + b) \tag{8}$$

The coefficients α_i are the solutions of the following problem:

$$\begin{cases} \max \sum_{i=1}^{k} \alpha_i - \frac{1}{2} \sum_{i=1}^{k} \sum_{j=1}^{k} \alpha_i \alpha_j y_i y_j \left[\phi(x_i)\phi(x_j)\right] \\ s.t: \\ 0 \le \alpha_i \le C \qquad \sum_{i=1}^{N} \alpha_i y_i = 0 \end{cases} \tag{9}$$

Support Vector machines have also been adapted to regression problems with the decision function [20]:

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) K(x_i, x_j) + b \tag{10}$$

The coefficients α_i, α_i^* represent the Lagrange multipliers, they are the solutions of the problem:

$$\max \tilde{L}(\alpha_i, \alpha_i^*) = \sum_{i=1}^{n} \left(\alpha_i (y_i - \varepsilon) - \alpha_i^* (y_i + \varepsilon)\right)$$

$$-\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) \tag{11}$$

Under constraints

$$0 \le \alpha_i \le C \quad 0 \le \alpha_i^* \le C \quad \sum_{i=1}^{n} (\alpha_i - \alpha_i^*) = 0 \tag{12}$$

Alternatively:

$$\bar{\alpha}, \bar{\alpha}^* = argmin_{\alpha,\alpha^*} \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) + \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) \tag{13}$$

$$\bar{\alpha}, \bar{\alpha}^* = argmin_{\alpha,\alpha^*} \frac{1}{2} \textstyle\sum_{i=1}^{n} \sum_{j=1}^{n} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \sum_{i=1}^{n} y_i (\alpha_i - \alpha_i^*) + \varepsilon \sum_{i=1}^{n} (\alpha_i + \alpha_i^*) \tag{14}$$

The solutions verifies Karush-Kuhn-Tucker condition [21]

$$\bar{\alpha}_i . \bar{\alpha}_i^* = 0 \qquad (i = 1, \dots n) \tag{15}$$

Support vectors are then the points where one of Lagrange multipliers is strictly positive.

Also, when $\varepsilon = 0$, our optimization problem can be written as:

$$\min_\beta \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_i \beta_j K(x_i, x_j) - \sum_{i=1}^{n} y_i \beta_i \tag{16}$$

Under constraints:

$$-C \leq \beta_i \leq C \quad ; \quad \sum_{i=1}^{n} \beta_i = 0 \tag{17}$$

In this study, in addition to using SVM for regression purposes using our processed data set, we used SVM for classification purposes in order to double check if the results we got using the regression SVM were accurate.
The target to be predicted in the ith day, using the classification SVM, refers to the sign of price change between two successive days and is calculated as follows:

$$target_i = Sign\ (close_{i+1} - close_i) \tag{18}$$

3. Results and analysis

3.1. Evaluation metrics

- Root mean Square Error

The performance of product values prediction is evaluated in terms of root mean-square error (RMSE) [22]. It can be written as follows:

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(y_i^{\,p} - y_i^{\,r})^2}{n}} \tag{19}$$

Where: $y_i^{\,p}$ is the predicted value using SVM and $y_i^{\,r}$ the real value of the product.

- F-measure

For action type prediction (binary classification), we will consider three evaluation measures: Recall, Precision and F-measure FM, defined as below [23]:

$$\mathrm{Recall} = \frac{a}{a+c} \; ; \quad \mathrm{Precision} = \frac{a}{a+b} \tag{20}$$

$$FM = \frac{2 \times \mathrm{Recall} \times \mathrm{Precision}}{\mathrm{Recall} + \mathrm{Precision}} \tag{21}$$

Where a is the number of true positives, b : the number of false positives and c is the number of false negatives. The overall results are expressed in terms of F-measure.
As the results of SVM approach rely largely on the choice of the adequate kernel, we will compare main kernel functions and consider the one that returns the best results.

3.2. Results analysis

- ‘2y3y5y’ product time series forecasting

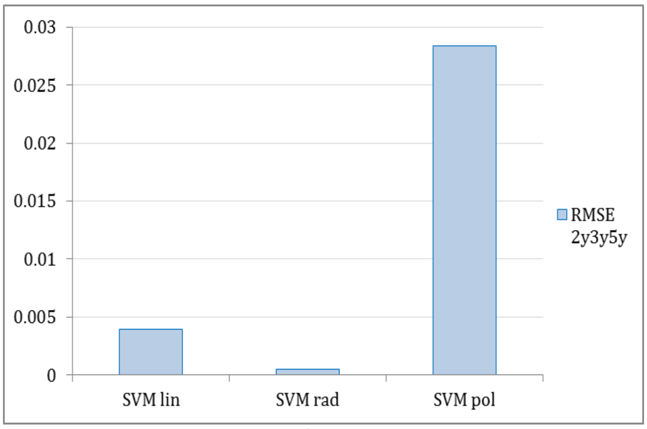

Fig. 3: RMSE while forecasting 2y3y5y prices

When applying Support Vector Machine to our processed data set, the optimal performance corresponds to the minimal root mean square error: 0.00051, which corresponds to the radial kernel. Moreover, evaluating only the nature of the price action (increase or decrease) is a crucial stage for predicting 2y3y5y, and is evaluated by computing the f-measure:

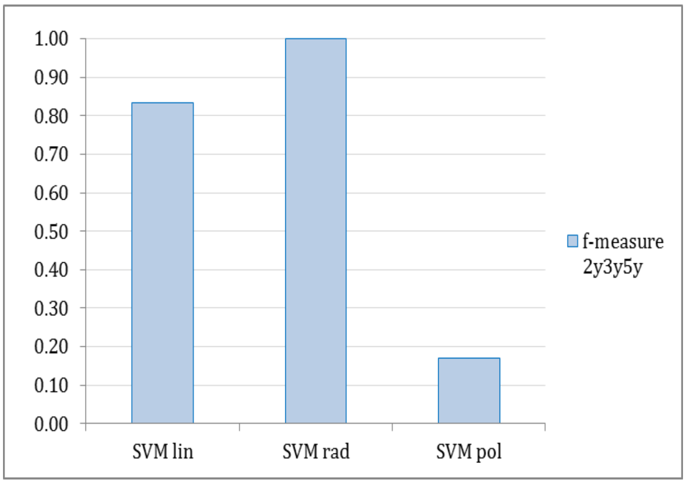

Fig. 4: f-measure when forecasting 2y3y5y prices

The radial kernel returned an f-measure=1. That means that the nature of variations of the product 2y3y5y has been correctly predicted by our model (increase or decrease).

- ‘2y5y10y’ product time series forecasting

We apply the proposed model to the second dataset: 2y5y10y:

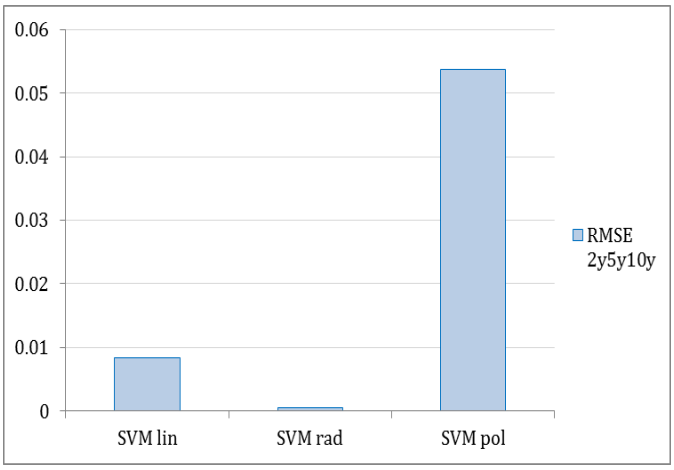

Fig. 5: RMSE while forecasting 2y5y10y prices

As it is the case for the first product, Support Vector Machines with radial kernel yielded the minimal root mean square error: 0.00050. The evaluation of the nature of the price action (increase or decrease) is detailed in the following figure:

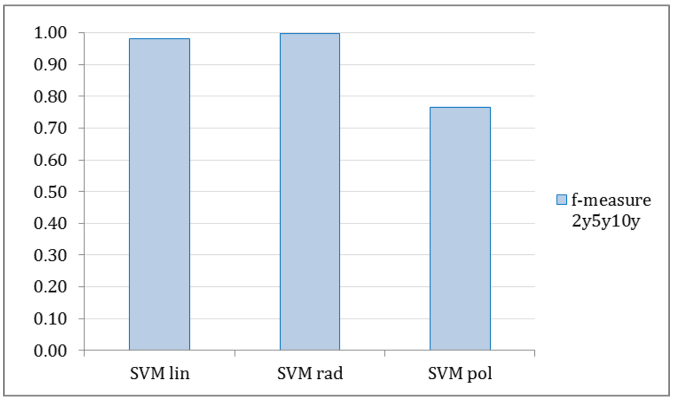

Fig. 6: f-measure while forecasting 2y5y10y prices

The maximal f-measure is still for the radial kernel, with a maximal value of: 99.851%. Only two instances having values: -7.25×10^{-4} and -2.5×10^{-5} were predicted as increasing actions.

4. Conclusion

Predicting fixed income market is very difficult due to its dynamic, non-linear and complex nature. However, machine learning techniques have recently proved effective in financial time series forecasting in general and in stocks forecasting in particular.

This study has proposed a novel hybrid model, proved to be robust in predicting future direction of butterfly products price movement. The model has combined both SVM and technical analysis for efficient financial price time series forecasting. Also, one should note that our model is not tied to interest rate swaps or to US fixed income market only, but it is transferrable to other classes of financial instruments that exhibit similar properties.

For future work, we could consider the forecasting of financial time series as a pure classification problem instead of a regression problem. This can be achieved by either assigning discrete values to the output (-1 and 1), depending on whether the price has increased or decreased, or by generating a continuous trading signal in range 0-1, that will reflect the price variation and that can be generated using fuzzy logic, combined with deep learning techniques.

References

[1] Cox, John C., Jonathan E. Ingersoll Jr, and Stephen A. Ross. "A theory of the term structure of interest rates." In Theory of Valuation, 2005, pp. 129-164.
[2] Adhikari, Ratnadip, and R. K. Agrawal. "A combination of artificial neural network and random walk models for financial time series forecasting." Neural Computing and Applications 24, no. 6 (2014): 1441-1449.
[3] Babu, C. Narendra, and B. Eswara Reddy. "A moving-average filter based hybrid ARIMA–ANN model for forecasting time series data." Applied Soft Computing 23 (2014): 27-38.
[4] Kanchymalay, Kasturi, N. Salim, Anupong Sukprasert, Ramesh Krishnan, and Ummi Raba'ah Hashim. "Multivariate Time Series Forecasting of Crude Palm Oil Price Using Machine Learning Techniques." In IOP Conference Series: Materials Science and Engineering, vol. 226, no. 1, p. 012117. IOP Publishing, 2017.
[5] Khashman A. and Nwulu N. I. Support Vector Machines versus Back Propagation Algorithm for Oil Price Prediction Liu D., Zhang H., Polycarpou M., Alippi C., He H. Adv. Neural Networks – ISNN 2011. ISNN 2011. Lect. Notes Comput. Sci. vol 6677. Springer, Berlin, Heidelb. 530-538
[6] González-Mancha, Juan Javier, Juan Frausto-Solís, Guadalupe Castilla Valdez, Jesús David Terán-Villanueva, and Juan Javier González Barbosa. "Financial time series forecasting using Simulated Annealing and Support Vector Regression." International Journal of Combinatorial Optimization Problems and Informatics 8, no. 2 (2017): 10-18.
[7] Bao, Wei, Jun Yue, and Yulei Rao. "A deep learning framework for financial time series using stacked autoencoders and long-short term memory." PloS one 12, no. 7 (2017): e0180944.
[8] Davig, Troy, and Taeyoung Doh. "Monetary policy regime shifts and inflation persistence." Review of Economics and Statistics 96, no. 5 (2014): 862-875.
[9] De Ketelaere, Bart, Mia Hubert, and Eric Schmitt. "Overview of PCA-based statistical process-monitoring methods for time-dependent, high-dimensional data." Journal of Quality Technology 47, no. 4 (2015): 318-335.
[10] Shlens, Jonathon. "A tutorial on principal component analysis." arXiv preprint arXiv:1404.1100 (2014).
[11] Ahmar, Ansari Saleh. "Sutte Indicator: A technical indicator in stock market." (2017).
[12] Cervelló-Royo, Roberto, Francisco Guijarro, and Karolina Michniuk. "Stock market trading rule based on pattern recognition and technical analysis: Forecasting the DJIA index with intraday data." Expert systems with Applications 42, no. 14 (2015): 5963-5975.
[13] Pring, Martin J. Martin Pring's Introduction to Technical Analysis. McGraw Hill Professional, 2015.
[14] Chang, C. C., & Lin, C. J. (2006). Training v-support vector classifiers: theory and algorithms. Training, 13(9).
[15] Chang, C. C., & Lin, C. J. (2011). LIBSVM: a library for support vector machines. ACM transactions on intelligent systems and technology (TIST), 2(3), 27.
[16] Gunn, S. R. (1998). Support vector machines for classification and regression. ISIS technical report, (14), pp. 85-86.
[17] Cristianini, N., & Shawe-Taylor, J. (2000). An introduction to support vector machines and other kernel-based learning methods. Cambridge university press.
[18] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9(3), 293-300.
[19] Smits, G. F., & Jordaan, E. M. (2002). Improved SVM regression using mixtures of kernels. In Neural Networks, 2002. IJCNN'02. Proceedings of the 2002 International Joint Conference on (Vol. 3, pp. 2785-2790). IEEE.
[20] Smola, A., & Vapnik, V. (1997). Support vector regression machines. Advances in neural information processing systems, (9), pp. 155-161.
[21] Müller, K. R., Smola, A. J., Rätsch, G., Schölkopf, B., Kohlmorgen, J., & Vapnik, V. (1997). Predicting time series with support vector machines. In International Conference on Artificial Neural Networks. Springer Berlin Heidelberg, pp. 999-1004.
[22] R. Darnag, E.L. M. Mazouz, A. Schmitzer, D. Villemin, A. Jarid, D. Cherqaoui, "Support vector machines: Development of QSAR models for predicting anti-HIV-1 activity of TIBO derivatives" European Journal of Medicinal Chemistry 45, p.p: 1590–1597, 2010.
[23] Powers, D.M.W., 2011. Evaluation: from Precision, Recall and F-measure to ROC, Informedness, Markedness and Correlation. Journal of Machine Learning Technologies, 2(1), 37-63.