International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

A Two-Stage Decision-Support Framework for Crop Recommendation and Harvest-Price Forecasting: A Case Study of Sehore District, Madhya Pradesh, India

Anas Shamshad[1], Sarah Shamshad[1], Aryan Mathur [1] Priyanshu Sahu[1], Aditya Sinha[1], Krish Singh Sengar[1], Pratosh Kumar Pal[1]

[1]School of Computing Science Engineering and Artificial Intelligence, VIT Bhopal University, Kothrikalan, Sehore, Madhya Pradesh - 466114, India

---

Abstract - Farmers in rainfed districts of central India must choose crops and plan sales under simultaneous agronomic and price uncertainty, yet most decision-support tools address only one of the two. This paper presents a two-stage framework for Sehore district, Madhya Pradesh, that couples unsupervised crop-suitability assessment with harvest-time price forecasting. In the first stage, soil (N, P, K, pH) and weather variables are grouped by K-means and Gaussian mixture model (GMM) clustering, and the two solutions are combined into an ensemble; on the weather data, the ensemble reached a silhouette score of 0.4885, a Davies–Bouldin index of 0.9814 and a Calinski–Harabasz score of 65.25, indicating moderate cluster separation. In the second stage, a stacked ensemble of Random Forest, Extra Trees, XGBoost and Ridge regressors forecasts market prices from historical mandi prices, minimum support price, season, year and weather features. On the held-out test set, the stacked model achieved an MAE of INR 243.6 per quintal, an RMSE of INR 308.0, an R2 of 0.967 and a MAPE of 7.37%, compared with an MAE of INR 251.6, an RMSE of INR 360.9 and an R2 of 0.947 for a Random Forest baseline. Both stages are served through a web interface that returns candidate crops and price forecasts from user-entered field conditions. The test set is small and the suitability stage has not been validated against agronomic ground truth, so the results are presented as a proof of concept; the validation needed before field deployment is outlined.

Key Words: *crop recommendation; price forecasting; K-means; Gaussian mixture model; stacking ensemble; decision support; Agmarknet; Sehore*

1. INTRODUCTION

Agriculture remains the main livelihood for a large share of India's rural population, and variable rainfall, soil fertility, and market prices shape its outcomes. Sehore district in Madhya Pradesh illustrates the problem. It has a subtropical climate with an annual rainfall of about 1,261 mm, predominantly medium dark black soils, and a Kharif–Rabi system in which soybean, paddy, tur and maize are followed by wheat and gram [1, 2]. Many farmers still depend on rainfall, and cropping decisions are commonly guided by past experience and generic advisories, which reflect neither field-level soil status nor the price a crop is likely to fetch at harvest.

Machine learning has been applied to both halves of this decision. Crop-recommendation systems select a suitable crop from soil nutrients (N, P, K), pH, temperature, humidity and rainfall, most often with supervised learners [3–5]. Separately, agricultural price forecasting has moved from statistical models toward machine-learning and deep-learning models that use weather and other exogenous variables [6–8]. The two lines of work are rarely combined, although a crop that is agronomically suitable may still be unprofitable, and a crop with a favourable price outlook may be poorly matched to local soil and climate. Integrated systems, such as the voice-enabled framework of Sindhur et al. for Karnataka [9], show the value of linking suitability and price, but they depend on labeled suitability data and on sequence models for the price series.

District-level deployment raises two practical difficulties. First, expert-verified suitability labels are scarce for a single district, making supervised classifiers hard to train and audit. Second, local price records are limited, which favours compact models over deep sequence architectures. This paper addresses both difficulties for Sehore district by (i) grouping soil and weather conditions with unsupervised clustering instead of labelled suitability data, and (ii) forecasting prices with a stacked ensemble of tree-based and linear regressors that uses lagged prices and the minimum support price (MSP) as predictors.

The contributions are: (1) a two-stage pipeline that links unsupervised crop-suitability assessment to harvest-price forecasting for a single district; (2) an ensemble of K-means and Gaussian mixture model (GMM) clusterings evaluated with three internal validity indices; (3) a stacked ensemble of Random Forest, Extra Trees, XGBoost and Ridge regressors for price forecasting, benchmarked against a Random Forest baseline; and (4) a working web prototype, together with a candid appraisal of its limitations. Section 2 reviews related work, Section 3 describes the study area and data, Section 4 presents the methodology, Section 5 reports the results, Section 6 discusses implications and limitations, and Section 7 concludes.

2. RELATED WORK

2.1 Crop recommendation

Supervised learning dominates the crop-recommendation literature. Pande et al. [3] and Gosai et al. [4] recommend crops from soil-nutrient and weather attributes with conventional classifiers, and Shastri et al. [5] combine gradient boosting with explainable artificial intelligence to make recommendations transparent. These systems are typically trained and evaluated on tabular datasets in which each record carries a crop label, and their reported accuracy depends on the quality and representativeness of those labels, which are hard to audit for a specific district. Where district-specific labels are unavailable, unsupervised methods offer an alternative: they group similar

soil and climate conditions so that crops can be associated with each group.

2.2 Agricultural price forecasting

Agricultural prices are seasonal, non-stationary, and sensitive to weather, so forecasting research has moved from ARIMA-type models toward machine-learning and deep-learning approaches. Manogna et al. [6] compared statistical, machine-learning, and deep-learning models on daily wholesale prices of 23 commodities (January 2010–June 2024) and found LSTM and GRU networks to be the most accurate. Nayak et al. [7] showed that adding precipitation and temperature as exogenous variables to NBEATSX and TransformerX improved forecasts of tomato, onion, and potato prices in Indian markets relative to statistical and machine-learning benchmarks. Graph-based models have also been explored: Özden and Bulut [8] applied a spectral temporal graph neural network to multivariate vegetable prices and reported it as the best-performing model against CNN, LSTM and Random Forest baselines. These results favour deep sequence models where long, regular series are available. Where records are short or irregular, tree ensembles with lagged and policy variables such as the MSP remain a practical alternative.

2.3 Integrated decision-support systems

Few studies couple suitability and price. Sindhur et al. [9] pair a Random Forest suitability classifier (98.5% reported accuracy) with an LSTM price forecaster and deliver the output through a Kannada voice interface for farmers in Karnataka. The present work follows the same suitability-then-price logic but differs in three respects: suitability is derived from unsupervised clustering rather than labeled data; price forecasting uses a stacked tree-and-linear ensemble with MSP and lagged-price predictors; and the study is confined to a single district, which allows local soil and weather characteristics to be represented explicitly.

3. Study Area and Data

3.1 Study area

Sehore district lies in central Madhya Pradesh and spans more than one agro-climatic zone [1, 2]. Its climate is subtropical, with an annual rainfall of about 1,261 mm, summer temperatures of roughly 25–45 °C and winter temperatures of 10–25 °C. The soils are predominantly medium dark black soils with medium nitrogen and phosphorus and medium-to-high potash, and sulphur, zinc, iron and manganese are deficient [1]. Soybean, paddy, tur and maize are the main Kharif crops, wheat and gram the main Rabi crops, and summer (Zaid) moong is expanding [1].

3.2 Data sources and variables

Three groups of data were assembled (Table 1). Market prices were obtained from Agmarknet, the Government of India's agricultural marketing portal [10]. Weather data were drawn from NASA POWER [11], Meteostat [12] and WeatherSpark [13]. Soil status enters the framework as N, P and K at low, medium and high levels and pH at acidic, neutral and alkaline values (Fig. 2a); the soil-data source is Sehore District Administration, Agriculture Department, supplemented by the ICAR–Central Research Institute for Dryland Agriculture (ICAR–CRIDA) Agriculture Contingency Plan for Sehore district, Madhya Pradesh. The MSP announced for each crop and marketing year was included as a policy variable in the price model.

Table -1: Data used in the study.

Module	Variables	Source	Records/period
Soil clustering	N, P, K levels; pH class	Sehore District Administration; ICAR–CRIDA	3 records (2023–24 to 2025–26)
Weather clustering	Temperature, humidity class, precipitation type, wind, pressure	NASA POWER; Meteostat; WeatherSpark	132 records (January 2015–December 2025)
Price forecasting	Crop, season, year, market price (INR per quintal), previous price and price change, MSP, weather features	Agmarknet; MSP announcements	66 records (2021–2026; 26 crops)

3.3 Preprocessing

Preprocessing followed a common sequence for both stages: (1) loading and merging of the source tables; (2) treatment of missing values, with interpolation of gaps in weather series and removal of incomplete price records; (3) scaling of numerical features so that distance-based clustering and linear models operate on comparable ranges; (4) encoding of categorical variables (season, crop, humidity and precipitation classes); and (5) a split into training and test sets at an 80:20 split, giving a test set of n = 14 observations.

4. METHODOLOGY

4.1 Framework overview

Fig. 1 summarises the framework. A user supplies soil and weather conditions, season and year through a web interface. After preprocessing, Stage 1 assigns the inputs to soil and weather clusters and returns candidate crops with suitability scores. Stage 2 forecasts the harvest-time price of each candidate crop. A decision layer then combines the two outputs (Section 4.4). The framework is designed to log user inputs and predictions so that models can be retrained as new data accumulate.

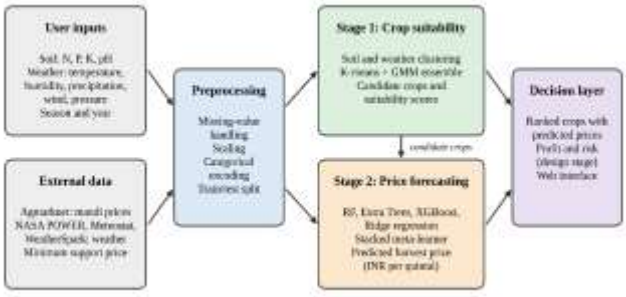


Fig -1: Architecture of the two-stage framework. GMM, Gaussian mixture model; RF, Random Forest; INR, Indian rupee.

4.2 Stage 1: crop suitability by clustering

Suitability is assessed by clustering rather than classification because labelled suitability records are not available for the district. Soil variables (N, P, K, pH) and weather variables are clustered separately. K-means [14] partitions scaled observations into k groups by minimising within-cluster squared distance, with k selected by the elbow criterion [15]. A Gaussian mixture model fitted by expectation–maximisation [16] provides a probabilistic alternative in which every observation has a posterior membership probability for each component. The two solutions are combined into an ensemble clustering.

The K-means and GMM cluster assignments were combined using majority voting between the two model outputs. Both the soil and weather K-means models used k = 3 clusters, while the corresponding GMM models used three components. The GMMs used a full covariance matrix. . A new field is assigned to a soil cluster and a weather cluster, and the crops associated with this pair are returned as candidates ranked by a suitability score. Crops are linked to the joint soil–weather cluster pair through a predefined cluster-to-crop mapping. After the soil and weather cluster labels are obtained, the corresponding pair is used as a key in the crop map to retrieve the candidate crops. The current prototype does not compute a separate numerical suitability score; instead, crops associated with the matched cluster pair are returned as candidate recommendations.
Clustering quality was assessed with three internal validity indices. The silhouette coefficient [17] of observation i is

$$s(i) = \frac{b(i)-a(i)}{\{a(i),b(i)\}} \quad (1)$$

where a(i) is the mean distance from i to the other members of its cluster and b(i) is its mean distance to the nearest other cluster; the mean over all observations lies between −1 and 1, and higher values are better.
The Davies–Bouldin index [18] averages, over clusters, the ratio of within-cluster scatter to between-cluster separation for the most similar neighbouring cluster, so lower values indicate better separation. The Calinski–Harabasz score [19] is the ratio of between-cluster to within-cluster dispersion, adjusted for degrees of freedom, and higher values indicate more compact and better-separated clusters.

4.3 Stage 2: price forecasting
For each candidate crop, the harvest-time price is predicted by regression. Predictors comprised crop identity, season, year, the applicable MSP, the previous market price and its change (Fig. 2b), and weather features. No weather variables were included in the final saved price-prediction model. The final predictors were crop identity, season, year, MSP, previous market price, and price change. Weather variables were used in the separate weather-clustering stage. Four base learners were trained: Random Forest [20], Extra Trees [21], XGBoost [22] and Ridge regression [23]. Their predictions were combined by stacked generalisation [24], in which a meta-learner combines the base learners' outputs. The base-model predictions were combined using Ridge regression as the meta-learner, with α = 5.0. The stacking ensemble used 5-fold cross-validation to generate the base-model predictions used by the meta-learner.

A Random Forest fitted on the same predictors served as the baseline. Models were implemented in Python with scikit-learn [25].
Performance on the held-out test set was measured by mean absolute error (MAE), root mean squared error (RMSE), the coefficient of determination (R2), median absolute error and mean absolute percentage error (MAPE) [26]:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \quad | y_i - \hat{y}_i | \quad (2)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (y_i - \hat{y}_i)^2} \quad (3)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \quad (y_i-\hat{y}_i)^2}{\sum_{i=1}^{n} \quad (y_i-\bar{y})^2} \quad (4)$$

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n} \quad \left| \frac{y_i-\hat{y}_i}{y_i} \right| \quad (5)$$

where n is the number of test samples, yi the observed price, ŷi the predicted price and ȳ the mean observed price. The mean squared error is the square of the RMSE, and the median absolute error is the median of |yi − ŷi|.

4.4 Decision layer
The decision layer combines suitability and price. For crop c, the expected profit per hectare is

$$\Pi_c = P_c \cdot Y_c - C_c \quad (6)$$

where Pc is the forecast price, Yc the expected yield per hectare from regional benchmarks and Cc the input cost (seed, fertiliser, irrigation and labour). Risk is characterised by the historical variability of the crop's price and the spread of its forecast, and candidate crops are ranked jointly by suitability, expected profit and risk. In the current prototype the interface returns cluster assignments, candidate crops and price forecasts (Fig. 2); the profit and risk computations are specified here as the intended design and have not yet been evaluated.

4.5 Implementation
The prototype is a Python web application built with Streamlit. Users enter N, P and K levels, pH, humidity, temperature, precipitation, wind, pressure, prediction year and season; the application returns the soil and weather cluster labels, the recommended crops, and a table of predicted prices together with the MSP and previous price used by the model (Fig. 2). The interface is intended for farmers and extension officers with limited technical support.

5. RESULTS

5.1 Crop-suitability clustering
Table 2 reports the validity indices for the weather clusters. The ensemble clustering obtained a silhouette score of 0.4885, a Davies–Bouldin index of 0.9814 and a Calinski–Harabasz score of 65.25. On the Kaufman–Rousseeuw scale [27], silhouette values between 0.26 and 0.50 indicate weak-to-moderate structure, so the clusters are informative but overlap. K-means obtained the highest silhouette score (0.4938), followed by the ensemble (0.4885) and the GMM (0.4824). These differences are of the order of 0.01, and the silhouette score alone does not separate the three approaches; the ensemble was retained in the framework for its balanced performance.

Table -2: Validity indices for the weather-data clusters.

Method	Silhouette	DBI	CH
K-means	0.4938	n/r	n/r
GMM	0.4824	n/r	n/r
Ensemble (K-means + GMM)	0.4885	0.9814	65.25

DBI, Davies–Bouldin index (lower is better); CH, Calinski–Harabasz score (higher is better); n/r, not reported..

5.2 Price forecasting
Table 3 shows the stacked model's performance. On the test set it achieved an MAE of INR 243.65 per quintal, an RMSE of INR 307.99, an R2 of 0.967 and a MAPE of 7.37%. Errors on

the training set were lower (MAE 111.02; R2 0.994; MAPE 2.94%), a gap expected for tree-based ensembles fitted to a small sample. The test median absolute error (170.64) was well below the mean (243.65), indicating that a few large errors dominate the error distribution.

Table -3: Performance of the stacked price-prediction model (errors in INR per quintal).

Metric	Training	Test
MAE	111.02	243.65
MSE	20,600.34	94,857.45
RMSE	143.53	307.99
R^2	0.9941	0.9671
Median absolute error	82.93	170.64
MAPE (%)	2.94	7.37

Table 4 compares the stacked model with the Random Forest baseline. Stacking reduced the MAE by 3.1% and the RMSE by 14.7%, and raised R2 from 0.947 to 0.967. The larger gain in RMSE than in MAE indicates that the ensemble mainly reduces large errors. Because the test set is small [n = 14], these differences have not been tested for statistical significance and should be read as indicative. Direct comparison with the studies reviewed in Section 2 is not meaningful, because they use different crops, horizons and error metrics.

Table -4: Random Forest baseline versus stacked model on the test set (errors in INR per quintal).

Metric	Random Forest	Stacked model	Change
MAE	251.56	243.65	−3.1%
RMSE	360.91	307.99	−14.7%
R^2	0.9471	0.9671	+0.0200

5.3 Prototype output

Fig. 2b shows an example query for summer 2026. The soil inputs were assigned to soil cluster 0 and the weather inputs to weather cluster 1, and the system returned five candidate crops: jowar (sorghum), sesame, arhar (tur), black gram (urad) and green gram (moong). Forecast prices ranged from INR 2,768 per quintal for jowar to INR 8,322 per quintal for green gram. The example demonstrates the end-to-end flow from user input to candidate crops with price forecasts; it is not evidence of agronomic accuracy, which requires validation against field records and expert assessment.



Fig -2: Web prototype for Sehore district. (a) Input form for soil and weather parameters. (b) Output for an example query (summer 2026): soil cluster, weather cluster, candidate crops and price forecasts in INR per quintal.

6. DISCUSSION

6.1 Interpretation

The results show that the two stages can be coupled in a lightweight system that runs on a district's own data. In Stage 1, unsupervised clustering yields structure that is moderate rather than sharp (silhouette 0.49), which is plausible for soil and weather variables that vary continuously and overlap between neighbouring conditions. In Stage 2, the high test R2 should be read with the predictors in mind: the MSP and the previous price are likely to carry much of the information about the next price of a crop, so the model may be learning mainly an adjustment around recent price levels. A naïve persistence baseline (previous price) was not evaluated and should be reported alongside the Random Forest and stacking results.

6.2 Practical implications

For farmers and extension officers, the system offers in one place the candidate crops and their expected prices, which generic advisories do not provide, and its modular design allows crops, districts and data sources to be added. Because the forecasts are point estimates, the interface should communicate uncertainty, for example through prediction intervals, before it is used to support financial decisions.

6.3 Limitations

Several limitations qualify the findings. First, the price model was evaluated on a small held-out set, so the metrics are sensitive to individual observations and no confidence intervals or significance tests are reported. Second, Stage 1 was evaluated only with internal validity indices on weather data; the crops associated with each cluster have not been validated against yield records, agronomic guidelines, or expert judgment, and a silhouette of 0.49 indicates overlapping clusters. Third, lagged prices and the MSP may inflate R2

under a random split, so time-ordered validation and a persistence baseline are needed to establish the added value of the model. Fourth, the gap between training and test error suggests some overfitting. Fifth, expected profit and risk have not been evaluated, and cost and yield inputs would come from regional benchmarks rather than farm records. Sixth, the prototype has not been tested in a controlled study with farmers, so no claim can be made about usability, decision confidence or income gains. Seventh, all results are for Sehore district, and transfer to other districts is untested. Finally, soil tests are infrequent, mandi records can be incomplete, and shocks such as droughts or lockdowns remain difficult to predict.

7. CONCLUSIONS

This study presented a two-stage framework that links unsupervised crop-suitability assessment with harvest-price forecasting for Sehore district. The clustering stage produced moderately separated groups (silhouette 0.4885 for the K-means and GMM ensemble), and the stacked ensemble of Random Forest, Extra Trees, XGBoost, and Ridge regressors improved on a Random Forest baseline (RMSE 307.99 versus 360.91; R2 0.967 versus 0.947) on a small test set. Both stages are available through a web prototype. The work is a proof of concept. Future work will (i) enlarge the price and soil datasets and evaluate with time-ordered cross-validation against persistence and sequence-model baselines; (ii) validate the suitability stage against yield records and expert assessment; (iii) add prediction intervals and evaluate the profit and risk layer with cost-of-cultivation data; and (iv) test the interface with farmers and extension officers, including local-language support.

ACKNOWLEDGEMENT

The authors thank Mr. Debraj Gogoi and Dr. Pradeep Kumar Kashyap for their constructive comments, and the faculty of the School of Computing Science and Engineering, VIT Bhopal University, for their support.

REFERENCES

1. Sehore District Administration. Agriculture. https://sehore.nic.in/?p=24754
2. ICAR–Central Research Institute for Dryland Agriculture. Agriculture Contingency Plan for District: Sehore, Madhya Pradesh. 2012. https://icar-crida.res.in/CP/madhya%20pradesh/MP16-Sehore-26.6.2012.pdf
3. Pande SM, Ramesh PK, Anmol A, Aishwarya BR, Rohilla K, Shaurya K. Crop recommender system using machine learning approach. In: 2021 5th International Conference on Computing Methodologies and Communication (ICCMC). Erode, India: IEEE; 2021. https://doi.org/10.1109/ICCMC51019.2021.9418351
4. Gosai D, Raval C, Nayak R, Jayswal H, Patel A. Crop recommendation system using machine learning. Int J Sci Res Comput Sci Eng Inf Technol. 2021;7(3):554–557. https://doi.org/10.32628/CSEIT2173129
5. Shastri S, Kumar S, Mansotra V, Salgotra R. Advancing crop recommendation system with supervised machine learning and explainable artificial intelligence. Sci Rep. 2025;15:25498. https://doi.org/10.1038/s41598-025-07003-8
6. Manogna RL, Dharmaji V, Sarang S. Enhancing agricultural commodity price forecasting with deep learning. Sci Rep. 2025. https://doi.org/10.1038/s41598-025-05103-z
7. Nayak GHH, Alam MW, Singh KN, Avinash G, Kumar RR, Ray M, Deb CK. Exogenous variable driven deep learning models for improved price forecasting of TOP crops in India. Sci Rep. 2024;14. https://doi.org/10.1038/s41598-024-68040-3
8. Özden C, Bulut M. Spectral temporal graph neural network for multivariate agricultural price forecasting. Cienc Rural. 2024;54(1). https://doi.org/10.1590/0103-8478cr20220677
9. Sindhur NM, Pavithra C, Muchikel N. A hybrid machine learning framework for optimizing crop selection via agronomic and economic forecasting. arXiv:2507.08832. 2025. https://arxiv.org/abs/2507.08832
10. Agricultural Marketing Information Network (Agmarknet), Government of India. https://agmarknet.gov.in
11. NASA Langley Research Center. Prediction Of Worldwide Energy Resources (POWER) Project. https://power.larc.nasa.gov
12. Meteostat. Weather and climate data. https://meteostat.net
13. WeatherSpark. Historical weather data. https://weatherspark.com
14. Lloyd SP. Least squares quantization in PCM. IEEE Trans Inf Theory. 1982;28(2):129–137. https://doi.org/10.1109/TIT.1982.1056489
15. Thorndike RL. Who belongs in the family? Psychometrika. 1953;18(4):267–276. https://doi.org/10.1007/BF02289263
16. Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J R Stat Soc Series B. 1977;39(1):1–38. https://doi.org/10.1111/j.2517-6161.1977.tb01600.x
17. Rousseeuw PJ. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. J Comput Appl Math. 1987;20:53–65. https://doi.org/10.1016/0377-0427(87)90125-7
18. Davies DL, Bouldin DW. A cluster separation measure. IEEE Trans Pattern Anal Mach Intell. 1979;PAMI-1(2):224–227. https://doi.org/10.1109/TPAMI.1979.4766909
19. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat. 1974;3(1):1–27. https://doi.org/10.1080/03610927408827101
20. Breiman L. Random forests. Mach Learn. 2001;45(1):5–32. https://doi.org/10.1023/A:1010933404324
21. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. Mach Learn. 2006;63(1):3–42. https://doi.org/10.1007/s10994-006-6226-1
22. Chen T, Guestrin C. XGBoost: a scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM; 2016. p. 785–794. https://doi.org/10.1145/2939672.2939785
23. Hoerl AE, Kennard RW. Ridge regression: biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55–67. https://doi.org/10.1080/00401706.1970.10488634
24. Wolpert DH. Stacked generalization. Neural Netw. 1992;5(2):241–259. https://doi.org/10.1016/S0893-6080(05)80023-1
25. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res. 2011;12:2825–2830.
26. Hyndman RJ, Koehler AB. Another look at measures of forecast accuracy. Int J Forecast. 2006;22(4):679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001
27. Kaufman L, Rousseeuw PJ. Finding Groups in Data: An Introduction to Cluster Analysis. Hoboken: Wiley; 1990. https://doi.org/10.1002/9780470316801