

<https://doi.org/10.69639/arandu.v12i3.1591>

Identificación de patrones de desempeño agronómico en el cultivo de arroz mediante un enfoque multivariante de aprendizaje automático no supervisado

Identification of agronomic performance patterns in rice cultivation using a multivariate unsupervised machine learning approach

Carlos Arturo Carvajal Chávez

ccarvajal@uagraria.edu.ec

<https://orcid.org/0000-0002-2781-6953>

Universidad Agraria Del Ecuador

Johanna Elisabeth Duchimaza Supliguicha

jduchimaza@uagraria.edu.ec

<https://orcid.org/0000-0002-9662-2212>

Universidad Agraria Del Ecuador

Erik Rolando Cedeño Anchundia

ecedennoa2@unemi.edu.ec

<https://orcid.org/0000-0003-4986-7652>

Universidad Estatal De Milagro

Artículo recibido: 18 agosto 2025 - Aceptado para publicación: 28 septiembre 2025

Conflictos de intereses: Ninguno que declarar.

RESUMEN

El arroz (*Oryza sativa* L.) constituye un cultivo esencial para la seguridad alimentaria mundial, al aportar alrededor del 21 % de la ingesta calórica diaria y sostener la dieta de más de dos tercios de la población global. No obstante, su productividad se ve amenazada por factores bióticos como enfermedades fúngicas (*Rhizoctonia solani*), plagas, y abióticos asociados al cambio climático, tales como sequías, inundaciones y salinidad de los suelos. Estas limitaciones, sumadas a la dependencia de agroquímicos, evidencian la necesidad de innovar en la evaluación del desempeño agronómico. Los métodos tradicionales, basados en análisis univariantes, muestran restricciones para capturar la complejidad de las interacciones genéticas, ambientales y de manejo. Frente a ello, las técnicas multivariantes y el aprendizaje automático no supervisado representan herramientas robustas para identificar patrones ocultos y clasificar genotipos en función de rendimiento, resistencia y eficiencia. En este estudio, se aplicaron análisis de componentes principales (PCA), biplots y clusterización en datos de germinación y desarrollo inicial de arroz en la provincia del Guayas, Ecuador. Los resultados mostraron correlaciones significativas entre variables fisiológicas y tres clústeres diferenciados de desempeño. Se confirma así que el uso de enfoques no supervisados constituye una estrategia eficaz para la selección y mejoramiento del arroz en escenarios de sostenibilidad y seguridad alimentaria.

Palabras clave: oryza sativa, desempeño agronómico, seguridad alimentaria, aprendizaje automático, clusterización

ABSTRACT

Rice (*Oryza sativa* L.) constitutes an essential crop for global food security, contributing around 21% of daily caloric intake and sustaining the diet of more than two-thirds of the world's population. However, its productivity is threatened by biotic factors such as fungal diseases (*Rhizoctonia solani*), pests, and abiotic factors associated with climate change, including droughts, floods, and soil salinity. These limitations, combined with the dependence on agrochemicals, highlight the need to innovate in the evaluation of agronomic performance. Traditional methods, based on univariate analyses, show limitations in capturing the complexity of genetic, environmental, and management interactions. In contrast, multivariate techniques and unsupervised machine learning represent robust tools to identify hidden patterns and classify genotypes according to yield, resistance, and efficiency. In this study, principal component analysis (PCA), biplots, and clustering were applied to germination and early development data of rice in the province of Guayas, Ecuador. The results showed significant correlations among physiological variables and three differentiated clusters of performance. This confirms that the use of unsupervised approaches constitutes an effective strategy for rice selection and improvement in scenarios of sustainability and food security.

Keywords: *Oryza sativa*, agronomic performance, food security, machine learning, clustering

Todo el contenido de la Revista Científica Internacional Arandu UTIC publicado en este sitio está disponible bajo licencia Creative Commons Attribution 4.0 International. 

INTRODUCCIÓN

El arroz (*Oryza sativa* L.) constituye uno de los cultivos más importantes a nivel mundial, no solo por su relevancia en la dieta de millones de personas, sino también por su rol estratégico en la seguridad alimentaria y en el desarrollo económico de numerosos países productores. Se estima que más de dos tercios de la población global dependen del arroz como alimento básico, siendo responsable de aproximadamente el 21 % de la ingesta calórica diaria en el mundo [1]. Su importancia es aún más evidente en Asia y África, donde el consumo per cápita lo posiciona como el principal componente de la dieta, mientras que en América Latina cumple un rol complementario, pero igualmente crucial en la nutrición de millones de familias. Esta centralidad ha convertido al arroz en un cultivo de interés estratégico para alcanzar los Objetivos de Desarrollo Sostenible (ODS), en particular los relacionados con hambre cero, seguridad alimentaria y producción sostenible [2].

A pesar de su relevancia, la productividad del arroz enfrenta desafíos crecientes. Factores bióticos, como el ataque de hongos patógenos entre ellos *Rhizoctonia solani*, causante del *sheath blight*, limitan severamente el rendimiento en zonas productivas críticas de Asia. De igual manera, los estreses abióticos vinculados al cambio climático, como sequías, inundaciones y variabilidad de temperaturas, amenazan la estabilidad de los sistemas de cultivo y ponen en riesgo los medios de vida de pequeños productores. Estas condiciones adversas se suman a la necesidad urgente de reducir el uso de agroquímicos, promover una agricultura sostenible y aumentar la resiliencia del cultivo mediante variedades mejoradas, capaces de adaptarse a entornos productivos cada vez más complejos [3,4].

Tradicionalmente, la mejora del desempeño agronómico del arroz ha dependido de metodologías convencionales de análisis experimental y comparaciones univariantes, las cuales, si bien útiles en contextos específicos, presentan limitaciones para capturar la complejidad de la interacción entre factores genéticos, ambientales y de manejo [5]. Estos enfoques tradicionales tienden a centrarse en pocas variables, lo que dificulta la identificación de patrones ocultos y reduce la capacidad predictiva en entornos heterogéneos. En consecuencia, se genera un vacío metodológico en la capacidad de anticipar el rendimiento del cultivo bajo escenarios de estrés múltiple, lo cual restringe la eficiencia en la toma de decisiones agronómicas.

Ante este panorama, la incorporación de metodologías multivariantes y técnicas de aprendizaje automático emerge como una alternativa innovadora y poderosa. En particular, el *machine learning* no supervisado se presenta como una herramienta clave para descubrir relaciones ocultas en grandes volúmenes de datos sin necesidad de etiquetas previas, permitiendo identificar patrones y agrupar genotipos de acuerdo con su desempeño agronómico global. El uso de estas herramientas ha mostrado avances significativos en otras áreas de la biotecnología y la

producción agrícola, evidenciando su potencial para transformar los enfoques de selección, evaluación y clasificación de cultivares de arroz.

De esta forma, el estudio de identificación de patrones de desempeño agronómico mediante un enfoque multivariante de aprendizaje automático no supervisado se justifica como una contribución innovadora en el campo de la agronomía, en tanto que ofrece una visión más integral y predictiva del cultivo. La integración de indicadores como rendimiento de grano, resistencia a plagas y enfermedades, y eficiencia en el uso de nutrientes y agua, constituye un marco analítico capaz de orientar nuevas estrategias de manejo y de mejoramiento genético [1]. Este enfoque, además, responde a la necesidad global de generar soluciones que aseguren la sostenibilidad de la producción de arroz, garantizando su papel como cultivo esencial para la seguridad alimentaria en el futuro cercano.

Importancia del arroz en la seguridad alimentaria

El arroz (*Oryza sativa L.*) es uno de los principales cultivos que sustentan la alimentación global, al aportar una fracción significativa de calorías a más de dos tercios de la población mundial. Su centralidad en la dieta lo convierte en un pilar de la seguridad alimentaria, especialmente en Asia, donde constituye la base nutricional cotidiana. En África y América Latina también ocupa un lugar destacado en los sistemas alimentarios, siendo fundamental para comunidades rurales que dependen de su producción y consumo [6].

Más allá de su función como alimento, el arroz sostiene economías nacionales y locales. La mayoría de su producción proviene de pequeños agricultores, cuyo sustento depende directamente del cultivo [7]. En este sentido, la estabilidad de la producción arroceras está directamente vinculada a la reducción de la pobreza, la mejora del bienestar rural y el cumplimiento de los Objetivos de Desarrollo Sostenible relacionados con hambre cero y producción sostenible [8].

El aprovechamiento de sus subproductos refuerza su relevancia. Residuos como la paja o la cascarilla se han utilizado para la producción de biofertilizantes, energía renovable y materiales con valor agregado, lo que amplía el impacto económico y ambiental del arroz en contextos de innovación agroindustrial [9]. Así, la importancia del arroz en la seguridad alimentaria radica no solo en su aporte calórico, sino también en su capacidad para sostener comunidades y dinamizar cadenas de valor diversificadas.

Retos actuales en la mejora de desempeño agronómico

Los sistemas de cultivo de arroz enfrentan un conjunto de desafíos que condicionan su sostenibilidad. Entre ellos se encuentran las enfermedades, como el *sheath blight* causado por *Rhizoctonia solani*, que en países asiáticos representa una seria amenaza para la productividad [10]. A ello se suma el impacto de plagas y la baja disponibilidad de variedades resistentes, lo que obliga a utilizar agroquímicos que generan dependencia económica y riesgos ambientales [11].

El cambio climático ha intensificado los problemas de estrés abiótico: sequías, inundaciones, variabilidad térmica y salinidad creciente en suelos costeros se han convertido en factores que limitan los rendimientos y reducen la resiliencia de los cultivos [12]. La degradación de suelos, la disminución de recursos hídricos y la presión por producir arroz con mayor calidad nutricional plantean retos adicionales para la investigación agronómica [13].

Superar estos desafíos requiere enfoques integrales. No basta con seleccionar genotipos de alto rendimiento, sino que es necesario avanzar en la obtención de variedades capaces de resistir condiciones adversas y de adaptarse a nuevos contextos ambientales [14]. En este sentido, la innovación tecnológica y el uso de metodologías analíticas avanzadas se perfilan como respuestas estratégicas a los retos actuales en la mejora del desempeño agronómico del arroz.

Limitaciones de análisis tradicionales

Los métodos convencionales de evaluación agronómica han estado basados en comparaciones univariantes, pruebas de campo y análisis estadísticos básicos. Aunque han permitido identificar diferencias entre variedades, presentan limitaciones para capturar la complejidad de las interacciones entre factores genéticos, ambientales y de manejo [15].

Estos enfoques no logran modelar adecuadamente las relaciones no lineales entre variables como biomasa, contenido de clorofila, resistencia a plagas y rendimiento, las cuales interactúan de manera dinámica en diferentes entornos productivos [16]. Además, la baja capacidad predictiva de los modelos tradicionales impide anticipar el desempeño del cultivo en condiciones distintas a las observadas en campo [17].

Otra limitación es que los métodos clásicos no integran información de diferentes escalas (molecular, fisiológica y productiva), lo que restringe la comprensión integral del desempeño agronómico. Así, se genera un vacío en la posibilidad de tomar decisiones estratégicas para programas de mejoramiento genético y manejo agronómico [18]. Esta insuficiencia justifica el paso hacia metodologías multivariantes y la adopción de enfoques computacionales más robustos.

Justificación del uso de *machine learning* no supervisado

El aprendizaje automático no supervisado constituye una alternativa metodológica innovadora para superar las limitaciones señaladas. A diferencia de los métodos tradicionales, permite identificar patrones ocultos en grandes volúmenes de datos sin necesidad de contar con etiquetas previas [19]. Esto es especialmente relevante en la agricultura, donde la interacción entre genotipo, ambiente y manejo genera bases de datos complejas y de alta dimensionalidad.

Herramientas como el análisis de conglomerados, los mapas autoorganizados y la reducción de dimensionalidad permiten agrupar genotipos de arroz según características comunes, facilitando la clasificación en función de su desempeño integral [20]. Esta capacidad de reconocer estructuras latentes ofrece ventajas tanto en investigación básica como en la toma de decisiones aplicada en campo.

Asimismo, el uso de *machine learning* no supervisado fomenta la generación de nuevas hipótesis, pues la identificación de grupos de variedades con comportamientos similares puede orientar programas de mejoramiento hacia objetivos específicos, como la tolerancia a sequía o la eficiencia en el uso de nutrientes [21]. En consecuencia, esta metodología no solo amplía la comprensión científica, sino que también fortalece la innovación en sistemas de producción de arroz.

Desempeño agronómico (rendimiento, resistencia, eficiencia)

El desempeño agronómico del arroz se explica por la interacción de múltiples variables. El rendimiento de grano sigue siendo el indicador principal, pero debe evaluarse junto con la resistencia a plagas y enfermedades, dado que la estabilidad productiva depende de la capacidad de las variedades para enfrentar estreses bióticos [22].

Otro componente esencial es la eficiencia en el uso de recursos, en particular agua y nutrientes, cada vez más limitados en contextos de presión climática y ambiental [23]. La investigación reciente también ha subrayado la importancia de factores fisiológicos, como la calidad de la semilla y la capacidad de adaptación a estreses múltiples, como determinantes del desempeño integral [24].

El análisis multivariante permite integrar estas dimensiones para construir perfiles robustos de desempeño, que clasifican genotipos de acuerdo con su resiliencia y potencial productivo [25]. En este marco, el uso de técnicas de *machine learning* no supervisado se posiciona como una herramienta esencial para identificar patrones que fortalezcan el rendimiento, la resistencia y la eficiencia del arroz, garantizando así su papel estratégico en la seguridad alimentaria global [26].

Análisis de componente principales

El análisis de componente principales en sus siglas en inglés PCA corresponde a un método de análisis factorial de aprendizaje automático no supervisado [27]. [28]y [29] define al proceso de obtención de los componentes principales como la extracción de los componentes ortogonales y descartando a través de esparsificación las variable irrelevantes, además. [30] menciona que el método PCA permite reducir la dimensionalidad del conjunto de variables evaluadas, concentrando la mayor varianza en nuevas variables independientes denominada Factores o componentes principales.

1. $PC\ 1 = \beta_{11}X_1 + \beta_{12}X_2 + \dots + \beta_{1n}X_n$
2. $PC\ 2 = \beta_{21}X_1 + \beta_{22}X_2 + \dots + \beta_{2n}X_n$
3. $PCn = \beta_{n1}X_1 + \beta_{n2}X_2 + \dots + \beta_{nn}X_n$

(1)

Para seleccionar los componentes o factores [31] sugiere que el procedimiento estándar para aplicar el método PCA requiere en primer lugar obtenido más de un componente, luego los componente a seleccionar al menos concentren el 10% de la varianza y por último se debe de obtener en la varianza acumulada al menos un 60%.

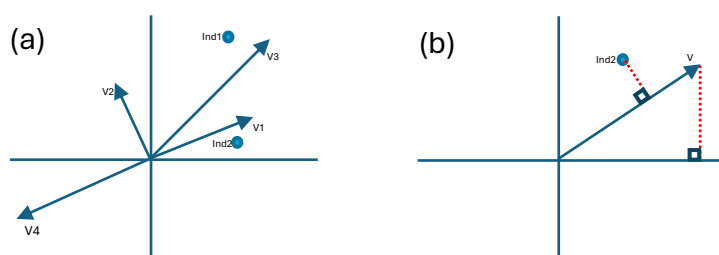
Técnica de análisis biplot

El biplot es una representación gráfica de una matriz $N \times M$, donde las filas representan individuos y las columnas representan variables [32]. Su estructura gráfica permite analizar datos que involucran más de una variable [33]. En el análisis multivariante, se reduce la dimensión de un conjunto de datos de N individuos y M variables proyectando los datos sobre un espacio ajustado denominado Euclidiano. Según [34], el biplot proporciona una herramienta funcional que puede representar el resultado de un análisis de componentes principales, mostrando valores y productos estadísticos como distancias entre unidades de datos y sus correspondientes agrupaciones, así como la varianza explicada y las correlaciones entre variables o entre individuos.

Por ejemplo, en una representación de 2 dimensiones en la Figura 1 se muestra con círculos de color naranja a los individuos o filas, y las líneas de color azul que inician desde el origen y terminan en punta de flecha representan a las variables o columnas.

Figura 1

Gráfico Biplot



La Figura 1, las gráficas (a) y (b), los ejes de los planos representan los componentes que tienen mayor representatividad para los individuos y variables, basándose en los resultados de un PCA. En la gráfica (a), cada vector representa una variable del estudio y su interpretación gráfica muestra el grado de correlación entre ellas. Por ejemplo, para las variables $V1$ y $V3$, la correlación es fuerte porque el ángulo entre los vectores es agudo, lo que indica una mayor correlación positiva. Cuanto más corto sea el ángulo, mayor será la correlación entre las variables. Si el ángulo entre dos variables es de 90° , como en el caso de $V1$ y $V2$, su correlación es nula. Por otro lado, cuando el ángulo es obtuso, como en el caso de $V4$ y $V1$, la correlación es inversa.

En la gráfica (b), el ángulo formado por el vector proyectado hacia los ejes proporciona información sobre cuál componente tiene una mayor correlación con la variable. Además, el ángulo formado por la proyección desde el individuo hacia una vector muestra cuál variable tiene una mayor correlación con ese individuo.

Análisis por clusterización

En estudios experimentales agronómicos que involucran múltiples variables fisiológicas y de crecimiento, como en el caso del establecimiento inicial del cultivo de arroz, resulta fundamental identificar patrones de comportamiento que no siempre son evidentes mediante

análisis univariantes. En este contexto, el análisis de clúster constituye una herramienta estadística adecuada para agrupar observaciones en función de su similitud multivariada, sin requerir información previa sobre la estructura de los datos.

La elección de este método se justifica por la necesidad de clasificar objetivamente los genotipos o tratamientos evaluados según su desempeño combinado en variables como los días para la aparición de radícula (V1), días para aparición de cotiledón (V2), dimensión de radícula (V3) y dimensión de cotiledón (V4). Dado que estas variables reflejan distintos aspectos del vigor y el establecimiento temprano de las plántulas, el análisis de clúster permite descubrir perfiles fisiológicos diferenciados que pueden orientar decisiones experimentales, de selección o de manejo.

Además, el análisis de conglomerados es especialmente útil en estudios exploratorios donde no existe una hipótesis previa sobre la cantidad o tipo de grupos esperados [35], lo que lo convierte en una técnica apropiada dentro del enfoque de aprendizaje automático no supervisado. Al basarse en distancias multivariadas y criterios de similitud interna, este análisis permite reducir la complejidad del conjunto de datos y generar una representación estructurada de las observaciones, facilitando su interpretación desde el punto de vista agronómico.

Para [36], el uso del análisis de clúster complementa el análisis de componentes principales (PCA), validando la existencia de agrupamientos consistentes entre observaciones y permitiendo la identificación de patrones de desempeño agronómico diferenciados. Esta combinación metodológica refuerza la robustez del enfoque multivariante adoptado y contribuye a una clasificación más precisa de los tratamientos según su comportamiento fisiológico inicial.

El presente trabajo busca responder la siguiente pregunta ¿Qué patrones de desempeño agronómico pueden identificarse en el cultivo de arroz utilizando métodos multivariantes de aprendizaje automático no supervisado? y planteamos la siguiente hipótesis:

H1: El uso de técnicas multivariantes de aprendizaje automático no supervisado permite identificar patrones diferenciados en el desempeño agronómico del arroz, que facilitan la clasificación de genotipos o tratamientos según sus características de rendimiento, resistencia y eficiencia en el uso de recursos.

Por lo tanto, el cultivo de arroz constituye uno de los pilares fundamentales de la seguridad alimentaria en América Latina y a nivel mundial, lo que ha motivado numerosos estudios orientados a optimizar su rendimiento y adaptabilidad, mostrando un interés creciente en la identificación de factores que influyen en el desempeño agronómico del arroz, tales como la productividad, la resistencia a condiciones de estrés biótico y abiótico, y la eficiencia en el uso de nutrientes y agua. Sin embargo, la mayoría de los enfoques tradicionales se basan en metodologías univariantes o en comparaciones directas entre tratamientos, lo que limita la capacidad de capturar la complejidad multivariada inherente a los sistemas agrícolas.

En este contexto, las técnicas de aprendizaje automático no supervisado, como el análisis de componentes principales (PCA) y los algoritmos de clustering, ofrecen herramientas robustas para descubrir patrones ocultos en grandes volúmenes de datos agronómicos. Estas metodologías permiten integrar simultáneamente múltiples variables y clasificarlas en grupos coherentes que reflejan la variabilidad real de los genotipos o tratamientos de arroz. La evidencia bibliográfica analizada refuerza la necesidad de aplicar enfoques de este tipo, dado que varios estudios reportan la dificultad de predecir el rendimiento agrícola únicamente con modelos convencionales.

Así, la hipótesis planteada se justifica en la convergencia de dos elementos: por un lado, la creciente disponibilidad de datos experimentales detallados sobre el arroz y, por otro, la capacidad del aprendizaje automático no supervisado para procesar dichos datos de manera más eficiente y objetiva. La identificación de patrones diferenciados en el desempeño agronómico no solo contribuye a una mejor clasificación de los genotipos, sino que también genera información estratégica para programas de mejoramiento genético y manejo agronómico, alineándose con las tendencias actuales de digitalización y análisis de datos en la agricultura.

METODOLOGÍA

Conjunto de dato

Los datos analizados fueron tomados de un diseño experimental explicativo diseñado para determinar la relación causal entre los tratamientos aplicados y el crecimiento vegetativo del cultivo. El estudio se realizó en una unidad agrícola de la región del Programa Interamericano del Estado Daulé, Provincia del Guayas, Ecuador, en condiciones de invernadero en su fase de germinación y desarrollo de la plántula, siendo la zona la principal productora de arroz en la región. La variedad utilizada fue *Oryza sativa* L. SFL 011, seleccionada por su representatividad en las prácticas agrícolas locales y su amplio uso comercial. El experimento empleó un diseño completamente aleatorizado (DCA) con cuatro tratamientos diferentes Ver Tabla 1, cada uno con cuatro réplicas tuvieron 50 réplicas, para un total de 200 unidades experimentales y las variables medidas son descritas en la Tabla 2.

Tabla 1

Tratamientos definidos para el experimento

Tratamiento	Descripción
T1	Producto a base de <i>Trichoderma</i>
T2	Sepa nativa <i>Trichoderma</i>
T3	Producto industrializado con funcionalidad de estimular en la fase de germinación
T4	Testigo

Tabla 2

Variables medidas en el experimento

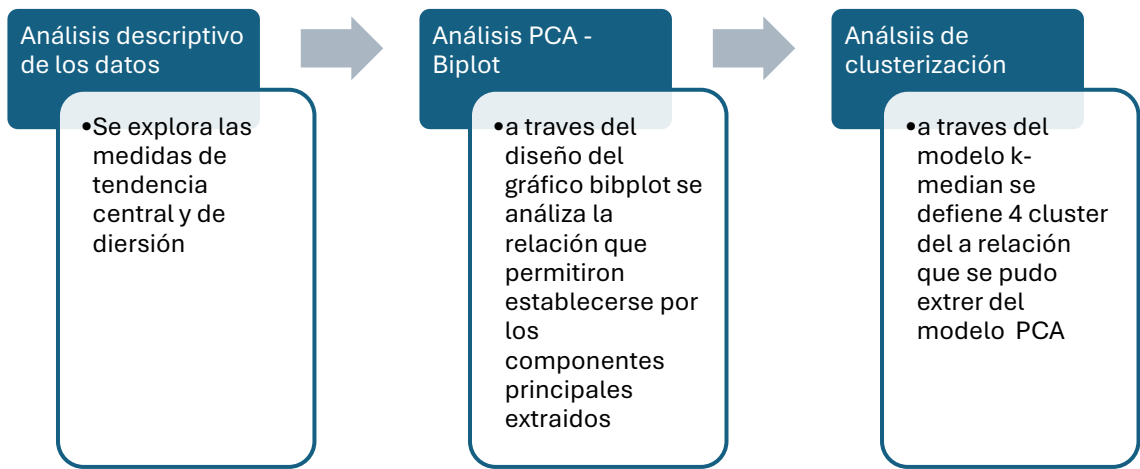
Variables	Descripción
V1 - Días para aparición de radícula	Se refiere al número de días transcurridos desde la siembra hasta la emergencia visible de la radícula, que es la primera raíz embrionaria que sale de la semilla. Es un indicador de la velocidad de germinación y del vigor inicial de la semilla.
V2 - Días para aparición de cotiledón	Corresponde al número de días desde la siembra hasta la emergencia del cotiledón (la primera hoja embrionaria de la plántula). Este parámetro refleja la rapidez del desarrollo inicial del brote y la capacidad de la semilla para establecerse en condiciones de campo o vivero.
V3 - Dimensión de radícula (cm)	Mide la longitud de la radícula principal en centímetros una vez que ha emergido. Esta variable es fundamental para evaluar la capacidad de exploración del suelo, el establecimiento inicial y el potencial de absorción de agua y nutrientes en las primeras fases de desarrollo.
V4 - Dimensión de cotiledón (cm)	Se refiere a la longitud del cotiledón en centímetros, una vez que este ha emergido. Es un indicador del crecimiento inicial aéreo de la plántula, vinculado a la fotosíntesis temprana y al vigor del establecimiento.

Análisis estadístico de los datos

Para el análisis de los datos se desarrolló el siguiente flujo de trabajo ver Figura 1:

Figura 2

Flujo de trabajo estadístico



RESULTADOS

Análisis descriptivo de los datos

El análisis descriptivo de las variables evaluadas permitió caracterizar el comportamiento del proceso de germinación y desarrollo inicial del arroz en términos de la aparición y crecimiento de estructuras básicas.

En relación con los días para aparición de radícula (V1), se observó un promedio de 5.37 días (DE = 1.55), con valores entre 1 y 9 días. La mediana fue de 5 días, lo que indica que en la mayoría de las plántulas la emergencia de la radícula ocurrió alrededor de la primera semana. La distribución mostró una ligera simetría (asimetría = -0.04), lo que confirma estabilidad en esta fase inicial.

En cuanto a los días para aparición de cotiledón (V2), el promedio fue de 8.36 días (DE = 1.54), con un rango entre 4 y 12 días. La mediana se situó en 8 días, evidenciando que la emergencia del cotiledón se da generalmente después de la radícula, con una leve tendencia hacia valores superiores (asimetría = 0.18). Estos resultados reflejan un desarrollo progresivo entre ambas estructuras.

Respecto a la dimensión de la radícula (V3), se obtuvo un promedio de 4.22 cm (DE = 0.75), con valores comprendidos entre 2.38 y 6.08 cm. La mediana (4.25 cm) y la baja asimetría (0.006) indican homogeneidad en el crecimiento radical, destacando que la mayoría de las plántulas alcanzaron longitudes cercanas a este valor.

Finalmente, la dimensión del cotiledón (V4) presentó un promedio de 2.59 cm (DE = 0.65), con un rango de variación entre 0.72 y 4.13 cm. La mediana fue de 2.71 cm, mostrando un crecimiento relativamente consistente. La distribución evidenció una ligera asimetría negativa (-0.22), lo que sugiere que algunos individuos presentaron longitudes menores al promedio.

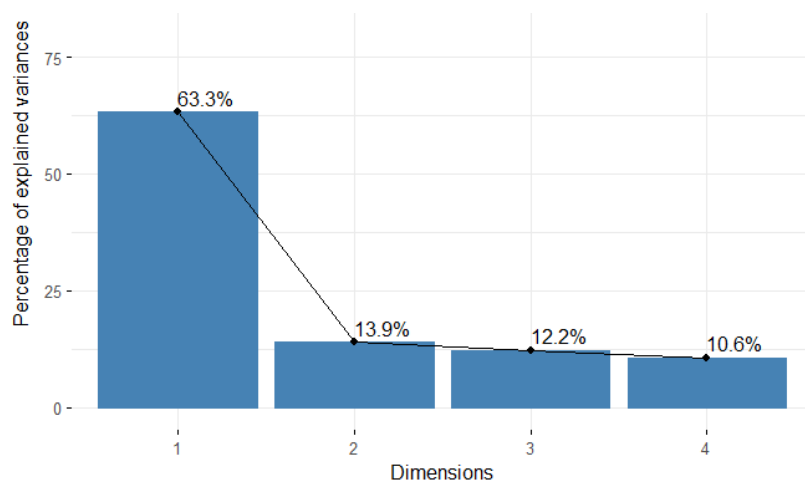
En conjunto, estos resultados reflejan un patrón coherente de germinación y desarrollo inicial, donde la radícula emerge primero y presenta un crecimiento más acelerado que el cotiledón, lo que es consistente con la literatura agronómica sobre el establecimiento temprano del arroz.

Análisis de componentes principales – Biplot

El Análisis de Componentes Principales se extrajeron cuatro componentes Ver Figura 2. Los dos primeros componentes concentran el 77.2 % de la varianza explicada.

Figura 3

Componentes Principales Extraídos



La Figura 3 se muestra el gráfico bibplot de los primeros componentes principales, donde se observa en el cuadrante I, que las variables V1 (Días para aparición de radícula) y la variable V2 (Días para aparición de cotiledón) tiene una alta correlación positiva.

La alta correlación positiva observada entre ambas variables sugiere que los factores fisiológicos, genéticos y ambientales que condicionan la aparición temprana de la radícula también influyen en la velocidad de emergencia del cotiledón. En términos agronómicos, esta relación indica que las semillas que inician rápidamente el proceso de germinación también tienden a completar más pronto su desarrollo inicial aéreo, lo que resulta en plántulas más uniformes y con mayor vigor.

Esta correlación es coherente con la secuencia natural de desarrollo en las gramíneas como el arroz, donde la emergencia de la radícula precede funcional y temporalmente al crecimiento del coleóptilo y cotiledón. Además, desde el punto de vista del diseño experimental y del mejoramiento genético, esta asociación sugiere que la medición de la aparición de la radícula (V1) puede actuar como un indicador temprano y predictivo del comportamiento de la plántula en etapas posteriores, simplificando evaluaciones en pruebas de vigor o eficiencia de tratamientos de pre-siembra.

En conjunto, esta alta correlación fortalece la interpretación de los datos y justifica su consideración conjunta en análisis multivariantes, especialmente cuando se busca comprender los patrones de establecimiento temprano del cultivo bajo distintas condiciones experimentales.

Además, en el cuadrante dos y tres encontramos en la variable V4(Dimensión de cotiledón (cm)) y la variable V3 (Dimensión de radícula (cm)) presentan una correlación buena.

La observación de una buena correlación positiva entre ambas variables indica que, en general, las plántulas que desarrollan raíces más largas también tienden a producir cotiledones de

mayor longitud, lo que refleja un vigor de crecimiento equilibrado entre los sistemas radicular y aéreo.

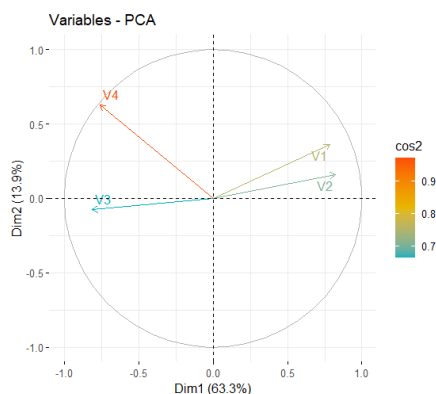
Desde el punto de vista fisiológico, esta relación es coherente, ya que el desarrollo de ambos órganos depende de la eficiencia metabólica de la semilla, la disponibilidad de reservas, y la capacidad de absorción y translocación de agua y nutrientes durante los primeros días después de la germinación. En semillas de arroz con buen estado fisiológico o bajo condiciones ambientales favorables, se espera un crecimiento armónico entre radícula y cotiledón, lo cual justifica la correlación observada.

Esta buena correlación también aporta valor desde el diseño experimental y el análisis multivariado, ya que permite interpretar el desempeño inicial de la plántula como una expresión conjunta del crecimiento radicular y foliar, favoreciendo la agrupación de genotipos o tratamientos con respuestas similares. Además, al estar relacionadas, estas variables pueden utilizarse de forma complementaria en modelos de análisis de componentes principales (PCA), identificación de patrones de vigor o clasificación temprana de materiales promisorios.

En síntesis, la correlación positiva entre V3 y V4 evidencia una respuesta fisiológica coordinada, útil para evaluar la calidad del establecimiento en estudios de germinación, priming, o comparación de tratamientos agronómicos.

Figura 4

Gráfico Biplot de las Variables de estudio



En la Figura 4 la relación entre las variables V1 y V2 y las observaciones en el primer cuadrante,

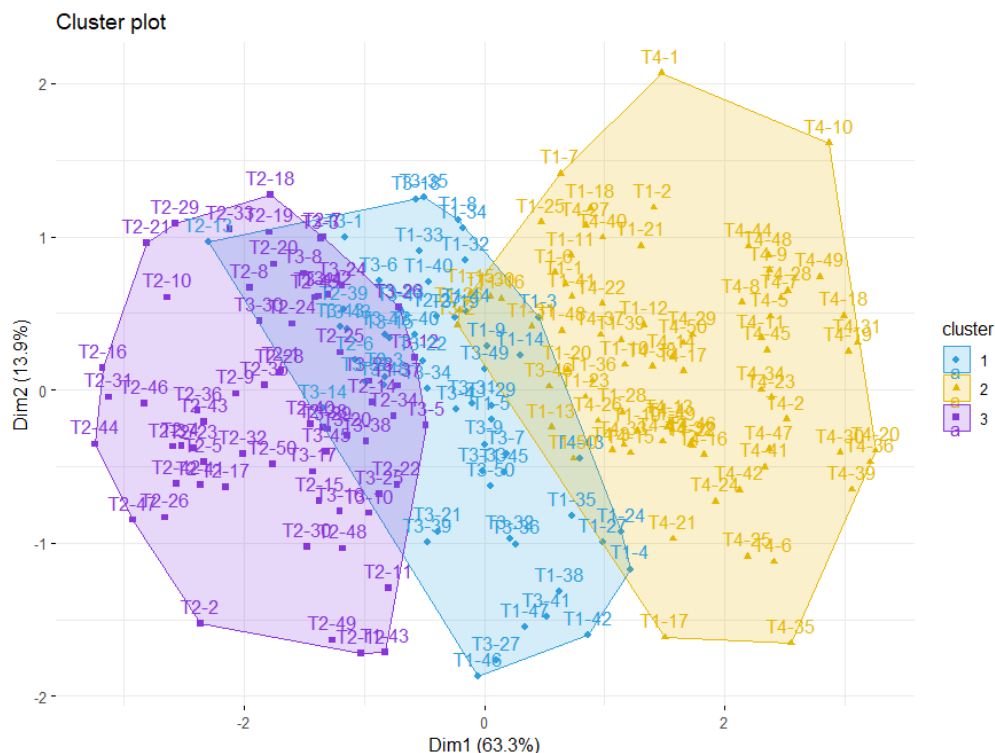
Análisis de clusterización

En la Figura 4 se identifican tres clústeres claramente definidos. El Clúster 1 agrupa observaciones asociadas predominantemente con la variable V3 (Dimensión de radícula), las cuales se localizan en el cuadrante inferior izquierdo de la representación en la Figura 3. Por su parte, el Clúster 2 concentra las observaciones vinculadas con las variables V1 (Días para aparición de radícula) y V2 (Días para aparición de cotiledón), lo que coincide con la agrupación mostrada en el plano factorial de la Figura 3. Finalmente, el Clúster 3 refleja la relación de un grupo de observaciones con la variable V4 (Dimensión de cotiledón), patrón que también guarda coherencia con su ubicación en la representación biplot de la Figura 3. Esta correspondencia entre

ambas figuras confirma la consistencia entre la clasificación obtenida por el análisis de conglomerados y la distribución espacial revelada en el análisis de componentes principales.

Figura 5

Gráfico de clusterización



DISCUSIÓN

El análisis de las variables evaluadas permitió caracterizar con precisión el desempeño inicial de las plántulas de arroz bajo diferentes condiciones experimentales. En particular, la fuerte correlación positiva observada entre V1 (días para aparición de radícula) y V2 (días para aparición de cotiledón) sugiere un patrón sincronizado de emergencia, en el cual las semillas con mayor rapidez para emitir la radícula también tienden a desarrollar precozmente el cotiledón. Este comportamiento es fisiológicamente coherente, dado que ambas estructuras representan fases consecutivas del proceso de germinación, y su emergencia refleja la eficiencia metabólica de la semilla y la calidad del ambiente inmediato.

Asimismo, la buena correlación entre V3 (dimensión de radícula) y V4 (dimensión de cotiledón) indica un crecimiento coordinado entre los sistemas radicular y aéreo durante las etapas iniciales del desarrollo. Esta relación sugiere que aquellas plántulas que alcanzan mayor longitud de radícula tienden también a desarrollar cotiledones más extensos, lo cual es un reflejo de vigor y balance fisiológico. Tales asociaciones son valiosas para interpretar el comportamiento general de las plántulas y para identificar materiales con potencial agronómico superior.

La aplicación del análisis de componentes principales (PCA) permitió reducir la dimensionalidad de los datos y visualizar la variación multivariable de las observaciones. En el

plano factorial se observaron agrupamientos que reflejan asociaciones definidas entre variables: un primer grupo relacionado con el desarrollo radicular (V3), un segundo con el tiempo de emergencia (V1 y V2), y un tercero con el crecimiento del cotiledón (V4). Estos patrones fueron reforzados por el análisis de conglomerados jerárquicos, en el cual se identificaron tres clústeres principales que coinciden con las agrupaciones reveladas en el PCA. Esta consistencia entre ambos métodos no supervisados valida la clasificación estructural de los datos y respalda la interpretación biológica de los resultados.

Los hallazgos obtenidos en este estudio permiten confirmar la hipótesis planteada: el uso de técnicas multivariantes de aprendizaje automático no supervisado permite identificar patrones diferenciados en el desempeño agronómico del arroz, que facilitan la clasificación de genotipos o tratamientos según sus características de rendimiento, resistencia y eficiencia en el uso de recursos. El enfoque aplicado, basado en PCA y análisis de conglomerados, fue eficaz para revelar la estructura latente del conjunto de datos, diferenciar grupos de observaciones con comportamientos agronómicos distintos y, por tanto, proveer una herramienta útil para la toma de decisiones experimentales y la selección de materiales en programas de mejoramiento o manejo.

En conjunto, los resultados demuestran que los métodos multivariantes no supervisados no solo permiten sintetizar grandes volúmenes de información, sino que también ofrecen una base objetiva para la clasificación agronómica de plántulas y la identificación de perfiles de respuesta que serían difíciles de detectar mediante análisis univariantes tradicionales.

CONCLUSIONES

El presente estudio permitió caracterizar de manera integral el desempeño agronómico inicial del arroz mediante el análisis de variables clave relacionadas con la emergencia y el desarrollo morfológico de las plántulas. A través de la evaluación conjunta de los días para aparición de radícula (V1), días para aparición de cotiledón (V2), dimensión de la radícula (V3) y dimensión del cotiledón (V4), fue posible identificar relaciones significativas que reflejan un crecimiento fisiológico coordinado y una respuesta agronómica diferenciada entre los tratamientos analizados.

Los resultados confirmaron la existencia de correlaciones positivas relevantes entre las variables V1–V2 y V3–V4, lo cual evidencia tanto la sincronización del proceso de germinación como el equilibrio en el desarrollo de los órganos vegetativos. Estas asociaciones fortalecen el valor de dichas variables como indicadores confiables de vigor y eficiencia de establecimiento en condiciones controladas.

La aplicación de técnicas multivariantes de aprendizaje automático no supervisado, como el análisis de componentes principales (PCA) y el análisis de conglomerados, resultó altamente efectiva para reducir la dimensionalidad del conjunto de datos, visualizar patrones estructurales y clasificar objetivamente las observaciones en grupos coherentes. Se identificaron tres clústeres

principales que reflejan perfiles diferenciados de respuesta agronómica, asociados al tiempo de emergencia, el desarrollo radicular y la expansión del cotiledón.

En consecuencia, se confirma la hipótesis planteada: el uso de métodos multivariantes no supervisados permite identificar patrones diferenciados en el desempeño agronómico del arroz, facilitando la clasificación de genotipos o tratamientos en función de variables fisiológicas clave. Esta aproximación analítica no solo aporta una comprensión más profunda del comportamiento inicial del cultivo, sino que también se proyecta como una herramienta útil para optimizar procesos de selección en programas de mejoramiento, evaluación de tratamientos pre-siembra o manejo agronómico en etapas tempranas.

Finalmente, se recomienda la aplicación de estos enfoques multivariantes en futuros estudios experimentales con mayor diversidad genética y en condiciones de campo, a fin de validar la robustez de los patrones identificados y ampliar su aplicabilidad en escenarios reales de producción.

REFERENCIA

1. Afzal, M.; Tan, X.; Fang, M.; Zeng, D.; Huang, M.; Chen, X.; Zhang, X.; Tan, Z. Chitosan-Silver Nanoparticles and *Bacillus Tequilensis* Modulate Antioxidant Pathways and Microbiome Dynamics for Sheath Blight Resistance in Rice. *Plant Stress* 2025, 17.
2. Tiwari, P.; Adil, M.; Park, K. Biopriming with Endophytes Improves Plant Resilience to Develop Climate-Smart, Futuristic Agricultural Crops. *World J. Microbiol. Biotechnol.* 2025, 41.
3. Syahrawati, M.; Saputra, N.; Trizelia, T.; Sulyanti, E. Entomopathogenicity of Endophytic Fungi Isolated from Payo Rice Variety in Kerinci, Indonesia. *Egypt. J. Biol. Pest Control* 2025, 35.
4. Rahma, H.; Trizelia, T.; Martinius, M.; Rozen, N.; Qalbina, F. Detection of Seed Borne Pathogens in Local Rice Varieties in West Sumatra Province. *AIP Conf. Proc.* 2025, 3277.
5. Singh, N.; Shahi, U.P.; Dhyani, B.P.; Singh, S.C. Effect of Various Rice Residue Decomposer on Nutrient Release Pattern from Treated Rice Residue in Light Textured Soil. *Natl. Acad. Sci. Lett.* 2025, 48, 409–412.
6. Song, J.; Guan, X.; Chen, L.; Han, Z.; Cui, H.; Ma, S. Cooperative Interplay Between PGPR and *Trichoderma Longibrachiatum* Reprograms the Rhizosphere Microecology for Improved Saline Alkaline Stress Resilience in Rice Seedlings. *Microorganisms* 2025, 13.
7. Parveen, F.; Hussain, A.; Saxena, A.; Husain, F.M.; Ashfaq, M. Optimized Bioethanol Production from Diverse Lignocellulosic Biomass Sources through Integrative Approach of Plant Cell Wall Degrading Enzyme Pretreatment. *Biomass Bioenergy* 2025, 197.
8. Prismantoro, D.; Chua, K.O.; Teo, K.W.E.; Chan, R.; Jefferson, T.A.; Mohd Suhaimi, N.S.; Mispan, M.S.; Miranti, M.; Doni, F. Whole Genome Sequence Data of *Trichoderma Yunnanense* Strain TM10, a Plant Growth-Promoting Fungus and Biocontrol Agent. *Data Brief* 2025, 58.
9. Lilai, S.A.; Hussein, J.M.; Kapinga, F.A.; Nene, W.A.; Temu, S.G.; Tibuhwa, D.D. Efficacy of Organic-Based Substrate Formulation of *Bacillus* Strains and *Trichoderma Asperellum* Against *Fusarium* Wilt of Cashew. *J. Phytopathol.* 2025, 173.
10. Modrzewska, M.; Popowski, D.; Błaszczyk, L.; Stepień, L.; Urbaniak, M.; Bryła, M.; Cramer, B.; Humpf, H.U.; Twarużek, M. Antagonistic Properties against *Fusarium Sporotrichioides* and Glycosylation of HT-2 and T-2 Toxins by Selected *Trichoderma* Strains. *Sci. Rep.* 2024, 14.
11. Dewi, R.S.; Sastranegara, M.H.; Ayunda, B.C. Preliminary Study of the Ability of Indigenous Fungi from River as an Alternative to Degrade Pesticides and Their Effect on Insect Larvae. *BIO Web Conf.* 2024, 123.

12. Intana, W.; Suwannarach, N.; Kumla, J.; Wonglom, P.; Sunpapao, A. Plant Growth Promotion and Biological Control against *Rhizoctonia Solani* in Thai Local Rice Variety “Chor Khing” Using *Trichoderma Breve* Z2-03. *J. Fungi* 2024, *10*.
13. Li, L.; Tian, J.; Huang, K.; Xue, X.; Chen, J.; Guan, F.; Zhang, T.; Sun, Y.; He, C.; Zeng, X. Metal-Binding Protein TaGlo1 Improves Fungal Resistance to Arsenite (AsIII) and Methylarsenite (MAsIII) in Paddy Soil. *Environ. Sci. Technol.* 2024, *58*, 7469–7479.
14. Duque, E.R.Y.; Aguirre, M.; Hood, N.C.; Hood, E.E. Specific Activity and Utility of Recombinant Cellobiohydrolase II (Cel6A) Produced in Maize Endosperm. *Transgenic Res.* 2024, *33*, 47–57.
15. Hao, H.; Yue, Y.; Wang, Q.; Xiao, T.; Zhao, Z.; Zhang, J.; Chen, H. Effects of the Rice-Mushroom Rotation Pattern on Soil Properties and Microbial Community Succession in Paddy Fields. *Front. Microbiol.* 2024, *15*.
16. Ichsan, C.N.; Salsabila, Y.A.; Mayani, N.; Jumini, null; Kurniawan, T.; Santi, I.V. Application of Soil Amendments and Paclobutrazol to Shorten Internode and Increase Rice Yields. *IOP Conf. Ser. Earth Environ. Sci.* 2024, *1297*.
17. Tufail, M.W.; Ali, S.A.; Ansari, M.D.; Sajjad, S.; Ali, P.K.; Anwar, S.; Zahoor, M.; Mansha, H. Effects of Bio-Fertilizers on the Seedling of Rice (*Oryza Sativa* L.) in in-Vitro and in-Vivo Conditions under Salt Stress. *Int. J. Recycl. Org. Waste Agric.* 2024, *13*.
18. Liu, Y.; Li, T.; Zhu, H.; Zhou, Y.; Shen, Q.; Liu, D. Cysteine Facilitates the Lignocellulolytic Response of *Trichoderma Guizhouense* NJAU4742 by Indirectly Up-Regulating Membrane Sugar Transporters. *Biotechnol. Biofuels Bioprod.* 2023, *16*.
19. Hoa, P.T.B.; Tue, N.H.; Trang, H.T.Q.; Thu, H.A.; Nhung, L.N.H.; Luong, N.N.; Huy, N.X.; Tien, N.Q.D.; Loc, N.H. Enhancement of Resistance against Fungal Pathogens in Peanut (*Arachis Hypogaea* L.) Cultivar L14 by Heterologous Expression of Gene Encoding Chitinase 42 kDa from *Trichoderma Asperellum* SH16. *South Afr. J. Bot.* 2023, *160*, 673–681.
20. Islam, M.R.; Chowdhury, R.; Roy, A.S.; Islam, M.N.; Mita, M.M.; Bashar, S.; Saha, P.; Rahat, R.A.; Hasan, M.; Akter, M.A. Native *Trichoderma* Induced the Defense-Related Enzymes and Genes in Rice against *Xanthomonas Oryzae* Pv. *Oryzae* (Xoo). *Plants* 2023, *12*.
21. Syamsafitri, null; Nurhayati, null; Kesuma, R.P.; Ningsih, S.S. The Increase of Organic Shallots (*Allium Cepa* Var *Ascalonicum* L.) Production through the Application of Compost on Inceptisol Soils. *Univers. J. Agric. Res.* 2023, *11*, 185–190.
22. Wang, Y.; Mridha, M.A.U. Editorial: Community Series in Plants and Microbial Communities: Diversity, Pathogens and Biological Control, Volume II. *Front. Microbiol.* 2023, *14*.

23. Mutlag, N.H.; Kermasha, H.S.N.; Majeed, A.M. Efficiency of *Trichoderma Harzianum* Fungus in Bioremediation of Nominee and Superflak Pesticides Residues in Rice Fields in Najaf-Iraq. *IOP Conf. Ser. Earth Environ. Sci.* 2023, 1215.
24. Wiyono, S.; Russianzi, W.; Mursyidi, H.H.; Khamidi, T. Physiological Characteristics and Effectiveness of Some *Trichoderma* Isolates against *Fusarium* Basal Rot of Shallot. *IOP Conf. Ser. Earth Environ. Sci.* 2023, 1133.
25. Delmo-Organo, N.; Granada, S.M.J.M.; Pineda, H.G.S.; Sandro, J.M.; Nguyen, V.H.; Gummert, M. Publisher Correction: Assessing the Potential of a *Trichoderma*-Based Compost Activator to Hasten the Decomposition of Incorporated Rice Straw (Scientific Reports, (2022), 12, 1, (448), 10.1038/S41598-021-03828-1). *Sci. Rep.* 2022, 12.
26. Klaram, R.; Jantasorn, A.; Dethoup, T. Efficacy of Marine Antagonist, *Trichoderma* Spp. as Halo-Tolerant Biofungicide in Controlling Rice Diseases and Yield Improvement. *Biol. Control* 2022, 172.
27. Chen, I.-C. Predicting Regional Sustainable Development to Enhance Decision-Making in Brownfield Redevelopment Using Machine Learning Algorithms. *Ecol. Indic.* 2024, 163, 112117, doi:10.1016/j.ecolind.2024.112117.
28. Bocci, L.; D'Urso, P.; Vicari, D.; Vitale, V. A Three-Way Approach for Defining Competitiveness Indexes of the European Regions (NUTS-2). *Soc. Indic. Res.* 2021, doi:10.1007/s11205-021-02839-9.
29. Yu, J.; Gómez-Corona, C.; Abdi, H.; Guillemot, V. Sparse Multiple Factor Analysis, Sparse STATIS, and Sparse DiSTATIS with Applications to Sensory Evaluation. *J. Chemom.* 2024, 38, e3443, doi:10.1002/cem.3443.
30. StataCorp LLC STATA Structural Equation Modeling Reference Manual Release. 2017, 15.
31. OECD Handbook on Constructing Composite Indicators Methodology and User Guide. OECD. 2008.
32. Gabriel, K.R. The Biplot Graphic Display of Matrices with Application to Principal Component Analysis. In *Biometrika*; 1971; Vol. 58, pp. 453-467.
33. ODOROFF, CH.L.; GABRIEL, K.R. Biplot in Biomedical Research. 1990, 0, 469-485.
34. GALINDO, M.P.; CUADRAS, C.M. Una Extensión Del Método Biplot a Su Relación Con Otras Técnicas. *Publ. Bioestad. Biomatemática* 1986.
35. Chikhi, N.F. Scientific Publications Clustering Using Textual and Citation Information. *Expert Syst. Appl.* 2024, 248, 123319, doi:10.1016/j.eswa.2024.123319.
36. Timbalari, C.; Neagoe (Barbu), A.L.; Cristian, M.-G. Clustering Corporate Governance and Leadership. A Bibliometric Analysis. *Stud. Bus. Econ.* 2024, 19, 255-275.