MACHINE LEARNING ALGORITHM PERFORMANCE IMPROVEMENT IN STOCK PREDICTION WITH ADAPTED NONNEGATIVE DISCRIMINATIVE FEATURE SELECTION (NDFS)

[1]Agbator, L. O. & [2]Agbi, A

[1]Computer Science Department, Edo State Polytechnic, Usen,
Benin city Nigeria. mail4agbator@gmail.com
&
[2]Office Technology and Management, Edo State Polytechnic, Usen,
Benin city Nigeria. Agbicity2003@yahoo.co.uk

KEYWORDS

Fundamental information, technical information, Feature selection, Spectral analysis, Laplacian graph

.

ABSTRACT

Feature selection algorithms lie at the heart of machine learning, playing a crucial role in identifying and prioritizing essential attributes within datasets. By refining sample attributes, these algorithms aim to elevate classification performance and pinpoint the most relevant features associated with distinct data classes. The primary objective is to optimize classification and prediction by selecting features based on their effectiveness. This study introduces a novel feature selection algorithm, Nonnegative Discriminative Feature Selection (NDFS), incorporating spectral analysis and multi-variate regression into stock price trend prediction. The chosen characteristics using NDFS are evaluated for classification prowess with three different classifiers. Furthermore, the performance of NDFS is benchmarked against three feature selection algorithms from existing literature. The findings indicate that NDFS emerges as a competitive technique capable of improving the efficiency of machine learning algorithms, particularly in stock price trend prediction. The adopted feature selection method performed well, achieving an accuracy of 85% in the Naïve Bayes model and an F1-score of 83%, indicating a balanced measure of precision and recall. The receiver operating characteristics curve reveals an optimal model with performance surpassing 80% when utilizing features selected by Nonnegative Discriminative Feature Selection.

INTRODUCTION

Stock market data has been characterized by researchers as chaotic and highly stochastic, leading to a significant level of uncertainty and making accurate stock price forecasts challenging (Alelyani, Tang, & Liu, 2019). Despite substantial progress in prediction and trend forecasting within this field by statisticians and computer scientists, achieving consistent accuracy remains difficult. The availability of powerful computing resources and machine learning algorithms has facilitated advancements, yet the unpredictable nature of the stock market persists (Sachdeva, et al, 2019). To address this challenge, this study aims to identify optimal features for stock trend classification through an effective feature selection technique (Xia, 2018). Acknowledging the market's erratic and unstable nature, the research leverages historical stock data, recognizing the hidden correlations. Raw historical data, including opening/closing prices, low/high prices, volume, and occasionally adjusted closing prices, form the basis. However, the accuracy of categorization and forecasting using raw data is limited to machine learning algorithms adept at interpreting time series data (Zheng, et al, 2020).

To overcome these limitations and enhance accuracy, the study introduces feature engineering, wherein various functions are applied to the raw historical data to unveil concealed relationships and patterns. Enhancing predictive performance is achieved by reducing feature dimensions through feature selection, thereby reducing calculation time and addressing overfitting concerns in classification problems (Hartman, & Hlinka, 2018). Predicting stock price movements is inherently challenging due to the non-linear and highly erratic nature of stock data. Feature engineering for stock prediction is widespread, but not all features are equally crucial. The goal of this study is to develop a novel feature selection algorithm that combines filter and embedded selection approaches to derive a feature subset, ultimately improving stock trend prediction accuracy. The current challenge lies in determining the most effective feature combination within a stock dataset. Numerous researchers have explored aspects of this issue, designing new features and employing diverse feature selection techniques. Some have focused on constructing subsets from available features to optimize predictive performance. Additionally, efforts have been made to enhance machine learning algorithms through ensemble approaches and different types of neural network models (Devi, & Seelammal, 2018). The collective aim is to identify a set of features that best contribute to accurate stock trend prediction (Alsubaie, Hindi, El, & Alsalman, 2019).

LITERATURE REVIEW

The domain of stock price prediction captivates both social scientists and information scientists due to its intrinsic associations across various data categories. Fundamental data, comprising financial ratios and declared assets from annual or quarterly financial statements, constitutes the first type of structured data in stock market research. The second category, technical information, involves analyzing market behavior and is considered by technical analysts as already integrated into demand and supply curves, reflecting traders' sentiments. Technical indicators, formed from past prices, capture recurring price patterns (Devi, & Seelammal, 2018). Numerous studies emphasize the role of effective feature selection algorithms in enhancing the accuracy of stock market price trend forecasts (Hu, et al, 2020). Over time, researchers have explored feature selection techniques from filter, wrapper, and embedding methods. This study delves into various feature selection methods, aiming to identify features for stock market price trend prediction using a novel algorithm specific to stock price prediction.

A comprehensive overview of state-of-art feature selection techniques including mathematical formulas and fundamental algorithms to facilitate understanding was done by Amir Moslemi in 2023. This survey includes different approaches to variable selection which can be grouped into five domains: A) information theory which covers multi-label neighborhood entropy, symmetrical uncertainty, Monte Carlo and Markov blanket, B) sparse representation learning which includes compressed sensing and dictionary learning, C) subspace learning which involves matrix factorization and matrix projection, D) reinforcement learning techniques and E) evolutionary computational algorithms including Genetic algorithm (GA), particle swarm optimization (PSO), Ant colony (AC) and Grey wolf optimization (GWO),. This survey helped us in this study to gain valuable insight and to acquire a deep understanding of feature selection techniques.

A related investigation in the domain of unsupervised classification was done by Jundong L. et al in 2019. They studied how to harness the tie-strength information embedded in the information network structure to facilitate the selection of relevant attributes belonging to different nodes. They adopted a principled unsupervised feature selection framework ADAPT to find informative features that can be used to regenerate the observed links and further characterize the adaptive neighborhood structure of the network. An effective optimization algorithm for the adopted ADAPT framework was also presented in the work. Experiments were performed on several real-world attributed networks and the results proved the superiority of the adopted ADAPT framework.

In the study done in the renewable energy domain, by Malakar, et al (2021, a special combination of deep-learning-based sequence model Bidirectional Gated Recurrent Unit (BGRU) with a new augmented and bidirectional feature representation was used for the first time in the solar energy domain to capture the complex weather conditions which lead to uncertainty in photovoltaic (PV) systems which often complicates solar energy prediction. The used BGRU network which proved to be more generalized as it can handle unequal lengths of forward and backward context, produced 59.21%, 37.47%, and 76.80% better prediction accuracy compared to traditional sequence-based, bidirectional models, and some of the established states-of-the-art models. The result of this study validates our concept and adopted work.

In a bid to achieve a balance between speed and accuracy. Garcia-Ramirez, I.-A. et al constructed a framework that consists of a novel combination of Approximated and Simulate Annealing versions of the Maximal Information Coefficient (MIC) to generalize the simple linear relation between features for the selection of relevant features in supervised datasets based on a cascade of methods where speed and precision are in mind. This process is performed in a series of steps by applying the MIC algorithms and cutoff strategies to remove irrelevant and redundant features to achieve balance between accuracy and speed. The adopted framework was tested in a series of experiments conducted on a large set of datasets from SPECTF Heart to Sonar data. The results show the balance of accuracy and speed that the adopted framework can achieve. This study also supports our effort to adopt an existing algorithm in the new domain because algorithms are mostly constructed to meet certain objectives and for specific domains.

In this study, the authors used the clustering part of Nonnegative discriminative feature selection, our adopted algorithm to handle the often redundant and noisy high-dimensional features in many image processing and pattern recognition problems. A novel unsupervised feature selection scheme, namely, nonnegative spectral analysis with constrained redundancy, by jointly leveraging nonnegative spectral clustering and redundancy analysis was adopted. The adopted method can directly identify a discriminative subset of the most useful and redundancy-constrained features. Nonnegative spectral analysis is developed to learn more accurate cluster labels of the input images, during which the feature selection is performed simultaneously. The joint learning of the cluster labels and feature selection matrix enables the algorithm to select the most discriminative features. Row-wise sparse models with a general ℓ 2, p -norm (0 <; p ≤ 1) are leveraged to make the adopted model suitable for feature selection and robust to noise. Extensive experiments on nine diverse image benchmarks, including face data, handwritten digit data, and object image data showed that the adopted method achieves encouraging experimental results relative to several representative algorithms. This algorithm was specifically constructed for this kind of data. Our study deploys the algorithm in a completely new domain with the inclusion of multivariate regression to further reduce the dimensionality of the feature space

The quest for a feature selection method contributing to heightened accuracy in stock market prediction has driven extensive experimentation with algorithms in the filter, embedded, and wrapper categories. Notably, wrapper algorithms exhibit low convergence rates, susceptibility to local optima, and reliance on machine learning algorithms for feature selection (Li, Du, & Nian, 2014). This study introduces a novel filter category feature selection algorithm previously unexplored in the stock market. For the first time, it will be compared with algorithms in the filter and embedded categories, to discover a new feature selection algorithm in the stock market domain that can significantly enhance prediction accuracy.

The Adopted Nonnegative Discriminative Feature Selection (NDFS)

This work introduces a feature selection approach for unsupervised learning, termed Nonnegative Discriminative Feature Selection (NDFS), based on spectral analysis. Utilizing cluster indicator labels derived from the intrinsic structure of stock data, NDFS identifies relevant features within the dataset. Spectral analysis-driven feature selection has found application in various domains, including image and microarray datasets. Experimental results indicate that features chosen by the NDFS algorithm enhance prediction performance compared to similar algorithms (Li, et al, 2012).

NDFS employs a constructed Laplacian graph in spectral analysis and a robust multivariate regression method to select pertinent features. The L2,1-norm regularized spectral regression model's sparse solution helps to reduce the selection of irrelevant features that could lead to overfitting in prediction tasks (Shi, & Liu, 2014). In the NDFS algorithm mechanism, spectral analysis learns cluster labels that effectively describe class labels, guiding the construction of the feature selection matrix. This process identifies features most relevant to class labels through a linear transformation between features and labels. The subsequent section employs specific notations to express mathematical aspects of the Nonnegative Discriminative Feature Selection model, supported by the following assumptions for clarity.

Let X∈Rnxm denote the data matrix from a dataset of n number of samples and m number of features in which xi∈Rd is the feature descriptor of the i-th sample. Suppose these n samples are sampled from c classes;

Y = [y1, · · · , yn]T∈ {0, 1}nxc (1)

where yi∈ {0, 1}cx1 is the cluster indicator vector for xi (Li et al., 2012), which approximates the scaled cluster indicator matrix F defined as

F = [F1, · · · , Fn]T=Y(YTY)-1/2 (2)

where Fi is the scaled cluster indicator of xi. and is the target or classes of the data sample (Li et al., 2012). The algorithm presented concurrently learns the scaled cluster indicator matrix and the feature selection matrix. The steps are as follows: First, the sample similarity matrix is established from dataset sample pairs utilizing the Gaussian radial basis (RBF) kernel function, a widely-used similarity measurement, particularly when class information is absent in the relation (Zhao & Liu, 2017).

Sij =exp (-|| xi - xj ||)2 2δ2 (3)

This pairwise sample similarity is extensively employed in both supervised and unsupervised learning to characterize relationships among samples. Its utility extends to accurately portraying cluster affiliations or class affiliations within the sample set. In the sample similarity relation mentioned earlier exp (·) is the exponential function, and δ is the parameter for controlling the width of the region of similarity. When the value is large, even points farther apart can be deemed similar, whereas a small value implies that points must be near to be considered similar. In this context, the objective is to guarantee that samples within the same cluster exhibit a large similarity value, while samples from different clusters possess a small similarity value.

LAPLACIAN GRAPH

Following the feature selection method, the subsequent step involves creating a Laplacian matrix derived from the similarity matrix. This Laplacian matrix is pivotal in generating eigenvectors, renowned for their remarkable properties that ensure consistency in spectral clustering (Zhao, & Liu, 2012). The spectrum of the Laplacian matrix provides insights into the structural information of the graph from which it originated and facilitates the assessment of feature relevance in spectral feature selection.

Multivariate Formulation for Spectral Feature Selection.

In Nonnegative Discriminative Feature Selection (NDFS), a spectral clustering criterion represented by $J(\hat{Y})$ is jointly optimized with the multivariate $L_{2,1}$ regression model of a feature selection matrix of the given sample data to obtain the objective function for Nonnegative Discriminative Feature Selection (NDFS). The joint optimization relation is given as:

$$Min(Y,W)\, J(\hat{Y}) + \alpha(\|X^T W - \hat{Y}\| + \beta\|W\|_{21})$$

(4)

subject to $I - I(I \quad I)$ where α and β are parameters, and $J(\hat{Y})$ is the objective function of spectral analysis which can be represented as:

$$MinTr\left(\hat{Y}^T L \hat{Y}\right) \qquad (5)$$

where $\hat{Y}$ is the cluster indicator matrix and *L* is the Laplacian matrix constructed from the similarity matrix as shown above. The optimization in spectral analysis is addressed through the eigen-decomposition of the Laplacian matrix *L*. To eliminate redundant features, the L2,1-norm regularization term is applied to ensure sparsity in the rows of matrix *W* (Li et al., 2012). The resulting sparse matrix *W* assesses the correlation between pseudo labels derived from the solution in the term $J(\hat{Y})$ and features from the data matrix. This joint minimization of the regression model and

L2,1-norm regularization term enhances the suitability for feature selection.

Methodology

To assess the effectiveness of the features selected by Nonnegative Discriminative Feature Selection (NDFS), three classification algorithms, Support Vector Machine (SVM), Naïve Bayes, and Neural Network were used to construct classification models. Additionally, two feature selection algorithms, Information Gain (from the filter category) and Recursive Feature Elimination (from the wrapper category), were used as comparative benchmarks to evaluate NDFS. The features chosen by NDFS were utilized to predict the stock dataset price trend, and the outcomes were compared with those of the selected competitors to confirm the anticipated higher prediction accuracy associated with features selected NDFS.

The prediction algorithm derives the target or class from the stock price using the following relation:

$$R_t = \frac{\text{closePrice}_t - \text{openPrice}_n}{\text{openPrice}_n} \quad ; |C_t| = \begin{cases} 1, & R_t > 0 \\ 0, & R_t \le 0 \end{cases}$$

Where Rt is the result from the expression for the trading day(t), Ct is the class or label for the same trading day, and closePricet is the close price for the trading day, for example, the first row in Table 1 (*) and openPricen is the first open price in a chosen number of days before the trading day. The next-day prediction will be one day before the trading day for example the second row in Table 1, the one-week prediction will be the seventh day

Table 1 *Apple historical datasets*

Table 1 *Apple historical datasets*

Date	Open	High	Low	Close
*10/25/2019	243.16	246.73	242.88	246.58
**10/24/2019	244.51	244.8	241.81	243.58
10/23/2019	242.1	243.24	241.22	243.18
10/22/2019	241.16	242.2	239.62	239.96
10/21/2019	237.52	240.99	237.32	240.51
10/18/2019	234.59	237.58	234.29	236.41
#10/17/2019	235.09	236.15	233.52	235.28
10/16/2019	233.37	235.24	233.2	234.37

Numerous studies on various feature selection algorithms have demonstrated that a successful feature selection method can significantly improve the prediction accuracy of stock market price trends (Hu, J., Li, Gao, & Zhang, 2020). This assertion is further corroborated by our work, where the accuracy and F1-score were evaluated in our experiments, as depicted in Table 2, particularly when utilizing all features without feature selection. In this study, the adopted feature selection strategy, Nonnegative Discriminative Feature Selection (NDFS), was assessed in terms of accuracy and F1-score, representing the harmonic mean of recall and precision values. As detailed in Table 2, the adopted feature selection method performed well, achieving an accuracy of 85% in the Naïve Bayes model and an F1-score of 83%, indicating a balanced measure of precision and recall. This balanced approach is justified by the relatively balanced nature of the stock dataset used in the experiment. The Apple dataset, with 2,458 instances, exhibits only a seven-unit difference between positive and negative outcomes, resulting in an equal weighting of precision and recall.

In comparison with selected benchmarks, our observations indicate that the adopted feature selection algorithm, Nonnegative Discriminative Feature Selection (NDFS), outperformed the two competitor feature selection algorithms, Information Gain and Recursive Feature Elimination, on average across the three classifiers employed.

Table 2 *Accuracy, precision, recal,l and f-measure from all features, and NDFS algorithm*

Information Gain				Recursive Feature Elimination			
Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
0.80	0.79	0.75	0.76	0.78	0.75	0.69	0.74
0.84	0.82	0.82	0.82	0.76	0.71	0.62	0.63
0.72	0.70	0.62	0.67	0.73	0.70	0.70	0.71
0.79	0.77	0.73	0.75	0.76	0.72	0.67	0.69

Table 3: *Accuracy, precision, recall, and f-measure from information gain, and NDFS recursive feature elimination algorithm*

Feature Selection Algorithm	**All features**				**NDFS**			
Classification Algorithm	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score
NB	0.54	0.53	0.12	0.19	0.85	0.86	0.80	0.83
SVM	0.50	0.46	0.45	0.45	0.83	0.81	0.83	0.82
NN	0.59	0.52	0.23	0.34	0.75	0.76	0.66	0.70
Average	0.54	0.50	0.26	0.32	0.81	0.81	0.76	0.78

The Receiver Operating Characteristics Area Under the Curve clearly showed that the three models—Naïve Bayes, Support Vector Machine, and Neural Network—performed well with the features selected by the three feature selection algorithms. This reinforces the existing claim by researchers that feature selection helps remove irrelevant features from datasets without incurring a loss of information, thereby improving prediction performance.

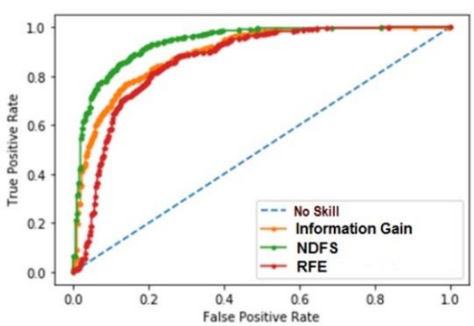

Figure 1*: ROC curve for all features and selected features*

Figure 1 demonstrates that including relevant and irrelevant features severely hampers model performance, to the extent that the model hardly deviates from a random estimation, indicating a lack of discernible skill or intelligence. This decline in performance is attributable to the presence of irrelevant features in the dataset. To showcase the impact of feature selection, the baseline classifier Naïve Bayes, which outperforms other classifiers in this study, was employed to illustrate the performance of a model using only selected relevant features. In Figure 1, the receiver operating characteristics curve reveals an optimal model with performance surpassing 80% when utilizing features selected by Nonnegative Discriminative Feature Selection, Information Gain, and Recursive Feature Elimination. Notably, Nonnegative Discriminative Feature Selection outperforms the two competing methods. The discernible conclusion is that feature selection significantly improves model performance, as evidenced by the stark contrast in performance observed in Figure 1 when using all features.

CONCLUSION

In this paper, we explore three feature selection methods from the filter and wrapper categories—Nonnegative Discriminative Feature Selection, Information Gain, and Recursive Feature Elimination—in predicting stock price trends. We introduce Nonnegative Discriminative Feature Selection as a novel method for feature selection in stock price prediction. Empirical results based on Apple stock data from the Nasdaq Stock Exchange indicate that our adopted method, Nonnegative Discriminative Feature Selection, enhances the accuracy of stock price trend prediction and outperforms selected competitors. In Figure 1, the receiver operating characteristics curve reveals an optimal model with performance surpassing 80% when utilizing features selected by Nonnegative Discriminative Feature Selection,

To validate the efficacy of the feature selection methods, we employed three models—Support Vector Machine, Naïve Bayes, and Neural Network—to assess their ability to select relevant features from the introduced fifty-two features with a binary class. Evaluation metrics such as accuracy, F1-Score, and the Receiver Operating Characteristics area under the curve confirmed that our adopted feature selection method surpasses the chosen competitors, As detailed in Table 2, the adopted feature selection method performed well, achieving an accuracy of 85% in the Naïve Bayes model and an F1-score of 83%, indicating a balanced measure of precision and recall.. The ROC AUC metric underscores the importance of feature selection in constructing optimal models, while also revealing that a model's performance can be significantly compromised in the presence of irrelevant features in datasets.

Future work will explore the application of Nonnegative Discriminative Feature Selection in other domains, comparing its performance with other competitors, and assessing how it can enhance accuracy and return on investment through a profitable trading strategy with the adopted model.

REFERENCES

Alelyani, S., Tang, J., & Liu, H. (2019). Feature Selection forClustering: A Review. Data Clustering, 2 9–60. https://doi.org/10.1201/9781315373515-2

Alsubaie, Y., Hindi, K. El, & Alsalman, H. (2019). Cost-Sensitive Prediction of Stock Price Direction: Selection of Technical Indicators. IEEE Access, 7, 146876–146892. https://doi.org/10.1109/ACCESS.2019.2945907

Amir M.,(2023) A tutorial-based survey on feature selection: Recent advancements on feature selection, Engineering Applications of Artificial Intelligence, Volume 126, Part D, 2023, 107136, https://doi.org/10.1016/j.engappai.2023.107136.

https://www.sciencedirect.com/science/article/pii/S0952197623013209

Devi, K. V., & Seelammal, C. (2018). Multi-criteria decision support for feature selection in network anomaly detection system. International Journal of Data Analysis Techniques and Strategies, 18(3), pp.334-350. https://doi.org/DOI: 10.1504/IJDATS.2018.094132

Garcia-Ramirez, I.-A.; Calderon-Mora, A.; Mendez-Vazquez, A.; Ortega-Cisneros, S.; ReyesAmezcua, I.(2022) A Novel Framework for Fast Feature Selection Based on Multi-Stage Correlation Measures. Mach. Learn. Knowl. Extr. 2022, 4, 131–149. https://doi.org/10.3390/ make4010007

Hartman, D., & Hlinka, J. (2018). Nonlinearity in stock networks. Chaos, 28(8). https://doi.org/10.1063/1.5023309

Hu, J., Li, Y., Gao, W., & Zhang, P. (2020). Robust multi-label feature selection with dual-graph regularization. Knowledge-Based Systems, 203, 106126. https://doi.org/10.1016/j.knosys.2020.106126

Jundong Li, Ruocheng Guo, Chenghao Liu, Huan Liu. (2019). Adaptive Unsupervised Feature Selection on Attributed Networks. In The 25th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’19), August 4–8, 2019, Anchorage, AK, USA. ACM, New York, NY, USA, . https://doi.org/10.1145/3292500.3330856

Li, M., Du, W., & Nian, F. (2014). An Adaptive Particle Swarm Optimization Algorithm Based on Directed Weighted Complex Network. 2014.

Li, Z., Yang, Y., Liu, J., Zhou, X., & Lu, H. (2012). Unsupervised feature selection using nonnegative spectral analysis. Proceedings of the National Conference on Artificial Intelligence, 2, 1026–1032.

Li Z & Tang, J. (2015) Unsupervised Feature Selection via Nonnegative Spectral Analysis and Redundancy Control, in IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5343-5355, Dec. 2015, doi: 10.1109/TIP.2015.2479560.

Malakar, S.; Goswami, S.; Ganguli, B.; Chakrabarti, A.; Roy, S.S.; Boopathi, K.; Rangaraj, A.G.(2021) A Novel Feature Representation for Prediction of Global Horizontal Irradiance Using a Bidirectional Model. Mach. Learn. Knowl. Extr. 2021, 3, 946–965. https://doi.org/10.3390/ make3040047

Sachdeva, A., Jethwani, G., Manjunath, C., Balamurugan, M., & Krishna, A. V. N. (2019). An Effective Time Series Analysis for Equity Market Prediction Using Deep Learning Model. 2019 International Conference on Data Science and Communication, IconDSC 2019, 1–5. https://doi.org/10.1109/IconDSC.2019.8817035

Shi, H., & Liu, X. (2014). Application on stock price prediction of Elman neural networks based on principal component analysis method. 2014 11th International Computer Conference on Wavelet Active Media Technology and Information Processing, ICCWAMTIP 2014, 411–414. https://doi.org/10.1109/ICCWAMTIP.2014.7073438

Xia, T., Sun, Q., Zhou, A., Wang, S., Xiong, S., Gao, S., Li, J., & Yuan, Q. (2018). Improving the performance of stock trend prediction by applying GA to feature selection. Proceedings - 8th IEEE International Symposium on Cloud and Services Computing, SC2 2018, 122–126. https://doi.org/10.1109/SC2.2018.00025

Zhao, Z. A., & Liu, H. (2012). Spectral Feature Selection for Data Mining. Taylor & Francis Group.

Zheng, W., Zhu, X., Wen, G., Zhu, Y., Yu, H., & Gan, J. (2020). Unsupervised feature selection by self-paced learning regularization. Pattern Recognition Letters, 132, 4–11. https://doi.org/10.1016/j.patrec.2018.06.029