Predictive Modelling of Local Currency Sovereign Debt yields using Machine Learning

Yang Zhou *

Department of Statistics, University of Warwick, Coventry, United Kingdom

* Corresponding Author Email: Sam.Zhou@warwick.ac.uk

Abstract. This paper explores applying Statistical Machine Learning (ML) models to forecast local currency 10-year bond yields in 17 Emerging Markets (EM) economies, using Economic & Financial Indicators as feature parameters. The models considered include Random Forest (RF), Gradient Boosting (GB), Lasso Regression, Ridge Regression, and Linear Regression (LR). The performance is assessed using three key metrics: Mean Squared Error (MSE), R2 on the testing set, and R2 on Cross-Validation (CV). The results show that RF and GB models generally outperform traditional linear models, showing lower MSE and higher R2 values, indicating better accuracy in their predictions. By exploring the practical application of those models in the context of EM economies, this paper contributes valuable insights into the predictive modelling of sovereign debt yields, which is crucial for investors and policymakers. The findings suggest that ML models, with their ability to capture non-linear relationships, offer a more reliable approach to predicting 10-year bond yields in EM economies.

Keywords: Machine learning, Bond Yield forecasting, Emerging markets, Sovereign debt, Macroeconomic forecasting.

1. Introduction

Investing in EM Rates involves navigating more uncertainties than Developed Markets (DM). EM countries face a greater risk of default and debt restructuring, driven by weaker fiscal conditions, which can significantly impact bond pricing and liquidity [1]. Additionally, political instability and geopolitical risks weigh more heavily on EM Rates, with sudden regime changes, sanctions, or conflicts causing sharp market sell-offs unrelated to global macro fundamentals. The goal of this research effort was to develop accurate statistical machine learning models to predict local currency 10-year bond yields across 17 EM countries, which could be used to identify trading signals.

There is an ever-growing literature on ML applications in the world of macro fixed income because of their capability to learn complex non-linear relationships. For instance, ML models have been effectively used in predicting real GDP growth [2] and financial crises [3]. However, sovereign debt yield prediction through applying ML techniques to Economic & Financial indicators has more commonly been focused on DM, such as the US [4, 5], and significantly less attention has been directed to predicting local currency sovereign debt yields in EM. Existing studies in EM often focus on understanding how macroeconomic factors influence bond yield, primarily using LR models [6], rather than employing ML techniques for more accurate forecasting. This highlights a gap in the literature on applying more advanced forecasting methods to EM sovereign debt markets. To expand on this growing literature, this paper’s contribution focuses on applying ML models to forecast local currency 10-year bond yields specifically in EM economies, filling a notable gap in the existing research. By showcasing the effectiveness of ML models in capturing the relationships between Economic & Financial indicators and Sovereign Debt yields, this study contributes to academic research and practical forecasting in global macro fixed-income markets.

2. Methodology

2.1. Linear Models

LR models the relationship between a target variable y and features x as a linear function [7]. The basic LR model is expressed as:

$$y = x^{\top}\beta + \epsilon \tag{1}$$

Where the coefficients to be estimated are denoted by β, and the error term is normally distributed with a mean of zero, denoted by ϵ.

Lasso Regression introduces an ℓ1 penalty term on the size of the coefficients [8], defined as:

$$\hat{\beta}^{\text{lasso}} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\top}\beta)^2 + \lambda\|\beta\|_1 \tag{2}$$

Where $\lambda > 0$ is the regularisation parameter [7]. Automatic feature selection can be achieved through the application of Lasso's ℓ1-penalty, which can shrink some of the coefficients exactly to zero [9]. K-fold CV was performed to select the optimal value of λ. In K-fold CV, the dataset is divided into K approximately equal-sized folds [2]. The Lasso model is trained on K-1 folds and validated on the remaining fold. This procedure is repeated K times so each fold is utilized to validate once [7]. The average validation error across all K iterations is computed for a range of λ values. Formally, the optimal regularization parameter $\lambda*$ minimizes CV error (λ), which is calculated as [7].

$$\lambda^* = \arg\min_{\lambda} CV(\lambda) = \arg\min_{\lambda} \frac{1}{K}\sum_{k=1}^{K} \text{MSE}_k(\lambda) \tag{3}$$

Where $\text{MSE}_k(\lambda)$ is the MSE on the k-th validation set. Once $\lambda*$ is determined, the Lasso model is retrained on the full dataset using this optimal $\lambda*$.

Ridge Regression introduces an ℓ2 penalty term on the size of the coefficients, defined as [7]:

$$\hat{\beta}^{\text{ridge}} = \arg\min_{\beta} \sum_{i=1}^{n} (y_i - x_i^{\top}\beta)^2 + \lambda\|\beta\|_2^2 \tag{4}$$

Where $\lambda > 0$ is a regularisation parameter that limits the amount of shrinkage. This penalty term shrinks the coefficients proportionally to address multicollinearity [10]. Like Lasso Regression, CV is performed to find the optimal $\lambda*$.

2.2. Machine Learning Models

RF, introduced by Breiman (2001), is an ensemble method that builds a number of decision trees on bootstrapped training samples and outputs the average prediction of the individual trees [11]. The mathematical idea behind RF can be understood through the Central Limit Theorem, which justifies the reduction in variance when the sample mean is taken as the final prediction [7]

Let $\hat{y}^{(b)}(x)$ represent the prediction of the b-th decision tree for input x, where $b \in \{1, 2, \ldots, B\}$ and B is the total number of trees in the forest [7]. The ultimate RF prediction is the mean of the predictions from all trees [7]:

$$\hat{y}_{\text{RF}}(x) = \frac{1}{B}\sum_{b=1}^{B} \hat{y}^{(b)}(x) \tag{5}$$

Assume that each tree's prediction $\hat{y}^{(b)}(x)$ is a random variable with mean $\mathbb{E}\left[\hat{y}^{(b)}(x)\right] = \mu(x)$ and variance $\text{Var}\left[\hat{y}^{(b)}(x)\right] = \sigma^2(x)$. If the tree predictions are uncorrelated (or weakly correlated), then the variance of the RF prediction decreases with the number of trees B according to:

$$\text{Var}(\hat{y}_{\text{RF}}(x)) = \frac{\sigma^2(x)}{B} \tag{6}$$

In building those decision tresses, only a subset of variables is used during each split, rather than greedily choosing the best split point in the construction of the tree [8]. The explanation for this is simple. If there is a very strong predictor among the complete set of predictors, the majority of decision trees will use this predictor in the initial split, resulting in highly correlated predictions. The RF model will not significantly reduce variance compared to a single tree [7].

GB, introduced by Friedman (2001) [12], is another powerful ensemble method. The fundamental concept involved constructing a model sequentially, where each decision tree (a weak learner) is trained to forecast the residual error of the current model [13]. Over time, these weak learners correct the mistakes of their predecessors, leading to a strong predictive model. GB has an error function:

$$L(y, F(x)) = \frac{1}{2}(y - F(x))^2 \tag{7}$$

It aims to model a function F (x) that approximates the relationship between input variables x ∈ Rd and a continuous target variable y ∈ R. The model F (x) is built to minimize a predefined loss function L (y, F (x)). The GB algorithm constructs the model F (x) through an iterative process. At each step m, the model is updated as [2]:

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{8}$$

Where $h_m(x)$ is the weak learner added at step m, and γ_m is the learning rate [2]. At each step, $h_m(x)$ is fitted to the residual [13]. The residuals at iteration m are defined as:

$$r_{i,m} = -\frac{\partial L(y_i, F_{m-1}(x_i))}{\partial F_{m-1}(x_i)} \tag{9}$$

This ensures that the new learner focuses on the errors made by the current model, leading to a reduction in the overall loss function.

2.3. Model Evaluation

To compare the performance of different regression models, the following metrics are used:

Testing set R2 measures the model's ability to generalise to unseen data [7]. It is calculated as:

$$R^2_{\text{test}} = 1 - \frac{\sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n_{\text{test}}} (y_i - \bar{y})^2} \tag{10}$$

Where y_i is the actual historical value; $\hat{y}_i$ is the predicted value from the model; $\bar{y}$ is the mean of the actual historical values. The observation number in the testing set is denoted as n_{test}. A value near 1 suggests a perfect fit, but a value near 0 indicates the model does not explain any variance in the target variable.

Cross-validated R2 is computed using K-fold C and calculated as [7]:

$$R^2_{\text{CV}} = \frac{1}{K}\sum_{k=1}^{K} R^2_{\text{test},k} \tag{11}$$

Where $R^2_{\text{test},k}$ is the R2 value calculated on the k-th test fold, and K is the number of folds.

MSE on the testing set measures the mean squared difference between the actual historical values y_i and the predicted values $\hat{y}_i$ [7]:

$$\text{MSE}_{\text{test}} = \frac{1}{n_{\text{test}}}\sum_{i=1}^{n_{\text{test}}} (y_i - \hat{y}_i)^2 \tag{12}$$

3. Results

3.1. Data preparation

To predict 10-year bond yields across 17 EM countries, the initial set of features for this analysis was drawn from a broad range of Economic & Financial indicators including Retail Sales, Consumer Price Index, Producer Price Index, Policy Rate, Unemployment Rate, Debt-to-GDP ratio, Fiscal and Budget Deficit as a percentage of GDP, 5-year/10-year Credit Default Swap (CDS) prices, Brent Oil Price, US 10-year Treasury yield, Consumer Confidence Index. The dataset was collected from

Bloomberg and Macrobond. The number of observations, along with the start and end years of the data collection period, is presented in Figure 1r.

To address the complexity that many features are not independent, interaction terms (e.g., Inflation × Policy Rate) were included to capture feature dependencies [7]. Furthermore, it is often unrealistic to assume a linear relationship between features and the target variable. For example, inflation's effect on bond yields might increase when inflation is above the country's inflation target. This relationship was better captured by introducing quadratic or cubic terms.

The dataset was divided into features X and target variable y. It was then split into training sets (70%) and testing sets (30%). The training sets were used to fit the model; the testing sets were held out to evaluate the model performance on unseen data. Hypothesis tests were performed to determine whether each feature has a statistically significant effect on the target variable (10-year bond yield).

- Null Hypothesis (H0): The feature's coefficient is zero
- Alternative Hypothesis (H1): The feature's coefficient is not zero

If the p-value was below 0.05, H0 was rejected, suggesting that the feature is statistically significant and should be included in the model [7].

LR was used as the baseline model for feature selection due to its ability to directly assess the statistical significance of each feature and the same set of selected features is subsequently used in other regression models. Separate models were created for each country to account for differences in economic structure and stages of the economic cycle, which cannot be captured effectively if a single global model was used.

3.2. Feature selection

LR was used as the baseline model for feature selection. For each country's model, features that are statistically significant are presented in Fig. 1.

	coef
const	4.1992
Central Bank Rate	0.6700
CPI inflation	0.5203
Unemployment Rate	-0.3339
10-year CDS	0.0165
Retail Sales(y0y%)	0.1192
Central Bank Rate ^ 2	-0.0211
Inflation & Retail Sales Interaction	-0.0157
Inflation & policy rate interaction	-0.0337

(a) BRL

	coef
const	2.1496
Central Bank Rate	0.5472
Central Bank Rate ^2	-0.0215
CPI Inflation ^ 2	0.0091
US 10-year bond yield ^2	0.1880
US inflation	0.3899
5Y CDS	0.0073
US inflation & rate interaction	-0.1222

(b) MXN

	coef
const	0.9382
Interest Rate	1.0238
Interest Rate ^ 2	-0.0262
CPI inflation	-0.2321
CDS	0.0098
US inflation	0.5842
Rate & inflation interaction	0.0128

(c) TRY

	coef
const	3.2786
CPI inflation	0.1231
10-year CDS	0.0094
US 10 year bond yield	0.4926
US inflation/interest rate interaction	0.0901

(d) COP

	coef
const	0.2938
10-year CDS	0.0079
US Inflation	0.2732
US Inflation ^ 2	0.0682
US 10 year bond yield	1.5183
US inflation & interest rate interaction	-0.2600

(e) CLP

	coef
const	2.9839
CPI inflation	0.2654
Brent Oil price	-0.0071
Central Bank Rate	-0.0865
US inflation & rate interaction	0.0403
10-year CDS	0.0070
US 10 year bond yield	0.5444

(f) PEN

	coef
const	4.5791
CPI inflation	-0.3456
10-year CDS	0.0128
US Inflation	0.6831
US 10 year bond yield ^ 2	0.1878
Local inflation & rates interaction	0.0245
US inflation & rate interaction	-0.1765

(g) ZAR

	coef
const	-0.5693
Central Bank Rate	0.2620
Unemployment Rate	0.0878
CPI inflation	0.1927
CDS 5Y Price	0.0095
US 10 year bond yield	0.7340
Inflation & Rates Interaction	-0.0178

(h) PLN

	coef
const	0.4421
US inflation	-0.1234
Central Bank Rate	0.5701
US inflation & rate interaction	0.0558
Central Bank Rate ^ 2	-0.0858
CPI inflation	0.0560
Current Account Balance (% of GDP)	-0.4034
Real GDP	-0.1532
US 10 year bond yield	0.5650
Current Account & GDP interaction	0.0274

(i) CZK

	coef
const	0.4682
Central Bank Rate	1.6117
Central Bank Rate ^ 2	-0.0767
PPI	0.0938
US Inflation	-0.2216
Inflation & Rate interactions	-0.0081

(j) RON

	coef
const	2.0095
Central Bank Rate	0.9984
Local rates & US rates interaction	-0.0653
US real interest rate	0.1616
Local inflation & rates interaction	-0.0173
PPI change	0.0735

(k) HUF

	coef
const	3.6763
US 10-year bond yied ^2	0.1638
US 10-year bond yied ^ 3	-0.0361
Oil Price	0.0100
Central Bank Rate	0.4192

(l) INR

	coef
const	0.6717
Central Bank Rate	0.9158
PPI Inflation	-0.2485
Consumer Confidence Index	0.0044
US Inflation	-0.2234
US Inflation ^ 2	0.0114
Local Inflation & Rate interaction	0.1299
US inflation & rate interaction	0.0301

(m) CNY

	coef
const	7.6552
Core CPI	-0.9094
US 10 year bond yield	1.0732
US Inflation	-0.1513
BoI Consumer Confidence	-0.0156
US 10 year bond yield ^ 2	-0.2631
5Y CDS price	0.0098
Core CPI & US rates interaction	0.3163

(n) IDR

	coef
const	1.1078
Central Bank Rate ^ 2	0.2052
Central Bank Rate ^ 3	-0.0232
CPI inflation	0.2472
Consumer Confidence Index	-0.0060
US Inflation	-0.2793
US 10 year bond yield	0.9786
Local rate & US rate interaction	-0.1048

(o) THB

	coef
const	1.9347
US Inflation	-0.0798
US Inflation ^ 2	0.0099
US 10 year bond yield	1.0699
US 10 year bond yield ^ 2	-0.1415
5y CDS price	0.0030

(p) MYR

	coef
const	-1.3714
Oil	-0.0388
Central Bank rate	0.9962
CPI Inflation	1.3200
US interest rate	1.2072
US inflation & oil price interaction	0.0046
US inflation & rate interaction	-0.1365
5y CDS Price	0.0109
Inflation & Rate Interaction	-0.2166

(q) PHP

Country	Start year	End year	Number of observations
Brazil	2012	2024	98
Mexico	2011	2023	107
Turkey	2011	2023	107
Colombia	2010	2024	120
Chile	2014	2024	76
Peru	2006	2024	163
South Africa	2001	2024	225
Poland	2003	2024	177
Czech Repub	2007	2024	144
Romania	2007	2024	135
Hungary	2007	2024	144
India	2012	2024	98
China	2012	2024	98
Indonesia	2006	2024	146
Thailand	2000	2024	252
Malaysia	2010	2024	138
Philippines	2009	2024	122

(r) Data collection

Figure 1. Feature Selection

Features included in the LR model varied across countries. For example, in Hungary, 10-year bond yield is significantly influenced by local inflation and the policy rate, whereas in Chile, 10-year bond yield is significantly influenced by US inflation and the 10-year US treasury yield. Those features are then used in all other regression models to provide us with the model comparison results in the next subsection.

3.3. Model comparison

In the following subsection, a detailed evaluation of different models' predictive capabilities based on the results obtained for each metric is provided. The best-performing model for each country, based on each specific metric (MSE, Test R2, and CV R2), is highlighted in red to indicate superior performance.

	Lasso_MSE	Ridge_MSE	Linear_MSE	RF_MSE	GB_MSE
Brazil	0.629676	0.584849	0.585392	0.404744	0.326144
Mexico	0.11435	0.10089	0.102529	0.136468	0.123865
Turkey	4.501791	2.336509	2.36755	1.627472	1.401192
Colombia	0.370629	0.326098	0.32609	0.235874	0.246524
Chile	0.167028	0.160838	0.161235	0.265625	0.271099
Peru	0.250946	0.23575	0.235272	0.162104	0.193852
South Africa	0.321995	0.283358	0.284319	0.336378	0.40726
Poland	0.259261	0.256713	0.256402	0.198676	0.151089
Czech Repub	0.15765	0.158388	0.160689	0.085083	0.074169
Romania	1.354921	1.34527	1.345767	0.705021	0.744294
Hungary	0.624406	0.570955	0.568379	0.231472	0.300085
India	0.09948	0.101283	0.102662	0.09526	0.053125
China	0.034205	0.03376	0.033921	0.026675	0.043518
Indonesia	0.482382	0.37659	0.375877	0.238473	0.252557
Thailand	0.334771	0.327305	0.327359	0.198348	0.218875
Malaysia	0.024459	0.025021	0.024896	0.031397	0.036518
Philippines	0.661375	0.474874	0.484123	0.599129	0.491277

(a) MSE Comparison

Country	Lasso_Test_R2	Ridge_Test_R2	Linear_Test_R2	RF_Test_R2	GB_Test_R2
Brazil	0.872226	0.881322	0.881212	0.90901	0.933819
Mexico	0.890753	0.903612	0.902047	0.896884	0.881663
Turkey	0.673827	0.83071	0.828461	0.898406	0.898478
Colombia	0.864096	0.880425	0.880428	0.917763	0.909603
Chile	0.831356	0.837606	0.837205	0.737849	0.726279
Peru	0.824606	0.835227	0.835561	0.88188	0.864511
South Africa	0.743344	0.774141	0.773375	0.778205	0.675381
Poland	0.916264	0.917087	0.917188	0.949492	0.951202
Czech Repub	0.933908	0.933599	0.932634	0.961947	0.968906
Romania	0.811715	0.813056	0.812987	0.902322	0.89657
Hungary	0.894076	0.903143	0.90358	0.96042	0.949094
India	0.805799	0.80228	0.799588	0.836363	0.896291
China	0.795114	0.797778	0.796814	0.829152	0.739333
Indonesia	0.75895	0.811815	0.812171	0.876176	0.873795
Thailand	0.794502	0.799085	0.799052	0.894872	0.865644
Malaysia	0.862656	0.859503	0.860203	0.839277	0.794942
Philippines	0.640658	0.741989	0.736964	0.751021	0.733077

(b) Test R2 Comparison

Country	Lasso_CV_R2	Ridge_CV_R2	Linear_CV_R2	RF_CV_R2	GB_CV_R2
Brazil	0.815519	0.854811	0.8545	0.841186	0.871483
Mexico	0.845827	0.866618	0.865868	0.919049	0.91959
Turkey	0.72692	0.857665	0.855465	0.92024	0.90298
Colombia	0.844557	0.860543	0.860716	0.910913	0.898818
Chile	0.827471	0.83407	0.834424	0.852321	0.834822
Peru	0.752627	0.756913	0.757195	0.816461	0.786828
South Africa	0.62983	0.654188	0.652508	0.822934	0.847936
Poland	0.888609	0.887989	0.888807	0.943292	0.947986
Czech Repub	0.908153	0.90862	0.908478	0.954678	0.954017
Romania	0.864881	0.866877	0.866912	0.911509	0.900417
Hungary	0.876693	0.879859	0.879748	0.947019	0.947916
India	0.820663	0.824701	0.823055	0.903826	0.891743
China	0.82357	0.826867	0.826772	0.839246	0.789522
Indonesia	0.766985	0.801729	0.801769	0.859485	0.854796
Thailand	0.775579	0.780284	0.780494	0.863031	0.814113
Malaysia	0.745946	0.731672	0.731789	0.838243	0.857533
Philippines	0.634105	0.74713	0.743746	0.829383	0.801338

(c) CV R2 Comparison

Figure 2. Performance of Different Models Across Countries

The models' performance in terms of MSE is shown in Figure 2a. The RF and GB models show meaningfully lower MSE values compared to linear models in most countries. RF models show particularly low MSE in countries such as Peru with 0.1621, Poland with 0.1987, Thailand with 0.1983, and Hungary with 0.2315, GB models performed best in countries such as Brazil, India, Czech Republic, with MSE values of 0.3261, 0.0531, and 0.0742, respectively. Nevertheless, Ridge shows the lowest MSE values in countries such as the Philippines with 0.4749, and South Africa with 0.2834, although the improvement is only marginal in comparison to LR.

The Test R2 results, shown in Figure 2b, illustrate that both RF and GB models perform well in terms of R2 on the test sets, indicating strong explanatory power. In several cases, RF and GB outperform traditional linear models by a significant margin. Notable examples include Hungary, where R2 values are 0.9604 for RF, 0.9491 for GB, and 0.7996 for LR; and Thailand, where RF achieves an R2 of 0.8949, GB an R2 is 0.8656, and LR an R2 of 0.7991. When comparing the linear models, Ridge Regression again tends to perform marginally better. For example, in Brazil, Ridge achieves an R2 of 0.8813, outperforming Lasso with an R2 of 0.8722 and LR with an R2 of 0.8812.

Finally, Figure 2c presents the results of CV R2. ML models consistently perform strongly across all countries during CV, further demonstrating their predictive power. Countries like Poland, the Czech Republic, and Hungary show very close CV R2 values for RF and GB, highlighting their comparable strengths. In Poland, RF achieves a CV R2 of 0.9433, GB a CV R2 of 0.9480, while in the Czech Republic RF achieves a CV R2 of 0.9547, GB a CV R2 of 0.9540.

4. Summary

Overall, the analysis suggests that Machine Learning models are highly effective in predicting Local Currency Sovereign Debt yields. In most countries, RF and GB outperform linear models, especially in countries such as Hungary, Thailand, and India, where they achieve significantly lower MSE and higher R2 scores. In certain cases, GB has a slight edge in predictive performance, such as in Brazil, Turkey, and India, while in others, RF performs marginally better. This suggests that both models are suitable for capturing the complexities in the dataset, and either could be favoured depending on the specific country or data characteristics. Among linear models, Ridge Regression occasionally yields competitive results, where it slightly outperforms the other linear models. Future research could extend this analysis by experimenting with more advanced ML models, such as deep learning methods, to further enhance predictive power. In conclusion, the recommendation is to consider both RF and GB as leading models for predicting local currency sovereign debt yields, as demonstrated in this comparative study.

References

[1] Clavé, Peter. Trading Fixed Income and FX in Emerging Markets: A Practitioner's Guide. Wiley, 2016: 13.

[2] Yoon Jaehyun. Forecasting of Real GDP Growth Using Machine Learning Models: Gradient Boosting and Random Forest Approach. Journal of Economic Forecasting, 2022, 25 (3): 50 - 68.

[3] Bluwstein Kristina, Jung Philip, Taylor Alan M. Credit Growth, the Yield Curve and Financial Crisis Prediction: Evidence from a Machine Learning Approach. Macroeconomic Policy Review, 2023, 47 (1): 15 - 28.

[4] Dias Daniel. Forecasting Long-Term Government Bond Yields: An Application of Statistical and AI Models. Working Paper Series, Instituto Superior de Economia e Gestão (ISEG), 2006: 2 - 10.

[5] Li Zhihua, Wang Bing, Chen Yiming. Incorporating Economic Indicators and Market Sentiment Effect into US Treasury Bond Yield Prediction. Financial Forecasting Journal, 2024, 32 (4): 67 - 80.

[6] Buzas Laszlo, Feyen Erik, Nie Oleksiy, Paladines Carolina. Analyzing the Determinants of EMDE Local-Currency Sovereign Yields: A Simple Comparative Approach. Risks, 2022, 10(8): 123 - 140.

[7] James Gareth, Witten Daniela, Hastie Trevor, Tibshirani Robert. An Introduction to Statistical Learning: With Applications in R. Springer, 2013: 327 - 351, 197 - 216, 59 - 83.

[8] Swaminathan Manohar. Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python. Apress, 2017: 156, 225 - 233.

[9] Tibshirani Robert. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society: Series B (Methodological), 1996, 58 (1): 267 - 288.

[10] Hoerl Arthur E, Kennard Robert W. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics, 1970, 12 (1): 55 - 67.

[11] Breiman Leo. Random Forests. Machine Learning, 2001, 45 (1): 5 - 32.

[12] Friedman Jerome H. Greedy Function Approximation: A Gradient Boosting Machine. Annals of Statistics, 2001, 29 (5): 1189 - 1231.

[13] Géron Aurélien. Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O’Reilly Media, 2017: 199 - 204, 37 - 42.