

Wheat yield prediction for multiple cultivars based on agroclimatic factors

Hoceme Degaichia^{1,§}

Touati Bakria¹

Ahcène Hakem¹

1 Centro de Investigación en Agropastoralismo (CRAPast). Djelfa, Argelia.

Autor para correspondencia: hoceme.degaichia@crapast.dz.

Abstract

This study aimed to evaluate the performance of four durum wheat cultivars Odysseo, Saragola, Irid and Maestrале using two machine learning techniques: classification and regression trees and random trees. Classification tree and regression analysis showed that mean annual temperature is the dominant factor influencing yield in all cultivars. For the Saragola, Irid and Maestrале cultivars, yield increased significantly when the mean annual temperature exceeded 17.25 °C, particularly when emergence density was optimal. In contrast, the Odysseo cultivar showed sensitivity to both average annual temperature and seeds per spike, with higher yields associated with an average annual temperature above 17.25 °C and seeds per spike above 33.6. The random tree analysis confirmed the importance of average annual temperature and emergence density, highlighting their strong predictive power. The models provided greater robustness and generalizability by reducing prediction variance, making them reliable tools for yield prediction. These findings highlight cultivar-specific responses to agroclimatic conditions, with Odysseo influenced by both mean annual temperature and seeds per ear, while Saragola, Irid and Maestrале demonstrate a critical interaction between mean annual temperature and emergence density. Integrating random tree models improves prediction accuracy and provides valuable information for developing precision agriculture strategies tailored to environmental conditions.

Keywords:

Triticum durum, decision tree analysis, machine learning, machine learning.



Introduction

Wheat (*Triticum durum*) is a staple crop of global importance, with its production significantly influenced by agroclimatic factors (Martínez-Moreno *et al.*, 2022). Understanding the relationship between environmental conditions and yield is essential for improving productivity and ensuring food security, especially in the face of climate variability (Shewry *et al.*, 2015). Key factors such as annual average temperature (AAT), precipitation, plant density, and seed characteristics play a crucial role in determining wheat yield (Kang *et al.*, 2020).

Traditional statistical methods, such as linear regression and generalized linear models, have been widely used to predict crop yields. However, these approaches often fall short in capturing complex, nonlinear relationships between multiple variables (Sharma *et al.*, 2021). Recent advances in machine learning (ML) provide more robust and adaptable models for analyzing such interactions. Decision tree-based models, including Classification and Regression Trees (C RT) and Random Forests (RF), are particularly well-suited for agricultural applications due to their ability to handle nonlinear relationships and rank variable importance (Breiman, 2001; Sarker *et al.*, 2020).

Despite the increasing use of ML models in agriculture, limited studies have focused on the comparative performance of C RT and RF for predicting wheat yield across multiple varieties. This study aims to address this gap by evaluating the predictive accuracy of these models for four wheat varieties, identifying the most influential agroclimatic factors, and establishing decision rules for yield optimization.

Material and methods

Source material and experimental treatments

Four durum wheat cultivars (Odysseo, Saragola, Irid and Maestrale) were selected for this study based on their agronomic performance and adaptability. These cultivars are commercially recognized for their high yield potential, grain quality and stress tolerance (De Vita *et al.*, 2007; Kabbaj *et al.*, 2017). Field experiments were conducted during the 2020 growing season across three different agroclimatic zones in Algeria: Annaba (Annaba), Coastal region with a humid Mediterranean climate; Ouled Rahmoune (Constantine), Semi-arid region with moderate rainfall; Oued Zenati (Guelma), Dry region with limited water availability.

Each experimental site covered an area of 2 500 m², and the trials were conducted using a randomized complete block design (RCBD) with three replications per cultivar. A seeding rate of 200 kg ha⁻¹ was employed to achieve adequate plant density, promoting uniform emergence and crop establishment. Basal fertilization was carried out using monoammonium phosphate (MAP) applied at a rate of 150 kg ha⁻¹ to provide essential nutrients for early growth. Additionally, crop protection measures included the application of fungicidal treatments such as Celest Xtra and Amistar Xtra, along with Acil, to safeguard the wheat plants against potential diseases and enhance crop performance.

Data collection

Agroclimatic and agronomic data were collected throughout the growing season, including annual average temperature (AAT) (°C), altitude, annual total precipitation (ATP) (mm), seeds per spike (count), emergence density (plants m⁻²), spike m⁻² (count), tiller per plant (count), thousand-kernel weight TKW (g), and practical wheat yield (q ha⁻¹), used as the target variable. Meteorological data were obtained from the National Meteorological Office (Algeria), while agronomic parameters were measured following standardized field and laboratory procedures (Blum, 2011; Joia *et al.*, 2025).

Predictive modeling approaches

Two machine-learning approaches were applied using IBM SPSS Modeler 18.0 to predict wheat yield: Classification and regression trees (C RT), a decision tree-based model that partitions data into homogeneous subsets based on the most significant variables (Breiman *et al.*, 1984); and Random trees regression (RT), an ensemble learning method that enhances predictive accuracy by averaging multiple decision trees (Liaw and Wiener, 2002). Model performance was assessed using root square error (RMSE), relative error (RE) and explained variance (EV) (Chlingaryan *et al.*, 2018).

Variable importance and decision tree interpretation

Feature importance was evaluated using Gini impurity (C RT) and permutation importance (RT). The generated decision trees were analyzed for each wheat cultivar to identify key thresholds influencing wheat yield (Hastie *et al.*, 2009).

Results

Agronomic performance of durum wheat cultivars

Results highlight significant variability in the agronomic performance of the four durum wheat cultivars across three distinct localities (Table 1). This variability is primarily attributed to environmental factors, particularly climatic conditions and agronomic practices, which are known to influence growth, yield and phenotypic traits of wheat cultivars (Kabbaj *et al.*, 2017; Royo *et al.*, 2020).

Table 1. Agronomic performance of durum wheat cultivars.

Locality	Cultivar	Emergence density (plants m ⁻²)	Tillers per plant	Spikes m ⁻²	Seeds per spike	TKW (g)	Practical Yield (q ha ⁻¹)
Annaba (Annaba)	Ir#d	280.5 ±6.36	3 ±0	392 ±4.24	35.4 ±1.56	47.5 ±0.71	51.5 ±4.95
	Maestrale	269.5 ±6.36	3 ±0	380 ±7.07	35.1 ±2.69	48 ±1.41	51.5 ±3.54
	Odysseo	282 ±0	3 ±0	395 ±0	33.6 ±3.39	49.5 ±0.71	52.5 ±4.95
	Saragola	276 ±4.24	3 ±0	387 ±7.07	33.5 ±2.12	48.5 ±0.71	54 ±5.66
Ouled	Ir#d	287 ±0	3.8 ±0	563 ±0	38.5 ±0	51 ±0	35 ±3.25
Rahmoune (Canstantine)	Maestrale	292 ±0	4.6 ±0	612 ±0	39.65 ±0	50.25 ±0	35 ±3.75
	Odysseo	284 ±0	3.75 ±0	526 ±0	37.75 ±0	49.6 ±0	28 ±2.4
	Saragola	278 ±0	3.86 ±0	535 ±0	36.5 ±0	49.85 ±0	27.9 ±3.39
Oued Znati	Ir#d	298 ±0	5 ±0	665 ±0	35.75 ±0	49.7 ±0	31.5 ±3.96
(Guelma)	Maestrale	297 ±0	5 ±0	789 ±0	36.5 ±0	48.3 ±0	40 ±4.81
	Odysseo	289 ±0	6 ±0	703 ±0	39.5 ±0	48.3 ±0	40 ±3.68
	Saragola	292 ±0	5 ±0	664 ±0	37.25 ±0	51 ±0	41.5 ±4.67

In Annaba, practical yields were highest for Saragola (54 ±5.66 q ha⁻¹), which is consistent with findings from previous studies indicating that this cultivar exhibits good adaptation to moderate conditions, particularly when temperature and soil moisture are adequate (Csépl# *et al.*, 2024). The TKW values for Odysseo (49.5 ±0.71 g) and IRID (47.5 ±0.71 g) suggest good grain filling potential, which is a desirable trait for yield improvement (Maccaferri *et al.*, 2011).

The Ouled Rahmoune locality demonstrated increased tillering and spike density across all cultivars, with Maestrale achieving the highest emergence density (292 ±0 plants m⁻²) and tillering rate (4.6 ±0 tillers per plant). This phenomenon can be attributed to favorable soil conditions that likely promoted tiller formation and spike emergence, as supported by Kabbaj *et al.* (2017), who

reported that improved soil fertility enhances tiller production and consequently increases yield. However, practical yields were lower compared to Annaba, with Odysseo and Saragola recording the lowest yields ($28 \pm 2.4 \text{ q ha}^{-1}$ and $27.9 \pm 3.39 \text{ q ha}^{-1}$, respectively). This suggests that yield potential may not solely depend on spike density but also on grain filling efficiency, which may have been compromised by suboptimal climatic conditions during the grain-filling period (Royo *et al.*, 2020).

Oued Znati exhibited the highest overall productivity, particularly for the Saragola cultivar, which achieved a practical yield of $41.5 \pm 4.67 \text{ q ha}^{-1}$ with a TKW of $51 \pm 0 \text{ g}$. This locality also demonstrated superior tillering ability and spike density for all cultivars, with Odysseo reaching $703 \pm 0 \text{ spikes m}^{-2}$ and 6 ± 0 tillers per plant. Moreover, the high TKW values observed in this locality are indicative of favorable conditions for grain filling, a critical determinant of yield (Kabbaj *et al.*, 2017).

Performance of the predictive model

The random trees regression model exhibited a strong predictive capability for wheat yield estimation, with an explained variance of 70.4%, suggesting that the selected agroclimatic and agronomic variables account for a substantial proportion of yield variability. The root mean square error (RMSE) was 7.395, indicating a moderate level of deviation between predicted and observed values. Furthermore, the relative error of 0.296 suggests a fairly reliable model performance (Table 2).

Table 2. Performance metrics of the random trees regression model for practical wheat yield prediction.

Model parameters	Input
Target variable	Practical wheat yield
Model generation method	Random Trees Regression
Number of predictor #nputs	7
Root mean square error (RMSE)	7.395
Relative error (RE)	0.296
Explained variance (EV)	0.704

These results demonstrate the robustness of machine learning techniques in agricultural yield prediction, aligning with previous studies highlighting the effectiveness of decision tree-based models for predicting crop responses to environmental factors (Chlingaryan *et al.*, 2018; López-Granados *et al.*, 2020).

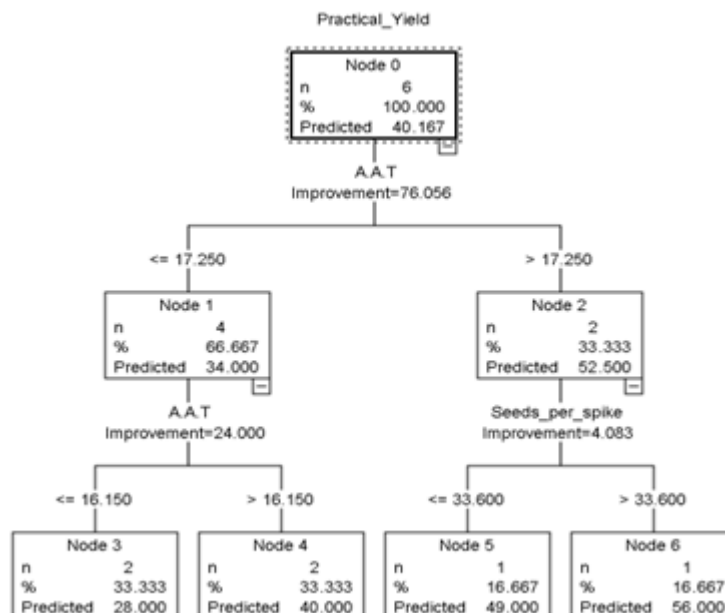
Agroclimatic and agronomic factors affecting wheat yield

The C RT analysis revealed that the AAT was the dominant variable influencing yield for the Saragola, Irid, and Maestrale cultivars, with emergence density also playing a significant role. In contrast, for the Odysseo cultivar, yield was mainly influenced by AAT and the number of seeds per spike.

Odysseo cultivar

For Odysseo, the C RT decision tree identified AAT as the primary determinant of yield variation. When $\text{AAT} \leq 16.15 \text{ }^{\circ}\text{C}$, the average yield was 28 q ha^{-1} , representing a significant reduction due to suboptimal temperature conditions. For an AAT between $16.15 \text{ }^{\circ}\text{C}$ and $17.25 \text{ }^{\circ}\text{C}$, yield increased to 34 q ha^{-1} , showing a positive impact of higher temperatures on grain development. When the AAT exceeded $17.25 \text{ }^{\circ}\text{C}$, yield reached 52.5 q ha^{-1} , if the seeds per spike exceeded 33.6. These findings suggest that Odysseo cultivar responds favourably to warmer temperatures, with yield improving as AAT increases above $17.25 \text{ }^{\circ}\text{C}$. The critical role of seed density further highlights the importance of optimizing spike fertility under varying temperature regimes (Figure 1).

Figure 1. Regression tree analysis for predicting wheat yield in Odysseo cultivar.

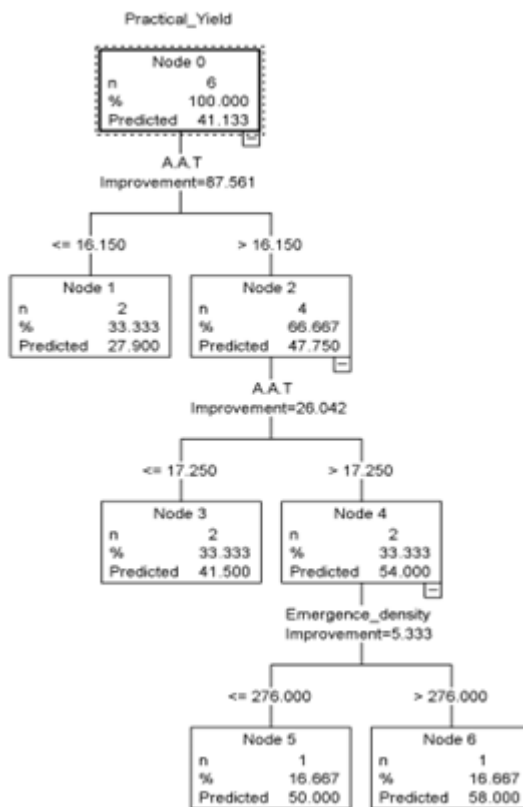


Saragola cultivar

For Saragola, yield was highly sensitive to AAT and emergence density. When AAT was below 16.15 °C, yield dropped to 38 q ha⁻¹, indicating a negative impact of lower temperatures on grain filling. When AAT exceeded 16.15 °C and emergence density was optimal, yield increased to 51 q ha⁻¹, demonstrating the combined effect of temperature and agronomic management on productivity. These results highlight that Saragola cultivar is less tolerant to low temperatures, requiring warmer conditions for optimal yield expression. This aligns with previous reports on durum wheat varieties that show reduced grain development under cooler climates (Ferrise *et al.*, 2019) (Figure 2).



Figure 2. Regression tree analysis for predicting wheat yield in Saragola cultivar.

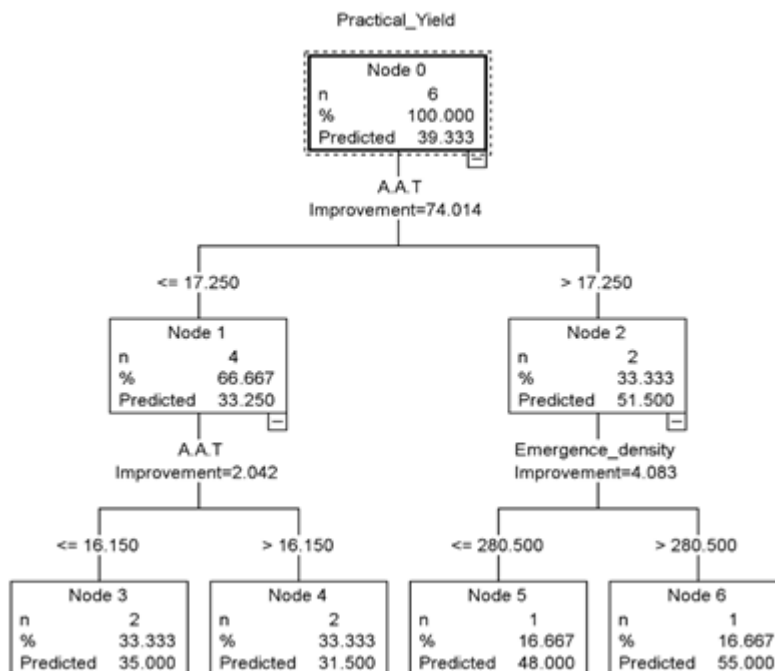


Irid cultivar

The Irid cultivar decision tree model identified AAT and emergence density as the key yield determinants. When AAT was below 17.25 °C, yield remained low, suggesting that Irid cultivar requires higher temperatures for grain development. When AAT exceeded 17.25 °C, yield increased significantly, particularly when the plant density was high. This behaviour indicates that Irid cultivar benefits from higher temperatures, but plant density also plays a crucial role in achieving high productivity. This finding is consistent with studies emphasizing the role of grain weight as a primary yield component in wheat (Lobell *et al.*, 2017) (Figure 3).



Figure 3. Regression tree analysis for predicting wheat yield in Irid cultivar.

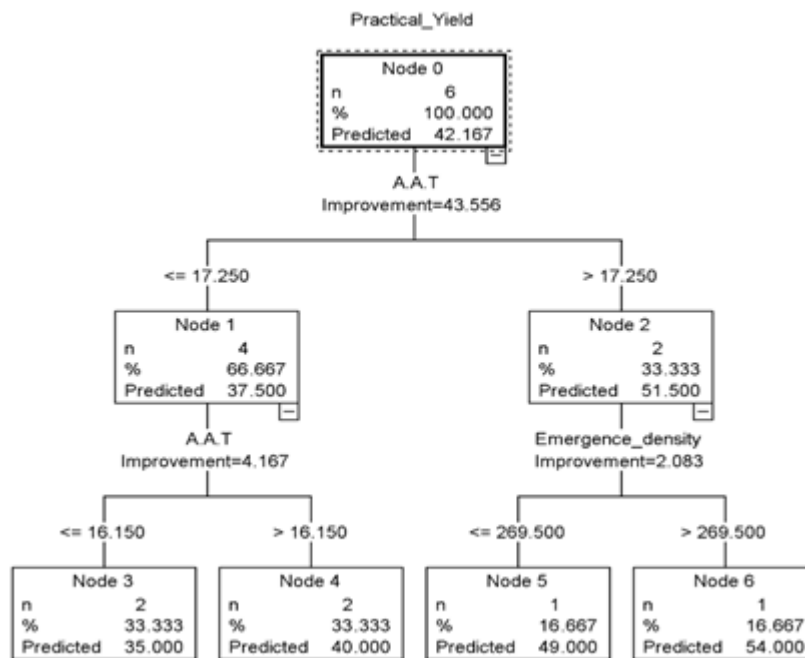


Maestrale cultivar

The C RT analysis for Maestrale cultivar indicated a strong dependency on AAT and emergence density. Yield remained low when AAT was below 17.25 °C, likely due to poor grain filling conditions. When AAT exceeded 17.25 °C, yield improved significantly, provided that emergence density was optimal. These results suggest that Maestrale requires both warm temperatures and adequate emergence density for optimal productivity. The interplay between temperature and plant density is well documented in wheat physiology, where poor emergence density can exacerbate the negative effects of suboptimal temperatures (Trnka *et al.*, 2021) (Figure 4).



Figure 4. Regression tree analysis for predicting wheat yield in Maestale cultivar.



Saragola and Irid cultivars are highly dependent on AAT and emergence density, with yield improving significantly, when AAT exceeds 16.15 °C and 17.25 °C, respectively, and emergence density is optimal. Maestrale cultivar exhibits similar behaviour to Irid, with yield enhancement linked to AAT exceeding 17.25 °C and favourable emergence density.

Conclusions

Odysseo cultivar demonstrated greater resilience to temperature fluctuations, particularly benefiting from higher temperatures. However, its productivity is strongly dependent on high emergence density, indicating the importance of dense and uniform sowing, especially in warmer regions.

Saragola, Irid, and Maestrale cultivars showed increased sensitivity to both temperature and emergence density, implying that these varieties require more precise seed rate calibration and adapted sowing schedules under changing climatic conditions to avoid yield penalties. Practical recommendations for improving wheat crop management include: 1) tailoring sowing density by cultivar and expected temperature regime: adopt higher seed rates for Odysseo in warm zones and fine-tune densities for other cultivars based on predictive emergence models; 2) integrating real-time agroclimatic data to adjust management practices, particularly in terms of sowing date and field preparation; and 3) employing site-specific management zones using decision rules derived from the models to optimize inputs (fertilizer, irrigation) where they will have the greatest effect on yield.

The comparison between C RT and RF models revealed their complementary strengths in predicting wheat yield and developing actionable decision rules. RF provided robust and generalizable insights due to its ensemble nature, making it a valuable tool for data-driven agronomic decision-making. Given the promising results, artificial intelligence (AI) tools, especially those based on machine learning and ensemble learning algorithms, offer significant potential for refining yield predictions and supporting adaptive agronomic decisions. AI-driven systems can dynamically integrate multi-source data (satellite, sensor, weather forecasts) to provide real-time, site-specific recommendations, fostering the transition toward precision agriculture and climate-resilient wheat production.

Acknowledgments

This research was funded by the General Directorate of Scientific Research and Technological Development (DGRSDT, for its French acronym)-Algeria and CRAPast, Algeria'. The authors express their gratitude to the Directorate General for Scientific Research and Technological Development (DGRSDT, for its acronym in French-Algeria) and to CRAPAST, Algeria.

Bibliography

- 1 Blum, A. 2011. Plant breeding for water-limited environments. Springer Science Business Media. ISBN: 978-1-4419-7490-7. 68-69 pp.
- 2 Breiman, L. 2001. Random Forests. Machine Learning. 45(1):5-32.
- 3 Chlingaryan, A.; Sukkarieh, S. and Whelan, B. 2018. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. Computers and Electronics in Agriculture. 151(1):61-69.
- 4 Csépl#, M.; Puskás, K.; Vida, G.; Mészáros, K.; Uhrin, A.; Tóth, V.; Ambrózy, Z.; Grausgruber, H.; Bonfiglioli, L.; Pagnotta, M. A.; Urbanavi####, I.; Mikó, P. and Bányai, J. 2024. Performance of a durum wheat diversity panel under different management systems. Cereal Research Communications. 52(1):489-502.
- 5 De Vita, P.; Mastrangelo, A. M.; Matteu, L.; Mazzucotelli, E. and Cattivelli, L. 2007. Genetic improvement effects on yield stability in durum wheat genotypes grown in Italy. Field Crops Research. 100(1):133-141.
- 6 Ferrise, R.; Moriondo, M. and Bindi, M. 2019. Climate change impacts on wheat and maize production in Europe and adaptation options. Agricultural Systems. 175(1):112-124.
- 7 Hastie, T.; Tibshirani, R. and Friedman, J. 2009. The elements of statistical learning: data mining, #nference, and prediction 2nd Ed. Springer Science Business Media. 745-763 pp.
- 8 Joia, S.; Siddiqi, R. A. and Singh, T. P. 2025. Study of physicochemical properties, amino acid composition and fatty acid profile of bitter himalayan wild apricot kernel: a step towards its valorization. Waste Biomass Valor. 16(1):1-12.
- 9 Kabbaj, H.; Sall, A. T.; Al-Abdallat, A.; Geleta, M.; Amri, A.; Filali-Maltouf, A.; Belkadi, B.; Ortiz, R. and Bassi, F. M. 2017. Genetic diversity within a global panel of durum wheat (*Triticum durum*) landraces and modern germplasm reveals the history of alleles exchange. Frontiers in Plant Science. 8(1):1-13.
- 10 Kang, Y.; Khan, S. and Ma, X. 2020. Climate change impacts on crop yield, crop water productivity and food security. A review. Progress in Natural Science: Materials International. 19(2):166-173.
- 11 Liaw, A. and Wiener, M. 2002. Classification and regression by random Forest. R News. 2(3):18-22.
- 12 Lobell, D. B. and Asseng, S. 2017. Comparing estimates of climate change impacts from process-based and statistical crop models. Environmental Research Letters. 12(1):1-12.
- 13 López-Granados, F.; Jurado-Expósito, M.; Peña, J. M. and Serrano, N. 2020. Precision agriculture for weed management. Weed Science. 68(3):171-190.
- 14 Maccaferri, M.; Sanguineti, M. C.; Demontis, A.; El-Ahmed, A.; Garcia-Moral, L.; Maalouf, F.; Nachit, M.; Nserallah, N.; Ouabbou, H.; Rhouma, S.; Royo, C.; Villegas, D. and Tuberosa, R. 2011. Association mapping in durum wheat grown across a broad range of water regimes. Journal of Experimental Botany. 62(2):409-438.
- 15 Martínez-Moreno, F.; Ammar, K. and Solís, I. 2022. Global changes in cultivated area and breeding activities of durum wheat from 1800 to date: a historical review. Agronomy. 12(5):1-17.

- 16 Royo, C.; Soriano, J. M. and Villegas, D. 2020. Assessing the adaptation of durum wheat genotypes to Mediterranean environments. *Euphytica*. 216(5):1-14.
- 17 Sarker, I. H.; El-Gayar, O. and Badar, M. A. 2020. Machine Learning: a review of applications in agriculture. *Agronomy*. 10(4), 563-582.
- 18 Sharma, R.; Shukla, S. and Shukla, A. K. 2021. Machine learning in agriculture: A review. *Journal of Applied and Natural Science*. 13(1):64-71.
- 19 Shewry, P. R. and Hey, S. J. 2015. The contribution of wheat to human diet and health. *Food and Energy Security*. 4(3):178-202.
- 20 Trnka, M.; Rötter, R. P.; Ruiz-Ramos, M.; Kersebaum, K. C.; Olesen, J. E.; Žalud, Z. and Semenov, M. A. 2021. Adverse weather conditions for European wheat production will become more frequent with climate change. *Nature Climate Change*. 11(7):675-680.



Predicción del rendimiento del trigo para múltiples cultivares basada en factores agroclimáticos

Journal Information
Journal ID (publisher-id): remexca
Title: Revista mexicana de ciencias agrícolas
Abbreviated Title: Rev. Mex. Cienc. Agríc
ISSN (print): 2007-0934
Publisher: Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias

Article/Issue Information
Date received: 01 October 2025
Date accepted: 01 January 2026
Publication date: 01 January 2026
Publication date: Jan-Feb 2026
Volume: 17
Issue: 1
Electronic Location Identifier: e3892
DOI: 10.29312/remexca.v17i1.3892

Categories

Subject: Artículos

Keywords:

Keywords:

Triticum durum

decision tree analysis

machine learning

precision agriculture

Counts

Figures: 4

Tables: 2

Equations: 0

References: 20