Predicting Video Game Sales Based on Machine Learning and Hybrid Feature Selection Method

1st Jianbin Li
School of Electrical and Computer Engineering
Nanfang College · Guangzhou
Guangzhou, China
leejianbin66@gmail.com

2nd Yufan Zheng
School of Electrical and Computer Engineering
Nanfang College · Guangzhou
Guangzhou, China
zhjpre@gmail.com

3rd Haoran Hu
School of Electrical and Computer Engineering
Nanfang College · Guangzhou
Guangzhou, China
huhaoran2@gmail.com

4th Junhui Lu
School of Electrical and Computer Engineering
Nanfang College · Guangzhou
Guangzhou, China
lukjunhui@gmail.com

5th Choujun Zhan
School of Electrical and Computer Engineering
Nanfang College · Guangzhou
Guangzhou, China
zchoujun2@gmail.com

2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE) | 978-1-6654-0553-9/21/$31.00 ©2021 IEEE | DOI: 10.1109/ISKE54062.2021.9755343

***Abstract*—Sales prediction plays a critical role in the rapid development of the internet. Unlike the goal of predicting physical items sales, the virtual items of video games needn't consider the relationship of inventory and demand, and it reflects consumer preferences to a certain extent. Predicting sales of the video game can pre-adjust sales strategies and development plans in advance. In the process of predicting, an excellent feature selection method can remove the features irrelevant to enhance the accuracy of the model. In this paper, we proposed a new hybrid feature selection method Pearson correlation coefficient - Random Forest Feature Selection(PCC-RFFS), and utilized 9 machine learning methods combined with PCC-RFFS to predict the sales of video games. For the filter-based stage, we used the absolute value of Pearson correlation coefficient and feature ranking technique, while the wrapper-based stage is based on Random Forest to measure the importance between features and target. The strategy of combination is in addition to both stages. Experiments on the real-world dataset from February 2006 to November 2016 show that the machine learning method combined PCC-RFFS outperforms the machine learning method combined Pearson correlation coefficient or Random Forest.**

***Index Terms*—Video game, Sales prediction, Machine learning, Feature selection.**

I. Introduction

Sales prediction plays a critical role with the rapid development of the internet. Accurately predicting sales not only maintains the equilibrium inventory and demand to keep minimizing the costs but also can pre-adjust strategy based on sales. Therefore, it applied many kinds of spheres, such as food sales [1], e-commerce sales [2], book sales [3], etc. The video game is an interactive game based on electronic equipment. The rapid development of video games and electronic equipment has led many companies to develop abundant and high-quality video games. In addition, due to the COVID-19 epidemic outbreak, many people stayed at home forcibly for a long time and played video games as their pastime activity [4]. Therefore, video games have become increasingly popular and recognized. In the video game market, the video game have a huge difference from general items, it doesn't need to pay too much attention to the relationship between inventory and demand, while the cost of the video game is mainly concentrated on the development state and early publicity state. Hence, accurately predicting video games sales has a vital effect, it can pre-adjust the sales strategy duly to maximize profits as much as possible, and the number of sales in the future can feedback consumer's popularity and demand for the video game, developers can specify development plans in advance. However, there is rarely research applied to the prediction of video game sales [5].

In recent years, due to high application value, many researchers have concentrated on studying sales prediction and proposed advanced predicting methods. The traditional sales prediction is based on statistics. ARIMA is a typical single-variable sales prediction method based on statistics [2] [6], which uses historical sales as features to predict future sales. However, this method ignores other related features, and it is difficult to capture changes in sales by using only a single variable as a feature to predict sales. Markov chain can consider the data at the current moment to predict the short-term future. It is often used for short-term sales forecasting [7].

With the improvement of computing power, more and more ensemble learning methods and deep learning methods are adopted [3] [8]. Because of good interpretability and excellent performance, ensemble learning has been applied to sales prediction [8] [1]. The excellent interpretability of the model means that businesses can make analyses according to the importance of the features to pre-adjust sales strategies. Wisesa et al. adopt Gradient Boosting Decision Tree(GBDT) to predict Business To Business sales [8]. Ming Gao et al. combined with Extreme learning machine(EXT), and many e-commerce related indicators as features to predict book sales [3]. Bohdan M. Pavlyshenko proposed a multilevel stacking approach by using different kinds of individual models and weights respectively to combine a new model. It can learn the characteristics of multiple models so that the model performance can be improved [9]. For sales prediction methods based on deep learning [10] [11], which performs well, especially LSTM. Due to its excellent structure, it can have a certain memory ability, which is very suitable for time-series sales data. Dai and Huang applied LSTM with hyperparameter search to predict sales of an open dataset and designed a piecewise loss function based on the difference of the true value and the predicted value [10]. Bandara et al. proposed a preprocessing framework by product grouping on the basis of sales condition and adopted LSTM to predict sales of e-commerce in a real-world online marketplace dataset [12].

Feature selection method focuses on reducing noise and decreasing feature dimensions to reduce training time and improve accuracy. Because of the abundant type of feature selection method [13], we focus on introducing methods of filter-based, wrapper-based, and hybrid. Most existing approaches of sales prediction utilized the feature selection method of correlation criteria [14] [8]. These methods can calculate the relationship between the feature and the target in some cases, but it is not ideal in other cases. For example, Pearson correlation coefficient feature selection(PCC) considers the linear relationship but ignores the non-linear relationship between the arbitrary two features. Filter-based is one of the classical feature selection methods. It utilizes feature ranking techniques and adapts features of the strongly related part by setting a threshold making feature scores of the algorithm above it. The Filter-based feature selection method includes correlation criteria and mutual information criteria. Correlation criteria feature selection(CRFS) is a common feature selection approach, correlation coefficient can obtain a score between features and target of linear/nonlinear correlation, such as Pearson, Spearman, and Kendall. Mutual information(MI) feature selection measures the relationship between two variables based on Shannon's entropy. It can reflect the linear and nonlinear importance of arbitrary two features [15] [16]. In [15], the authors assimilated greed selection into MI, which considered the mutual information of output class and already-selected features. In [16], the authors normalized the score of MI to reduce bias and also proposed a method that combines with a genetic algorithm to overcome the limitations of incremental search. Unlike

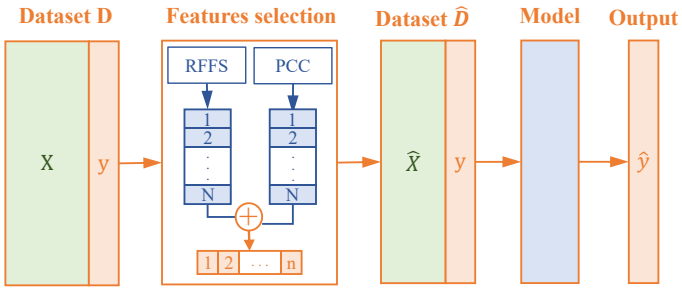


Fig. 1. The workflow of the proposed sales prediction of the video game.

filter-based feature selection does not consider measure feature important through estimator, wrapped-based feature selection directly uses the performance of the final model to be used as the evaluation criterion for feature subsets. The purpose of wrapped-based feature selection is to select a given model to train by subset which is most conducive to its performance. The performance of this method is better than that of the filter-based, but it takes a relatively larger time costs. [17] proposed a method of combining genetic algorithm and KNN to identify whether there are insignificant features or not. In [18], the author proposed a wrapper-based method based on KNN. This method used the distance matrix update parameter to speed up the execution time without losing accuracy. [19], [20] used the Random Forest Feature Selection(RFFS) method. The former is compared with standard chemometric methods, while the latter cost-sensitive is considered based on the original. The hybrid feature selection method is very popular in recent years because it can combine the advantages of different methods to remove redundant features. In [21] combined with PCC and MI, which calculates two correlations and sets a gating unit outputs two scores in proportion. Among the hybrid feature selection method, the more commonly used is to combine wrapper-based and filter-based to propose a new and excellent feature selection method. In [22], wrapper-based and filter-based are used as part of the features to select features and use a firefly algorithm to optimize.

In this paper, we proposed a feature selection method Pearson correlation coefficient - Random Forest Feature Selection(PCC-RFFS) and sales prediction of video games using PCC-RFFS, figure 1 illustrates the workflow. Firstly, the dataset is first passed through the feature selection method. In PCC-RFFS, the values of PCC and RFFS are calculated separately. The top n feature importance score is taken as the final selected feature to obtain a new dataset after adding the correlation of PCC and RFFS. After that, the new dataset is reconstructed through PCC-RFFS, and input into the machine learning model. Finally, for each sample, we output the predicted value by model. The main contribution of this work is summarized as follows:

- A new feature hybrid feature selection method PCC-RFFS considering the advantage of the filter-based and wrapper-based is proposed.

- We collected video game data about game information from 1970 to November 2020 and game sales from 1970 to 2018 and constructed a dataset of sales of the video game from 2004 to 2018.
- Extensive experiments were conducted on a real-world dataset and utilized 9 machine learning methods with 3 feature selection methods to predict sales of the video game, the best performance of R^2 is up to 0.585.

II. DATA DESCRIPTION AND PROBLEM DEFINITION

A. Data Description

VGChartz is a video game sales tracking website, which was launched in 2005 by Brett Walton [23]. VGChartz records the sales information of video games in Japan, Europe, and America from 1970 to 2018. We collected and sorted out the sales data of 37,841 games and 17 gaming platforms from 1970 to 2018 and collected more than 35,324 video game information from 1970 to 2020. The annual released number and sales of video games can reflect the overall development trend of the entire video game industry. Figure 2 shows the changes in the number of video games released each year in Japan, Europe, and America from 1970 to 2020. On the whole, the annual number of video game releases in these countries has shown similar trends. It can observe from 1970 to 1985, the video game market in Europe and the two countries is in its early stages of development, in which no more than 300 video games are released each year. Since 1985, video games have begun to develop rapidly, and the growth trend has become increasingly large after 2004. Figure 3 shows the

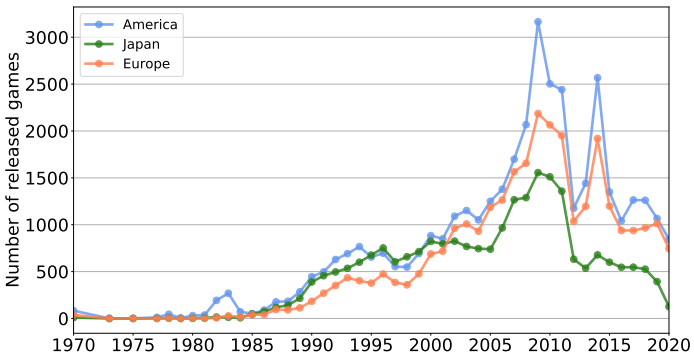


Fig. 2. Number of video games released per year from America, Japan, Europe

changes in the annual sales of video games in Japan, Europe, and America from 2004 to 2018. On the whole, the annual sales of electronic games in America and Europe have similar trends, while the annual sales of electronic games in Japan are in a trend of steady development, and the sales of video games in Japan do not exceed 500 million per year. Before 2006, sales in Europe and the two countries were on a steady upward trend. After 2006, the sales of video games in America countries and Europe began to increase rapidly. The sales of software video games in America reached their peak in 2011, and it reached 1,975,967,739. After a period of steady growth from 2008 to 2016, the sales of video games in Europe reached their peak in 2016, when the sales of video games reached 1,831,909,701. It can also be observed that after 2017, the sales of video games in Europe and the two countries have begun to show a downward trend. Through the analysis of the annual number of video game releases and sales, it can be found that the video game market began to develop rapidly after 2004. After 10 years, the video game market began saturation phenomenon has led to a downward trend in the number and sales and released games each year.

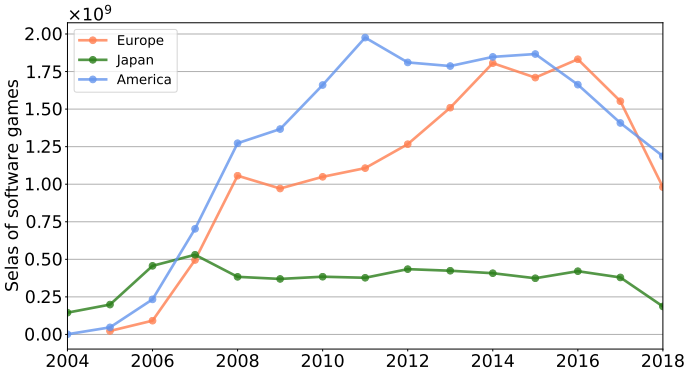


Fig. 3. Sales of video games per year from America, Japan, Europe

B. Problem Definition

For the relevant staff to be more fully prepared, the problem of video game sales prediction is a regression task that aims to predict sales of video games after 8 weeks. Here we give a formulation of this problem as below. Formally, let $X = \{X_1, X_2, ..., X_N\} \in \mathbb{R}^{i \times N}$ denoted as feature space, each $X_m \in X$, $X_m = \{x_1, x_2, ..., x_i\} \in \mathbb{R}^i$ denoted as arbitrary feature, where i represents the length of a time series and $i = 8$ in our study. For improving accuracy and reduce training time, let $\hat{X} = \{X_1, X_2, ..., X_n\} \in \mathbb{R}^{i \times n}$ denoted as new feature space after utilizing the feature selection method to reduce noise and decrease dimension. our target is to predict sales of video games after 8 weeks by the model f which utilizes past data of video game including sales, information, and the sales of predicting next 8 weeks represents as

$$\hat{y} = f(\hat{X}, \theta), \tag{1}$$

where θ is a set of model parameters, which are learned in the various model.

III. METHODOLOGY

A. Pearson Correlation Coefficient

PCC is commonly used to obtain the linear relationship between each feature and target, ρ ranges from -1 to 1 where $\rho < 0$ means that the feature and target is an inverse relation, and $\rho > 0$ means a correlated relation. ρ can be defined as

$$\rho = \frac{\sum_T^{t=1}(X_t - \bar{X})(Y_t - \bar{Y})}{\sqrt{\sum_T^{t=1}(X_t - \bar{X})^2}\sqrt{\sum_T^{t=1}(Y_t - \bar{Y})^2}} \tag{2}$$

where $\bar{X}$,$\bar{Y}$ is various X, Y average value respectively

B. Random Forest Feature Selection

The impurity are used by RFFS usually to calculated feature importance. Formally, RF is composed of decision trees M and nodes N of each $m \in M$, each node n_i has a number of samples w_i of arriving at n_i and its impurity is denoted as C_i, the change of impurity(p_{ij}) after adding feature j to node i is

$$p_{ij} = w_i C_i - w_{(j,l)} C_{(j,l)} - w_{(j,r)} C_{(j,r)} \tag{3}$$

where $C_{(i,l)}, C_{(i,r)}$ are the impurity values of the left and right subtrees of node i, $w_{(i,l)}, w_{(i,r)}$ are the number of samples of the left and right subtrees to reach node i respectively. The important coefficient between each feature and target can be represented as

$$\hat{p} = \frac{p_{ij}}{\sum_{k=1}^{N} p_{ik}}. \tag{4}$$

C. Pearson Correlation Coefficient - Random Forest Feature Selection

PCC-RFFS is a hybrid feature selection method combined PCC and RFFS, figure 1 illustrates its detailed process. Firstly, calculate the PCC correlation and RFFS importance coefficient between each feature and target respectively to obtain the set $P = \{\rho_1, \rho_2, \ldots, \rho_N\}$ and the set $\hat{P} = \{\hat{p}_1, \hat{p}_2, \ldots, \hat{p}_N\}$. Because of $\rho \in [-1, 1]$ and $\hat{p} \in [0, 1]$, we adopt function of absolute value for PCC, which can be on the same order of magnitude, each score

$$z = |\rho| + \hat{p} \tag{5}$$

represents the importance of each two arbitrary features. Then the set of features important score $Z = \{z_1, z_2, \ldots, z_N\}$. Ultimately, we output subset $\hat{D}$ of the top n of Z as the final dataset by ranking techniques.

IV. Experiment Result

We established a video game sales research dataset containing 5087 video games from February 2006 to November 2016 based on VGChartz. Table IV shows each feature in the research dataset. To establish a sales prediction model for the after 8 weeks, we divided the dataset with 53 features into a training set (79.8%) from February 2006 to December 2015 and a testing set(20.2%) from January 2016 to December 2018. After that, figure 4 illustrates the distribution of results of three different feature selection methods, to better analyze the influence of feature selection methods on modeling, group 1 and group 2 used the features for coefficients of the ranking of top 5 and 10 as input to the model respectively. Then 9 machine learning models were adopted for data-driven modeling: Adaboost(Ada), Catboost(CAT), Decision Tree(DT), Extreme learning machine(EXT), Gradient Boosting Decision Tree(GBDT), K-Nearest Neighbor(KNN), LightGBM(LGB), Random Forest(RF), Xgboost(XGB). In order to evaluate the effect of each prediction model, we used 4 regression evaluation indicators to evaluate the model: Mean Absolute Error (MAE), Root Mean Square Error(RMSE), Coefficient

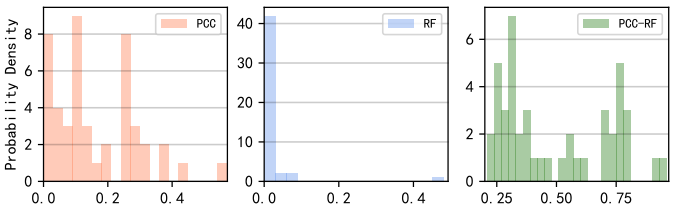


Fig. 4. Probability density distribution of different feature selection methods

Features	Description
Console	Game hardware or software equipment
Weekly sales on hardware	Sales of weekly based on hardware equipment
Weekly sales on software	Sales of weekly based on software equipment
Sales on hardware	Total sales based on hardware from the platform
Sales on software	Total sales based on software from the platform
Platform weekly sales on hardware	Total weekly sales on hardware equipment from the platform
Platform weekly sales on software	Total weekly sales on software equipment from software
Whole sales	Total sales of the game
Weekly sales	Weekly sales of the game
Week count	Number of the weeks after the game released
Year	Year of record
Week	Week of record
Day	Day of record
Pid	Unique ID of the platform
Genre	Game genre
Release type	The game was released in the first half or the second half of the year
Pos	Sales ranking

of Determination(R^2), and Mean Absolute Percentage Error (MAPE).

$$MAE = \frac{1}{T} \sum_{t=1}^{T} |\hat{y}_t - y_t|, \tag{6}$$

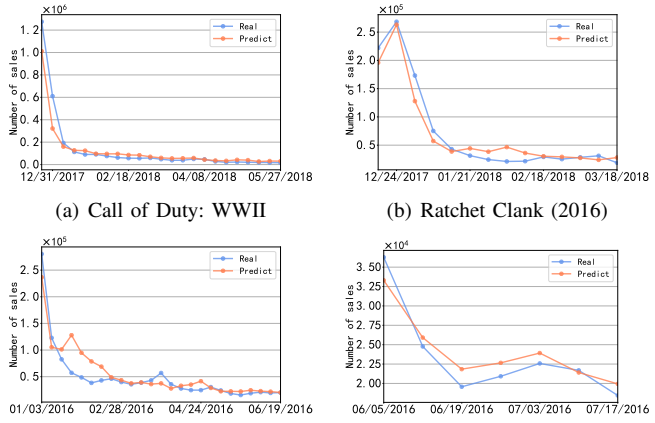
$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (\hat{y}_t - y_t)^2}, \tag{7}$$

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}} = 1 - \frac{\sum_{t=1}^{T} (y_t - \hat{y}_t)^2}{\sum_{t=1}^{T} (y_t - \overline{y})^2}, \tag{8}$$

$$MAPE = \frac{100\%}{T} \sum_{t=1}^{T} \left| \frac{\hat{y}_t - y_t}{y_t} \right|, \tag{9}$$

where y represents the true value of the sample, and $\hat{y}$ represents the predicted value of the sample. Let y represents the average of the true values of T samples. y_t and $\hat{y}_t$ respectively represented as the true value and predicted value of the t sample.

Table I illustrates the best results of each feature selection method after establishing prediction models. It is worth noting that the results after utilizing RCC-RF in group 1 and group 2 are the best. In Group 1, compared with two feature selection methods, the model Ada with PCC-RFFS has the best effect. R^2 increases 0.1575 and 0.1635, and RMSE decreases 12,301 and 12,742, respectively. In Group 2, the best performance of the model LGB with PCC-RFFS outperforms others. R^2



(a) Call of Duty: WWII (b) Ratchet Clank (2016)

(c) Fallout 4 (d) Assassin's Creed Origins

Fig. 5. Results of game sales prediction

increased 0.1317 and 0.0587, and RMSE also decreased 11,058 and 4,940, respectively. The best performance of the prediction model with RFFS has better performance than the best performance of the prediction model with PCC. By comparing group 1 and group 2, it is observed that in the establishment of a model for sales prediction, the use of PCC-RFFS for feature selection and LGB for data modeling in Group 2 has the best prediction results, R^2 can reach 0.585, and RMSE is 81,173. Figure 5 shows its result in predicting the sales of different games.

TABLE I
THE BEST RESULT OF DIFFERENT GROUPS USING 9 MACHINE LEARNING METHODS

Group	Group 1			Group 2		
	PCC5	RF5	PCC-RFFS5	PCC10	RF10	PCC-RFFS10
Model	Ada	GBDT	Ada	Ada	Ada	LGB
RMSE	93474	93915	81173	86439	80321	75381
MAE	39492	37240	34101	34280	32469	33741
R^2	0.3619	0.3559	0.5188	0.4543	0.5288	0.585
MAPE	54.5174	46.8986	44.5724	42.6346	43.5326	49.6132

V. CONCLUSION

With the market saturation of video games gradually, sales prediction of video games can promote pre-adjust sales strategy and plan of game development for the company, which make the company more adaptable to market changes. In this paper, we proposed a new hybrid feature selection method Pearson Correlation Coefficient-Random Forest Feature Selection(PCC-RFFS) to predict sales of the video game by 9 machine learning methods. We designed 2 groups of experiments using different feature selection methods and compared the best experimental results. The results can be summarized as follows: i) The performance of the model using machine learning with PCC-RFFS outperforms the model using machine learning with Pearson correlation coefficient(PCC) or with Random Forest Feature selection(RFFS). ii) When selecting a few of the features, our approach still has high performance and exceeds PCC and RFFS by more than 0.15 on R^2. iii) We used real-world data VGChartz to predict the sales of video games, and the performance of R^2 is up to 0.585. However, our proposed sales prediction approach still has limitations. For sales prediction, one of the reasons why the prediction results are not excellent is that a large number of sales of item have similar trend like figure 5, they are a process of rapid decline over time, and they have a huge number of sales are concentrated in the game released early stage. In the future, we will adopt the segment prediction method to predict the game released early and late stage. Moreover, we will utilize more feature selection methods as a baseline to compare their performance, and also will adopt more advanced algorithms to make the models have a better ability to predict sales.

ACKNOWLEDGMENT

This work was supported by Science and Technology Program of Guangzhou, China (201904010224), and Natural Science Foundation of Guangdong Province, China (2020A1515010761).

REFERENCES

[1] P. Meulstee and M. Pechenizkiy, "Food sales prediction:" if only it knew what we know"," in *2008 IEEE International Conference on Data Mining Workshops*. IEEE, 2008, pp. 134–143.

[2] M. Li, S. Ji, and G. Liu, "Forecasting of chinese e-commerce sales: an empirical comparison of arima, nonlinear autoregressive neural network, and a combined arima-narnn model," *Mathematical Problems in Engineering*, vol. 2018, 2018.

[3] M. Gao, W. Xu, H. Fu, M. Wang, and X. Liang, "A novel forecasting method for large-scale sales prediction using extreme learning machine," in *2014 Seventh International Joint Conference on Computational Sciences and Optimization*. IEEE, 2014, pp. 602–606.

[4] M. Á. López-Cabarcos, D. Ribeiro-Soriano, and J. Piñeiro-Chousa, "All that glitters is not gold. the rise of gaming in the covid-19 pandemic," *Journal of Innovation & Knowledge*, vol. 5, no. 4, pp. 289–296, 2020.

[5] J. Marcoux and S.-A. Selouani, "A hybrid subspace-connectionist data mining approach for sales forecasting in the video game industry," in *2009 WRI World Congress on Computer Science and Information Engineering*, vol. 5. IEEE, 2009, pp. 666–670.

[6] P. Ramos, N. Santos, and R. Rebelo, "Performance of state space and arima models for consumer retail sales forecasting," *Robotics and computer-integrated manufacturing*, vol. 34, pp. 151–163, 2015.

[7] W. Ching, S. Zhang, and M. Ng, "On multi-dimensional markov chain models," *Pacific Journal of Optimization*, vol. 3, no. 2, pp. 235–243, 2007.

[8] O. Wisesa, A. Adriansyah, and O. I. Khalaf, "Prediction analysis sales for corporate services telecommunications company using gradient boost algorithm," in *2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP)*. IEEE, 2020, pp. 101–106.

[9] B. Pavlyshenko, "Using stacking approaches for machine learning models," in *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*. IEEE, 2018, pp. 255–258.

[10] Y. Dai and J. Huang, "A sales prediction method based on lstm with hyper-parameter search," in *Journal of Physics: Conference Series*, vol. 1756, no. 1. IOP Publishing, 2021, p. 012015.

[11] Y. Kaneko and K. Yada, "A deep learning approach for the prediction of retail store sales," in *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)*. IEEE, 2016, pp. 531–537.

[12] K. Bandara, P. Shi, C. Bergmeir, H. Hewamalage, Q. Tran, and B. Seaman, "Sales demand forecast in e-commerce using a long short-term memory neural network methodology," in *International conference on neural information processing*. Springer, 2019, pp. 462–474.

[13] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014.

[14] Z. Xia, S. Xue, L. Wu, J. Sun, Y. Chen, and R. Zhang, "Forexgboost: passenger car sales prediction based on xgboost," *Distributed and Parallel Databases*, vol. 38, pp. 713–738, 2020.
[15] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on neural networks*, vol. 5, no. 4, pp. 537–550, 1994.
[16] P. A. Estévez, M. Tesmer, C. A. Perez, and J. M. Zurada, "Normalized mutual information feature selection," *IEEE Transactions on neural networks*, vol. 20, no. 2, pp. 189–201, 2009.
[17] J. Leng, C. Valli, and L. Armstrong, "A wrapper-based feature selection for analysis of large data sets," 2010.
[18] A. Wang, N. An, G. Chen, L. Li, and G. Alterovitz, "Accelerating wrapper-based feature selection with k-nearest-neighbor," *Knowledge-Based Systems*, vol. 83, pp. 81–91, 2015.
[19] B. H. Menze, B. M. Kelm, R. Masuch, U. Himmelreich, P. Bachert, W. Petrich, and F. A. Hamprecht, "A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data," *BMC bioinformatics*, vol. 10, no. 1, pp. 1–16, 2009.
[20] Q. Zhou, H. Zhou, and T. Li, "Cost-sensitive feature selection using random forest: Selecting low-cost subsets of informative features," *Knowledge-based systems*, vol. 95, pp. 1–11, 2016.
[21] R. Guha, K. K. Ghosh, S. Bhowmik, and R. Sarkar, "Mutually informed correlation coefficient (micc)-a new filter based feature selection method," in *2020 IEEE Calcutta Conference (CALCON)*. IEEE, 2020, pp. 54–58.
[22] Z. Hu, Y. Bao, T. Xiong, and R. Chiong, "Hybrid filter–wrapper feature selection for short-term load forecasting," *Engineering Applications of Artificial Intelligence*, vol. 40, pp. 17–27, 2015.
[23] B. Walton, "Vgchartz data," https://www.vgchartz.com/, accessed August 4, 2021.