International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 07 Issue: 04 | Apr 2020 www.irjet.net p-ISSN: 2395-0072

Feature Extraction using Machine Learning Technique with financial Time Series Prediction

Assistant Prof. Anjali Sanjivanrao More[1], Deepa Sunil Ranaware[2], Bhakti Dattatraya Wamane[3], Gouri Shivaji Salunkhe[4]

[1]Assistant Professor of Computer Engineering &SavitribaiPhule Pune University
[2,3,4]Pursuing Bachelor of Computer Engineering &SavitribaiPhule Pune University
[1-4]Suman Ramesh Tulsiani Technical Campus Faculty of Engineering Kamshet, Pune, India

---

Abstract-*News has been a really important supply for several monetary statistic predictions supported elementary analysis. However, digesting a huge quantity of reports and knowledge revealed on the internet to predict a market is going to be heavy. Here show a topic model supported latent Dirichlet allocation (LDA) to urge options from a mixture of text, particularly news articles and monetary statistics, denoted as monetary LDA (FinLDA). The choices from FinLDA are served as extra input options for any machine learning algorithm to spice up the prediction of the monetary statistic.* The proposed system *provide posterior distributions employed in chemist sampling for two variants of the FinLDA and propose a framework for applying the FinLDA during a text and processing for monetary statistic prediction. Here* The proposed system *use SVM classification to urge results. The experimental results show that the choices from the FinLDA by trial and error add worth to the prediction and provides higher results than the comparative options alongside topic distributions from the common LDA.*

Keywords: Data mining, data preparation, data processing, feature extraction, financial time series, latent Dirichlet allocation, news, prediction, stock market, SVM.

1. INTRODUCTION

The efficient market hypothesis (EMH) developed by Fama advised that worth changes instantly answer new info, which they area unit unpredictable. Consequently , historical knowledge cannot be accustomed to build profitable predictions. However, several approaches are accustomed to predict monetary market movement, crashes or booms, and also the prediction remains the subject of active continued analysis. Technical and basic analyses area unit utilized by investors to predict monetary time evolution, like stock costs. Technical analysts believe that historical market knowledge, primarily worth and volume, offer options for worth prediction. Worth and volume could also be extended to tons of complicated indices, like relative strength index (RSI), Accumulation/ Distribution generator (A/D), etc. Technical analysis focuses on victimization ways to extract different info from the historical worth and volume. In distinction, varied knowledge sources could also be employed in basic analysis; they'll be any info a couple of company or its sector, e.g., income quantitative relation, come on assets (ROA), etc., or economics, e.g., America gross national product, America shopper indicant (CPI), etc. moreover, the basic knowledge could also be unstructured matter knowledge, e.g., international news articles, messages during a very net board, public company disclosures, etc., from that area unit tougher to extract info. consequently, monetary models for stock prediction area unit typically supported numerical technical and basic knowledge and specialize in modeling to spice up the results, e.g., ARCH models, GARCH models, machine learning algorithms, etc. However, once text mining had emerged and become sensible to extract info from the text, the monetary analysis took the unstructured matter info into account tons of typically.

2. LITERATURE SURVEY

[1] Stefan Feuerriegel, AntalRatku, Dirk Neumann"Analysis of How Underlying Topics in Financial News Affect Stock Prices Using Latent Dirichlet Allocation", 16 Sept.2018, IEEE paper.

a. Methodology: Companies listed on the stock markets are typically obliged to publicly disclose any information that might have a significant influence on their stock prices. This transparency regulation is intended to ensure that all market participants have access to the same information. The corresponding press releases are one of the most reliable news sources concerning a company's operations.

b. Findings and Application: Interestingly, even though the researcher has investigated the timing of releases, research has invested little effort into examining the underlying news topics. In this paper, The proposed system analyze the effects of topics found in such corporate press releases on stock market returns in the German market. The proposed

system determine the topic of ad hoc announcements by using Latent Dirichlet Allocation.

c. Remark(Future and conclusion): Effectively, The proposed system succeed in extracting 40 topics. As hypothesized, the effect of these topic groups differs greatly from each other. Some topics have no resulting effect on abnormal returns of stocks, whereas other topics, such as drug testing, exhibit a large effect.

[2] C.-F. Tsai and Y.-C.Hsiao"Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches".

a. Methodology: To effectively predict the stock price for investors could even be a crucial research problem. In literature, processing techniques are applied to stock (market) prediction.

b. Findings and Application: Feature selection, a pre-processing step of knowledge mining, aims at filtering out unrepresentative variables from a given dataset for effective prediction. This paper aims to mix multiple feature selection methods to spot more representative variables for better prediction.

c. Remark(Future and conclusion): Three feature selection methods, which are Principal Component Analysis (PCA), Genetic Algorithms (GA) and decision trees (CART), are used. the mixture methods to filter unrepresentative variables are supported by the union, intersection, and multi-intersection strategies.

[3] Gabriel Pui Cheong Fung, Jeffrey Xu Yu, and WaiLam"News Sensitive Stock Trend Prediction", in Proc. Pacific-Asia Conf. Knowl. Berlin, Germany: Springer-Verlag, 2002,

a. Methodology: Stock market prediction with data processing techniques is one of the foremost important issues to be investigated. during this paper, The proposed system present a system that predicts the changes in the stock trend by analyzing the influence of non-quantifiable information (news articles).

b.Findings and Application: Especially, the proposed system investigate the immediate impact of stories articles on the statistic supported the Efficient Markets Hypothesis. Several data processing and text mining techniques are utilized in a completely unique way. A replacement statistical-based piecewise segmentation algorithm is proposed to spot trends on the statistic. The segmented trends are clustered into two categories, Rise and Drop, consistent with the slope of trends and therefore the coefficient of determination. The proposed system propose an algorithm, which is named guided clustering, to filter news articles with the assistance of the clusters that The proposed system have obtained from trends.

c. Remark(Future and conclusion): The proposed system also propose a replacement differentiated weighting scheme that assigns higher weights to the features if they occur within the Rise (Drop) news-article cluster but don't occur in its opposite Drop (Rise).

[4] L. J. Cao and F. E. H. Tay"Support Vector Machine With Adaptive Parameters in Financial Time Series Forecasting", Nov. 2003, IEEE paper.

a. Methodology: A novel sort of learning machine called support vector machine (SVM) has been receiving increasing interest in areas starting from its original application in pattern recognition to other applications like regression estimation thanks to its remarkable generalization performance. This paper deals with the appliance of SVM in financial statistic forecasting.

b. Findings and Application: The feasibility of applying SVM in financial forecasting is first examined by comparing it with the multilayer back-propagation (BP) neural network and therefore the regularized radial basis function (RBF) neural network. The variability in the performance of SVM with reference to the free parameters is investigated experimentally. Adaptive parameters are then proposed by incorporating the non-stationary of monetary statistics into SVM. Five real futures contracts collated from the Chicago Mercantile Market are used because of the data sets.

c. Remark(Future and conclusion): Simulation shows that among the three methods, SVM outperforms the BP neural network in financial forecasting, and therefore there is comparable generalizations performance between SVM and the regularized RBF neural network. Furthermore, the free parameters of SVM have an excellent effect on the generalization performance.

[5] Francis E.H. Tay, Lijuan Cao" Application of support vector machines in financial time series forecasting", Aug. 2001.

a. Methodology: This paper deals with the appliance of a completely unique neural network technique, support vector machine (SVM), in financial statistic forecasting. the target of this paper is to look at the feasibility of SVM in financial statistic forecasting by comparing it with a multi-layer back-propagation (BP) neural network.

b. Findings and Application: Five real futures contracts that are collated from the Chicago Mercantile Market are used because of the data sets. The experiment shows that SVM outperforms the BP neural network supported the standards of normalized mean square error (NMSE), mean absolute error (MAE), directional symmetry (DS) and weighted directional symmetry (WDS).

c. Remark(Future and conclusion): since there are no structured thanks to choosing the free parameters of SVMs, the variability in performance with reference to the free parameters is investigated during this study.

3. PROPOSED SYSTEM

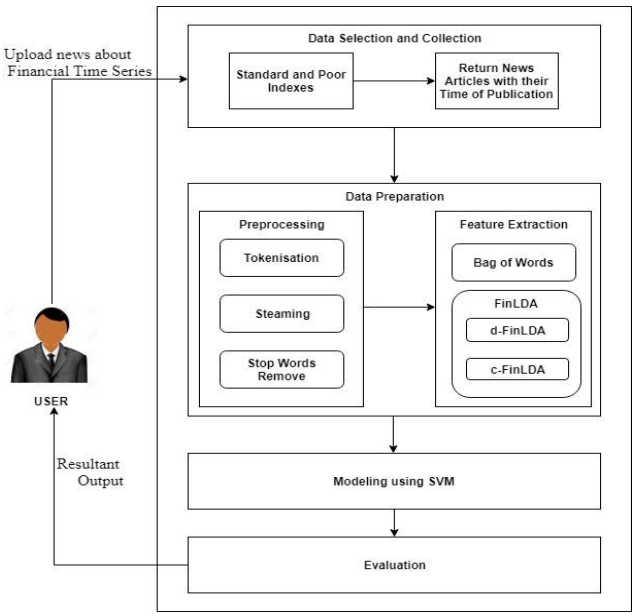
The proposed system propose a replacement domain-specific topic model, the FinLDA model, incorporating changes in money statistics into the common Latent Dirichlet Allocation to return up with a replacement set of latent topics related to the changes in an exceedingly statistic. Here The proposed system have a bent to delineate 2 variant of FinLDA: 1) distinct FinLDA (d-FinLDA) uses, as input, the movements that area unit changes classified into a definite set of values, whereas 2) continuous FinLDA(c-FinLDA) uses real numbers or actual variations. The proposed system have a bent to provided posterior distributions utilized in chemist sampling for parameter estimation and reasoning in topic modeling with FinLDA. The proposed system have a bent to the thought of FinLDA to be a feature extraction in processing. As a result, this text focuses on the feature extraction in an exceeding information preparation part, however, The proposed system have a bent to still would really like the other phases in processing to urge the last word results. Consequently, The proposed system have a bent to provide the tiny print of a framework for applying FinLDA in text and processing for money statistic prediction.

4. ADVANTAGES:

1. Empirically add value to the prediction.

2. Give better results

5. SYSTEM ARCHITECTURE:



6. ALGORITHM:

6.1 LDA:

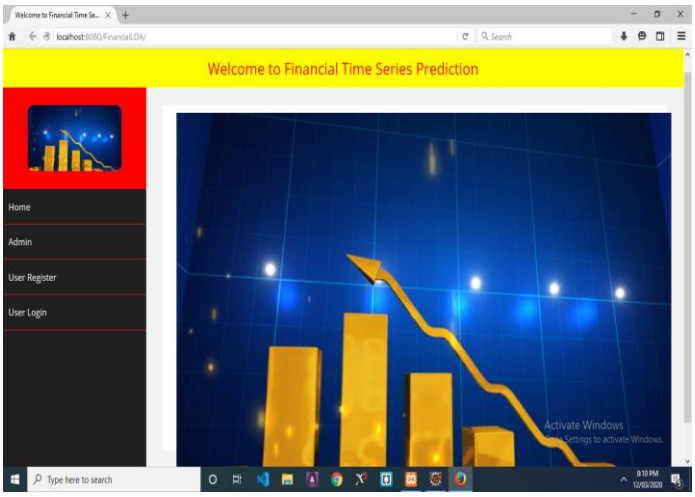
- Compute the d-dimensional mean vectors for the different classes from the dataset.
- Compute the scatter matrices.
- Compute the eigenvectors (e1,e2,...,ed) and corresponding eigenvalues (λ1,λ2,...,λd) for the scatter matrices.
- Sort the eigenvectors by decreasing eigenvalues and choose k eigenvectors with the largest eigenvalues to form a d×k dimensional matrix W.

6.2 SVM :

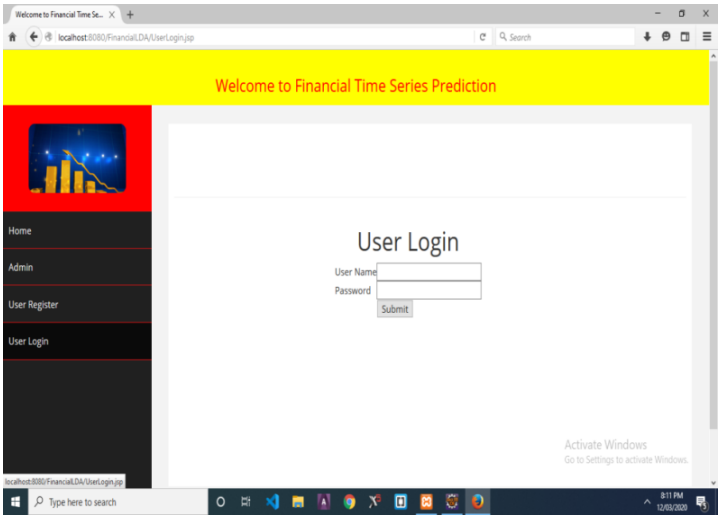
- It should separate the two classes A and B very well so that the function defined by:
- f(x) = a.x + b is positive if and only if x E A
- f(x) 0 if and only if x E B
- It exists as far away as possible from all the observations (robustness of the model). Given that the distance from an observation x to the hyperplane is a.x+ b/a.

7. RESULT AND SCREENSHOT

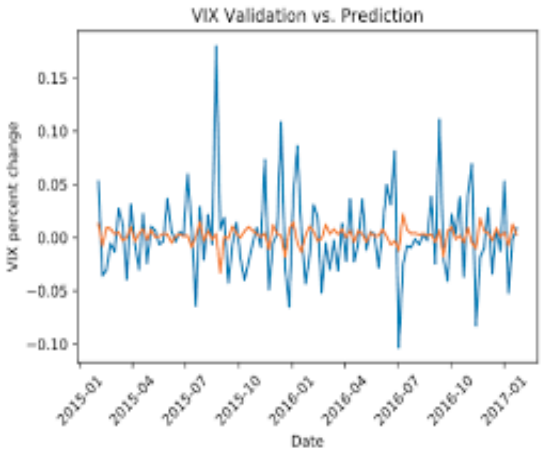
1. GUI



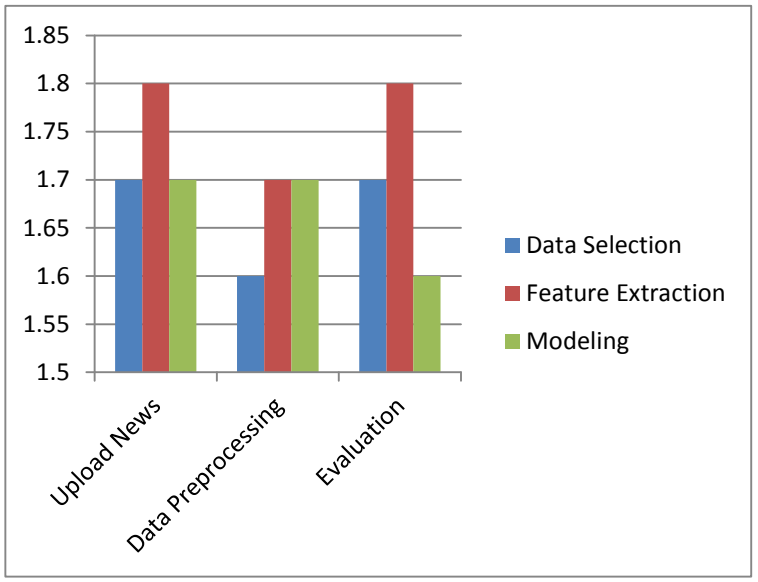
2. User Login and Uploading news



3.Graph and Result Analysis



	Data Selection	Feature Extraction	Modeling
Upload News	1.7	1.8	1.7
Data Preprocessing	1.6	1.7	1.7
Evaluation	1.7	1.8	1.6



8. CONCLUSION:

The proposed system introduced FinLDA to extract higher options from news articles for the prediction. The extracted options are often employed in any machine learning rule to predict monetary results. In our experiment, parameters of our 2 FinLDA variants (one with separate knowledge and thus the various with continuous variables describing changes) were calculable by mistreatment each news articles from Reuters and normal & Poor's five hundred Index data and thus the ultimate outputs from {the 2|the 2} FinLDA variants were used as input options in two standard machine learning algorithms, i.e., SluzhbaVneshneyRazvedki and BPNN, to validate the advantage of the choices from FinLDA once examination with different options. Though adding FinLDA resulted in just minor changes with SluzhbaVneshneyRazvedki, FinLDA gave some price to the prediction. Additionally, BPNN was considerably higher with FinLDA showing 5- to 6-fold drops in MSE in predictions. Consequently, our options from FinLDA through empirical observation provide price another to the prediction after they're employed in each BPNN and SluzhbaVneshneyRazvedki. As here is that the first article within which The proposed system tend to in theory established and formalized the FinLDA, The proposed

system tend to so center on the rationale of FinLDA in information preparation part and conducted the initial experiment to point the benefits of the choices from FinLDA applied in 2 standard machine learning algorithms.

9. REFERENCES:

[1] S. Feuerriegel, A. Ratku, and D. Neumann, ``Analysis of how underlying topics in financial news affect stock prices using latent Dirichlet allocation,'' in Proc. 49th Hawaii Int. Conf. Syst. Sci. (HICSS), 2016, pp. 10721081.

[2] C.-F. Tsai and Y.-C. Hsiao, ``Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches,'' Decis. Support Syst., vol. 50, no. 1, pp. 258269, 2010.

[3] L. J. Cao and F. E. H. Tay, ``Support vector machine with adaptive parameters in financial time series forecasting,'' IEEE Trans. Neural Netw., vol. 14, no. 6, pp. 15061518, Nov. 2003.

[4] G. P. C. Fung, J. X. Yu, and W. Lam, ``News sensitive stock trend prediction,'' in Proc. Pacific-Asia Conf. Knowl. Discovery Data Mining. Berlin, Germany: Springer-Verlag, 2002, pp. 481493.

[5] F. E. Tay and L. Cao, ``Application of support vector machines in financial time series forecasting,'' Omega, vol. 29, no. 4, pp. 309317, Aug. 2001.

[6] Y. Guo, S. Han, C. Shen, Y. Li, X. Yin, and Y. Bai, ``An adaptive SVR for high-frequency stock price forecasting,'' IEEE Access, vol. 6, pp. 1139711404, 2018.

[7] T. Chen and C. Guestrin, ``XGBoost: A scalable tree boosting system,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), San Francisco, CA, USA, 2016, pp. 785794.

[8] D. Wang, Y. Zhang, and Y. Zhao, ``LightGBM: An effective miRNA classification method in breast cancer patients,'' in Proc. Int. Conf. Comput. Biol. Bioinf. (ICCBB).NewYork,NY, USA:ACM, 2017,pp.

[9] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, ``CatBoost: Unbiased boosting with categorical features,'' in Proc. Adv. Neural Inf. Process. Syst. Curran Associates, 2018,.

[10] A. V. Dorogush, V. Ershov, and A. Gulin, ``CatBoost: Gradient boosting with categorical features support,'' Oct. 2014.

[11] I. Kaastra and M. Boyd, ``Designing a neural network for forecasting financial and economic time series,'' Neurocomputing, vol. 10, no. 3, pp. 215236, 1996.

[12] R. Pardo and R. Pardo, The Evaluation and Optimization of Trading Strategies (Wiley Trading), 2nd ed. Hoboken, NJ, USA: Wiley, 2008.

[13] J. D. Kelleher, B. M. Namee, and A. D'Arcy, Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies. Cambridge, MA, USA: MIT Press, 2015.

[14] R. Wirth and J. Hipp, ``CRISP-DM: Towards a standard process model for data mining,'' in Proc. 4th Int. Conf. Practical Appl. Knowl. Discovery Data Mining, 2000, pp. 111.

[15] H. Drucker, C. J. C. Burges, L. Kaufman, A. J. Smola, and V. Vapnik, ``Support vector regression machines,'' in Proc. Adv. Neural Inf. Process. Syst., vol. 9. Cambridge, MA, USA: MIT Press, 1997, pp. 155161.