



Optimization and Validation of Two Machine Learning Algorithms for Accurate Prediction of Irrigated Wheat (*Triticum aestivum* L.) Yield and Identification of its Influential Factors in Khorasan Razavi Province

M. Jahan ^{1*}

1- Department of Agrotechnology, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran

(*- Corresponding author's Email: jahan@ferdowsi.um.ac.ir)

Received: 04 May 2025

Revised: 09 July 2025

Accepted: 22 July 2025

Available Online: 22 July 2025

How to cite this article:

Jahan, M. (2025). Optimization and Validation of Two Machine Learning Algorithms for Accurate Prediction of Irrigated Wheat (*Triticum aestivum* L.) Yield and Identification of its Influential Factors in Khorasan Razavi Province. *Iranian Journal of Field Crops Research*, 23(4), 511-543. (In Persian with English abstract) <https://doi.org/10.22067/jcesc.2025.93323.1395>

Introduction

This study undertook a detailed comparison of two supervised machine-learning algorithms—Random Forest (RF) and eXtreme Gradient Boosting (XGBoost)—to predict irrigated wheat (*Triticum aestivum* L.) yield across 20 counties in Razavi Khorasan Province. Both models were trained on 70 % of the dataset (years 1383–1402) and tested on the remaining 30 %. Hyperparameter tuning was performed via a grid search coupled with five-fold cross-validation.

Materials and Methods

The two algorithms were initially trained and optimized using 70% of the data (approximately 14 years per county) and subsequently tested on the remaining 30% (approximately 6 years). Hyperparameter tuning was performed through a grid search combined with five-fold cross-validation. Hyperparameter Tuning was performed using Gridsearch over key parameters (e.g., number of trees $n_{\text{estimators}}$, maximum depth max_depth , learning rate for XGBoost) with 5-fold cross-validation. Afterwards, both models were evaluated and validated using RMSE (Root Mean Squared Error), R^2 (Coefficient of Determination), MAE (Mean Absolute Error), Willmott's d (Index of Agreement). Two machine learning algorithms, Random Forest and XGBoost, were developed to predict wheat performance at the district level. The performance values were initially predicted numerically using regression and then divided into three distinct classes using statistical percentiles: Low: 0th to 33rd percentile; Medium: 33rd to 66th percentile; High: 66th to 100th percentile. This classification was based on both the actual and predicted values for each district. The results of this classification are presented in the form of a confusion matrix for each model. Using three indices—Aridity Index, Seasonal Intensity of Temperature, and Growing Degree Days—districts in the province were clustered into three climatic zones. Then, wheat performance in these zones was analyzed based on two models. In this study, the machine learning algorithms Random Forest and XGBoost were implemented in Python 3.3 using the Scikit-learn library. DataFrame preparation and the clustering of the province's counties into climatic zones were carried out using R 4.3.2 and ArcGIS 10.8.2.

Results and Discussion

Algorithms Performance: On average, RF reduced prediction error by ~19 % (395 vs. 492 kg ha⁻¹ RMSE) and achieved a slightly higher agreement with observed yields ($d=0.467$ vs. 0.419). Random forest showed Best



Authors retain the copyright. This is an open access article distributed under [Creative Commons Attribution 4.0 International License \(CC BY 4.0\)](https://creativecommons.org/licenses/by/4.0/).

 <https://doi.org/10.22067/jcesc.2025.93323.1395>

RF Performance for Khalilabad ($RMSE = 141.19 \text{ kg ha}^{-1}$, $MAE = 125.68 \text{ kg ha}^{-1}$, $d = 0.37$) and Bardeskan ($RMSE = 187.69 \text{ kg ha}^{-1}$, $d = 0.80$); Worst RF Performance was obtained for Quchan ($RMSE = 667.65 \text{ kg ha}^{-1}$, $d = 0.36$) and Torbat-e Heydarieh ($RMSE = 581.33 \text{ kg ha}^{-1}$, $d = 0.47$). Best XGBoost Performance was obtained for Torbat-e Jam ($RMSE = 269.36 \text{ kg ha}^{-1}$, $d = 0.77$), Neyshabur ($RMSE = 377.91$, $d = 0.78$). Worst XGBoost Performance resulted for Quchan ($RMSE = 943.99 \text{ kg ha}^{-1}$, $d = 0.25$) and Torbat-e Heydarieh ($RMSE = 786.20 \text{ kg ha}^{-1}$, $d = 0.39$). In 6 out of 20 counties (Khaf, Mahvelat, Kalat-e Nader, Kashmar, Chenaran, Fariman) both RF and XGBoost performed nearly identical errors ($\Delta RMSE < 15 \text{ kg ha}^{-1}$), indicating similar predictive power under those local conditions.

Feature Importance: Daily Minimum Temperature (Tmin): Ranked #1 in RF's importance list; ranked #3 in XGBoost. Seasonal Tmin (TminGS): Consistently #3 in both models. Other Key Predictors: Precipitation over the growing season (Prec, PGS), Growing Degree Days (GDDGS), and Evapotranspiration (ETGS) all contributed substantially, though with 4th–8th ranks markedly lower in RF than in XGB.

Classification of Yield into Three Performance Classes: Using percentile thresholds—Low (0–33rd), Medium (34–66th), High (67–100th)—the models were also evaluated as classifiers. Low-Performing Counties Needing Intervention Seven counties (35 % of the sample)- Quchan, Torbat-e Heydarieh, Sarakhs, Kalat-e Nader, Gonabad, Neyshabur, Taybad- fell in the Low performance class across both models. These areas should be prioritized for targeted agronomic management and resource allocation.

Cluster-Based Insights: Counties were grouped into three agro-climatic clusters: Very Dry & Hot (4 counties): RF outperformed XGB in all (e.g., Bardeskan, Khalilabad, Mahvelat, Sarakhs). Semi-humid Cooler (6 counties): Mixed results-RF won in 4 (Chenaran, Kalat-e Nader, Mashhad, Nishapur); XGB was slightly better in 2 (Fariman, Quchan). Warm Semi-arid (10 counties): RF superior in 7; equivalence in Taybad & Torbat-e Heydarieh; XGB edged ahead only in no counties here.

Conclusion

Overall, the Random Forest model showed better results for predicting wheat yield in Razavi Khorasan province, especially in most counties. Although the XGBoost model has higher potential for modeling complex patterns, Random Forest performed more accurately in conditions of greater data dispersion. In conclusion, the results of this study emphasize that by utilizing climatic and agricultural data, machine learning algorithms can be optimized not only to achieve high accuracy but also to provide a clear interpretation of the contribution of each variable in wheat yield using the SHAP tool.

Keywords: Cross-validation, Grid search, Random Forest, XGBoost, Yield class

بهینه‌سازی و اعتبارسنجی دو الگوریتم یادگیری ماشین برای پیش‌بینی دقیق عملکرد گندم آبی (*Triticum aestivum* L.) و تعیین عوامل مؤثر بر آن در استان خراسان رضوی

محسن جهان^{۱*}

تاریخ دریافت: ۱۴۰۴/۰۲/۱۴

تاریخ پذیرش: ۱۴۰۴/۰۴/۳۱

چکیده

در این تحقیق به مقایسه دو الگوریتم یادگیری ماشین، رندوم‌فارست و ایکس‌جی‌بوست، در پیش‌بینی عملکرد گندم (*Triticum aestivum* L.) آبی در ۲۰ شهرستان استان خراسان رضوی پرداخته شد. دو الگوریتم ابتدا با ۷۰ درصد داده‌ها آموزش دیده و بهینه‌سازی شدند و سپس با ۳۰ درصد داده‌ها مورد آزمون قرار گرفتند. بهینه‌سازی دو الگوریتم به کمک جستجوی شبکه‌ای با اعتبارسنجی متقابل پنج مرحله‌ای انجام گرفت. در ادامه، هر دو الگوریتم ارزیابی و تعیین اعتبار شدند. از نظر معیارهای ارزیابی، مانند جذر میانگین مربعات خطا (RMSE)، ضریب تبیین (R^2)، میانگین قدر مطلق خطا (MAE) و شاخص توافق ویلموت (d)، مدل رندوم‌فارست در بسیاری از شهرستان‌ها عملکرد بهتری داشت، درحالی‌که ایکس‌جی‌بوست در مناطقی با توزیع پیچیده‌تر داده‌ها (مانند مشهد) توانست به‌طور مؤثرتری عمل کند. مقایسه نتایج مدل‌ها در شهرستان‌های مختلف نشان می‌دهد که مدل رندوم‌فارست در بسیاری از شهرستان‌ها دقت بالاتری داشت. برای مثال، در شهرستان‌هایی مانند خلیل‌آباد، سبزوار، تربت‌جام، رشت‌خوار، نیشابور و گناباد، مدل رندوم‌فارست خطای کمتری نسبت به ایکس‌جی‌بوست داشت. از سوی دیگر، مدل ایکس‌جی‌بوست در بعضی مناطق خاص مانند مشهد، بردسکن، درگز، سرخس و تایباد عملکرد بهتری از خود نشان داد که احتمالاً به ویژگی‌های خاص این شهرستان‌ها مربوط می‌شود. برای شهرستان‌های خاف، مه‌ولات و کلات نادر، هر دو مدل یکسان عمل کردند. تحلیل شیب نشان داد که در هر دو مدل، دمای کمینه روزانه نقش برجسته‌ای در پیش‌بینی عملکرد گندم ایفا کرد. در مدل رندوم‌فارست، دمای کمینه روزانه رتبه اول را داشت، درحالی‌که در مدل ایکس‌جی‌بوست این ویژگی به‌طور معمول از اهمیت کمتری برخوردار بود. سایر ویژگی‌ها مانند بارندگی و درجه روز رشد نیز به‌شکل قابل توجهی در دقت مدل‌ها تأثیرگذار بودند. مدل‌ها همچنین در طبقه‌بندی عملکرد گندم با استفاده از ماتریس درهم‌ریختگی ارزیابی شدند. مدل ایکس‌جی‌بوست به‌طور کلی دقت بالاتری در طبقه‌بندی کلاس‌های عملکرد داشت، درحالی‌که رندوم‌فارست در شبیه‌سازی دقیق‌تر داده‌های واقعی و کاهش خطاهای پیش‌بینی موفق‌تر عمل کرد. به‌طور کلی، مدل رندوم‌فارست برای پیش‌بینی عملکرد گندم در استان خراسان رضوی نتایج بهتری در بیشتر شهرستان‌ها نشان داد. اگرچه مدل ایکس‌جی‌بوست پتانسیل بالاتری برای مدل‌سازی الگوهای پیچیده دارد، اما رندوم‌فارست در شرایط پراکندگی بیشتر داده‌ها دقیق‌تر عمل کرد. نتایج این مطالعه تأکید کرد که با بهره‌گیری از داده‌های اقلیمی و زراعی می‌توان الگوریتم‌های یادگیری ماشین را بهینه کرده و علاوه بر دقت بالا، به کمک ابزار شیب، تحلیلی شفاف و کاربردی از نقش هر متغیر در عملکرد گندم ارائه نمود.

واژه‌های کلیدی: ایکس‌جی‌بوست، تعیین اعتبار متقابل، جستجوی شبکه‌ای، رندوم‌فارست، کلاس عملکرد

مقدمه

و امنیت غذایی را به یکی از مهم‌ترین چالش‌های قرن بیست‌ویکم تبدیل کرده است (Aslan, Sabanic, & Aslan, 2024; Moreno Sanchez, Acosta Mesa, Trueba Espinosa, Ruiz Castilla, & Garcia Lamont, 2025). در این شرایط، صرفاً افزایش تولید کافی نیست، بلکه بهینه‌سازی مصرف نهاده‌ها و بهبود بهره‌وری، کاهش ضایعات و پیش‌بینی دقیق عملکرد محصولات کشاورزی نیز از اولویت‌های اساسی سیاست‌گذاران و پژوهشگران است (Arun & Ghimire, 2019; Jhajharia, Mathur, Jain, & Nijhawan,

افزایش بی‌سابقه جمعیت جهان (پیش‌بینی شده است ۹/۷ میلیارد نفر تا سال ۲۰۵۰) همراه با کاهش منابع آب و خاک و تشدید تغییرات اقلیمی، فشار مضاعفی بر نظام تولید مواد غذایی وارد می‌آورند

۱- گروه آگروتکنولوژی، دانشکده کشاورزی، دانشگاه فردوسی مشهد، مشهد، ایران
(Email: jahan@ferdowsi.um.ac.ir)
* - نویسنده مسئول: <https://doi.org/10.22067/jcsc.2025.93323.1395>

۲۰۲۳). گندم (*Triticum aestivum* L.)، مهم‌ترین غله جهان، با سهم حدود ۲۱ درصد از تولید کالری جهانی، یکی از محورهای اصلی تأمین غذا محسوب می‌شود. برآورد می‌شود که تا سال ۲۰۵۰، تقاضای جهانی برای گندم تا ۶۰ درصد افزایش یابد، این در حالی است که نوسانات اقلیمی روند کاهشی عملکرد سالانه این محصول را تشدید می‌کنند (Hatfield & Prueger, 2015).

تغییرات اقلیمی تأثیر عمیقی بر عملکرد محصولات زراعی دارند. دما و بارش به‌تنهایی حدود ۳۰ تا ۴۰ درصد از واریانس تغییرات عملکرد سالیانه را توضیح می‌دهند (Ray, Gerber, Macdonald, & West, 2015). اثر منفی افزایش دما بر عملکرد گندم، جو و ذرت در مقیاس جهانی به اثبات رسیده است. گزارش‌های هیئت بین‌الدول تغییر اقلیم در چرخه ارزیابی هفتم (AR7) و به‌روزرسانی‌های منتشرشده پس از گزارش ششم ارزیابی (AR6) در سال ۲۰۲۳ منتشر شده است، نشان می‌دهد که گرمایش جهانی تا سال ۲۰۲۰ به ۱/۱ درجه سانتی‌گراد بالاتر از سطح پیش‌صنعتی رسیده است و پیش‌بینی می‌شود تا سال ۲۰۳۰، حتی در سناریوهای خوش‌بینانه، این افزایش ادامه یابد و به بیش از ۱/۵ درجه سانتی‌گراد نزدیک شود یا از آن فراتر رود (Asseng, Ewert, Martre, Potter, & Lobell et al., 2015; IPCC, 2022). وقوع و تکرار پدیده‌های شدید مانند خشکی‌های طولانی‌مدت، سیل و سرمازدگی، به‌خصوص در مناطق خشک و نیمه‌خشک، چالش‌های عدیده‌ای را برای تولید گندم به وجود آورده است که می‌تواند به‌سرعت زیربنای کشاورزی را تضعیف کند (Asseng et al., 2015; Raymundo, Asseng, Robertson, et al., 2018). مدل‌سازی‌های اقلیمی آینده نیز کاهش عملکرد محصولات زراعی را پیش‌بینی و تأیید می‌کنند.

در ایران، مطالعات متعددی کاهش میزان بارش‌های فصلی، به‌ویژه در فصول بهار و پاییز را گزارش کرده‌اند که به‌نوبه خود بحران کم‌آبی و تنش گرمایی را تشدید نموده است، همچنین مدل‌سازی‌های اثر تغییر اقلیم بر سیستم‌های زراعی کشور در سال‌های گذشته نشان‌دهنده روند کاهشی در عملکرد گندم بوده و پیش‌بینی‌های آن‌ها مبنی بر اینکه تا افق ۱۴۰۰ خورشیدی، کاهش عملکرد در برخی مناطق تا حدود ۴۵ درصد نیز می‌رسد، تحقق یافته است (Koocheki, Nassiri-Mahallati, Kamali, & Shahandeh, 2006; Javadi, Ghahremanzadeh, Sassi, Javanbakht, & Hayati, 2024). نوسانات مکانی و زمانی عملکرد گندم در سطح جهان و ایران نه‌تنها ناشی از تغییرات اقلیمی، بلکه نتیجه تفاوت در دسترسی به منابع آبی، مدیریت زراعی و ادوات کشاورزی است (Farhadi, Bannayan, Fallah, & Jahan, 2024; Roell, Beucher, Moller, Greve, & Greve, 2020). به‌عنوان مثال، در یک مطالعه پنلی در سه شهرستان استان خراسان رضوی، بیش از ۴۰ درصد تغییرات عملکرد

گندم با متغیرهای اقلیمی توضیح داده شد (Farhadi et al., 2024). پیش‌بینی دقیق عملکرد به کشاورزان کمک می‌کند تا منابعی مانند آب، کود و نیروی کار را به‌طور مؤثرتری به کار گیرند. پیش‌بینی عملکرد محصول از قبل، به برنامه‌ریزی بازاریابی و توزیع نیز کمک می‌کند و به ذی‌نفعان - از کشاورزان گرفته تا تولیدکنندگان مواد غذایی و خرده‌فروشان - اجازه می‌دهد تا برای نوسانات عرضه و تقاضا، کاهش ضایعات و اطمینان از در دسترس بودن مواد غذایی آمادگی پیدا کنند (Aslan et al., 2024). با پیچیده‌تر شدن روابط میان متغیرهای اقلیمی، خاکی و مدیریتی، نیاز به روش‌های پیشرفته‌ای برای مدل‌سازی و پیش‌بینی عملکرد گندم احساس می‌شود. اگرچه امروزه مدل‌های فرآیند-محور با امکانات گسترده‌ای در دسترس هستند، اما نیاز به تنظیم پارامترها، واسنجی و اعتبارسنجی این مدل‌ها تحت شرایط خاص و در نهایت وجود خطا در خروجی‌ها، همچنان موضوعی چالش برانگیز می‌باشد (Dhillon et al., 2023; Krishnadoss & Ramasamy, 2024). یادگیری ماشین به‌عنوان ابزاری نوظهور، به‌ویژه الگوریتم جنگل تصادفی که با ترکیب نتایج چندین درخت تصمیم، توانایی بالایی در شناسایی الگوهای غیرخطی و چندمتغیری دارد، در سال‌های اخیر محبوبیت یافته است (Nayak et al., 2022; Jhajharia et al., 2023; Khodjaev, Bobojonov, Kuhn, & Glauben, 2025). مطالعات متعددی گزارش کرده‌اند که مدل‌های مبتنی بر جنگل تصادفی، نسبت به مدل‌های آماری کلاسیک مانند رگرسیون خطی چندان، به‌طور محسوسی عملکرد بهتری در پیش‌بینی عملکرد گندم ارائه می‌دهند. برای مثال، جذر میانگین مربعات خطا مدل رندوم‌فارست در یک مطالعه منطقه‌ای ۱۶۳/۹۰ کیلوگرم بر هکتار (عملکرد گندم زمستانه) گزارش شد که در مقایسه با مدل رگرسیون خطی که ۶۵۳/۳۹ کیلوگرم بر هکتار بود، حاکی از دقت بیشتر آن می‌باشد (Shen et al., 2024). افزون‌بر این، روش‌های نوینی مانند رندوم‌فارست خودآموز با امکان گسترش مجموعه‌های داده برچسب‌دار و بهبود دقت پیش‌بینی نشان داده‌اند که می‌توان بدون نیاز به داده‌های گسترده اولیه، دقت مدل در پیش‌بینی عملکرد گندم زمستانه را تا حدود ۱۰ تا ۱۵ درصد ارتقاء داد (Shen et al., 2024). در پژوهشی دیگر، به‌کارگیری پهپاد در برداشت داده‌ها از مزرعه گندم به‌منظور تهیه دیتاست برای پیش‌بینی عملکرد گندم نشان داد که مدل ماشین بهینه‌سازی‌شده در مقایسه با مدل‌های خطی سنتی، دقت پیش‌بینی را به‌میزان ۷ تا ۱۲ درصد افزایش داد (Khodjaev et al., 2025). همچنین پژوهش‌های اخیر بر اهمیت تنظیم آبرپارامترها با استفاده از تکنیک‌هایی مانند بهینه‌سازی شبکه‌ای و آزمون‌های پسا آزمون مانند کراس ولیدیشن تأکید دارند که منجر به بهبود ۷ تا ۱۲ درصدی دقت مدل‌های جنگل تصادفی می‌شود.

جدول ۱- ویژگی‌های دو الگوریتم یادگیری ماشین رندوم فارست و ایکس‌جی‌بوست
Table 1- Comparison of the characteristics of random forest and XGBoost algorithms

ویژگی/الگوریتم Feature / algorithm	رندوم فارست Random Forest	ایکس‌جی‌بوست XGBoost
نحوه یادگیری Learning approach	مدل‌ها به صورت موازی Bagging ساخته می‌شوند (نوع و ترکیب مدل‌ها از یکدیگر مستقل هستند) Models are built in parallel using Bagging (each model is independent)	مدل‌ها به صورت زنجیره‌ای Boosting ساخته می‌شوند (هر مدل، خطای مدل قبلی را اصلاح می‌کند) Models are built sequentially using Boosting (each model corrects the errors of the previous one)
سرعت یادگیری Training speed	روی CPU سریع Fast on CPU	روی GPU خیلی سریع Very fast on GPU
دقت نهایی Final accuracy	خوب Good	معمولاً بهتر، مخصوصاً روی داده‌های پیچیده Usually better, especially on complex data
کنترل روی بیش‌برازش Overfitting control	خوب Good	عالی (به دلیل داشتن پارامترهایی مثل regularization) Excellent (thanks to parameters like regularization)
قابلیت کار با داده‌های ناقص Handling missing data	ندارد Not supported	دارد Supported

(۱۴۰۱). این شرایط، از یک سو، نشان از آسیب‌پذیری بالای این زیست‌بوم کشاورزی در برابر تغییرات اقلیمی دارد و از سوی دیگر، تنوع خرداقلیم و تفاوت مدیریتی گسترده در سطح شهرستان‌های این استان، ضرورت بررسی عمیق عوامل مؤثر بر عملکرد گندم آبی و مطالعه عمیق چگونگی تأثیر متغیرهای اقلیمی بر آن‌ها را دو چندان می‌کند.

با توجه به اینکه نوسانات چشمگیر بارش، تلفات گسترده در سال‌های پرتش و کاهش شدید عملکرد طی دوره زمانی ۱۴۰۲-۱۳۸۳ در برخی شهرستان‌ها نیازمند تحلیل عمیق‌تری است، این پژوهش با هدف اصلی شناسایی و تحلیل روند بلندمدت عملکرد گندم آبی در ۲۰ شهرستان خراسان رضوی و توسعه مدل‌های پیش‌بینی دقیق مبتنی بر «یادگیری ماشین» طراحی و اجرا شد. برای این منظور، مجموعه‌ای از شاخص‌های اقلیمی و زراعی مؤثر بر عملکرد گندم از پایگاه‌های داده و منابع علمی منتشرشده استخراج و محاسبه گردید. سپس مدل‌های پیش‌بینی عملکرد گندم مبتنی بر دو الگوریتم رندوم‌فارست و ایکس‌جی‌بوست، برای هر شهرستان آموزش و توسعه داده شد و سپس دقت و اعتبار آن‌ها مورد ارزیابی قرار گرفت. در آخر، اهمیت نسبی هر متغیر تعیین و راهکارهای مدیریتی سازگار با تغییرات آب‌وهوایی ارائه گردید. نتایج این مطالعه می‌تواند مبنایی برای طراحی سیاست‌های مدیریتی سازگار با تغییرات اقلیمی و بهینه‌سازی تولید گندم در خراسان رضوی فراهم آورد.

جنگل تصادفی به دلیل توانایی بالای خود در مدل‌سازی روابط غیرخطی، مقاومت در برابر بیش‌برازش و ارائه شاخص‌های اهمیت متغیرها، به طور گسترده‌ای در مطالعات کشاورزی مورد استفاده قرار گرفته است (Roell et al., 2020; Pang, Chang, & Chen, 2022).

الگوریتم دیگری که در یادگیری ماشین و مسائل رگرسیون و کلاسیک‌بندی، توانایی بالایی دارد و به عنوان یک مدل قدرتمند در پیش‌بینی‌هایی همچون موضوع مقاله حاضر به شمار می‌رود، الگوریتم ایکس‌جی‌بوست^۱ است. ایده اصلی در این الگوریتم این است که به جای ساخت یک مدل بزرگ، چند مدل ساده (درخت تصمیم) را پشت‌سرهم می‌گذارد، به طوری که هر مدل بعدی، خطای مدل قبلی رو جبران می‌کند. هر دو الگوریتم از خانواده درخت تصمیم^۲ بوده و در عمل برای اهداف مشابه به کار گرفته می‌شوند. خلاصه‌ای از توانایی‌ها و مقایسه این دو الگوریتم در **جدول ۱** ارائه شده است.

استان خراسان رضوی، با اختصاص بیش از ۴۸ درصد از اراضی کشاورزی به گندم آبی، یکی از مهم‌ترین قطب‌های تولید این محصول در ایران است. داده‌های آماری مرکز جهاد کشاورزی نشان می‌دهد که در سال زراعی ۱۴۰۰-۱۳۹۹، کاهش بارندگی تا ۳۸ درصد نسبت به متوسط بلندمدت، به همراه سرمازدگی منجر به خسارت‌های گسترده در بیش از ۴۷ هزار هکتار از اراضی آبی و بیش از ۹۶ هزار هکتار از اراضی دیم شد (سازمان تحقیقات کشاورزی،

1- XGBoost (Extreme Gradient Boosting)
2- Decision Tree

مواد و روش‌ها

منطقه مورد مطالعه

منطقه مورد مطالعه در این تحقیق، استان خراسان رضوی با مساحتی حدود ۱۱۸۸۵۴ کیلومترمربع است که در مختصات جغرافیایی ۳۳ درجه و ۵۲ دقیقه تا ۳۷ درجه و ۴۲ دقیقه عرض شمالی و ۵۶ درجه و ۱۹ دقیقه تا ۶۱ درجه و ۱۶ دقیقه طول شرقی قرار دارد. این استان دارای اقلیم متنوعی از نیمه‌خشک تا خشک کوهستانی بوده و در ارتفاع متوسطی حدود ۹۰۰ تا ۱۲۰۰ متر از سطح دریا واقع شده است. این پژوهش در سطح ۲۰ شهرستان استان خراسان رضوی شامل مشهد، نیشابور، سبزوار، تربت حیدریه، کاشمر، قوچان، تربت جام، تایباد، گناباد، درگز، سرخس، فریمان، خواف، رشتخوار، بردسکن، چناران، بجستان، خلیل آباد، کلات نادر و مه‌ولات صورت گرفته است.

داده‌ها و اطلاعات مورد نیاز

اطلاعات اقلیمی: داده‌های روزانه دما (حداکثر، حداقل و میانگین)، بارندگی، تبخیر-تعرق، شاخص فصلی بودن درجه حرارت، درجه روز رشد و شاخص خشکی برای هر شهرستان در بازه زمانی ۱۳۸۳ تا ۱۴۰۲ از ایستگاه‌های هواشناسی هر شهرستان تهیه شد و در موارد اندکی که خلأ داده وجود داشت و یا داده‌ها ناقص بودند، به‌منظور تکمیل و بازسازی داده‌ها، از پایگاه‌های معتبر و اعتبارسنجی شده جهت شبیه‌سازی داده‌های هواشناسی در سطح استان نظیر APHRODITE, ERA5, IMERG, AgMERRA استفاده و در قالب فایل‌های سالانه پردازش شدند (Yaghoubi, Bannayan, & Asadi, 2020; Farhadi et al., 2024; Soltani, Jahan, Yaghoubi, 2025). هر شهرستان براساس داده‌های آب‌وهوایی خود و موقعیت جغرافیایی در یک فایل جداگانه ذخیره شده و سپس با داده‌های عملکرد گندم ادغام شد.

اطلاعات عملکرد: داده‌های عملکرد گندم در واحد کیلوگرم بر هکتار برای هر شهرستان طی بازه زمانی ذکرشده از پایگاه آمار و اطلاعات سازمان جهاد کشاورزی استان خراسان رضوی، مراجعه حضوری به سازمان جهاد کشاورزی و سایر منابع معتبر گردآوری شد. براساس این داده‌ها و نیز برخی منابع علمی منتشرشده، مجموعه‌ای از متغیرهای مهم اقلیمی و زراعی استخراج و محاسبه شدند که به‌عنوان متغیرهای مستقل در مدل مورد استفاده قرار گرفتند (Farhadi et al., 2024). این متغیرها در جدول ۲ آورده شده‌اند.

این متغیرها در قالب یک دیتافریم منظم ترکیب شده و مراحل آماده‌سازی داده‌ها شامل نرمال‌سازی (براساس امتیاز زد)، کنترل کیفیت و ادغام با داده‌های عملکرد روی آن‌ها انجام گرفت.

تجزیه و تحلیل داده‌ها

۱- پیش‌بینی عملکرد گندم به کمک الگوریتم‌های یادگیری ماشین

آموزش و تست مدل: در این مطالعه برای تحلیل اثر متغیرهای اقلیمی بر عملکرد گندم، از دو الگوریتم مستقل شامل رندوم‌فارست و ایکس‌جی‌بوست که هر دو از روش‌های یادگیری ماشین قدرتمند برای مدل‌سازی روابط پیچیده میان متغیرهای مستقل (ورودی‌ها) و متغیر وابسته (خروجی) محسوب می‌شوند در قالب سه رهیافت شامل پیش‌بینی کمی عملکرد گندم، پیش‌بینی طبقه‌ای عملکرد گندم، و پیش‌بینی عملکرد گندم در قالب ناحیه‌بندی اقلیمی استفاده شد. قابل یادآوری است که از آنجایی که این دو الگوریتم نسبت به داده‌های استاندارد نشده، مقاوم هستند (برخلاف شبکه‌های عصبی مصنوعی که هنگام کار با آن‌ها باید داده‌ها استاندارد شوند)، بنابراین از داده‌های دارای مقیاس واقعی استفاده شد (به‌جز در مورد نشان دادن اهمیت نسبی ویژگی‌ها). قبل از اجرای فرآیند یادگیری، داده‌ها به دو قسمت تقسیم شدند، ۷۰ درصد داده‌ها جهت آموزش مدل و ۳۰ درصد برای تست مدل اختصاص یافتند.

تحلیل با رندوم‌فارست: این الگوریتم با ایجاد مجموعه‌ای از درختان تصمیم‌گیری و تلفیق نتایج آن‌ها از طریق رأی‌گیری یا میانگین‌گیری، پیش‌بینی نهایی را انجام می‌دهد. رگرسور اختصاصی این الگوریتم، رندوم‌فارست رگرسور نام دارد. ویژگی‌های برجسته رندوم‌فارست شامل موارد زیر است: توانایی کشف روابط پیچیده و غیرخطی میان متغیرهای اقلیمی و عملکرد محصول؛ مقاومت در برابر بیش‌برازش^۱ به‌ویژه در حضور متغیرهای متعدد؛ ارائه رتبه‌بندی اهمیت متغیرها^۲ که در تحلیل نقش عوامل اقلیمی بسیار ارزشمند است؛ دقت بالای پیش‌بینی حتی در صورت وجود داده‌های ناقص یا پرنویز؛ کارایی در ساختارهای داده‌ای پنلی و سری زمانی بدون نیاز به فرض‌های محدودکننده.

1- Overfitting

2- Feature Importance

جدول ۲- متغیرهای اقلیمی و زراعی مورد استفاده در الگوریتم‌های پیش‌بینی عملکرد گندم در استان خراسان رضوی
Table 2- Climatic and agronomic variables used in wheat yield prediction algorithms in Khorasan Razavi province

شرح Description	متغیر اقلیمی / زراعی AgroClimatic variable
میانگین دمای روزانه در طول فصل رشد (درجه سانتی‌گراد) Average daily temperature during the growing season (°C)	Tmean (°C)
دمای بیشینه روزانه (درجه سانتی‌گراد) Daily maximum temperature (°C)	Tmax (°C)
دمای کمینه روزانه (درجه سانتی‌گراد) Daily minimum temperature (°C)	Tmin (°C)
بارندگی طی فصل رشد (میلی‌متر) Precipitation during the growing season (mm)	Prec (mm)
تبخیر و تعرق طی فصل رشد (میلی‌متر) Evapotranspiration during the growing season (mm)	ETGS (mm)
شاخص خشکی Aridity index	AI
تبخیر و تعرق (میلی‌متر) Evapotranspiration (mm)	ET (mm)
مجموع بارندگی در کل فصل رشد (میلی‌متر) Total precipitation during the growing season (mm)	PGS (mm)
تعداد روزهای بارندگی در فصل رشد Number of rainy days during the growing season	NPGS (روز)
دمای بیشینه در فصل رشد (درجه سانتی‌گراد) Maximum temperature during the growing season (°C)	TmaxGS (°C)
دمای کمینه در فصل رشد (درجه سانتی‌گراد) Minimum temperature during the growing season (°C)	TminGS (°C)
دمای میانگین در فصل رشد (درجه سانتی‌گراد) Average temperature during the growing season (°C)	TmeanGS (°C)
تعداد روزهای با دمای بیش از ۳۰ درجه سانتی‌گراد Number of days with temperature above 30°C	NDO30
طول فصل رشد (روز) Length of the growing season (days)	GSL
شدت فصلی بودن درجه حرارت Temperature seasonality	TS
درجه روز رشد Growing degree days (GDD)	GDD
درجه روز رشد طی فصل رشد (درجه سانتی‌گراد روز) Growing degree days during the growing season (°C·days)	GDDGS (°C day)

طول فصل رشد، از ۱۵ آبان تا ۱۵ خرداد سال بعد در نظر گرفته شد.

The growing season was considered to extend from November 6 to June 5 of the following year.

فراهم می‌سازد. تنها عاملی که به کارگیری رویکرد اول را می‌تواند محدود کند، این است که زمانی که تعداد پارامترها زیاد باشد، سرعت اجرای آن روی رایانه‌های ضعیف، کند خواهد بود. در این پژوهش، مدل‌های رندوم فارست و ایکس‌جی‌بوست با استفاده از این دو روش، بهینه‌سازی و با اعتبارسنجی متقابل پنج‌تایی^۷ تعیین اعتبار شدند. در **جدول ۳**، این دو رویکرد با یکدیگر مقایسه شده‌اند.

تعیین اعتبار^۸ و ارزیابی دقت مدل^۹: برای ارزیابی عملکرد مدل‌های یادگیری ماشین، فرآیند تعیین اعتبار یا اعتبارسنجی نقش کلیدی ایفا می‌کند. در این فرآیند، مدل‌ها روی داده‌هایی که در آموزش شرکت نداشته‌اند مورد آزمایش قرار می‌گیرند تا میزان تعمیم‌پذیری آن‌ها بررسی شود. در فرآیند مدل‌سازی عملکرد گندم، تعیین اعتبار نقشی حیاتی در اطمینان از قابلیت تعمیم مدل به داده‌های جدید ایفا می‌کند. یکی از مهم‌ترین جنبه‌های اعتبارسنجی، ارزیابی دقت است که با استفاده از معیارهایی شامل ضریب تبیین، جذر میانگین مربعات خطا، میانگین مطلق خطا و شاخص توافق ویلموت (دی)، کیفیت پیش‌بینی‌های مدل را کمی می‌سازد. بنابراین، ارزیابی دقت را می‌توان بخشی از فرآیند جامع‌تر تعیین اعتبار دانست که امکان مقایسه دقیق بین مدل‌ها و انتخاب گزینه بهینه را فراهم می‌سازد. بدین ترتیب، دقت مدل‌ها برای هر ناحیه اقلیمی به‌صورت مجزا سنجیده شد تا میزان اطمینان‌پذیری آن‌ها در مناطق با شرایط متفاوت اقلیمی بررسی گردد. این تحلیل، مبنایی برای انتخاب مدل نهایی و تفسیر نتایج نهایی پیش‌بینی فراهم ساخت.

برای ارزیابی دقت مدل‌های طبقه‌بندی، از ماتریس درهم‌ریختگی استفاده شد که امکان تحلیل دقیق عملکرد مدل را فراهم می‌سازد. این ماتریس، تعداد پیش‌بینی‌های درست و نادرست را برای هر کلاس به‌صورت مجزا نشان می‌دهد و شامل چهار مقدار کلیدی است: مثبت واقعی، منفی واقعی، مثبت نادرست، و منفی نادرست. با استفاده از این مقادیر، معیارهای مهمی نظیر دقت کلی یا صحت، بازیابی یا حساسیت، امتیاز اف-وان و دقت مثبت پیش‌بینی محاسبه می‌شوند (Ting, 2011). این تحلیل کمک می‌کند تا نقاط ضعف مدل در شناسایی کلاس‌ها به‌ویژه در داده‌های نامتوازن مشخص شود. در این مطالعه، از ماتریس درهم‌ریختگی برای مقایسه عملکرد الگوریتم‌های رندوم فارست و ایکس‌جی‌بوست در دسته‌بندی شهرستان‌های با سه کلاس عملکرد بالا، متوسط و پایین استفاده شد. برای این منظور، مراحل زیر انجام گرفت.

به‌منظور ارزیابی عملکرد مدل رندوم فارست، از شاخص‌های آماری نظیر ضریب تبیین^۱ (R^2)، جذر میانگین مربعات خطا^۲، میانگین مطلق خطا^۳ و شاخص توافق ویلموت^۴ استفاده گردید. جهت تحلیل عمیق‌تر نتایج، مجموعه‌ای از نمودارهای گرافیکی شامل تحلیل اهمیت نسبی متغیرها، درخت‌های تصمیم‌گیری همراه با شاخص‌های ارزیابی مدل برای هر شهرستان ترسیم شد. این نمودارها درک دقیق‌تری از ترتیب و نقش متغیرهای اقلیمی و زراعی در تبیین عملکرد گندم فراهم کرده و به تفسیر روابط حاکم در هر شهرستان و به‌دنبال آن هر ناحیه اقلیمی کمک می‌کنند.

شایان ذکر است که برای هر شهرستان، یک مدل اختصاصی آموزش داده شد و پس از ارزیابی دقیق با داده‌های جدید، عملکرد آن با استفاده از شاخص‌های آماری تأیید شد. این مدل‌های آموزش‌دیده در قالب فایل‌هایی با فرمت pk1 ذخیره و مستندسازی گردید تا در کاربردهای آتی (جهت پیش‌بینی) مورد استفاده قرار گیرند.

تحلیل با ایکس‌جی‌بوست: کلیه مراحل انجام‌گرفته در الگوریتم رندوم فارست، با داده‌های ورودی یکسان، با الگوریتم ایکس‌جی‌بوست نیز انجام گرفتند. رگرسور اختصاصی این الگوریتم، ایکس‌جی‌بوست رگرسور نام دارد.

۲- پیش‌بینی طبقه عملکرد گندم در شهرستان‌های استان خراسان رضوی

به‌منظور طبقه‌بندی (کلاس‌بندی) عملکرد گندم در هر شهرستان، از دو الگوریتم رندوم فارست و ایکس‌جی‌بوست استفاده شد، با این تفاوت که در این رهیافت، مدل‌ها بهینه‌سازی نشدند (به‌دلیل ماهیت طبقه‌ای داده‌ها در این رویکرد).

بهینه‌سازی مدل‌ها: جهت بهینه‌سازی ابرپارامترهای مدل‌های یادگیری ماشین و یافتن بهترین ترکیب ممکن برای هر الگوریتم، از دو رویکرد جستجوی شبکه‌ای با اعتبارسنجی متقابل^۵ و جستجوی تصادفی با اعتبارسنجی متقابل^۶ بهره‌گیری شد (تعداد فولد، ۵ و $n_estimator$ برابر با ۳۰۰). رویکرد اول تمام ترکیب‌های ممکن از پارامترهای الگوریتم (مثلاً عمق درخت، تعداد درخت، نرخ یادگیری و غیره) را تولید کرده و آن‌ها را تست می‌کند و در ادامه بهترین آن‌ها را براساس دقت (مثلاً R^2) برای متغیر هدف (مثلاً عملکرد گندم) پیدا می‌کند، درحالی‌که رویکرد دوم با انتخاب تصادفی اما هدفمند، امکان بهینه‌سازی سریع‌تر را در مسائل با فضای پارامتر گسترده

- 1- Coefficient of determination (R^2)
- 2- Root Mean Square Error (RMSE)
- 3- Mean Absolute Error (MAE)
- 4- Willmott's Agreement Index (d)
- 5- GridSearch Cross Validation
- 6- RandomizedSearch Cross Validation

7- 5-Fold Coss-Validation

8- Validation

9- Model Evaluation

جدول ۳- مقایسه دو رویکرد بهینه‌سازی الگوریتم‌های یادگیری ماشین در پیش‌بینی عملکرد گندم

Table 3- Comparison of two optimization approaches for machine learning algorithms in wheat yield prediction

ویژگی/الگوریتم	جستجوی تصادفی	جستجوی شبکه‌ای
Feature / algorithm	Randomized search	Grid search
دقت جستجو	متوسط، (تعدادی ترکیب تصادفی)	بالا، (تمام ترکیب‌های ممکن)
Search accuracy	Moderate (a number of random combinations)	High (all possible combinations)
سرعت	سریع‌تر	کند، به‌خصوص با پارامتر زیاد
Speed	Faster	Slower, especially with many parameters
استفاده بهینه از منابع	بله	نه زیاد
Efficient use of resources	Yes	Not much
برای تست پارامترهای زیاد در زمان محدود	خیلی خوب	کند
For testing many parameters in limited time	Very good	Slow
برای تست پارامترهای کم در زمان کافی	خیلی خوب	عالی
For testing few parameters with enough time	Very good	Excellent

۱- جمع زدن پیش‌بینی‌ها درون هر خوشه. ۲- ایجاد بردار برای نمونه واقعی و پیش‌بینی شده ($y_true_clusterN$) شامل برچسب‌های واقعی عملکرد مثلاً high/low و $y_pred_clusterN$ برای پیش‌بینی‌های مدل. ۳- ایجاد کانفیوژن ماتریکس برای هر خوشه.

تحلیل و ارزیابی نقش متغیرها براساس چارچوب SHAP

به‌منظور تبیین نقش نسبی هر یک از ویژگی‌های ورودی در پیش‌بینی عملکرد گندم، از چارچوب تفسیری SHapley Additive (SHAP) exPlanations استفاده شد. روش SHAP یکی از قدرتمندترین ابزارهای تفسیر مدل‌های یادگیری ماشین است که با بهره‌گیری از مفاهیم نظریه بازی شاپلی، سهم مشارکتی هر ویژگی را در خروجی مدل به‌صورت منصفانه و دقیق برآورد می‌کند. این رویکرد، برخلاف برخی روش‌های سنتی که صرفاً به ارزیابی حضور یا عدم حضور یک متغیر بسنده می‌کنند، قادر است که تأثیر هر متغیر را در بسترهای مختلف داده و در سطح محلی (برای یک نمونه خاص) و سراسری (میانگین مقادیر شیب برای همه نمونه‌ها به دست می‌آید و اهمیت کلی ویژگی‌ها مشخص می‌شود) مورد تحلیل قرار دهد. در نتیجه، تحلیل مبتنی بر شیب، تصویری جامع از نحوه و شدت تأثیرگذاری هر متغیر در سراسر فضای داده‌ها ارائه می‌دهد (Remman & Lekkas, 2021; Kim & Kim, 2022). مهم‌ترین دلایل بهتر بودن شیب سایر روش‌های مشابه به قرار زیر است: سازگار بودن با هر نوع مدل مانند مدل‌های دیپ، درختی، و خطی. نشان دادن سهم منصفانه هر ویژگی (طبق قضایای ریاضی)؛ برخلاف روش‌هایی مثل اهمیت برپایه جابه‌جایی، اثر تعاملی و همبستگی بین ویژگی‌ها را نیز در نظر می‌گیرد. درحالی‌که اهمیت برپایه جابه‌جایی با جابه‌جا کردن مقادیر هر ویژگی در داده‌های آزمون و سنجش تغییر در عملکرد مدل، میزان اهمیت هر ویژگی را برآورد می‌کند، روش شیب برپایه نظریه بازی‌های مشارکتی، سهم منصفانه و دقیق هر ویژگی را

در پیش‌بینی خروجی مدل - هم به‌صورت محلی برای هر نمونه و هم به‌صورت کلی - اندازه‌گیری می‌کند. برخلاف اهمیت برپایه جابه‌جایی که ممکن است در حضور ویژگی‌های همبسته دچار خطا یا تفسیر گمراه‌کننده شود، شیب توانایی جداسازی و تحلیل دقیق اثرات فردی و تعاملی ویژگی‌ها را دارد. به همین دلیل، در تحلیل‌های مبتنی بر مدل‌های پیچیده مانند رندوم‌فارست و ایکس‌جی‌بوست، استفاده از رهیافت شیب توصیه می‌شود. برای نمایش نتایج این تحلیل، از نمودارهای ویولن استفاده شد. این نمودارها توزیع مقادیر شیب مرتبط با هر ویژگی را نشان می‌دهند و از طریق شکل ترکیبی خود (شامل چگالی احتمال و آمار توصیفی)، هم جهت اثرگذاری (مثبت یا منفی بودن شیب)، به این معنی که ویژگی‌های با مقادیر شیب مثبت باعث افزایش مقدار پیش‌بینی شده می‌شوند و ویژگی‌های با مقادیر شیب منفی باعث کاهش مقدار پیش‌بینی شده می‌شوند) و هم شدت نسبی آن را بر خروجی مدل به تصویر می‌کشند.

نمودار دیگری که شیب تولید می‌کند، نمودار تصمیم^۱ نام دارد که نشان می‌دهد چگونه هر ویژگی بر پیش‌بینی مقدار متغیر وابسته اثر می‌گذارد. آخرین نموداری که به کمک شیب برای هر شهرستان ایجاد شد، نمودار وابستگی جزئی^۲ است که رابطه مستقیم بین هر ویژگی و متغیر وابسته (عملکرد گندم در هر شهرستان) را به تصویر می‌کشد. استفاده از این نمودارها، ضمن تقویت شفافیت مدل‌های یادگیری ماشین پیچیده، زمینه را برای تفسیر علمی خروجی مدل و شناسایی عوامل کلیدی مؤثر بر عملکرد گندم فراهم می‌سازد؛ موضوعی که برای تدوین راهبردهای مدیریتی و سیاست‌گذاری کشاورزی از اهمیت بالایی برخوردار است.

۳- پیش‌بینی عملکرد گندم براساس خوشه‌بندی شهرستان‌ها در ناحیه‌های اقلیمی: برای شناسایی نواحی اقلیمی-

1- Decision Plot

2- Partial Dependence Plot

مبنای تحلیل‌های بعدی در مطالعه حاضر قرار گرفته و نواحی اقلیمی-کشاورزی استان را شناسایی کرده‌اند. توصیف هر ناحیه اقلیمی مطابق طبقه‌بندی اقلیمی شهرستان‌ها براساس سیستم کوپن-گایگر و آمار روزشده سایت کلایمیت دیتا صورت گرفت (Peel, Finlayson, & McMahon, 2007 <http://en.climate-data.org/asia/iran/razavi-khorasan-2216>). همچنین به‌منظور یادگیری بهتر مدل‌ها، شماره ناحیه اقلیمی هر شهرستان در یک ستون مستقل قرار گرفت. مزیت افزودن این متغیر در کنار سایر متغیرها این است که مدل، تفاوت اقلیم‌ها را یاد می‌گیرد، ولی همچنان از تمام جزئیات در سطح شهرستان بهره می‌برد، به عبارت دیگر، تفکیک‌پذیری مدل بهبود می‌یابد.

نرم‌افزارهای مورد استفاده

در این پژوهش، الگوریتم‌های یادگیری ماشین، رندوم‌فارست و ایکس‌جی‌بوست، با استفاده از زبان برنامه‌نویسی پایتون 3.3 Python و کتابخانه Scikit-learn پیاده‌سازی شدند. آماده‌سازی دیتافریم‌ها و خوشه‌بندی شهرستان‌های استان در قالب ناحیه‌های اقلیمی، با استفاده از زبان برنامه‌نویسی R 4.3.2 و نرم‌افزار ArcGIS 10.8.2 انجام گرفت. از آنجایی که هر دو الگوریتم مورد استفاده در این پژوهش درخت‌محور هستند، و درخت‌ها به مقیاس داده حساس نیستند، و می‌توانند با داده‌های خام نیز کار کنند، بنابراین نیازی به نرمال‌سازی یا استانداردسازی داده‌ها نیست. برای اجرای رندوم‌فارست در مواردی که مقیاس داده‌ها خیلی متفاوت باشد (مثلاً اگر بازه داده‌های تبخیر و تعرق بین صفر تا پنج و درجه روز رشد بین ۱۰۰۰ تا ۳۰۰۰ باشد)، پیشنهاد شده در صورت امکان داده‌ها استاندارد شوند، اگرچه که اجباری در این کار نیست. در یک مورد خاص (نمودارهای شیب و بولونی)، داده‌ها جهت استفاده از ابزار شیب استاندارد (نرمال) شده و سپس خروجی‌ها بر این مبنا رسم شدند.

نتایج و بحث

تحلیل، مقایسه و ارزیابی مدل‌های پیش‌بینی عملکرد گندم آبی در سطح استان خراسان رضوی طی دو رهیافت مستقل انجام گرفت. در رهیافت اول، دو مدل رندوم‌فارست و ایکس‌جی‌بوست، آموزش داده و بهینه شدند، سپس روی داده‌هایی که در آموزش شرکت نداشتند، تست شدند.

کشاورزی مشابه (همگن) در استان خراسان رضوی، از روش خوشه‌بندی سلسله‌مراتبی^۱ استفاده شد. این روش برای تحلیل سه شاخص اقلیمی-کشاورزی شامل شاخص خشکی، شدت فصلی بودن درجه حرارت و درجه روز رشد به کار گرفته شده است. خوشه‌بندی سلسله‌مراتبی به دلیل توانایی آن در شناسایی الگوهای پیچیده و توزیع نواحی با ویژگی‌های مشابه، انتخابی مناسب برای تحلیل داده‌های توزیع‌شده از چندین شاخص اقلیمی-کشاورزی است. این روش به‌ویژه در این حالت که داده‌ها از چندین شاخص مختلف استخراج شده‌اند، دقت بالاتری در شناسایی نواحی با ویژگی‌های مشابه ارائه می‌دهد. در پژوهش حاضر، ابتدا شاخص‌های اقلیمی-کشاورزی در محیط GIS درون‌یابی شدند و سپس مجموعه‌ای از نقاط نمونه‌ای (تصادفی) در سراسر منطقه استخراج شدند (۳۵۰ نقطه). این نقاط به‌عنوان ورودی تحلیل خوشه‌بندی در نرم‌افزار R و با استفاده از بسته‌های factoextra و cluster پردازش شدند. مراحل تحلیل به‌ترتیب زیر انجام شد.

- پیش‌پردازش داده‌ها با حذف مقادیر نامعتبر و استانداردسازی (scale) شاخص‌های ورودی.
- محاسبه ماتریس فاصله با استفاده از روش فاصله اقلیدسی^۲ بین نقاط نمونه‌ای.
- اجرای خوشه‌بندی سلسله‌مراتبی و گروه‌بندی نقاط مشابه در خوشه‌های مجزا با استفاده از روش نسخه بهبودیافته از روش Ward (Ward.D2) در بسته cluster.
- انتخاب تعداد بهینه خوشه‌ها (k) با استفاده از شاخص Silhouette. با افزایش آستانه شباهت (که باعث ایجاد خوشه‌های بزرگ‌تر و عمومی‌تر می‌شود)، شاخص سیلوئت کاهش می‌یابد. با گرفتن خروجی‌های متعدد، در نهایت براساس شاخص سیلوئت ۰/۴۱ (برابر با نزدیکی^۳ یا شباهت^۴ ۶۰ درصد) خوشه‌بندی انجام گرفت. مقدار ۰/۴۱ بهترین توازن بین تفکیک و همگنی خوشه‌ها را ایجاد کرد.
- انتساب خوشه‌ها به داده‌ها (برچسب‌گذاری) براساس نتایج خوشه‌بندی، به هر نقطه نمونه‌ای.

در نهایت، نتایج حاصل از خوشه‌بندی به‌فرمت CSV ذخیره شده و برای تحلیل فضایی به محیط ArcGIS منتقل شدند. در این مرحله، داده‌های نقطه‌ای خوشه‌بندی‌شده با استفاده از روش کریجینگ ساده درون‌یابی شده و به یک سطح پیوسته تبدیل شدند. سپس، با طبقه‌بندی نقشه درون‌یابی‌شده، محدوده‌های اقلیمی متناظر با هر خوشه به‌صورت پلی‌گون استخراج شدند. این خروجی‌های پلی‌گونی

- 1- Hierarchical Clustering
- 2- Euclidean Distance
- 3- Distance
- 4- Similarity

جدول ۴- معیارهای ارزیابی و تعیین اعتبار دو الگوریتم رندوم فارست و ایکس‌جی‌بوست در پیش‌بین عملکرد گندم آبی در استان خراسان رضوی، قبل و بعد از بهینه‌سازی

Table 4- Evaluation and validation metrics for random forest and XGBoost algorithms in predicting irrigated wheat yield in Khorasan Razavi province, before and after optimization

معیار ارزیابی / وضعیت مدل ModelStatus/evaluation metric	RMSE	R ²	MAE	d
رندوم فارست (پیش از بهینه‌سازی) RF (befor optimization)	399.53	0.134	326.24	0.46
رندوم فارست (پس از بهینه‌سازی) RF (after optimization)	395.13	0.136	331.50	0.47
ایکس‌جی‌بوست (پیش از بهینه‌سازی) XGB (befor optimization)	499.93	0.063	413.02	0.46
ایکس‌جی‌بوست (پس از بهینه‌سازی) XGB (after optimization)	492.21	0.074	408.09	0.42

سیاه رنگ داخل ویولون‌ها، نشان‌دهنده مقدار اهمیت هر ویژگی در نمونه‌های مختلف (مثلاً در تکرارهای cross-validation یا در بین زیربخش‌های داده) هستند. وقتی این نقاط در یک متغیر خاص، به‌جای تمرکز در یک محدوده باریک، در کل محور y پخش شده‌اند، نشان می‌دهد که اهمیت آن متغیر در شرایط مختلف (یا نمونه‌های مختلف) مدل، نوسان زیادی دارد، به عبارت دیگر، مدل نسبت به این ویژگی در برخی موارد آن را خیلی مهم و در موارد دیگر کم‌اهمیت ارزیابی کرده، که می‌تواند ناشی از شرایط اقلیمی متغیر، اثرات فصلی، یا تعاملات پیچیده با سایر متغیرها باشد.

متغیرهایی مانند طول فصل رشد، شدت فصلی بودن درجه حرارت و دمای بیشینه روزانه در سمت چپ نمودار قرار دارند و دارای بیشترین میزان اهمیت میانگین هستند. پهنای زیاد در ناحیه بالایی این ویولون‌ها نشان می‌دهد که مقدار اهمیت این ویژگی‌ها در بین نمونه‌های مختلف به‌طور گسترده در سطح بالایی متمرکز است. در مقابل، ویژگی‌هایی مانند شاخص خشکی، بارندگی، و بارندگی طی فصل رشد با تراکم در سطوح پایین‌تر و میانه پایین‌تر، نقش کم‌تری در پیش‌بینی عملکرد مدل داشته‌اند. این نشان می‌دهد که در اغلب تکرارهای مدل، این ویژگی‌ها تأثیر نسبتاً کم‌تری در خروجی نهایی داشته‌اند. پراکندگی نقاط سیاه نشان‌دهنده نوسان اهمیت در بین نمونه‌ها است. وجود نقاط با پراکندگی عمودی زیاد مثل تعداد دفعات وقوع بارندگی طی فصل رشد و دمای کمینه طی فصل رشد، نشان‌دهنده وابستگی متغیر به موقعیت مکانی یا شرایط خاص داده‌ها است، درحالی‌که متغیرهایی با توزیع فشرده‌تر مثل طول فصل رشد و دمای بیشینه طی فصل رشد نشان‌دهنده ثبات بالاتر در اهمیت هستند.

در رهیافت دوم، همین دو مدل، داده‌های عملکرد گندم را کلاس‌بندی کرده و سپس عملکرد دو مدل بر روی نتایج تحلیل شد. در جدول ۴، معیارهای ارزیابی دو مدل در رهیافت اول، قبل و بعد از بهینه‌سازی نشان داده شده‌اند.

بهینه‌سازی مدل رندوم فارست، جذر میانگین مربعات خطا را به مقدار چهار واحد کاهش داد و ضریب همبستگی را به مقدار دو هزارم افزایش داد. بهینه‌سازی مدل ایکس‌جی‌بوست، تمام معیارهای ارزیابی (به جز شاخص توافق ویلموت) را نسبت به مدل بهینه‌نشده، بهبود داد. بنابراین، در ادامه نتایج مربوط به الگوریتم‌های بهینه‌سازی شده (به غیر از رویکرد طبقه‌بندی) آورده می‌شوند.

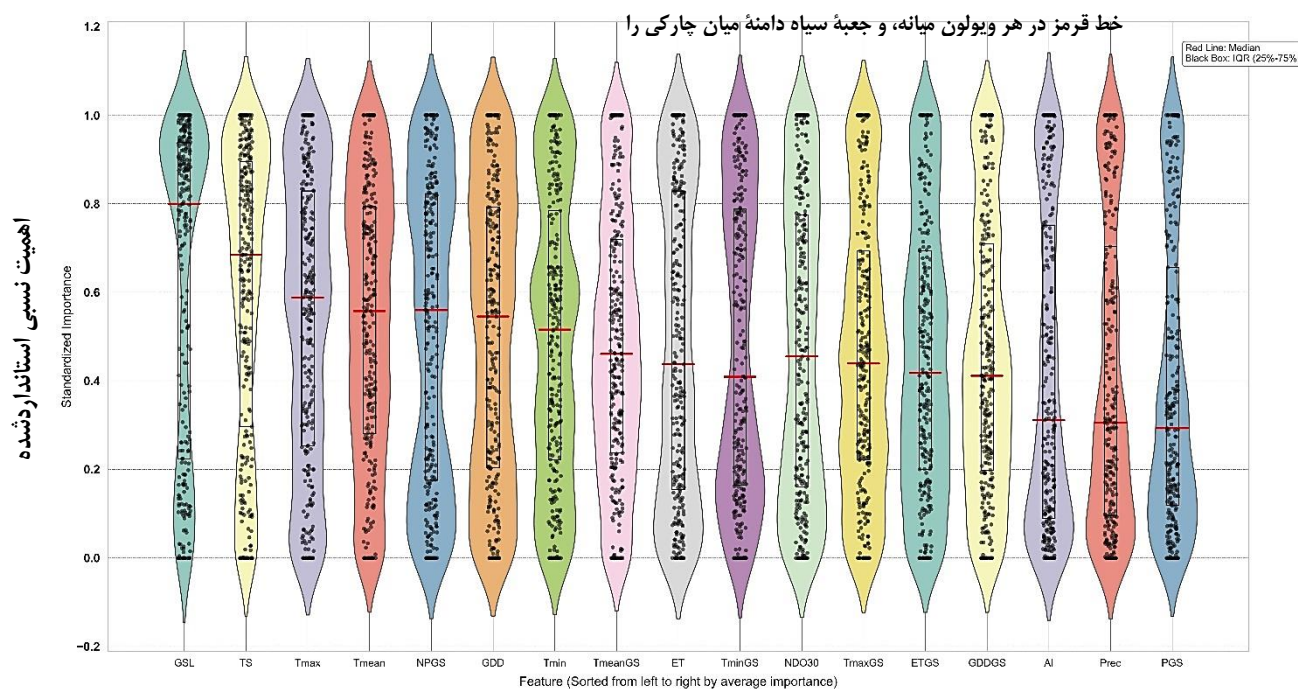
تحلیل ویژگی‌های مؤثر در دو الگوریتم برای شهرستان‌ها

برای درک بهتر از رفتار مدل‌های یادگیری ماشین و بررسی دلایل پیش‌بینی‌های انجام‌شده، نمودارهای شیب یا ویولونی اهمیت ویژگی‌ها (فیچر ایمپورتنس) (شکل ۱ آ و ب)، نمودارهای ضرایب همبستگی بین ۱۰ ویژگی برتر (شکل ۳ آ و ب) برای هر شهرستان به تفکیک و برای هر دو مدل رندوم فارست و ایکس‌جی‌بوست تولید شدند.

توزیع مقادیر شیب (اهمیت نسبی ویژگی‌ها) در مدل رندوم فارست

شکل ۱، توزیع اهمیت استاندارد شده ویژگی‌های اقلیمی و فیزیولوژیکی مؤثر بر عملکرد گندم را براساس مدل رندوم فارست نشان می‌دهد. ویژگی‌ها براساس میانگین اهمیت از چپ به راست مرتب شده‌اند و هر نمودار ویولون، تراکم و پراکندگی مقادیر اهمیت را برای یک ویژگی خاص به تصویر می‌کشد. خط قرمز نشان‌دهنده میانه^۱ و جعبه سیاه بیانگر فاصله بین چارک اول و سوم^۲ است. نقاط

1- Median
2- IQR



ویژگی‌ها (از چپ به راست براساس اهمیت مرتب شده‌اند)

Distribution, Median (Red Line), and IQR (Black Box) of Standardized Feature Importances in Random Forest Model

شکل ۱- نمودارهای ویولونی توزیع مقادیر شیپ (اهمیت ویژگی‌ها) در الگوریتم رندوم فارست جهت پیش‌بینی عملکرد گندم در استان خراسان رضوی. (شرح کامل ویژگی‌ها در جدول ۲ بخش مواد و روش‌ها آورده شده است)

Figure 1- Violin plots of SHAP value distributions (feature importance) in the Random Forest algorithm for predicting wheat yield in Khorasan Razavi province. (A complete description of the features is provided in Table 2 in the materials and methods section)

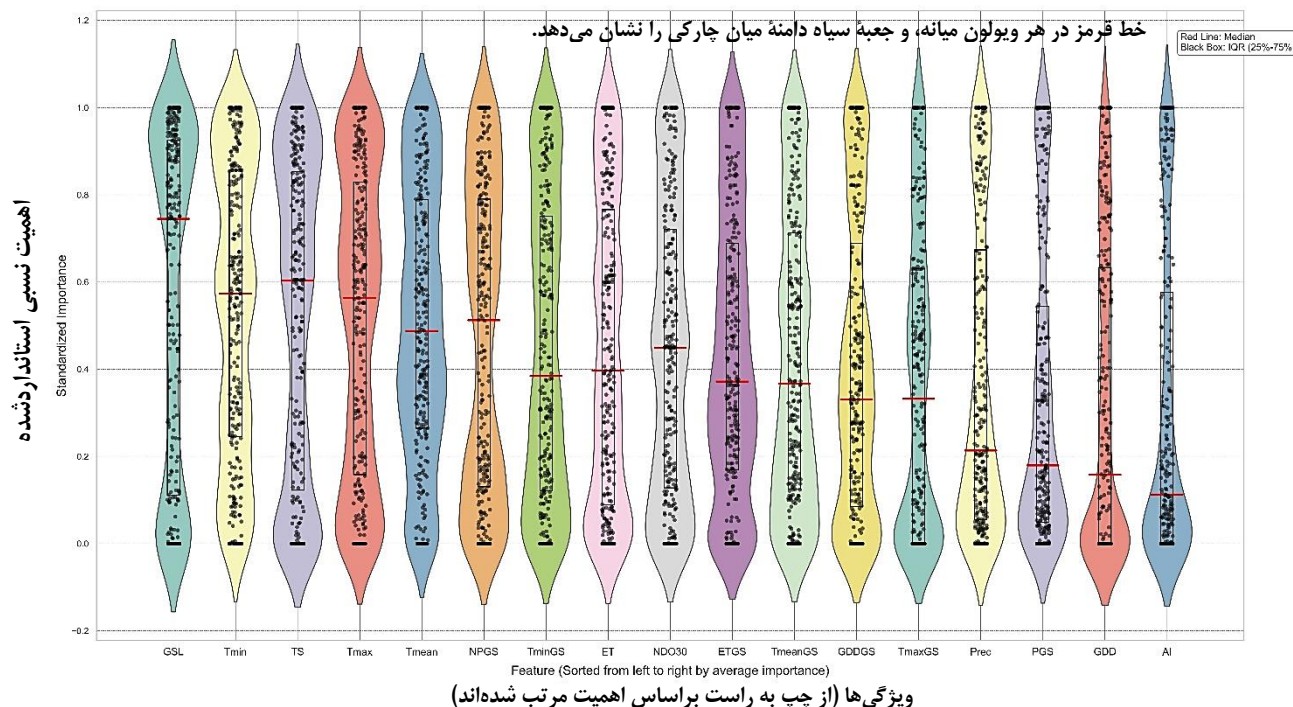
این ویژگی‌ها در بین نمونه‌ها است. این امر می‌تواند حاکی از حساسیت مدل نسبت به شرایط متغیر محیطی یا تعاملات زمانی آن‌ها باشد. مقادیر میانه برای بسیاری از متغیرهایی که در ناحیه وسط نمودار واقع شده‌اند، مثل تبخیر و تعرق طی فصل رشد، دمای میانگین طی فصل رشد یا درجه روز رشد طی فصل رشد نزدیک به ۰/۵ تا ۰/۶ قرار دارد که نشان از نقش میان‌رده آن‌ها در تصمیمات مدل دارد. تحلیل توزیع مقادیر شیپ، می‌تواند به شناسایی عوامل کلیدی تأثیرگذار بر عملکرد گندم در مدل‌های یادگیری ماشین کمک کرده و مبنایی برای تصمیم‌گیری در برنامه‌ریزی‌های زراعی فراهم آورد.

مقایسه مدل رندوم فارست و ایکس‌جی‌بوست در ارزیابی اهمیت ویژگی‌ها

در هر دو مدل، متغیر "طول فصل رشد" بیشترین اهمیت را دارد، که نشان‌دهنده تأثیر کلیدی این متغیر در پیش‌بینی عملکرد گندم است.

توزیع مقادیر شیپ (اهمیت نسبی ویژگی‌ها) در مدل ایکس‌جی‌بوست

شکل ۲، توزیع اهمیت استاندارد شده متغیرها را در مدل ایکس‌جی‌بوست نمایش می‌دهد. در این نمودار نیز مانند شکل قبلی، خط قرمز نشان‌دهنده میانه و جعبه سیاه نشان‌دهنده فاصله بین چارک اول و سوم است. ویژگی‌ها براساس میانگین اهمیت مرتب شده‌اند. متغیرهایی مانند طول فصل رشد، دمای کمینه روزانه، شدت فصلی بودن درجه حرارت و دمای بیشینه روزانه در سمت چپ نمودار و دارای بیشترین اهمیت میانگین هستند، به‌ویژه متغیر طول فصل رشد مجدداً بالاترین سطح اهمیت را نشان می‌دهد، با توزیع متراکم در نواحی بالای محور، که نشان از تأثیر پایدار آن در بین نمونه‌های مختلف دارد. متغیرهایی نظیر شاخص خشکی، درجه روز رشد و بارندگی طی فصل رشد در سمت راست قرار دارند و با تمرکز داده‌ها در بخش‌های پایین‌تر، از نظر مدل اهمیت نسبتاً کمی دارند. در برخی متغیرها مانند دمای کمینه طی فصل رشد، تبخیر و تعرق یا تعداد روزهای با دمای بالای ۳۰ درجه سانتی‌گراد، پراکنش عمودی قابل توجه نقاط درون نمودار ویولن، بیانگر نوسانات بالای اهمیت



Distribution, Median (Red Line), and IQR (Black Box) of Standardized Feature Importances in XGBoost Model

شکل ۲- نمودارهای شیب یا ویولونی اهمیت ویژگی‌های به کار گرفته شده در مدل ایکس‌جی‌بوست جهت پیش‌بینی عملکرد گندم در استان خراسان رضوی (شرح کامل ویژگی‌ها در جدول ۲ بخش مواد و روش‌ها آورده شده است)

Figure 2- Violin plots of SHAP value distributions (feature importance) in the XGBoost algorithm for predicting wheat yield in Khorasan Razavi province. (A complete description of the features is provided in Table 2 in the Materials and Methods section)

حرارت و دمای کمینه روزانه، دمای میانگین روزانه و طول فصل رشد اهمیت و تأثیر زیادی در پیش‌بینی‌های کلیه شهرستان‌ها دارد، درحالی‌که ویژگی‌ای مانند شاخص خشکسالی اهمیت و تأثیر کمتری دارند (در هر دو مدل). به نظر می‌رسد که ویژگی شدت فصلی بودن درجه حرارت، شاید به دلیل ارتباط قوی‌ای که با شرایط آب‌وهوایی یا ویژگی‌های محیطی دارد، تأثیر بیشتری در پیش‌بینی‌ها داشته باشد. از سوی دیگر، اهمیت کمتر ویژگی شاخص خشکسالی، احتمالاً به علت نزدیک بودن مقدار عددی آن برای تمام شهرستان‌ها به دلیل واقع شدن در گستره‌ای بزرگ‌تر از یک منطقه خشک و نیمه‌خشک است. به کمک نمودارهای شیب (اهمیت ویژگی‌ها) می‌توان نسبت به شناسایی و تعیین اولویت متغیرهای اقلیمی و زراعی و در ادامه برنامه‌ریزی و سیاست‌گذاری اقدام نمود.

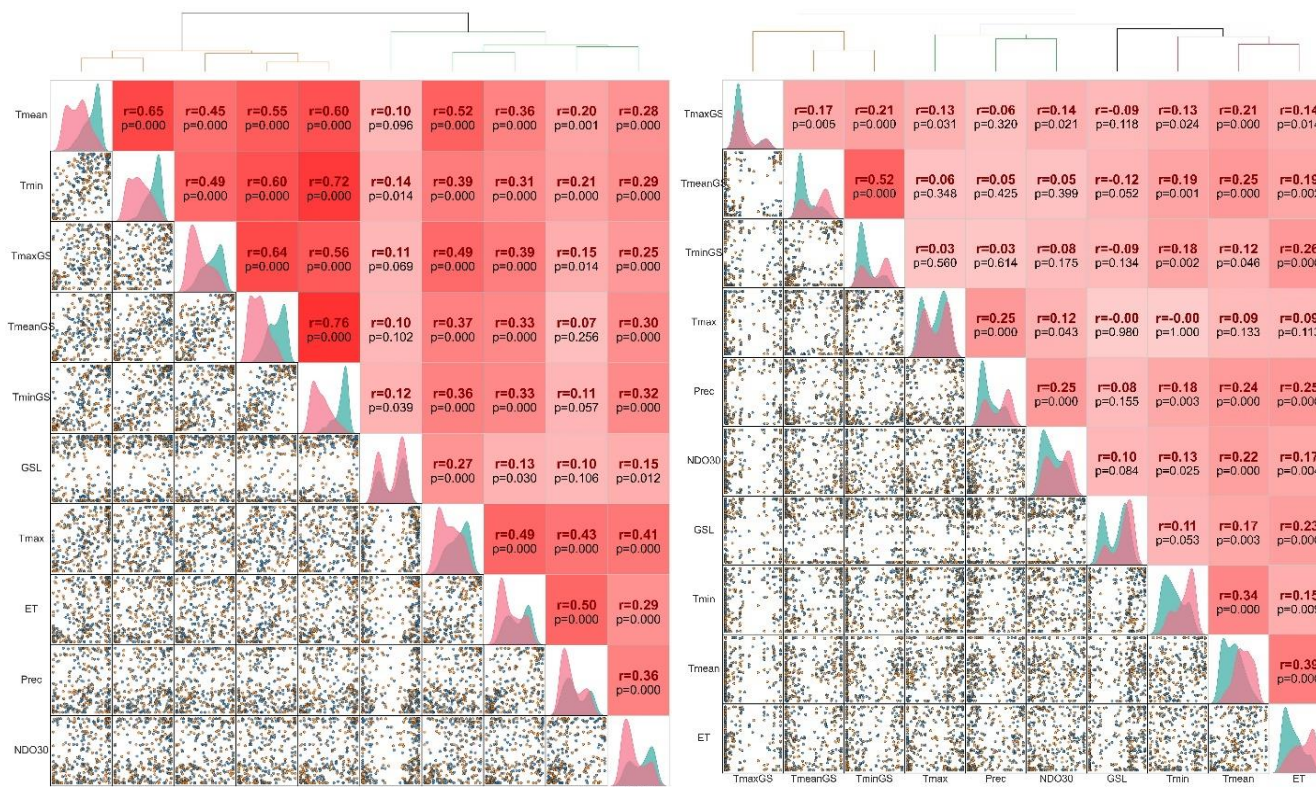
برآورد تبخیر و تعرق مرجع روزانه با استفاده از روش جنگل تصادفی بهینه‌شده با الگوریتم ژنتیک در منطقه آذربایجان شرقی نشان داد که الگوریتم جنگل تصادفی نسبت به روش‌های تجربی در ایستگاه‌های مورد مطالعه، عملکرد بهتری داشت (Monavar Sabegh, Zare Haghi, Samadianfard, Neishabouri, & Mikaeili, 2023).

- در مدل رندوم‌فارست، متغیر دمای بیشینه روزانه نیز در میان متغیرهای بسیار مهم قرار داشت، اما در مدل ایکس‌جی‌بوست به جایگاه پایین‌تری سقوط کرد، درحالی‌که دمای کمینه روزانه نقش مهم‌تری در مدل ایکس‌جی‌بوست ایفا می‌کند.
 - مدل ایکس‌جی‌بوست تنوع بیشتری در پراکندگی اهمیت ویژگی‌ها نشان می‌دهد، به‌ویژه در متغیرهای با اهمیت میانه، که می‌تواند نشانگر قدرت تفکیک دقیق‌تر این مدل در شرایط پیچیده‌تر باشد.
 - مقایسه نمودارها نشان می‌دهد که مدل رندوم‌فارست اهمیت نسبتاً متوازی برای متغیرها در نظر گرفته، درحالی‌که مدل ایکس‌جی‌بوست تفاوت‌های بیشتری در اهمیت ویژگی‌ها قائل شده و نسبت به برخی متغیرهای خاص، حساسیت و تمایز بالاتری نشان داده است، به عبارت دیگر، تفکیک دقیق‌تری بین متغیرهای با اهمیت بالا و پایین ایجاد کرده است.
- اهمیت هر ویژگی در مدل نشان می‌دهد که مدل چقدر به آن ویژگی برای پیش‌بینی‌ها اعتماد کرده است. به عبارت دیگر، این ویژگی‌ها تأثیر زیادی بر پیش‌بینی‌ها دارند و بر توان و دقت مدل‌های مختلف تأثیرگذار هستند. برای مثال، ویژگی شدت فصلی بودن درجه

ضرب همبستگی بین ویژگی‌های با اهمیت نسبی بالا در دو مدل

شکل ۳، ماتریس همبستگی ۱۰ ویژگی برتر انتخاب شده توسط الگوریتم رندوم فارست را برای تمام شهرستان‌ها نشان می‌دهد و با بهره‌گیری از خوشه‌بندی سلسله‌مراتبی، ویژگی‌های مشابه را در کنار هم قرار داده است. این شکل نشان می‌دهد که بین دمای بیشینه طی فصل رشد، دمای میانگین طی فصل رشد، دمای میانگین روزانه و دمای کمینه روزانه همبستگی‌های مثبت و نسبتاً قوی وجود

دارد. به‌ویژه بیشترین ضریب همبستگی بین دمای میانگین طی فصل رشد و دمای کمینه طی فصل رشد مشاهده شد ($r = 0.76, p < 0.001$)، و پس از آن، رابطه بین دمای کمینه روزانه و دمای کمینه فصل رشد با ضریب $r = 0.72$ قرار دارد. خوشه‌بندی ویژگی‌ها نشان می‌دهد که دمای میانگین طی فصل رشد، دمای بیشینه طی فصل رشد و دمای کمینه طی فصل رشد در یک خوشه، و دما کمینه روزانه و دمای میانگین روزانه در خوشه‌ای مجزا قرار گرفته‌اند که بیانگر تفاوت در الگوهای هم‌تغییری آن‌ها است.



RF Model Top 10 Features - Correlation, Distribution and Clustering

XGBoost Model Top 10 Features - Correlation, Distribution and Clustering

شکل ۳- ضرایب همبستگی، پی ویو و توزیع چگالی و خوشه‌بندی ۱۰ ویژگی برتر مؤثر بر عملکرد گندم در استان خراسان رضوی در دو الگوریتم رندوم فارست (سمت چپ) و ایکس جی بوست (سمت راست). (شرح کامل ویژگی‌ها در جدول ۲ بخش مواد و روش‌ها آورده شده است)

Figure 3- Correlation coefficients, p-values, density distributions, and clustering of the top ten most influential features on wheat yield in Khorasan Razavi province using the Random Forest algorithm (left) and XGBoost algorithm (right). (A complete description of the features is provided in table 2 in the materials and methods section)

همبستگی قابل توجه مربوط به رابطه بین تبخیر و تعرق طی فصل رشد و دمای میانگین روزانه بود ($r = 0.39, p < 0.01$) که از نظر فیزیولوژیکی نیز قابل پیش‌بینی است.

همچنین دمای کمینه روزانه در هر دو مدل، همبستگی بالایی با سایر متغیرهای دمایی نشان داد؛ موضوعی که بر نقش کلیدی این متغیر در عملکرد مدل‌ها دلالت دارد. با توجه به این یافته‌ها، توجه به دمای کمینه روزانه در تصمیم‌گیری‌های مرتبط با مدیریت زراعی

در مدل ایکس‌جی‌بوست، بالاترین ضریب همبستگی میان دو ویژگی دمای کمینه طی فصل رشد و دمای میانگین طی فصل رشد مشاهده شد ($r = 0.52, p < 0.01$). این رابطه پیش‌تر در مدل رندوم‌فارست نیز به دست آمده بود، اما با شدت بیشتری ($r = 0.72, p < 0.01$). در میان سایر متغیرهای دمایی در مدل ایکس‌جی‌بوست، همبستگی مثبت و نسبتاً قوی بین دمای میانگین روزانه و دمای کمینه روزانه دیده شد ($r = 0.34, p < 0.01$). دومین ضریب

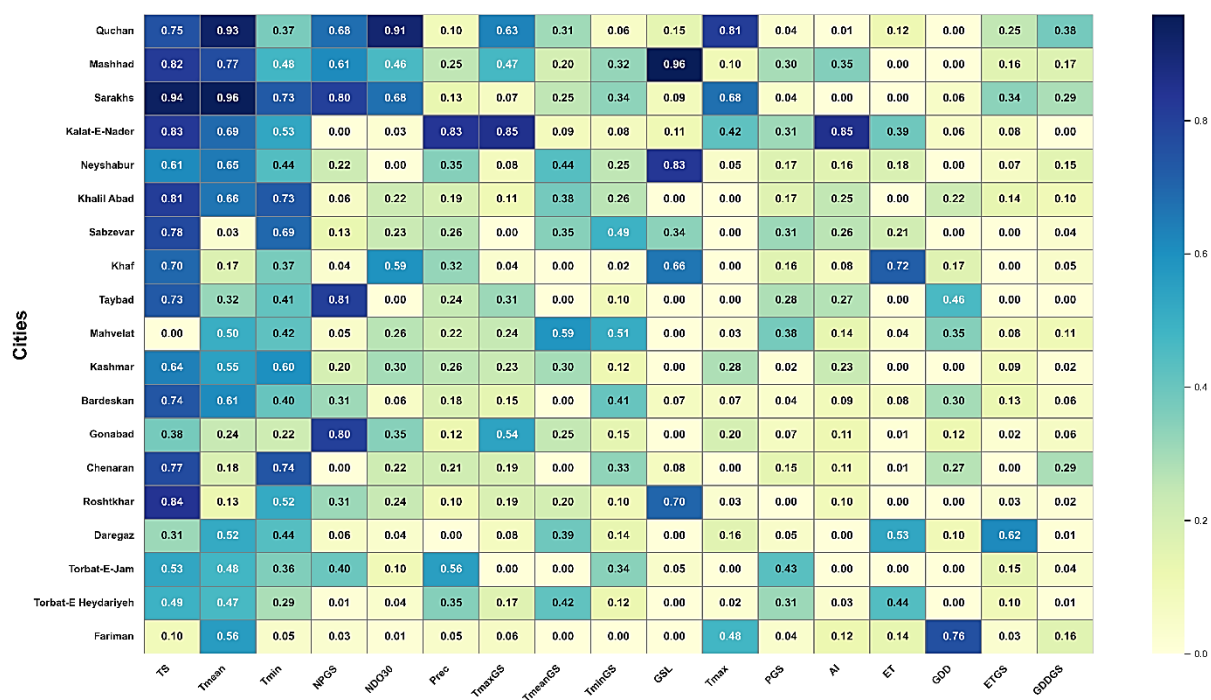
گندم آبی در استان خراسان اهمیت ویژه‌ای دارد.

اهمیت نسبی ویژگی‌های مورد استفاده در هر الگوریتم در پیش‌بینی عملکرد گندم

به منظور تحلیل عمیق‌تر و مختص هر شهرستان، ویژگی‌های تأثیرگذار در پیش‌بینی عملکرد گندم به ترتیب اولویت آن‌ها در هر شهرستان، برای هر دو مدل در شکل‌های ۴ و ۵ آورده شده است. داده‌های این شکل‌ها مقایسه بین شهرستان‌ها و شناسایی ویژگی‌های مشترک و همچنین ویژگی‌های ایجادکننده تفاوت را ممکن می‌سازد. این ویژگی‌ها در مدل‌های مختلف تأثیرگذار هستند و به میزان زیادی پیش‌بینی‌ها را تحت تأثیر قرار می‌دهند. در رندوم‌فارست، پنج ویژگی مهم به ترتیب عبارت بودند از: دمای میانگین روزانه، تعداد روزهای با دمای بالای ۳۰ درجه سانتی‌گراد،

دمای بیشینه روزانه، شدت فصلی بودن درجه حرارت و تعداد دفعات وقوع بارندگی طی فصل رشد. پنج ویژگی برتر در مدل ایکس‌جی‌بوست به شرح زیر بودند: دمای کمینه روزانه، تعداد دفعات وقوع بارندگی طی فصل رشد، بارندگی طی فصل رشد و دمای بیشینه طی فصل رشد.

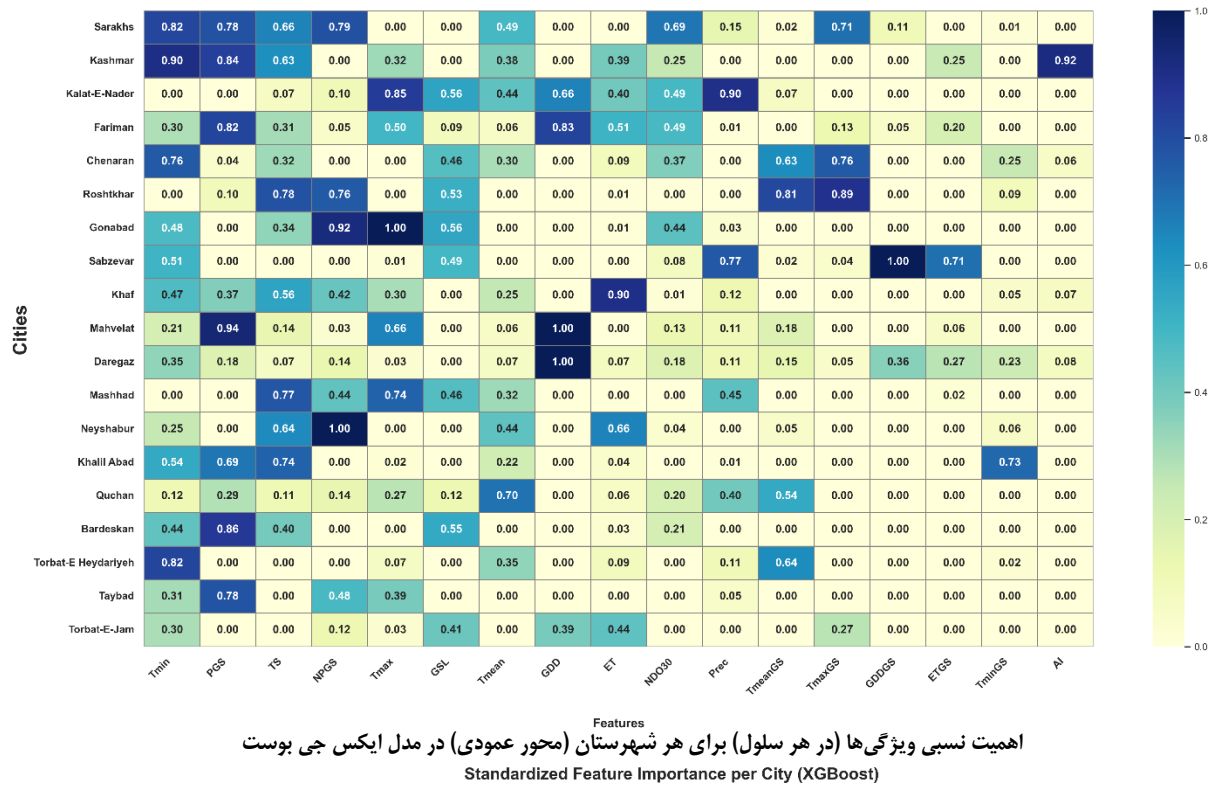
در پهنه‌بندی احتمال رخداد بیماری فوزاریوم گندم با استفاده از روش جنگل تصادفی و داده‌های حاصل از اینترنت اشیاء، پنج شاخص با درجه اهمیت نسبی زیاد تشخیص داده شدند (مثل تعداد ساعت‌هایی که دما بین ۱۵ تا ۳۰ درجه سانتی‌گراد و رطوبت نسبی بیشتر از ۸۰ درصد است). نتایج ارزیابی این مدل حاکی از کارایی آن در پیش‌بینی بیماری فوزاریوم گندم بود (Khodabandehloo, Azadbakht, Radiom, Ashourloo, & Alimohammadi, 2021).



اهمیت نسبی ویژگی‌ها (در هر سلول) برای هر شهرستان (محور عمودی) در مدل رندوم فارست
Standardized Feature Importance per City (Random Forest)

شکل ۴- اهمیت نسبی ویژگی‌های مؤثر بر پیش‌بینی عملکرد گندم در هر شهرستان توسط مدل رندوم‌فارست (بر مبنای داده‌های استاندارد شده).
(شرح کامل ویژگی‌ها در جدول ۲ بخش مواد و روش‌ها آورده شده است)

Figure 4- Relative importance of features influencing wheat yield prediction in each county using the Random Forest model (based on standardized data). (A complete description of the features is provided in Table 2 in the Materials and Methods section)



اهمیت نسبی ویژگی‌ها (در هر سلول) برای هر شهرستان (محور عمودی) در مدل ایکس جی بوست
Standardized Feature Importance per City (XGBoost)

شکل ۵- اهمیت نسبی ویژگی‌های مؤثر بر پیش‌بینی عملکرد گندم در هر شهرستان توسط مدل ایکس جی بوست (بر مبنای داده‌های استاندارد شده). (شرح کامل ویژگی‌ها در جدول ۲ بخش مواد و روش‌ها آورده شده است)

Figure 5- Relative importance of features influencing wheat yield prediction in each county using the XGBoost model (based on standardized data). (A complete description of the features is provided in Table 2 in the Materials and Methods section)

پس نمی‌توان یک درخت واحد را به‌عنوان "درخت نهایی" انتخاب کرد، چون همه درخت‌ها در کنار هم نتیجه کلی را تعیین می‌کنند. بنابراین، در مواقعی که نیاز باشد تصویری از ساختار تصمیم‌گیری درخت‌ها ارائه گردد، معمولاً تصویر درخت‌های فردی (برای یک درخت واحد) نمایش داده می‌شود، زیرا نمایش همه درخت‌ها به‌طور مستقیم از طریق یک نمودار دشوار و بی‌معنا می‌شود. (اگر تعداد درخت‌ها زیاد باشد و در حال بررسی آخرین درخت، یک ساختار درختی ناقص یا خیلی ساده در خروجی دیده شود، دلیل آن ممکن است این باشد که آن درخت به‌دلیل نمونه‌گیری تصادفی، اطلاعات مفیدی تولید نکرده باشد. درخت‌های رندوم‌فارست برخلاف درخت‌های تصمیم منفرد، ممکن است به‌دلیل نمونه‌گیری تصادفی ویژگی‌ها^۳ در برخی درخت‌ها مسیرهای خیلی کوتاه ایجاد کنند، مثلاً حالتی که در یک درخت، فقط یک یا دو ویژگی مؤثر بوده‌اند و بقیه مسیر تقسیم ساخته نشده باشد). حالت دیگر، برگ نهایی زودرس^۴ نامیده می‌شود به این معنی که در بعضی درخت‌ها (به‌ویژه در رندوم‌فارست) ممکن است برخی گره‌ها زودتر به برگ تبدیل شوند،

تحلیل بصری ویژگی‌های مؤثر مدل‌ها در سطح شهرستان‌ها برای کلیه شهرستان‌ها نمودارهای ساختار درختی، شیب، فایل‌های پی‌کی‌آل و معیارهای ارزیابی دقت مدل پیش‌بینی‌کننده محاسبه و تولید شدند. به‌دلیل تعداد زیاد این نمودارها، در اینجا تنها به ارائه نمودارهای مربوط به شهرستان مشهد به‌عنوان نمونه اکتفا می‌شود.

مصورسازی ساختار درختی^۱ مدل ایکس جی بوست و رندوم‌فارست برای شهرستان مشهد

در مدل‌های رندوم‌فارست به‌دلیل اینکه چندین درخت تصمیم^۲ به‌طور همزمان برای پیش‌بینی ساخته می‌شوند و نیز هر درخت تصمیمات را به‌طور مستقل براساس ویژگی‌های داده‌ها می‌گیرد، بنابراین نتیجه‌گیری‌های مدل، تجمعی از همه درخت‌ها است، نه یک درخت واحد. به عبارت دیگر، درخت‌های مدل به‌طور تصادفی ساخته می‌شوند و با ایجاد تنوع، قدرت پیش‌بینی مدل بیشتر می‌شود،

3- Feature bagging
4- Early Leaf Node

1 Tree Visualization
2 decision trees

مدل به‌صورت تجمیعی از تمام درخت‌ها حاصل می‌شود، اما نمایش درخت نهایی می‌تواند دید مناسبی از مسیرهای تصمیم‌گیری مدل نسبت به متغیرهای ورودی فراهم کند (مانند آستانه‌های بحرانی و اثرات متقابل ویژگی‌ها). در مدل رندوم‌فارست، متغیر تعداد دفعات وقوع بارندگی طی فصل رشد، نقش اساسی در پیش‌بینی عملکرد گندم دارد. در مدل رندوم‌فارست، متغیر دوم باز هم ماهیت رطوبتی دارد (تبخیر و تعرق طی فصل رشد).

در مدل ایکس‌جی‌بوست، متغیر دمای کمینه روزانه اولین متغیر تفکیک‌کننده بوده که دارای بالاترین میزان $gini$ یا برتری است و نقش اساسی در کاهش خطای مدل ایفا می‌کند. این موضوع نشان می‌دهد که در چارچوب این مدل، دما نقش محوری در پیش‌بینی عملکرد گندم داشته است. ویژگی دوم در مدل ایکس‌جی‌بوست، دوباره همان متغیر دمایی است (دمای میانگین روزانه).

رفتار مدل در تفکیک داده‌ها: در مدل رندوم‌فارست، از نه نمونه، هشت نمونه (یعنی $8/8$ درصد) به شاخه دوم منتقل می‌شوند و سپس دو و شش نمونه به برگ‌های بعدی و در نهایت یک نمونه در برگ نهایی باقی می‌ماند، یعنی مدل در تقسیم‌های عمیق‌تر فقط روی نمونه‌های کمتری تمرکز کرده است. این موضوع گاهی به بیش‌برازش^۶ رندوم‌فارست منجر می‌شود. در مقابل، مدل ایکس‌جی‌بوست با پوشش یکنواخت‌تر و تقسیم‌بندی‌های هدفمندتر، متغیرهایی مانند دمای کمینه روزانه، دمای میانگین روزانه و طول فصل رشد را به کار گرفته که هم متنوع‌تر هستند و هم هر یک براساس مقدار $gini$ ، سهم مشخصی در کاهش خطای کلی داشته‌اند. ساختار مدل، تعداد و مقدار برگ و تفسیر عملکرد: در رندوم‌فارست، مقدار نهایی معادل $4294/5$ واحد عملکرد گندم (کیلوگرم در هکتار) برای یک نمونه خاص است. در ایکس‌جی‌بوست، مقدار نهایی در برگ‌ها معمولاً مقدار اصلاح‌شده یا باقی‌مانده است، نه مقدار نهایی عملکرد. این موضوع باعث می‌شود که تفسیر مستقیم خروجی درخت سخت‌تر باشد، ولی مدل دقت بیشتری در مجموع خواهد داشت، چراکه این مقادیر به‌صورت تجمعی در بوسترها لحاظ می‌شوند. در مدل رندوم‌فارست ساختار ساده‌تر و تفسیر مستقیم‌تری ارائه می‌شود، ولی با خطر بیش‌برازش در برگ‌هایی با تعداد کم مواجه است. اگرچه مدل ایکس‌جی‌بوست پیچیده‌تر به نظر می‌رسد و معمولاً عمق بیشتری در شاخه‌های مرتبط با ویژگی‌ها ایجاد می‌کند، اما تقسیم‌بندی‌های آن براساس کاهش خطای مؤثرتر، بهره‌گیری از متغیرهای متنوع و مشارکت تدریجی در تقویت مدل نهایی، قدرت بیشتری در مدل‌سازی دارد.

زیرا همه نمونه‌ها در آن گره از یک کلاس هستند، یا به حداقل تعداد نمونه برای تقسیم^۱ یا حداقل نمونه برای برگ^۲ رسیده‌اند. در این صورت، آن گره دیگر تقسیم نمی‌شود، و به‌صورت برگ (حتی در عمق کم) باقی می‌ماند. این موضوع طبیعی و بخشی از فرآیند یادگیری درخت است. راه حل این مشکل، استفاده از نمودارهای اهمیت ویژگی^۳ است که می‌توانند به‌صورت گرافیکی تأثیر هر ویژگی بر پیش‌بینی مدل را نشان دهند (شکل ۷). با این حال، برخی محققان پیشنهاد کرده‌اند که نمایش آخرین درخت نیز به دلیل ارائه اطلاعات و ویژگی‌های نهایی شناسایی‌شده، می‌تواند مفید باشد (شکل ۸).

در مدل ایکس‌جی‌بوست، چون درخت‌ها را به‌طور تدریجی و گام‌به‌گام اضافه می‌کند، اولین درخت، بیشترین تأثیر را بر پیش‌بینی‌های مدل دارد، زیرا بقیه درخت‌ها در مسیر اصلاح خطاهای قبلی عمل می‌کنند. مفهوم بوس‌تینگ^۴ در ایکس‌جی‌بوست و دیگر مدل‌های مبتنی بر بوس‌تینگ، به همین مفهوم اشاره دارد، یعنی مدل به تدریج (درخت به درخت) یاد می‌گیرد و درخت‌های بعدی بیشتر به اصلاح پیش‌بینی‌های غلط (خطا) در درخت‌های قبلی می‌پردازند. درخت‌های ابتدایی (از نظر رتبه) اطلاعات بیشتری ارائه می‌دهند. درخت اول اطلاعات اولیه‌ای از روابط پیچیده داده‌ها به مدل می‌دهد و می‌توان آن را به‌صورت بصری بررسی کرد. بنابراین، مزایای نمایش اولین درخت در ایکس‌جی‌بوست به‌شرح زیر هستند: نمایش ساده‌تر و قابل درک‌تر از تصمیمات مدل، به‌خصوص درخت اول که بیشترین تغییر را ایجاد می‌کند؛ می‌تواند به‌عنوان یک نقطه شروع خوب برای توضیح مدل و نحوه تصمیم‌گیری‌های اولیه باشد؛ به‌ویژه برای درک ویژگی‌های کلیدی که در درخت اول مورد توجه قرار گرفته‌اند، می‌تواند مفید باشد.

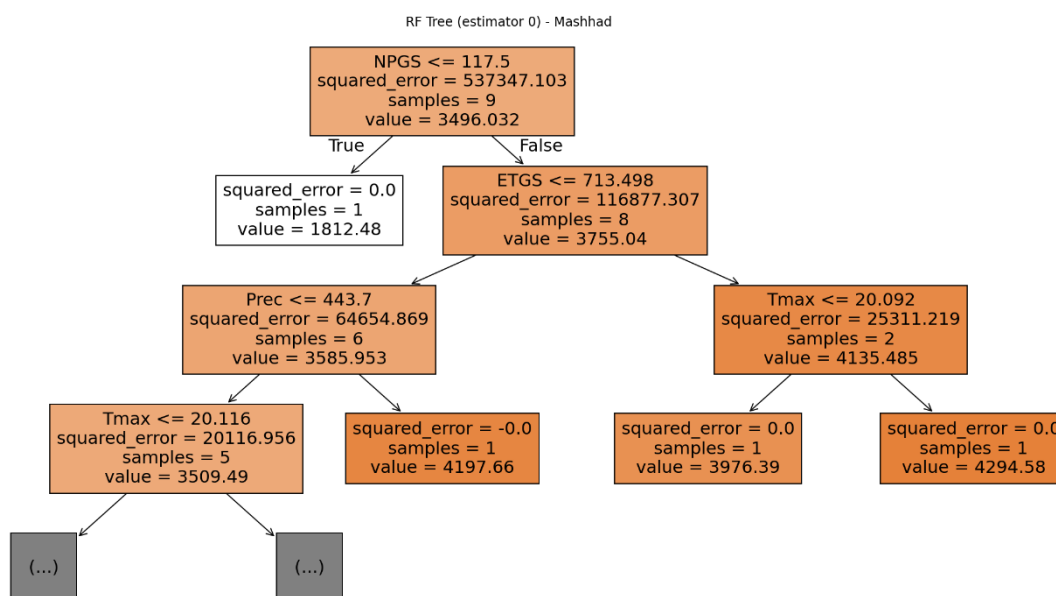
در نمودارهای درختی نهایی، ساختار تصمیم‌گیری مدل‌ها براساس شاخص‌های اصلی، از جمله ویژگی‌هایی نظیر میزان بارندگی، دمای هوا، تاریخ کاشت، و غیره نمایش داده شده است. مدل ایکس‌جی‌بوست و رندوم‌فارست هر دو درخت‌هایی با ساختار نسبتاً ساده برای شهرستان مشهد تولید کردند، ایکس‌جی‌بوست با چهار سطح و رندوم‌فارست با چهار سطح (شکل ۶ و ۶ ب)، اما تفاوت‌هایی اساسی در روش شناختی، ساختار و نوع متغیرهای به‌کاررفته و نحوه تقسیم نمونه‌ها دارند که عبارتند از:

متغیرهای تأثیرگذار اولیه: به‌منظور درک ساختار تصمیم‌گیری مدل رندوم‌فارست، درخت نهایی از میان درخت‌های ساخته‌شده در مدل بهینه‌شده برای هر شهرستان ترسیم شد. هرچند تصمیم نهایی

- 1- Min_samples_split
- 2- Min_samples_leaf
- 3- Importances of Features
- 4- Boosting

5 Gain

6- Overfitting



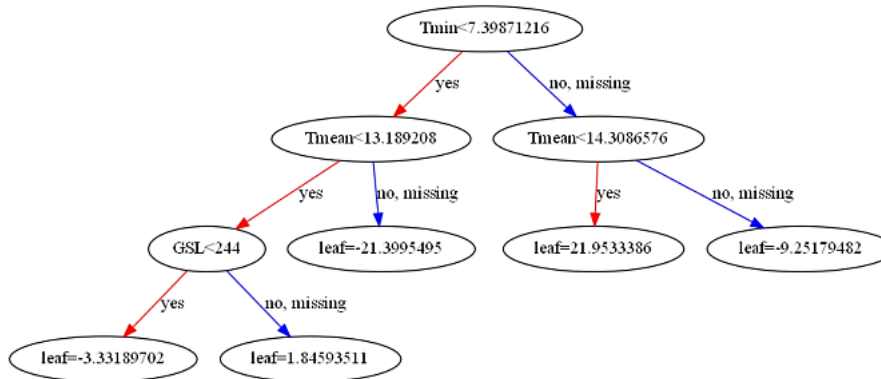
شکل ۶a- آخرین نمودار درختی تولیدشده توسط مدل رندوم فارست برای شهرستان مشهد

Figure 6A- The final decision tree generated by the Random Forest model for Mashhad County

NPGS: تعداد دفعات وقوع بارندگی طی فصل رشد، ETGS: تبخیر و تعرق طی فصل رشد، PREC: بارندگی، Tmax: دمای بیشینه روزانه

NPGS: Number of rainy days during the growing season, ETGS: evapotranspiration during the growing season, PREC: precipitation, () and Tmax: daily maximum temperature

XGBoost Tree 0 - Mashhad



شکل ۶b- اولین ساختار درختی تولیدشده توسط مدل ایکس جی بوست برای شهرستان مشهد

Figure 6B- The first tree structure generated by the XGBoost model for Mashhad County

Tmin: دمای کمینه روزانه، Tmean: دمای میانگین روزانه، GSL: طول فصل رشد

Tmin: Daily minimum temperature, Tmean: average daily temperature, and GSL: growing season length

مدل‌های خطی سنتی برآورد می‌کند (Everingham, Sexton, Skocaj, & Inman-Bamber, 2016).

به‌طور کلی، ساختارهای درختی بر قابلیت تحلیل و تفسیر نتایج مدل‌ها بسیار تأثیرگذار هستند و می‌توانند در مطالعات آینده برای شفاف‌سازی تصمیم‌های مدل به کار روند.

در نهایت، بسته به هدف مطالعه (تفسیر یا دقت پیش‌بینی)، هر یک از دو مدل مزایای خاص خود را دارند، ولی برای مدل‌سازی دقیق‌تر و شناسایی متغیرهای تأثیرگذار کلیدی، تفسیرهای ایکس جی بوست می‌توانند اطلاعات عمیق‌تری فراهم می‌کنند. گزارش شده است که رندوم فارست عملکرد نیشکر را بهتر از

تحلیل نمودارهای اهمیت نسبی ویژگی‌ها (نمودارهای شیب ویولونی) مدل رندوم‌فارست برای شهرستان مشهد

در مدل رندوم‌فارست برای شهرستان مشهد، ویژگی‌هایی که بیشترین تأثیر را در پیش‌بینی عملکرد گندم دارند به ترتیب از بیشترین اهمیت به کمترین اهمیت عبارتند از:

۱. T_{min} : دمای حداقل شبانه نقش مهمی در شدت تنفس نگهداری، تسهیم و توزیع فرآورده‌های فتوسنتزی تولیدشده طی مرحله روشنائی، تعادل هورمونی و غیره دارد. در دوره‌های شبانه، دمای پایین‌تر می‌تواند شدت تنفس را کاهش داده و سبب ذخیره بیشتر انرژی در گیاه شود (Hatfield & Prueger, 2015).

۲. TS : این ویژگی دومین رتبه اهمیت را در بین ویژگی‌ها داشت و تأثیر بسزایی در پیش‌بینی‌ها دارد. همچنین نقش برجسته‌ای در تغییرات دمایی و تنش دمایی (در طول فصل رشد گندم) دارد.

۳. T_{meanGS} : میانگین دمایی مطلوب سبب تعادل بین فتوسنتز و تنفس شده و به رشد بهینه کمک می‌کند (Asseng et al., 2015).

۴. T_{minGS} : دماهای پایین‌تر می‌توانند تأثیر منفی بر جوانه‌زنی و مراحل اولیه رشد داشته باشند، ولی در مواردی نیز از تنش گرمایی جلوگیری می‌کنند.

۵. PGS : مقدار بارندگی در طول فصل رشد رتبه پنجم اهمیت در بین ویژگی‌های مورد بررسی را به خود اختصاص داد. تأمین آب کافی و منظم در دوره رشد برای تأمین آب و جذب عناصر غذایی حیاتی است (Si, Qin, Liang, Duan, & Gao, 2023).

۶. $NPGS$: توزیع زمانی بارش حتی مهم‌تر از مقدار آن است، زیرا دفعات بارندگی در تثبیت ظرفیت ذخیره رطوبت در خاک نقش کلیدی دارد.

۷. GDD : این شاخص، جمع دماهای مفید برای رشد است و با مراحل فنولوژیک گیاه ارتباط مستقیم دارد (Vieira, Bradford, Osterholzer, Pierce, & Cockrell, et al., 2025).

۸. $Prec$: بارش‌ها بر آب‌رسانی به گیاه و تأثیر مستقیم بر فرآیندهای رشدونمو و در نهایت تعیین میزان عملکرد نهایی تأثیر مستقیم دارند. تأمین نیاز آبی گیاه، بر مراحل حساس رشد از جمله خوشه‌دهی و پر شدن دانه تأثیر می‌گذارد. یافته‌ها تأکید می‌کنند که حتی در شرایط آبیاری، مدیریت مناسب زمان‌بندی و میزان آبیاری (بارندگی) برای بهینه‌سازی جذب عناصر غذایی و افزایش عملکرد گندم ضروری است (Wang

(Li, 2002).

۹. T_{mean} : دمای میانگین روزانه در طول دوره رشد تأثیر زیادی بر عملکرد گندم دارد. این ویژگی به‌ویژه در نواحی با تغییرات دمایی زیاد مهم است.

۱۰. $NDO30$: تعداد روزهایی که دما بیش از ۳۰ درجه سانتی‌گراد است، نشان‌دهنده شرایط دمایی بحرانی برای گیاه می‌باشد (Farhadi et al., 2024).

۱۱. T_{max} : افزایش بیش از حد این دما، فرآیند فتوسنتز را مختل و تعرق را تشدید می‌کند که منجر به افت عملکرد می‌شود.

۱۲. $ETGS$: یکی از شاخص‌های کلیدی روابط آبی بوده که ارتباط نزدیکی با نیاز آبی گیاه دارد.

۱۳. AI : شاخص خشکی، بر عملکرد تأثیر منفی داشته است که این موضوع بسیار بدیهی است و توسط متخصصان کشاورزی برای ارزیابی شرایط عمومی یک ناحیه برای فعالیت‌های کشاورزی استفاده می‌شود.

۱۴. T_{maxGS} : باعث بروز تنش گرمایی در طول فصل رشد شده و رشد سلولی و پر شدن دانه را مختل می‌کند.

۱۵. ET : این شاخص در مناطق گرم و خشک، معیار مناسبی برای بررسی شدت نیاز آبی و خطر تنش خشکی است.

۱۶. $GDDGS$: مشابه GDD ، این ویژگی در تحلیل مراحل نموی از نظر زمان وقوع و طول دوره هر مرحله نموی عامل تعیین‌کننده‌ای به شمار می‌رود، به‌طور خاص نشان‌دهنده تأثیر گرمای تجمعی در فصل رشد بر پیشرفت فنولوژیک است.

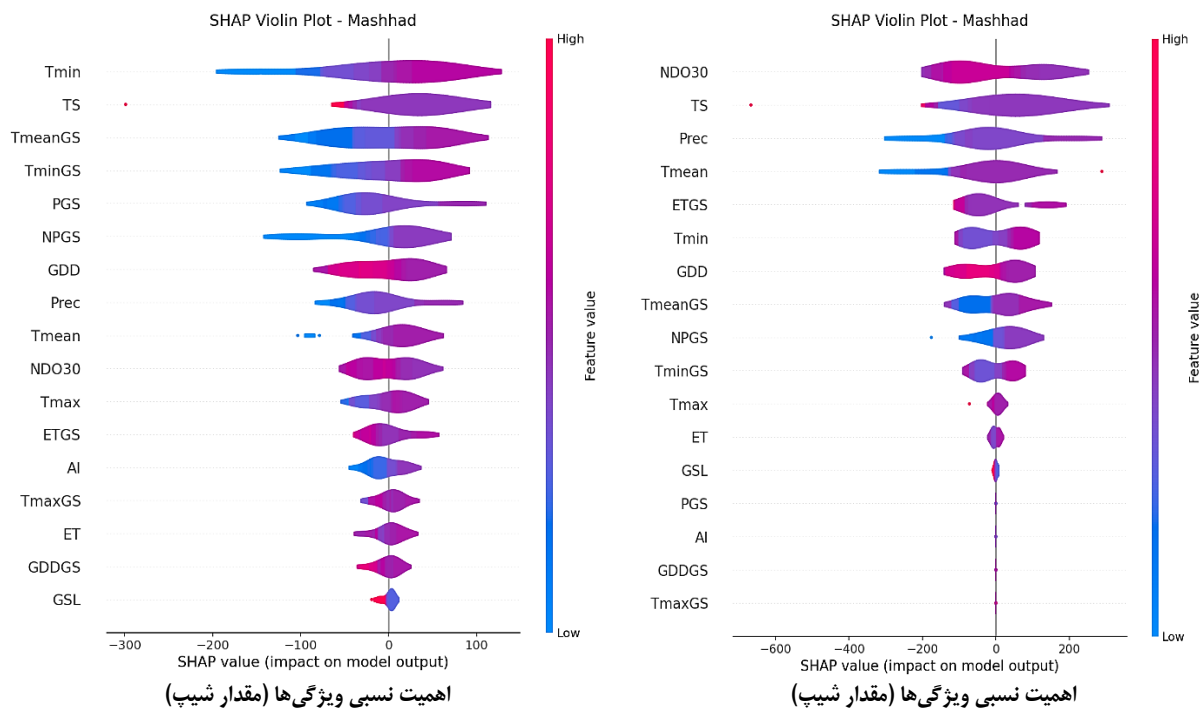
۱۷. GSL : رابطه مستقیمی با فرصت زمانی برای فتوسنتز و پر شدن دانه دارد؛ طولانی‌تر بودن این فصل معمولاً عملکرد بالاتر را سبب می‌شود.

ویژگی‌های با رتبه‌های یک تا پنج، به‌ویژه در مناطق با تغییرات دمایی زیاد، شرایط رشد گیاه را تحت تأثیر قرار می‌دهند و مدل به‌خوبی اهمیت آن‌ها را پیش‌بینی کرده است (Taiz, Zeiger, Moller, & Murphy, 2018; Slafer, Savin, Sadras, & Calderini, 2023). نتایجی مشابه با نتایج مطالعه حاضر توسط سلیمان نژاد (2018) گزارش شده است مبنی بر اینکه الگوریتم ناپارامتریک رندوم‌فارست روش مناسبی در برآورد مشخصه‌های سه مقطع، تاج‌پوشش و تراکم جنگل‌های زاگرس با استفاده از تصاویر OLI ماهواره لندست بود. در پژوهشی که با دو رهیافت ارزیابی میدانی و سنجش از دور انجام گرفت، بین نتایج حاصل از به‌کارگیری دو مدل رگرسیون خطی چندگانه و جنگل تصادفی در پیش‌بینی نیتروژن کل خاک (به عنوان مدل میدانی)، و سنجش از دور اختلاف معنی‌داری مشاهده نشد (Sadeghi & Ahmadi Nadoushan, 2021).

تشخیص داده شد. در رتبه دوم شدت فصلی بودن درجه حرارت، در رتبه سوم بارندگی و در رتبه چهارم مشابه با مدل رندوم فارست، متغیر دمایی میانگین روزانه (در رندوم فارست دمای میانگین طی فصل رشد است) قرار گرفت. این ویژگی‌ها ارتباط مستقیمی با رشد گیاهان به‌ویژه در مناطق گرم و خشک دارند و به همین دلیل در مدل ایکس‌جی‌بوست، به‌خصوص در مناطقی که گندم به‌میزان بیشتری تحت تأثیر گرما و خشکی که به‌طور مستقیم رشد و فیزیولوژی گیاه را تحت تأثیر قرار می‌دهند، اهمیت بیشتری پیدا کرده‌اند.

تحلیل نمودارهای اهمیت نسبی ویژگی‌ها (نمودارهای شیب ویولونی) مدل ایکس‌جی‌بوست برای شهرستان مشهد

داده‌های حاصل از مدل ایکس‌جی‌بوست برای شهرستان مشهد، ترتیب اهمیت متغیرها را متفاوت با مدل رندوم فارست نشان داد، با بیان این نکته مهم که در عین حال یک متغیر مهم این مدل در گروه پنج متغیری مهم مدل رندوم فارست نیز حضور دارند (به عبارت دیگر، اهمیت نسبی شدت فصلی بودن درجه حرارت را به‌خوبی تشخیص داده است) (شکل ۷). در این مدل، اولین و مهم‌ترین ویژگی، تعداد روزهای با دمای بالاتر از ۳۰ درجه سانتی‌گراد



شکل ۷- نمودارهای شیب (ویولونی) اهمیت نسبی ویژگی‌ها در مدل‌های بهینه‌شده رندوم فارست (سمت چپ) و ایکس‌جی‌بوست (سمت راست) در پیش‌بینی عملکرد گندم در شهرستان مشهد. (شرح کامل ویژگی‌ها در جدول ۲ بخش مواد و روش‌ها آورده شده است)

Figure 7- SHAP (violin) plots of the relative importance of features in the optimized Random Forest model (left) and XGBoost model (right) for predicting wheat yield in Mashhad County. (A complete description of the features is provided in Table 2 in the Materials and Methods section)

بارندگی در مدل ایکس‌جی‌بوست، می‌تواند نقش تعیین‌کننده عوامل دمایی مشترک در هر دو مدل را تحت‌الشعاع قرار دهد. به عبارت دیگر، عدم قطعیت موجود در مدل‌ها و تفاوت شرایط واقعی با آنچه که در تئوری فرض می‌شود، در نظر گرفته شود و توجه داشت که معنی‌داری آماری، و معنی‌داری بیولوژیکی دو مفهوم کاملاً متفاوت هستند.

بارندگی در هر دو مدل اهمیت دارد، اما در جایی که این ویژگی به‌عنوان یک عامل کلیدی در پیش‌بینی‌ها دیده شود، تأثیرات آن در مدل ایکس‌جی‌بوست ممکن است بیشتر از مدل رندوم فارست باشد.

نکته قابل توجه اینکه در هر دو مدل، متغیرهای با ماهیت دمایی حائز بیشترین اهمیت نسبی تشخیص داده شده‌اند. ویژگی‌های بعدی (از رتبه دهم تا هفدهم) نسبت به مدل رندوم فارست، اهمیت بسیار کمتری را به خود اختصاص داده‌اند. در تحلیل و تفسیر نتایج این چینی، باید دقت شود تا از دلیل تراشی بی‌مورد و توجیه‌های غیرمنطقی پرهیز گردد. برای مثال، اهمیت عامل بارندگی، در مدل ایکس‌جی‌بوست در رتبه سوم تعیین شده است. همین متغیر، در مدل رندوم فارست در رتبه هشتم قرار گرفته است (پراکنش مقادیر شیب آن در هر دو مدل تقریباً یکسان است). تأکید بیش از حد بر اهمیت

کمینه روزانه (آبی‌رنگ) باعث کاهش عملکرد شده‌اند (سمت چپ خط متوسط عملکرد یا مقدار پایه ۳۶۰۰ کیلوگرم در هکتار قرار دارند؛ به عبارت دیگر، شیب این مقدار پایین منفی بوده، یعنی باعث کاهش عملکرد نسبت به میانگین مدل شده) که بیانگر حساسیت گیاه به دماهای پایین در مراحل ابتدایی رشد است.

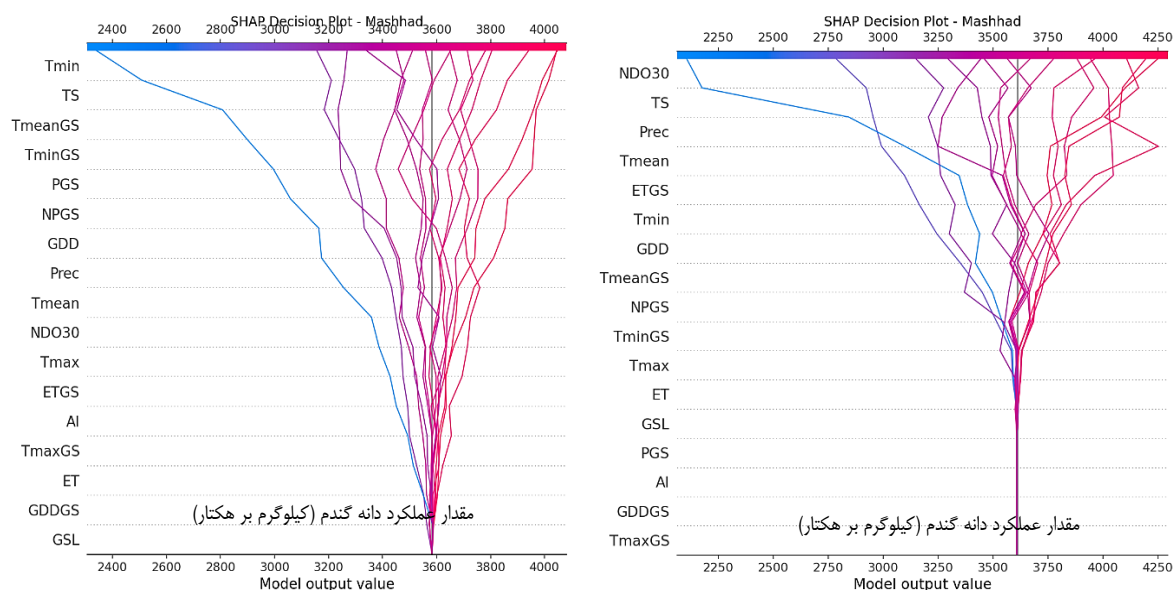
برای سایر ویژگی‌ها، به‌طور مشابه می‌توان تحلیل ارائه داد. به‌طور کلی، مدل رندوم‌فارست توانست رابطه پیچیده‌ای بین متغیرهای اقلیمی و عملکرد گندم استخراج کند. نمودار شیب تصمیم‌نشان می‌دهد که دمای کمینه روزانه، شدت فصلی بودن درجه حرارت، دمای مینگین طی فصل رشد، دمای میانگین روزانه و بارندگی طی فصل رشد مهم‌ترین نقش را در در پیش‌بینی عملکرد برای مشهد دارند. این تحلیل برای کشاورزان و سیاست‌گذاران می‌تواند در بهینه‌سازی تاریخ کاشت و آبیاری مؤثر باشد. نمودار تصمیم در الگوریتم ایکس‌جی‌بوست مشهد در سمت راست شکل ۸ نشان داده شده است.

شکل ۹ که از کاربردی‌ترین خروجی‌های شیب محسوب می‌شود، نمودار وابستگی جزئی یا دی‌دی نام دارد که اثر مستقیم و جداگانه هر ویژگی بر عملکرد گندم را نمایش می‌دهد، درحالی‌که سایر ویژگی‌ها ثابت نگه داشته شده‌اند. در ساختار این نمودار، روی محور افقی مقدار ویژگی (مثلاً دمای میانگین روزانه، دمای میانگین طی فصل رشد یا غیره)، و روی محور عمودی، مقدار پیش‌بینی‌شده عملکرد (توسط مدل) قرار می‌گیرند. خطوط نمودار، منحنی پاسخ مدل نسبت به تغییرات هر ویژگی را نشان می‌دهد. در مدل رندوم‌فارست (شکل بالایی)، با افزایش تعداد دفعات وقوع بارندگی از ۱۱۲ تا ۱۳۰ مرتبه، مقدار عملکرد گندم در مشهد با شیب زیادی افزایش می‌یابد و به حدود ۳۶۰۰ کیلوگرم در هکتار می‌رسد، ولی پس از آن با افزایش دفعات بارندگی تا حدود ۱۷۰، این شیب افزایشی کاهش یافت و به حالت تقریباً اشباع درآمد (یعنی افزایش بیشتر این ویژگی تأثیر محسوسی بر عملکرد ندارد). افزایش دمای کمینه روزانه، از ۶ تا ۸/۵ درجه سانتی‌گراد سبب افزایش یکنواخت عملکرد گندم شد. این روند افزایشی نشان می‌دهد که در شرایط آب‌وهوایی مشهد، گندم نسبت به افزایش دمای کمینه تا این بازه پاسخ مثبت دارد. با این توضیح که محدوده ۷ تا ۷/۵ درجه سانتی‌گراد، نوسانات جزئی وجود داشت که احتمالاً به اثرات پیچیده تعامل بین ویژگی‌ها یا نویز داده مربوط است. افزایش شاخص خشکی در بازه ۸/۵ تا حدود ۱۰/۵، اثر منفی چندانی ندارد، حتی گاهی عملکرد اندکی بیشتر هم می‌شود، اما پس از حدود ۱۰/۵، روند تغییر ناگهانی شده و با افزایش مقدار شاخص شدت فصلی بودن درجه حرارت به بیش از ۱۱/۵، عملکرد گندم حدود ۳۰۰ کیلوگرم در هکتار افت می‌کند که این یک نقطه آستانه بحرانی است که از آن به بعد خشکی به‌طور محسوس به عملکرد آسیب می‌زند.

تحلیل نمودارهای شیب تصمیم و وابستگی جزئی مدل رندوم‌فارست برای شهرستان مشهد

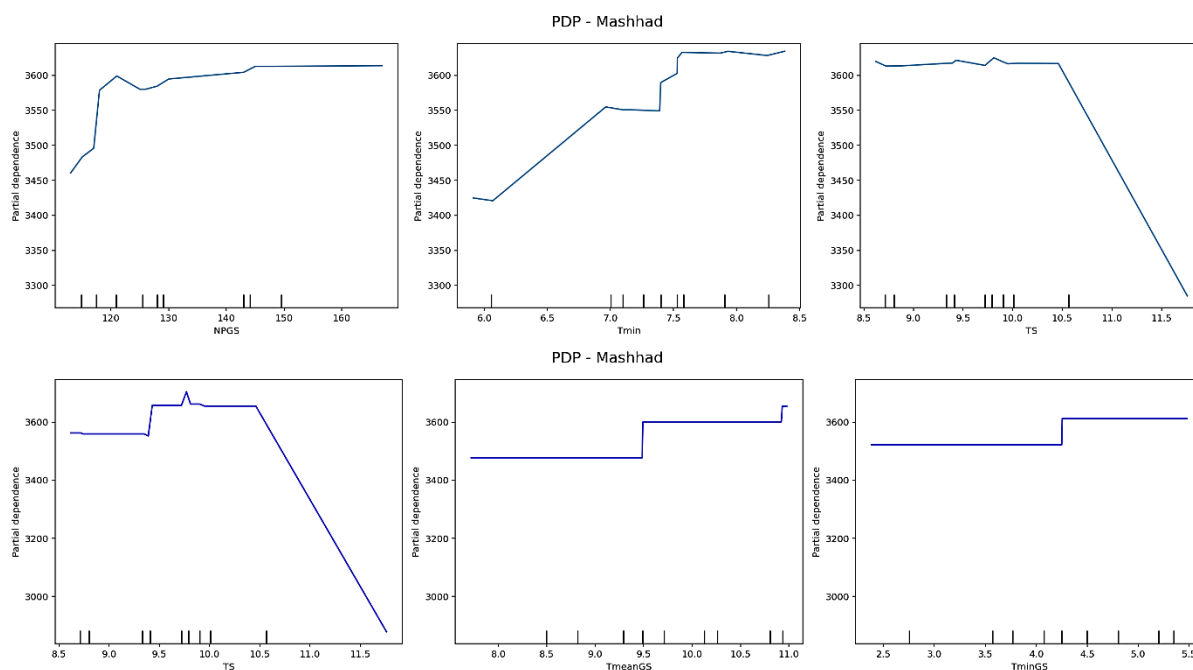
شکل ۸ یک نمودار تصمیم مبتنی بر مقادیر شیب برای مدل رندوم‌فارست (سمت چپ) و ایکس‌جی‌بوست (سمت راست) است که برای پیش‌بینی عملکرد گندم در شهرستان مشهد مورد استفاده قرار گرفته است. مقادیر شیب، ابزار قدرتمندی برای درک نحوه اثرگذاری هر ویژگی ورودی بر پیش‌بینی مدل هستند. در این تصویر، تغییرات مقادیر شیب برای هر ویژگی نسبت به مقادیر آن ویژگی در نمونه‌های مختلف مشاهده می‌شود. در این نمودار، محور افقی نمایانگر مقدار شیب است، که میزان تأثیر هر ویژگی را بر پیش‌بینی عملکرد نشان می‌دهد (مثلاً افزایش یا کاهش عملکرد). محور عمودی ویژگی‌های مدل را فهرست کرده است (مانند دمای کمینه روزانه، شدت فصلی بودن درجه حرارت، دمای میانگین طی فصل رشد و غیره). هر خط بیانگر یک نمونه (سال-مکان) در اینجا منظور عملکرد گندم در یک سال خاص در مشهد است) است. هر کدام از این خط‌های رنگی هنگام عبور از یک ویژگی (مثلاً دمای کمینه روزانه)، به سمت چپ یا راست منحرف می‌شوند: اگر به سمت چپ برود، یعنی شیب منفی بوده و عملکرد کاهش می‌یابد. اگر خط به سمت راست خط پایه منحرف شود، یعنی شیب مثبت بوده و عملکرد افزایش یافته است. در ویژگی‌هایی که قدرت توضیح‌دهندگی بالایی دارند، این انحراف‌ها زیاد و شدید هستند. اگر خطوط در یک ویژگی به‌صورت قابل توجهی از هم جدا بشوند (باز شوند)، یعنی آن ویژگی اثر زیادی در خروجی مدل دارد. اگر خطوط تقریباً صاف یا هم‌راستا باقی بمانند، آن ویژگی اثر چندانی در مدل نداشته است. اثر یک ویژگی همیشه خطی یا یکنواخت نیست؛ یعنی مقدار بیشتر یک ویژگی لزوماً همیشه به معنای بهتر یا بدتر بودن نیست، بلکه ممکن است از یک نقطه به بعد، اثر آن ویژگی کاهش پیدا یافته یا حتی برعکس شود. این پدیده در مدل‌های پیچیده مثل رندوم‌فارست یا ایکس‌جی‌بوست به‌وضوح دیده می‌شود و می‌تواند در تحلیل شیب یا نمودارهای تصمیم مشخص شود.

همان‌طور که در **شکل ۸** (سمت چپ) مشاهده می‌شود، دمای کمینه روزانه در رأس ویژگی‌های مؤثر قرار دارد، به عبارت دیگر، اولین ویژگی واردشونده در مدل است؛ یعنی مدل رندوم‌فارست در ابتدا تصمیم خود را براساس مقدار دمای کمینه روزانه در هر نمونه شکل می‌دهد. در شکل ۸ خط‌هایی که از ویژگی دمای کمینه روزانه عبور می‌کنند، به‌شدت به چپ یا راست منحرف شده‌اند (قرمزها به راست (عملکرد بالا) یعنی دمای کمینه روزانه بالا باعث افزایش عملکرد شده؛ آبی‌ها به چپ (عملکرد پایین) یعنی دمای کمینه روزانه پایین باعث افت عملکرد شده‌اند. بسیاری از نمونه‌ها، مقادیر پایین دمای



شکل ۸- نمودارهای شیب تصمیم برای پیش بینی عملکرد گندم در شهرستان مشهد توسط مدل رندوم فارست (سمت چپ) و ایکس جی بوست (سمت راست)

Figure 8- Decision SHAP plots for predicting wheat yield in Mashhad county by the Random Forest model (left) and XGBoost model (right)



شکل ۹- نمودار شیب وابستگی جزئی (دی پی دی) برای مدل رندوم فارست (بالا) و ایکس جی بوست (پایین) در پیش بینی عملکرد گندم مشهد

Figure 9- SHAP Partial Dependence (DPD) plots for the Random Forest model (top) and XGBoost model (bottom) in predicting wheat yield in Mashhad

هستند. همچنین وجود خشکی شدید (شاخص خشکی) یک تهدید محسوب می شود.

یافته‌ها تأیید می‌کنند که تعادل در گرما (دمای کمینه روزانه)، شاخص خشکی و شرایط مناسب آبی (تعداد دفعات وقوع بارندگی طی فصل رشد) برای دستیابی به عملکرد بهینه گندم در مشهد ضروری

نتیجه‌گیری

در مطالعات متعدد بیان شده است که مدل ایکس‌جی‌بوست به ویژگی‌های خاص‌تر و پیچیده‌تر مانند درجه روز رشد توجه دارد، و می‌تواند ویژگی‌ها و برهمکنش‌های بیشتری را مدل‌سازی کند، اما برای شهرستان مشهد این چنین نبود. این نمودارها به کارشناسان زراعت کمک می‌کنند تا دلیل تصمیم‌گیری مدل‌ها را بهتر درک کنند و از آن برای بهینه‌سازی کشت و مدیریت مزرعه در سال‌های آینده بهره ببرند. نمودارهای تکمیلی برای سایر شهرستان‌ها تولید شده و بنا به درخواست در دسترس علاقمندان قرار می‌گیرد.

گزارش شده است که الگوریتم رندوم‌فارست در ارزیابی توان اکولوژیک کاربری جنگلداری در حوزه آبخیز کن (استان تهران)، توان بسیار بالایی در پیش‌بینی مناطق مستعد از خود نشان داد و متغیرهای عمق خاک، بارندگی در فصل رویش، ارتفاع، بافت خاک، شیب و جهت به‌ترتیب دارای اهمیت از زیاد به کم بودند (Oghnoum, Fegghi, Makhdoum, Moghaddamnia, & Etemad, 2019). در به‌کارگیری مدل رگرسیونی جنگل تصادفی به‌منظور برآورد زیست‌توده روی زمین جنگل با استفاده از تصاویر ماهواره‌ای، مشخص شد که شاخص‌های RVI، GNDVI، NDVI و DVI اهمیت بیشتری در ارائه مدل برآورد زیست‌توده داشتند. ضریب تبیین این مدل ۰/۸۰ و جذر میانگین مربعات خطا برابر با ۲۸/۷ گزارش شد (Gheysarbeigi, Pir Bavaghar, & Valipour, 2024).

ارزیابی و تعیین اعتبار

به‌منظور تحلیل عمیق‌تر و مقایسه دقیق‌تر دقت دو الگوریتم در پیش‌بینی عملکرد گندم، چهار معیار ارزیابی مدل‌ها به تفکیک برای ۱۹ شهرستان، و همچنین میانگین معیارهای ارزیابی دو الگوریتم، در **جدول ۵** ارائه شده‌اند. ترتیب شهرستان‌ها بر مبنای معیارهای ارزیابی، در هر دو الگوریتم و نیز میانگین آن‌ها، از زیاد به کم، مرتب شده‌اند.

جذر میانگین مربعات خطا: هرچه این مقدار کمتر باشد، مدل پیش‌بینی دقیق‌تری دارد. درصد تفاوت برای RMSE نشان‌دهنده این است که کدام مدل به‌طور متوسط خطای بزرگ‌تری دارد. برای بیشتر شهرستان‌ها، مدل ایکس‌جی‌بوست خطای بالاتری دارد. در این بین، چند مورد وجود دارد که این مدل عملکرد بهتری نشان داده است (مثل شهرستان‌های خاف، مشهد، رشتخوار و تایباد).

○ **ضریب تبیین:** این مقدار نشان‌دهنده میزان تطابق پیش‌بینی‌ها با داده‌های واقعی است و هرچه این مقدار به یک نزدیک‌تر باشد، مدل بهتر عمل کرده است. درصد تفاوت در R^2 نشان‌دهنده این است که مدل‌ها چقدر در تطبیق داده‌ها موفق بوده‌اند. مدل رندوم‌فارست در شش شهرستان مقدار ضریب

تبیین بالاتری دارد. این موضوع نشان می‌دهد که این مدل در بیشتر موارد توانسته است توضیح‌دهنده بهتری برای داده‌ها باشد. وجود مقادیر صفر در ستون‌های ضریب تبیین، در نتیجه نرمال‌سازی این مقادیر ضریب تبیین حاصل شده است. نرمال‌سازی ضرایب تبیین در برخی موارد خاص مثل برخی مدل‌های رگرسیون غیرخطی یا رندوم‌فارست، امری معمول است.

○ **میانگین قدر مطلق خطا:** نشان‌دهنده خطای مطلق پیش‌بینی است، هرچه کمتر باشد، مدل بهتر است. برای MAE، در ۱۵ شهرستان مدل رندوم‌فارست عملکرد بهتری دارد که نشان‌دهنده این است که این مدل توانست پیش‌بینی‌های دقیق‌تری انجام دهد.

○ **شاخص توافق ویلموت نیز** برای اندازه‌گیری تطابق پیش‌بینی‌ها با مقادیر واقعی استفاده می‌شود و هرچه به یک نزدیک‌تر باشد، مدل بهتر است. شاخص d برای ارزیابی دقت مدل در شبیه‌سازی داده‌ها مفید است. رندوم‌فارست در ۱۵ شهرستان شامل بردسکن، سرخس، تربت جام، رشتخوار، نیشابور، مه ولات، کاشمر و گناباد امتیاز بالاتری دارد که نشان‌دهنده دقت بیشتر در پیش‌بینی است.

تعیین الگوریتم برتر در هر شهرستان براساس معیارهای ارزیابی و رتبه‌بندی شهرستان‌ها: مدل بهتر برای هر شهرستان با استفاده از میانگین چهار معیار ارزیابی هر مدل تعیین شد (**جدول ۶**). این مقایسه نشان می‌دهد که کدام مدل در پیش‌بینی دقیق‌تر عملکرد گندم در هر شهرستان عمل کرده است. طبق تحلیل انجام‌شده، مدل رندوم‌فارست به‌عنوان مدل برتر برای اکثر شهرستان‌ها شناخته شده است. این موضوع به‌وضوح در شهرستان‌هایی که رندوم‌فارست امتیاز بهتری نسبت به ایکس‌جی‌بوست کسب کرده است، مشاهده می‌شود. مدل ایکس‌جی‌بوست در چند شهرستان خاص مانند بردسکن، مشهد، درگز و سرخس، امتیاز بهتری نسبت به رندوم‌فارست داشته است. این تفاوت‌ها ممکن است به ویژگی‌های خاص داده‌های این شهرستان‌ها، مانند توزیع متفاوت داده‌ها یا پیچیدگی‌های خاص این مناطق برگردد. برای شش شهرستان شامل کاشمر، چناران، خاف، مه ولات و کلات نادر، رتبه‌ها در هر دو مدل مشابه بود (۳۱/۵ درصد کل شهرستان‌ها) (**جدول ۶**). در هفت شهرستان شامل خایا آباد، سبزوار، تربت جام، رشتخوار، نیشابور، گناباد و تربت حیدریه، مدل رندوم‌فارست امتیازهای بیشتری نسبت به مدل ایکس‌جی‌بوست داشت. شهرستان‌های رتبه‌های اول تا پنجم براساس میانگین معیارهای ارزیابی دو الگوریتم، همان شهرستان‌های رتبه‌های یک تا پنج در هر دو مدل بود (**جدول ۶**).

جدول ۵- چهار معیار ارزیابی دو الگوریتم رندوم فارست و ایکس جی بوست در پیش‌بینی عملکرد گندم در استان خراسان رضوی به تفکیک شهرستان‌ها

Table 5- Four evaluation metrics of the Random Forest and XGBoost algorithms in predicting wheat yield in Khorasan Razavi province, broken down by county

شهرستان County	R ² (XGB)	R ² (RF)	RMSE (XGB)	RMSE (RF)	MAE (XGB)	MAE (RF)	d (XGB)	d (RF)
تربت جام Torbat-E-Jam	0.13	0.34	269.36	312.61	211.48	308.78	0.77	0.71
نیشابور Neyshabur	0	0.33	377.91	343.82	298.37	232.39	0.78	0.59
مه ولات Mahvelat	0	0.16	481	344.22	441.6	329.94	0.71	0.69
فریمان Fariman	0	0.07	205.41	314.79	175.52	253.8	0.62	0.75
بردسکن Bardeskan	0	0.05	387.03	187.69	341.11	149.23	0.6	0.8
چناران Chenaran	0	0.05	410.79	281.64	317.07	202.42	0.59	0.62
کاشمر Kashmar	0	0	472.61	217.93	424.28	164.5	0.51	0.56
گناباد Gonabad	0	0	683.93	444.72	579.88	372.6	0.46	0.55
تایباد Taybad	0	0	455.4	521.03	380.93	473.99	0.44	0.52
مشهد Mashhad	0	0	223.05	454	167.03	350.58	0.43	0.51
تربت حیدریه Torbat-E Heydariyeh	0	0	786.2	581.33	685.32	505.85	0.39	0.47
درگز Daregaz	0	0	678.69	432.65	484.36	388.38	0.38	0.44
سبزوار Sabzevar	0	0	465.75	345.44	311.6	234.95	0.35	0.44
خلیل آباد Khalil Abad	0	0	247.96	141.19	221.54	125.68	0.33	0.37
قوچان Quchan	0	0	943.99	667.65	797.74	622.39	0.25	0.36
رشتخوار Roshtkhar	0	0	515.13	352.85	415.8	285.01	0.2	0.18
سرخس Sarakhsh	0	0	663.58	622.78	551.74	532.25	0.08	0.11

کلات نادر Kalat-E-Nader	0	0	464.9	609.82	413.58	450.16	0.06	0.1
خاف Khaf	0	0	619.28	331.28	534.82	315.66	0.02	0.1
میانگین Mean	0.0068	0.0526	492.21	395.13	408.09	331.5	0.4194	0.4668

توضیح: مقایسه هر دو مدل در شهرستان‌ها براساس معیارهای ارزیابی مقدار R^2 و d مدل رندوم‌فارست انجام گرفته است (مقادیر شاخص‌ها صعودی و ترتیب اهمیت شهرستان‌ها نزولی مرتب شده است).

Note: The comparison of both models in the counties is based on the evaluation metrics of R^2 and d for the Random Forest model (the index values are ordered in ascending order, and the counties are arranged in descending order of importance)

مدل در پیش‌بینی داده‌ها ست. رتبه‌بندی براساس ضریب تبیین و شاخص توافق کمک می‌کند که مدل‌هایی که به‌ترتیب بیشتری دقت پیش‌بینی را دارند و عملکرد خوبی در مقایسه با داده‌های واقعی دارند، در صدر قرار بگیرند. به‌منظور تصمیم‌گیری بهتر در خصوص استفاده از نوع مدل در هر شهرستان، در ستون آخر زیر عنوان مدل برتر، شهرستان‌ها براساس میانگین هر چهار معیار ارزیابی در هر دو الگوریتم، مرتب شدند. با این روش، با استفاده از میانگین مدل‌های دقیق‌تر و معتبرتر برای هر شهرستان، شهرستان‌ها از بهترین تا ضعیف‌ترین در ستون مجزا در جدول ۶ قرار گرفتند.

مقایسه و ارزیابی مدل‌های یادگیری ماشین در طبقه‌بندی عملکرد گندم با استفاده از معیارهای دقت و بازیابی

ابتدا داده‌های خام هر شهرستان به تفکیک سال خوانده شده و مدل‌های ذخیره‌شده با استفاده از کتابخانه جولیپ^۱ بارگذاری شدند. داده‌های آب‌وهوایی، اقلیمی و اکوفیزیولوژیک به‌عنوان ورودی مدل‌ها استفاده گردید و پس از آموزش، عملکرد مدل‌ها با شاخص‌های مختلف ارزیابی شد. در این مطالعه، دو الگوریتم یادگیری ماشین شامل رندوم‌فارست و ایکس‌جی‌بوست برای پیش‌بینی عملکرد گندم در سطح شهرستان‌ها توسعه داده شد. مقدار عملکرد ابتدا به‌صورت عددی با استفاده از رگرسیون پیش‌بینی گردید و سپس با استفاده از صدک‌های آماری به سه کلاس مجزا تقسیم شد:

متوسط‌های جذر میانگین مربعات خطا و میانگین قدر مطلق خطا در مدل برنده عموماً کمتر است. برای مثال، در شهرستان بردسکن، جذر میانگین مربعات خطا در مدل رندوم‌فارست برابر با $۱۸۷/۶$ و در ایکس‌جی‌بوست برابر با $۲۰۵/۴$ است که به‌وضوح اختلاف را نشان می‌دهد. در شهرستان‌هایی که ایکس‌جی‌بوست برتر بوده (مثل مشهد، رتبه ششم در برابر رتبه چهاردهم در رندوم‌فارست، مقدارهای آر ام اس ای تقریباً برابر و میانگین قدر مطلق خطا در مدل ایکس‌جی‌بوست اندکی کمتر است ($۴۱۰/۷$ در برابر ۴۵۴) که حاکی از پیش‌بینی کمتر خطا توسط ایکس‌جی‌بوست در داده‌های آن شهرستان است. شاخص توافق ویلموت که شباهت پیش‌بینی به مشاهدات را می‌سنجد، در مدل برنده (رندوم‌فارست) بالاتر است. در مه ولات مقدار d برای رندوم‌فارست برابر $۰/۶۹$ و برای ایکس‌جی‌بوست برابر $۰/۵۱$ است؛ این تفاوت نشان می‌دهد که رندوم‌فارست در آن ناحیه خطای پیش‌بینی کمتر داشته و الگوهای مشاهداتی را بهتر بازتولید می‌کند. در مناطقی که پراکندگی داده زیاد است (نقاطی با جذر میانگین مربعات خطای بالا شامل همه شهرستان‌ها به‌جز خاف، مشهد، تایباد، و سرخس) رندوم‌فارست معمولاً پایدارتر است و خطای کمتری ثبت می‌کند. ایکس‌جی‌بوست در مناطقی با توزیع پیچیده‌تر و نویز کمتر (مثل مشهد، درگز، تایباد و سرخس) توانست کمی بهتر یا برابر با رندوم‌فارست عمل کند. این تحلیل نشان می‌دهد که علی‌رغم پتانسیل بالای ایکس‌جی‌بوست در موارد خاص، مدل رندوم‌فارست در اغلب شهرستان‌ها نتایج بهتری از نظر دقت پیش‌بینی (R^2)، کاهش خطا ($RMSE/MAE$) و شباهت به مشاهدات (d Willmott's) ارائه کرده است. این تفاوت در عملکرد ممکن است ناشی از نحوه یادگیری و ویژگی‌های مدل‌ها باشد که در برخی موارد، رندوم‌فارست قادر است روابط پیچیده‌تری را شبیه‌سازی کند (Khodjaev et al., 2025). به‌منظور بهبود کلی، توصیه می‌شود بر روش‌های ترکیبی یا تنظیم دقیق پارامترها تمرکز شود تا بیش‌برازش در ایکس‌جی‌بوست و کم‌برازش احتمالی در رندوم‌فارست کاهش یابد.

مقادیر ضریب تبیین برابر با صفر، به‌طور طبیعی به‌دلیل ضعف

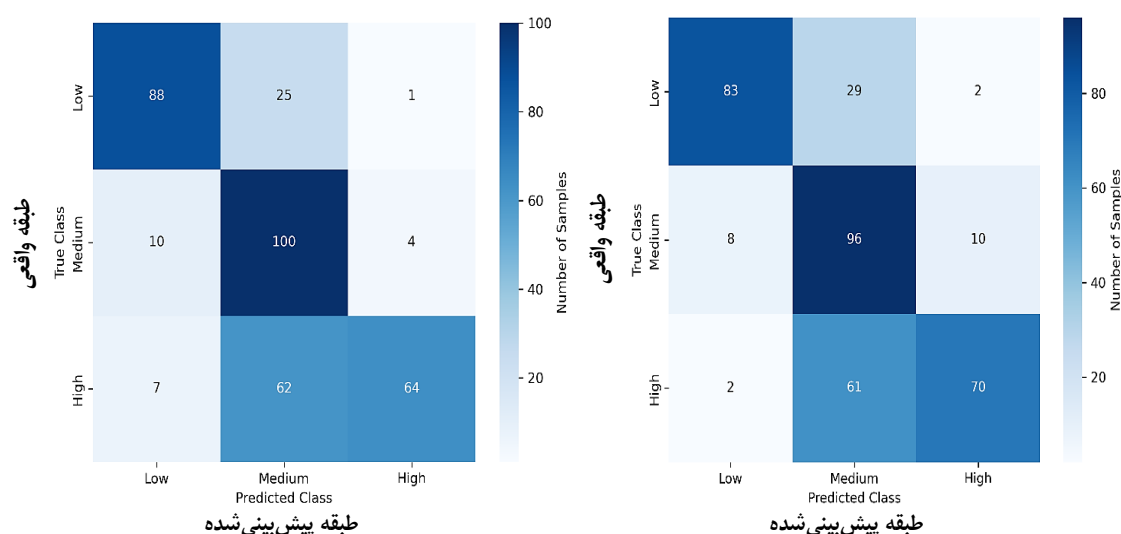
جدول ۶- رتبه‌بندی شهرستان‌ها از نظر دقت و توان مدل بهینه شده براساس معیارهای ارزیابی و تعیین اعتبار در دو مدل رندوم فارست، ایکس‌جی‌بوست و میانگین آن‌ها (رتبه در سمت چپ شهرستان آورده شده)

Table 6- Ranking of counties based on the accuracy and performance of the optimized model according to evaluation metrics and validation in the two models, Random Forest, XGBoost, and their average (rank listed on the left side of the county)

رتبه شهرستان در مدل رندوم فارست	رتبه شهرستان در مدل ایکس‌جی‌بوست	میانگین رتبه شهرستان‌ها
RF_rank_counties	XGB_rank_counties	Avg_rank_counties
خلیل آباد	خلیل آباد	خلیل آباد
1_Khalil Abad	2_Khalil Abad	1_Khalil Abad
بردسکن	بردسکن	بردسکن
2_Bardaskan	1_Bardaskan	2_Bardaskan
کاشمر	کاشمر	کاشمر
3_Kashmar	3_Kashmar	3_Kashmar
چناران	چناران	چناران
4_Chenaran	4_Chenaran	4_Chenaran
فریمان	فریمان	فریمان
5_Fariman	5_Fariman	5_Fariman
سبزوار	سبزوار	سبزوار
7_Sabzevar	8_Sabzevar	6_Sabzevar
خاف	خاف	خاف
10_Khaf	10_Khaf	7_Khaf
مشهد	مشهد	مشهد
12_Mashhad	6_Mashhad	8_Mashhad
تربت جام	تربت جام	تربت جام
8_Torbat-E-Jam	12_Torbat-E-Jam	9_Torbat-E-Jam
رشتخوار	رشتخوار	رشتخوار
9_Roshtkhar	13_Roshtkhar	10_Roshtkhar
مه ولات	مه ولات	مه ولات
11_Mahvelat	11_Mahvelat	11_Mahvelat
درگز	درگز	درگز
14_Daregaz	9_Daregaz	12_Daregaz
تایباد	تایباد	تایباد
15_Taybad	7_Taybad	13_Taybad
نیشابور	نیشابور	نیشابور
6_Neyshabur	15_Neyshabur	14_Neyshabur
گناباد	گناباد	گناباد
13_Gonabad	17_Gonabad	15_Gonabad
کلات نادر	کلات نادر	کلات نادر
16_Kalat-E-Nader	16_Kalat-E-Nader	16_Kalat-E-Nader
سرخس	سرخس	سرخس
18_Sarakhs	14_Sarakhs	17_Sarakhs
قوچان	قوچان	قوچان
19_Quchan	18_Quchan	18_Quchan
تربت حیدریه	تربت حیدریه	تربت حیدریه
17_Torbat-E Heydariyeh	19_Torbat-E Heydariyeh	19_Torbat-E Heydariyeh

ضعیف: (Low) صدک ۰ تا ۳۳؛ متوسط: (Medium) صدک ۳۳ تا ۶۶؛ خوب: (High) صدک ۶۶ تا ۱۰۰
این طبقه‌بندی براساس مقادیر واقعی و همچنین مقادیر پیش‌بینی شده برای هر شهرستان انجام شده است. نتایج این طبقه‌بندی در قالب ماتریس درهم‌ریختگی (confusion matrix) برای هر مدل ارائه گردید (شکل ۱۰).
در ماتریس ایکس‌جی‌بوست، مقادیر قطر اصلی (مقدارهای صحیح) بسیار بیشتر هستند و در نتیجه ماتریس متراکم‌تر روی قطر است. در ماتریس رندوم‌فارست، خطاهای بیشتری در خارج از قطر

اصلی دیده می‌شود که به معنی طبقه‌بندی اشتباه بالاتر است. مقایسه این دو مدل براساس امتیاز کلی مدل‌ها که نشان‌دهنده عملکرد کلی هر مدل در پیش‌بینی عملکرد گندم برای شهرستان‌های مختلف است، صورت می‌گیرد. این امتیاز به‌طور خاص به شبیه‌سازی عملکرد مدل در مجموع شهرستان‌ها توجه دارد و براساس معیارهای مشخصی مانند دقت کلی، دقت مثبت یا طبقه‌ای، بازیابی و امتیاز اف وان مدل در پیش‌بینی شهرستان‌های خاص محاسبه می‌شود (جدول ۷).



شکل ۱۰- ماتریس درهم‌ریختگی الگوریتم رندوم‌فارست (سمت راست) و ایکس‌جی‌بوست (سمت چپ) در طبقه‌بندی عملکرد گندم در استان خراسان رضوی

Figure 8- Confusion matrix of the Random Forest algorithm (right) and XGBoost algorithm (left) in classifying wheat yield in Khorasan Razavi province

جدول ۷- خلاصه دقت (accuracy)، دقت مثبت (precision)، بازیابی (recall) و امتیاز اف-وان برای دو الگوریتم طبقه‌بندی عملکرد گندم در استان خراسان رضوی

Table 7- Summary of accuracy, precision, recall, and F1 score for the two algorithms in classifying wheat yield in Khorasan Razavi province

مدل Model	دقت Accuracy (%)	دقت مثبت Precision (%)	بازیابی Recall (%)	F1-score (%)
رندوم‌فارست Random Forest	68.9	78.0	68.9	67.7
ایکس‌جی‌بوست XGBoost	69.8	73.8	69.8	66.8

بالاتری را نسبت به رندوم‌فارست ارائه داده است (۶۹/۸ درصد در برابر ۶۸/۹ درصد)، که نشان‌دهنده توانایی بیشتر این مدل در شناسایی صحیح کلاس‌های عملکرد می‌باشد. این نتیجه منطقی به نظر می‌رسد، زیرا ایکس‌جی‌بوست با بهره‌گیری از روش‌های بوستینگ،

اعتبارسنجی: هر دو مدل ایکس‌جی‌بوست و رندوم‌فارست در پیش‌بینی عملکرد شهرستان‌ها قابل اعتماد هستند، با این تفاوت که ایکس‌جی‌بوست به‌طور کلی کلاس‌های عملکرد را با دقت بالاتری تفکیک کرده است. با این حال، مدل ایکس‌جی‌بوست به‌طور کلی دقت

توانایی بیشتری در مدل سازی الگوهای پیچیده و اصلاح خطاهای مدل های ضعیف تر دارد.

مقایسه شاخص های کلیدی نیز تفاوت هایی را بین این دو مدل نشان می دهد:

مقایسه دقیق تری از معیارهای دقت، دقت طبقه ای و بازیابی برای هر مدل می تواند به درک بهتر نقاط قوت و ضعف هر مدل کمک کند. در این تحقیق، عملکرد بیشتر شهرستان ها با ایکس جی بوست پیش بینی شده اند، اما در برخی موارد خاص، رندوم فارست عملکرد بهتری نشان داده است که می تواند به نوع داده ها و ویژگی های آن ها بستگی داشته باشد.

۱. **دقت یا صحت:** ایکس جی بوست با دقت کلی ۶۹/۸ درصد عملکرد بهتری نسبت به رندوم فارست (۶۸/۹ درصد) دارد. فقط درصد درست پیش بینی شده ها را می سنجد و اگر داده ها نامتوازن باشند (مثلاً کلاس های زیاد با هم فرق داشته باشند)، می تواند گمراه کننده باشد.

۲. **دقت مثبت یا دقت طبقه ای:** برخلاف دقت کلی، در این شاخص، رندوم فارست عملکرد بهتری از خود نشان داده است (۷۸ درصد در برابر ۷۳/۸ درصد). این بدان معنا است که رندوم فارست نسبت به ایکس جی بوست در کاهش خطاهای مثبت کاذب^۱ موفق تر بوده است. دقت مثبت، دقت پیش بینی کلاس مثبت را نشان می دهد (چند درصد پیش بینی های مثبت درست بودند).

۳. **بازیابی:** درصد نمونه های مثبت واقعی که مدل به درستی شناسایی کرده است. این معیار نشان می دهد که مدل چه مقدار از کل نمونه های مثبت را شناسایی کرده است. به عبارت دیگر، بازیابی به طور خاص به درصدی از نمونه های مثبت واقعی که به درستی توسط مدل شناسایی شده اند، اشاره دارد. به عبارت ساده تر، بازیابی به مدل می گوید که از تمامی نمونه های که باید به کلاس مثبت تعلق داشته باشند، چه مقدار توسط مدل به درستی شناسایی شده اند.

فرمول بازیابی به صورت رابطه (۱) است:

$$Recall = TP / (TP + FN) \quad (1)$$

که در آن، P^2 : نشان دهنده تعداد نمونه هایی که مدل به درستی به عنوان مثبت شناسایی کرده است.

FN^3 تعداد نمونه هایی که مدل به اشتباه به عنوان منفی شناسایی کرده است، در حالی که باید مثبت می بودند. در نتیجه، بازیابی بالا به معنای این است که مدل به خوبی قادر است نمونه های مثبت را

شناسایی کند، اما ممکن است همراه با اشتباهاتی (FN) باشد. به این ترتیب، بازیابی می تواند در برخی موارد، اهمیت بیشتری از دقت بالاتر داشته باشد، به خصوص زمانی که اشتباه در شناسایی نمونه های مثبت پیامدهای جدی تری داشته باشد.

هر دو مدل در بازیابی کلاس های صحیح عملکرد مشابهی دارند، اما ایکس جی بوست با مقدار ۶۹/۸ درصد اندکی بهتر از رندوم فارست (۶۸/۹ درصد) عمل کرده است. این شاخص نشان دهنده توانایی مدل در شناسایی نمونه های واقعی هر کلاس است.

۴. **امتیاز اف-وان:** ترکیبی از دقت مثبت و بازیابی است که میانگینی متوازن از این دو معیار را می دهد و در مسائل کلاس بندی مفید واقع می شود. رندوم فارست با مقدار ۶۷/۷ درصد اندکی بهتر از ایکس جی بوست (۶۶/۸ درصد) بود که بیانگر تعادل بهتر بین دقت و بازیابی در این مدل است. بهترین شاخص است وقتی که داده ها نامتوازن باشند یا هم خطاهای مثبت و هم منفی اهمیت داشته باشند.

به طور کلی، اگرچه ایکس جی بوست از نظر دقت کلی کمی برتر است، اما رندوم فارست در دقت طبقه ای و امتیاز اف-وان عملکرد قابل توجهی دارد. این تفاوت ها می توانند ناشی از ویژگی های داده ها، نوع توزیع کلاس ها، یا حساسیت مدل ها نسبت به ویژگی های خاص باشند. بنابراین، انتخاب نهایی مدل باید با در نظر گرفتن هدف های مدل، نوع کاربرد و پیامدهای هر نوع خطا انجام شود. به عنوان مثال، در شرایطی که هدف کاهش هشدارهای اشتباه یا پیش بینی های نادرست مثبت است، مدلی با دقت بالاتر مانند رندوم فارست ترجیح داده می شود. اما اگر هدف شناسایی کامل تر تمامی موارد واقعی باشد، مدلی با بازیابی بالاتر مانند ایکس جی بوست می تواند مفیدتر باشد. افزون بر این، میزان پیچیدگی مدل، زمان آموزش و نیاز به تنظیم پارامترها نیز باید در تصمیم گیری لحاظ شود. رندوم فارست عموماً ساختار ساده تری دارد و به تنظیمات کمتری نیاز دارد، در حالی که اگرچه ایکس جی بوست پیچیده تر و زمان برتر است، اما پتانسیل بیشتری برای استخراج الگوهای دقیق تر در داده های پیچیده دارد. در نهایت، ارزیابی عملکرد مدل ها باید همیشه در بستر کاربرد واقعی و با توجه به نیازهای تصمیم گیری انجام گیرد، چرا که یک مدل از نظر آماری بهتر ممکن است در عمل کاربرپذیری کمتری داشته باشد.

در شکل ۱۱، اف وان-اسکور دو مدل برای تمام شهرستان ها نشان داده شده، تا تحلیل عمیق تر را ممکن سازد. مطابق این شکل، مدل رندوم فارست در پیش بینی عملکرد گندم شهرستان های بردسکن، چناران، درگز، گناباد، کلات نادر، کاشمر، خاف، خلیل آباد، مه ولات، مشهد، نیشابور، رشتخوار، سبزوار، سرخس و تربت جام بهتر از ایکس جی بوست عمل کرده است. مدل ایکس جی بوست در پیش بینی عملکرد گندم تنها دو شهرستان فریمان (با اختلاف سه کیلوگرم) و

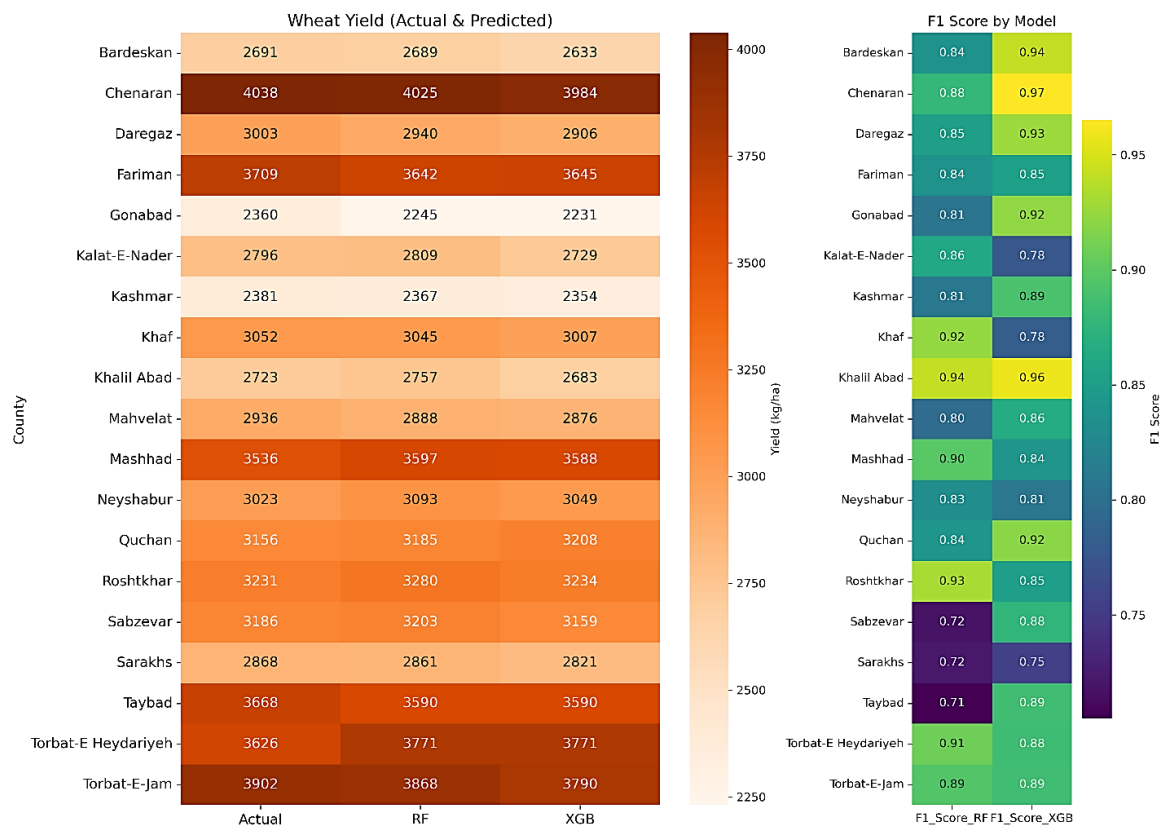
1- False Positives

2- True Positives

3- False Negatives

مدل بسیار نزدیک بوده‌اند و برای مدل ایکس‌جی‌بوست اندکی بیشتر بوده است، به عبارت دیگر پیش‌بینی‌های این مدل به مقادیر واقعی عملکرد نزدیک‌تر بودند.

قوچان بهتر از رندم فارست عمل کرد. برای دو شهرستان تربت حیدریه و تایباد پیش‌بینی دو مدل عیناً یکسان بود. مقایسه مقادیر اف-وان اسکور (شکل ۱۱) نشان می‌دهد که این مقادیر برای هر دو



شکل ۱۱- مقادیر عملکرد گندم واقعی و پیش‌بینی شده توسط رندوم فارست و ایکس‌جی‌بوست در استان خراسان رضوی به همراه امتیاز اف-وان برای هر مدل و شهرستان

Figure 9- Actual and predicted wheat yield values by Random Forest and XGBoost in Khorasan Razavi province, along with the F1 score for each model and county

اشتباهات «Medium→High» و «Low→High» توانست مرزهای میانی را دقیق‌تر تشخیص دهد؛ ولی در عوض اندکی اشتباهات شدید (High→Low) را افزایش داده است. به‌طور خاص، در مدل رندوم‌فارست موارد متعددی وجود داشت که شهرستان‌های با عملکرد ضعیف به‌اشتباه در کلاس متوسط یا حتی خوب قرار گرفتند. این موضوع در ایکس‌جی‌بوست با شدت کمتری مشاهده شد و نشان‌دهنده قدرت بیشتر این مدل در یادگیری الگوهای پیچیده‌تر از داده‌ها است. مدل رندوم‌فارست مناسب شرایطی است که داده‌ها کمی نویزدار یا پیچیده باشند، چون درخت‌های تصادفی می‌توانند روی ویژگی‌های متفاوت تمرکز کنند. مدل ایکس‌جی‌بوست به‌خاطر سازوکار بوستینگ، توانست بهتر الگوهای دقیق‌تر را یاد بگیرد و رکورد بهتری ثبت کند.

تحلیل خطا و محدودیت‌های دو الگوریتم در طبقه‌بندی عملکرد گندم

در هر دو مدل، برجسب‌های واقعی و پیش‌بینی شده در سه کلاس پایین، متوسط و بالا طبقه‌بندی شده‌اند. نتایج بیشتر به‌منظور مقایسه دو مدل در جدول ۸ آورده شده است. در رندوم‌فارست، بیشترین اشتباهات مربوط به جابه‌جایی از «High» به «Medium» و از «Low» به «Medium» است و تنها دو مرتبه برجسب «High» مستقیماً به «Low» یا برعکس تغییر یافت که نشان‌دهنده پراکندگی نسبتاً زیاد در مرزهای مجاور کلاس‌ها است. خطاهای طبقه‌بندی به نسبت زیاد دیده می‌شود. به‌ویژه در برخی سال‌ها، پیش‌بینی از Medium به Low یا از High به Medium تغییر کرده است. پراکندگی بین کلاس‌های پیش‌بینی شده نسبت به واقعیت زیاد است. ایکس‌جی‌بوست با بهبود جزئی در دقت کلی و کاهش چشمگیر

جدول ۸- تحلیل خطاهای دو الگوریتم رندوم فارست و ایکس جی بوست در پیش بینی طبقه عملکرد گندم در استان خراسان رضوی
Table 8- Error analysis of the two algorithms, Random Forest and XGBoost, in predicting wheat yield classification in Khorasan Razavi province

معیار / الگوی اشتباه Criteria/ Error pattern	رندوم فارست Random Forest	ایکس جی بوست XGBoost
دقت کلی Accuracy	68.98%	69.81%
میانگین امتیاز اف-وان F1 Score	0.678	0.668
میانگین دقت مثبت Precision	0.780	0.738
میانگین بازیابی Recall	0.690	0.698
تعداد اشتباه High → Medium Error	61	62
تعداد اشتباه Low → Medium Error	29	25
تعداد اشتباه Medium → High Error	10	4
تعداد اشتباه Medium → Low Error	8	10
تعداد اشتباه High → Low Error	2	7
تعداد اشتباه Low → High Error	2	1

- خوشه اول (جنوب و شرق استان؛ اقلیم بسیار خشک با تغییرات دمایی زیاد و دوره رشد بسیار گرم و طولانی) شهرستان های بردسکن، خلیل آباد، مهولات و سرخس قرار دارند که در همه این شهرستان ها مدل رندوم فارست نسبت به ایکس جی بوست عملکرد بهتری ارائه داده است.
- خوشه دوم (مرکز استان؛ اقلیم نیمه خشک تا نیمه مرطوب با تغییرات دمایی کم و دوره رشد خنک تر و کوتاه تر) شامل شش شهرستان چناران، فریمان، کلات نادر، مشهد، نیشابور و قوچان است که در چهار شهرستان چناران، کلات نادر، مشهد و نیشابور، رندوم فارست عملکرد برتر داشته است. در دو شهرستان فریمان و قوچان، ایکس جی بوست با اختلاف اندک (حدود سه کیلوگرم در فریمان) بهتر عمل کرده است.
- خوشه سوم (غرب و جنوب غرب استان؛ اقلیم نیمه خشک گرم با تغییرات دمایی متوسط و دوره رشد نسبتاً گرم و طولانی) شامل نه شهرستان درگز، گناباد، کاشمر، خاف، رشتخوار، سبزوار، تایباد، تربت جام و تربت حیدریه است که در هفت شهرستان (درگز، گناباد، کاشمر، خاف، رشتخوار، سبزوار و تربت جام)، رندوم فارست دقت بالاتری نسبت به ایکس جی بوست نشان داد. در دو شهرستان تایباد و تربت حیدریه، هر دو مدل دقیقاً نتایج یکسانی ارائه داده اند. با توجه به نتایج خوشه بندی ناحیه های اقلیمی (سه خوشه) و ادغام آن با کلاس بندی عملکرد دو مدل در پیش بینی عملکرد گندم، و همچنین توجه به این نکته که مدل رندوم فارست در ۸۰ درصد کل شهرستان ها پیش بینی دقیق تری از مدل ایکس جی بوست ارائه داد، این مدل به عنوان گزینه مناسب تر برای تحلیل های آتی در حوزه پیش بینی عملکرد گندم بر مبنای ناحیه اقلیمی پیشنهاد می شود.

نکات مهم و پیشنهاد: ایکس جی بوست: کاهش قابل توجه «Medium→High» و «Low→High»؛ اندکی دقت کلی بالاتر؛ افزایش اشتباهات شدید «High→Low». رندوم فارست: امتیاز اف-وان و دقت مثبت بالاتر؛ اشتباهات شدید کمتر؛ پراکندگی بیشتر در اشتباهات بین کلاس های مجاور. اگر اولویت، تفکیک دقیق مرزهای میانی (کاهش خطاهای جابه جایی مجاور) و حداکثر دقت کلی است، ایکس جی بوست مناسب تر است. اگر اولویت، دستیابی به امتیاز اف-وان و دقت مثبت بالاتر و کاهش اشتباهات شدید باشد، رندوم فارست گزینه بهتری است. نتایج حاکی از آن است که در طبقه بندی کلاس های عملکرد گندم، مدل رندوم فارست با مدل ایکس جی بوست تفاوت محسوسی ندارد، بنابراین بهبود عملکرد ایکس جی بوست به دلیل خود مدل است نه تغییر در توزیع داده ها. به طور کلی، اگر هدف طبقه بندی بهتر عملکرد گندم آبی در شرایط اقلیمی استان خراسان رضوی به سه طبقه (کلاس) باشد، ایکس جی بوست انتخاب بسیار بهتری نسبت به رندوم فارست است و در صورت وجود منابع محاسباتی کافی، به کارگیری آن توصیه می شود.

۳- تحلیل عملکرد مدل ها بر اساس خوشه های مکانی شهرستان ها

برای بررسی دقیق تر عملکرد مدل های یادگیری ماشین در مناطق مختلف، شهرستان ها بر اساس شباهت های مکانی و اقلیمی به سه خوشه تقسیم شدند (جدول ۹). این خوشه بندی به کمک الگوریتم های تحلیل خوشه ای و برپایه ویژگی های مکانی انجام شده است (مطالعات اقلیمی با استفاده از نقشه های جهانی به روز شده کوپن-گایگر (Peel et al., 2007) بر این تفکیک اقلیمی تأکید دارند):

جدول ۹- مدل برتر در پیش‌بینی عملکرد گندم در استان خراسان رضوی بر مبنای خوشه‌بندی شهرستان‌ها در ناحیه‌های اقلیمی سه گانه
Table 9- The best model for predicting wheat yield in Khorasan Razavi province based on the clustering of counties in the three climatic regions

شهرستان County	شماره خوشه Cluster No	مدل برتر Better model
نیشابور Neyshabur	2	رندوم فارست RF
چناران Chenaran	2	رندوم فارست RF
قوچان Quchan	2	ایکس جی بوست XGB
گناباد Gonabad	3	رندوم فارست RF
تربت جام Torbat-e-Jam	3	رندوم فارست RF
سبزوار Sabzevar	3	رندوم فارست RF
رشتخوار Roshtkhar	3	رندوم فارست RF
درگز Daregaz	3	رندوم فارست RF
بردسکن Bardaskan	1	رندوم فارست RF
خلیل آباد Khalil Abad	1	رندوم فارست RF
مه ولات Mahvelat	1	رندوم فارست RF
سرخس Sarakhs	1	رندوم فارست RF
فریمان Fariman	2	ایکس جی بوست XGB
کلات نادر Kalat-E-Nader	2	رندوم فارست RF
مشهد Mashhad	2	رندوم فارست RF
تربت حیدریه Torbat-E Heydarieh	3	Equal
تایباد Taybad	3	Equal
خاف Khaf	3	رندوم فارست RF
کاشمر Kashmar	3	رندوم فارست RF

۱: اقلیم بسیار خشک با تغییرات دمایی زیاد و دوره رشد بسیار گرم و طولانی.

۲: اقلیم نیمه‌خشک تا نیمه‌مرطوب با تغییرات دمایی کم و دوره رشد خنک‌تر و کوتاه‌تر.

۳: اقلیم نیمه‌خشک گرم با تغییرات دمایی متوسط و دوره رشد نسبتاً گرم و طولانی.

1: Hyper-arid climate with high temperature variations and a very hot and long growing season.

2: Semi-arid to semi-humid climate with low temperature variations and a cooler and shorter growing season.

3: Warm semi-arid climate with moderate temperature variations and a relatively warm and long growing season.

نتیجه گیری

رندوم فارست می تواند تصمیم گیری های بهتری فراهم کند.

سپاسگزاری

بخشی از هزینه های این تحقیق توسط معاونت محترم پژوهش و فناوری دانشگاه فردوسی مشهد، در قالب طرح پژوهشی به شماره ۲/۶۴۲۶۴ مصوب ۱۴۰۴/۰۲/۲۱ تأمین شده است که به این وسیله قدردانی می شود. از جناب آقای مهندس مهدی سلطانی، دانش آموخته کارشناسی ارشد اگرواکولوژی که بخش عمده داده های خام مورد نیاز را در اختیار قرار دادند، سپاسگزارم.

دو الگوریتم پس از بهینه سازی با رگرسیونهای مربوط به خود، تفاوت قابل توجهی از نظر توان و دقت پیش بینی نداشتند، با این حال با توجه به عملکرد اندکی دقیق تر رندوم فارست، این مدل می تواند گزینه ای مناسب تر برای استفاده در سیستم های تصمیم گیری در حوزه کشاورزی باشد، به ویژه در شرایطی که شناسایی شهرستان های با عملکرد بالا (برای سرمایه گذاری یا الگوبرداری) یا شهرستان های با عملکرد پایین (برای مداخلات حمایتی) اهمیت داشته باشد، مدل

References

1. Arun, G., and Ghimire, K. (2019). Estimating post-harvest loss at the farm level to enhance Food Security: A Case of Nepal. *International Journal of Agriculture, Environment and Food Sciences*, 3(September), 127–136. available from <https://doi.org/10.31015/jaefs.2019.3.3>
2. Aslan, M. F., Sabanci, K., & Aslan, B. (2024). Artificial intelligence techniques in crop yield estimation based on Sentinel-2 data: A comprehensive survey. *Sustainability*, 16(18), 8277. <https://doi.org/10.3390/su16188277>
3. Asseng, S., Ewert, F., Martre, P., Rötter, R. P., Lobell, D. B., Cammarano, D., Kimball, B.A., Ottman, M.J., Wall, G.W., White, J.W., Reynolds, M. P., Alderman, P.D., Prasad, P.V.V., Aggrawal, P.K., Anothai, J., Basso, B., Biernath, C., Challinor, A. J., De Sanctis, G., Doltra, J., Fereres, E., Garcia-Vila, M., Gayler, S., Hoogenboom, G., & Zhu, Y. (2015). Rising temperatures reduce global wheat production. *Nature Climate Change*, 5(2), 143–147. <https://doi.org/10.1038/nclimate2470>
4. Dhillon, M. S., Dahms, T., Kuebert-Flock, C., Rummler, T., Arnault, J., Steffan-Dewenter, I., & Ullmann, T. (2023). Integrating Random Forest and crop modeling improves the crop yield prediction of winter wheat and oil seed rape. *Frontiers in Remote Sensing*, 3, 1010978. <https://doi.org/10.3389/frsen.2022.1010978>
5. Everingham, Y., Sexton, J., Skocaj, D., & Inman-Bamber, G. (2016) Accurate prediction of sugarcane yield using a Random Forest algorithm. *Agronomy for Sustainable Development*, 36(2), 1–9. <https://doi.org/10.1007/s13593-016-0364-z>
6. Farhadi, M., Bannayan, M., Fallah, M. H., & Jahan, M. (2024). Identification of climatic and management factors influencing wheat's yield variability using AgMERRA dataset and DSSAT model across a temperate region. *Discover Life*, 54(8). <https://doi.org/10.1007/s11084-024-09651-8>
7. Gheysarbeigi, S., Pir Bavaghar, M., & Valipour, A. (2024). Forest aboveground biomass estimation using satellite imagery and Random Forest regression model. *Geography and Environmental Sustainability*, 14(1), 85-100. <https://doi.org/10.22126/GES.2024.9971.2715>
8. Hatfield, J. L., & Prueger, J. H. (2015). Temperature extremes: Effect on plant growth and development. *Weather and Climate Extremes*, 10, 4–10. <https://doi.org/10.1016/j.wace.2015.08.001>
9. IPCC. (2022). *Sixth Assessment Report*. Available Online at: https://www.ipcc.ch/site/assets/uploads/2022/04/AR6_Factsheet_April_2022.pdf
10. Javadi, A., Ghahremanzadeh, M., Sassi, M. Javanbakht, O., & Hayati, B. (2024). Impact of climate variables change on the yield of wheat and rice crops in Iran (Application of Stochastic Model based on Monte Carlo Simulation). (2024). *Computational Economics*, 63, 983–1000. <https://doi.org/10.1007/s10614-023-10389-0>
11. Jhajharia, K., Mathur, P., Jain, S., & Nijhawan, S. (2023). Crop yield prediction using machine learning and deep learning techniques. *Procedia Computer Science*, 218, 406-417. <https://doi.org/10.1016/j.procs.2023.01.023>
12. Khodabandehloo, E., Azadbakht, M., Radiom, S., Ashourloo, D., & Alimohammadi, A. (2021). Prediction of wheat fusarium head blight severity by using Random Forest. *Iranian Remote Sensing & GIS*, 13(4), 1-44. (in Persian with English abstract). <https://doi.org/10.52547/gisj.13.4.1>
13. Khodjaev, S., Bobojonov, I., Kuhn, L., & Glauben, T. (2025). Optimizing machine learning models for wheat yield estimation using a comprehensive UAV dataset. *Modeling Earth Systems and Environment*, 11, 15. <https://doi.org/10.1007/s40808-024-02188-9>
14. Kim, Y., & Kim, Y. (2022). Explainable heat-related mortality with random forest and SHapley Additive exPlanations (SHAP) models. *Sustainable Cities and Society*, 79, 103677. <https://doi.org/10.1016/j.scs.2022.103677>
15. Koocheki, A., Nassiri-Mahallati, M., Kamali, Gh., Shahandeh, Gh. (2006). Potential Impacts of Climate Change on Agroclimatic Indicators in Iran, *Arid Land Research and Management* 20(3):245-259. <https://doi.org/10.1080/15324980600705768>
16. Krishnadoss, N., & Ramasamy, L. K. (2024). Crop yield prediction with environmental and chemical variables

- using optimized ensemble predictive model in machine learning. *Environmental Research Communication*, 6(10), 101001. <https://doi.org/10.1088/2515-7620/ad7e81>
17. Monavar Sabegh, S., Zare Haghi, D., Samadianfard, S., Neishabouri, M. R., & Mikaeili, F. (2023). Estimation of daily reference evapotranspiration using Random Forest optimized by genetic algorithm. *Water and Soil Science*, 33(4), 33-53. <https://doi.org/10.22034/ws.2021.48756.2449>
18. Moreno Sánchez, J. C., Acosta Mesa, H. G., Trueba Espinosa, A., Ruiz Castilla, S., & García Lamont, F. (2025). Improving wheat yield prediction through variable selection using Support Vector Regression, Random Forest, and Extreme Gradient Boosting. *Smart Agricultural Technology*, 10, 100791. <https://doi.org/10.1016/j.atech.2025.100791>
19. Nayak, H. S., Silva, J. V., Parihar, C. M., Krupnik, T. J., Sena, D. R., Kakraliya, S. K., Jat, H. S., Sidhu, H. S., Sharma, P.C., Jat, M.L., & Sapkota, T.B. (2022). Interpretable machine learning methods to explain on-farm yield variability of high productivity wheat in Northwest India. *Field Crops. Research.*, 2022, 287, 108640. <https://doi.org/10.1016/j.fcr.2022.108640>
20. Oghnoum, M., Feghhi, J., Makhdoum, M., Moghaddamnia, A., & Etemad, V. (2019). Land capability evaluation of afforestation using Random Forest algorithm (Kan Watershed, Tehran). *Journal of Forest Research and Development*, 5(3), 387-403.
21. Pang, A., Chang, M. W., & Chen, Y. (2022). Evaluation of Random Forest for regional and local-scale wheat yield prediction in Southeast Australia. *Sensors*, 22, 717. <https://doi.org/10.3390/s22030717>
22. Peel, M. C., Finlayson, B. L., & McMahon, T. A. (2007). Updated world map of the Köppen–Geiger climate classification. *Hydrology and Earth System Sciences*, 11(5), 1633–1644. <https://doi.org/10.5194/hess-11-1633-2007>
23. Ray, D. K., Gerber, J. S., Macdonald, G. K., & West, P. C. (2015). Climate variation explains a third of global crop yield variability. *Nature Communications*, 6, 1–9. <https://doi.org/10.1038/ncomms6989>
24. Raymundo, R., Asseng, S., Robertson, R., Petsakos, A., Hoogenboom, G., Quiroz, R., Hareau, G., & Wolf, J., (2018). Climate change impact on global potato production. *European Journal of Agronomy*, 100, 87-98. <https://doi.org/10.1016/j.eja.2017.11.008>
25. Remman, S. B., Lekkas, A. M. (2021). *Robotic lever manipulation using hindsight experience replay and shapley additive explanations*. European Control Conference (ECC). Cornell University, USA. <https://doi.org/10.23919/ecc54610.2021.9654850>
26. Roell, Y. E., Beucher, A., Møller, P. G., Greve, M. B., & Greve, M. H. (2020). Comparing a Random Forest based prediction of winter wheat yield to historical yield potential. *Agronomy*, 10(3), 3. <https://doi.org/10.3390/agronomy10030395>
27. Sadeghi, M., & Ahmadi Nadoushan, M. (2021). Modeling soil nitrogen using Remote Sensing, Regression and Random Forest models. *Journal of Water and Soil Resources Conservation (WSRCJ)*, 11(2), 97-111.
28. Shen, Y., Mercatoris, B., Liu, Q., Yao, H., Li, Z., Chen, Z., & Wang, W. (2024). Use Self-Training Random Forest for Predicting Winter Wheat Yield. *Remote Sensing*, 16(24), 4723. <https://doi.org/10.3390/rs16244723>
29. Si, Z., Qin, A., Liang, Y., Duan, A., & Gao, Y. (2023). A review on regulation of irrigation management on wheat physiology, grain yield, and quality. *Plants*, 12(4), 692. <https://doi.org/10.3390/plants12040692>
30. Slafer, G. A., Savin, R., Sadras, V. O., & Calderini, D. F. (2023). Wheat yield improvement: Physiological and agronomic basis. *Field Crops Research*, 291, 108757.
31. Soleimannejad, L., Bonyad, A. E., Naghdi, R., & Latifi, H. (2018). Classification of quantitative attributes of Zagros forest using Landsat 8-OLI and Random Forest algorithm (Case study: Protected area of Manesht forests). *Journal of Forest Research and Development*, 4(4), 415-434.
32. Soltani, M., Jahan, M., & Yaghoubi, F. (2025). Evaluation of power and accuracy of AgMERRA and ERA5 dataset to simulate long term data for temperature and radiation in grat Khorasan province. (Under referee process)
33. Taiz, L., Zeiger, E., Moller, I. M., & Murphy, A. (2018). *Fundamentals of Plant Physiology*. New York, USA: Oxford University Press. ISBN 978160535790
34. Ting, K. M. (2011). Confusion Matrix. In C. Sammut & G. I. Webb, (eds) *Encyclopedia of Machine Learning*. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_157
35. Vieira, H. V., Bradford, B. Z., Osterholzer, A., Pierce, E. S., Cockrell, D., Peairs, F., Frost, K., Groves, R., & Nachappa, P. (2025). A new growing degree-day phenology model for wheat stem sawfly (Hymenoptera: Cephidae) in Colorado wheat fields. *Plos One*, 20(4), e0320497. <https://doi.org/10.1371/journal.pone.0320497>
36. Wang, Z., & Li, S. (2002). Effects of water deficit and supplemental irrigation at different growing stages on uptake and distribution of nitrogen, phosphorus, and potassium in winter wheat. *Journal of Plant Nutrition and Fertilizers*, 8(3), 265–270. <https://dx.doi.org/10.11674/zwzf.2002.0302>
37. Yaghoubi, F., Bannayanm, M., & Asadi, G. (2020). Performance of predicted evapotranspiration and yield of rainfed wheat in the northeast Iran using gridded AgMERRA weather data. *International Journal of Biometeorology*, 64, 1519–1537. <https://doi.org/10.1007/s00484-020-01931-y>