

## Técnicas de Minería de datos aplicados a la agricultura: Estado del Arte y análisis bibliométrico

<http://doi.org/10.53358/ideas.v6i1.944>

**Ana C. Umaquina Criollo**

Universidad Técnica del Norte, Facultad de Ingeniería en Ciencias Aplicadas  
acumaquina@utn.edu.ec

*Fecha de envío, agosto 7/2023 - Fecha de aceptación, septiembre 13/2023 - Fecha de publicación, enero 19/2024*

### Resumen:

En esta investigación, se presenta un análisis bibliométrico de 106 artículos de revistas y estado del arte indexados en Scopus, junto con un análisis sistemático de 83 artículos seleccionados. Se identifican áreas de estudio que incluye la predicción de rendimiento y crecimiento de cultivos, la detección de enfermedades en plantas, análisis de agua y suelo, relacionados con diferentes tipos de cultivo como: cereales (arroz, cebada, maíz, trigo, soya); frutas (manzana, pepino); legumbres (alfalfa, frejol, cacahuate); tubérculos, entre otros. Se examinan variables climáticas, suelo, agua, condiciones topográficas, edafológicas y técnicas de minería de datos como, Redes Neuronales, Deep Learning, segmentación, reglas de asociación y clasificación, entre otras, para optimizar el uso de recursos y tomar decisiones agrícolas basadas en datos. Además, se destacan los desafíos y oportunidades en esta área de investigación, así como las perspectivas futuras para el desarrollo de soluciones de minería de datos avanzadas en el contexto agrícola. Este análisis contribuye a una mejor comprensión de cómo la minería de datos está transformando el sector agrícola, comunidad académica y científica, con el fin de impulsar la eficiencia, la sostenibilidad y la toma de decisiones informadas en la producción de alimentos.

**Palabras clave:** Minería de datos, análisis de datos agrícolas, Detección de enfermedades en plantas, agricultura.

### Abstract:

This research presents a bibliometric analysis of 106 journal and state-of-the-art articles indexed in Scopus and a systematic analysis of 83 selected papers. Areas of study are identified that include the prediction of crop yield and growth, the detection of plant diseases, and water and soil analysis related to different types of crops such as cereals (rice, barley, corn, wheat, soybeans); fruits (apple, cucumber); legumes (alfalfa, beans, peanuts); tubers, among others. Climatic variables, soil, water, topographic and edaphological conditions, and data mining techniques such as Neural Networks, Deep Learning, segmentation, association, and classification rules, among others, are examined to optimize the use of resources and make agricultural decisions based on data. In addition, the challenges and opportunities in this research area are highlighted as the future perspectives for developing advanced data mining solutions in the agricultural context. This analysis contributes to a better understanding of how data mining is transforming the farm sector academic and scientific community to drive efficiency, sustainability, and informed decision-making in food production.

**Keywords:** Data mining, agricultural data analysis, plant disease detection, agriculture.

*Autor de correspondencia:*

Ana Cristina Umaquina, [acumaquina@utn.edu.ec](mailto:acumaquina@utn.edu.ec)

## Introduction.

Si bien, los estudios de las Naciones Unidas para la agricultura y la alimentación estima que, para el año 2050 se requerirá del 70% más de alimentos en relación con el año 2006 [1], varios países líderes en investigación agrícola impulsan constantemente la investigación e implementación de la tecnología en este campo, con el fin de tecnificar los cultivos [2] [3].

El sector agrícola a nivel mundial es uno de los sectores en los cuales la toma de decisiones y la estimación precisa del rendimiento y la optimización de los cultivos dependen en gran medida de diversas variables, incluyendo factores ambientales, características del suelo [4] [5], la aplicación de productos químicos, y el control de la cantidad de agua [6]. Este sector desempeña un papel fundamental en la cadena de suministro de la industria, ya que los productos agrícolas se utilizan como materia prima en otras industrias [7]. Además, la calidad de los productos agrícolas influye directamente en las transacciones económicas, lo que repercute en la economía tanto del sector agrícola como en la de las personas relacionadas directa o indirectamente con él, como es el caso de los agronegocios [8]. Varios sectores como la economía, industria, producción, investigación, academia, emprendimiento, política, administración, deporte y otros, dependen directa o indirectamente del sector agrícola como fin, materia prima o derivados, por lo que la optimización, precisión, calidad son parámetros, que gracias al crecimiento de la ciencia de datos, permite incrementar la eficiencia o aportar a la productividad.

Otro aspecto de interés es la gestión de emergencias agrícolas [9] para mitigar las enfermedades en los cultivos, clasificación de muestras de cultivos [10], aguas servidas [11].

La aplicación de técnicas de aprendizaje automático o machine learning, ha revolucionado diversos sectores en los últimos años, y, la agricultura no es la excepción [12]: Agricultura Inteligente o Smart Agriculture [13], combinándose con IoT [14]. En un mundo en constante cambio y con una creciente demanda de alimentos, es fundamental buscar soluciones innovadoras que permitan optimizar la producción agrícola, mejorar la eficiencia y enfrentar los desafíos asociados a la gestión de los cultivos.

En este contexto, el machine learning ha demostrado ser una herramienta poderosa y prometedora para la agricultura moderna. Al aprovechar el poder del análisis de datos y la inteligencia artificial, esta disciplina ofrece una serie de beneficios que van desde la optimización de la producción hasta la toma de decisiones informadas. En esta discusión, examinaremos la relevancia de las técnicas de machine learning en la agricultura y su impacto en la transformación de este sector de vital importancia para la humanidad.

El propósito de esta investigación es realizar una revisión adecuada de literatura, para identificar las técnicas, el aporte, el área de aprendizaje automático en la búsqueda de obtener información útil para la toma de decisiones desde el enfoque de machine learning.

## **Materiales y métodos**

### **Metodología de investigación y análisis estadístico**

En esta investigación se adopta un enfoque cuantitativo, exploratorio y documental. Se utiliza la base de datos Scopus para identificar las técnicas de aprendizaje de máquina aplicadas en el sector de la agricultura. El objetivo principal es explorar y analizar el uso de dichas técnicas en el contexto agrícola, mediante el cumplimiento del protocolo de inclusión de publicaciones registradas en la base de datos Scopus, desde el año 2013 al 2023, de tipo artículo de revista y estados del arte (review) en idioma inglés.

### **Definiendo palabras clave**

En este estudio, se seleccionaron como palabras clave principales (i) "data mining techniques", (ii) "agriculture", para abarcar el uso de técnicas de minería de datos aplicadas en el ámbito de la agricultura.

### **Definiendo Ecuación de búsqueda**

Considerando las áreas de la presente investigación, la ecuación de búsqueda seleccionada fue:

```
( TITLE-ABS-KEY ( "data mining techniques" ) AND TITLE-ABS-KEY ( agriculture ) ) AND ( LIMIT-TO ( PUBYEAR , 2023 ) OR LIMIT-TO ( PUBYEAR , 2022 ) OR LIMIT-TO ( PUBYEAR , 2021 ) OR LIMIT-TO ( PUBYEAR , 2020 ) OR LIMIT-TO ( PUBYEAR , 2019 ) OR LIMIT-TO ( PUBYEAR , 2018 ) OR LIMIT-TO ( PUBYEAR , 2017 ) OR LIMIT-TO ( PUBYEAR , 2016 ) OR LIMIT-TO ( PUBYEAR , 2015 ) OR LIMIT-TO ( PUBYEAR , 2014 ) OR LIMIT-TO ( PUBYEAR , 2013 ) ) AND ( LIMIT-TO ( DOCTYPE , "ar" ) OR LIMIT-TO ( DOCTYPE , "re" ) ) AND ( LIMIT-TO ( LANGUAGE , "English" ) )
```

### **Análisis bibliométrico**

El análisis bibliométrico permite realizar un mapeo científico de las publicaciones [15] en las diferentes áreas de estudio, permite identificar de manera cuantitativa y descriptiva el interés y evolución que tiene una determinada área de estudio en el tiempo.

El objetivo que persigue esta sección es, identificar a los autores con mayores publicaciones en el área de minería de datos enfocado en la agricultura, por año, autores más citados, universidades, revistas y países con mayores publicaciones en el área de estudio.

A continuación, se describe la información de los artículos recopilados detallados en la Tabla 1.

**Tabla 1: Levantamiento de principal información de los artículos estudiados.**

Descripción	Resultados
<b>Principal información</b>	
Rango de años de estudio	2013:2023
Tipo de publicaciones	80
Artículos	106
Tasa de crecimiento anual %	4.14
Promedio documentos	3.99
Promedio de citación por artículo	13.71
Referencias	4298
<b>Contenido de documentos:</b>	
Palabras clave Plus (ID)	760
Palabras clave Autores (DE)	386
<b>Autores</b>	
Autores	361
Autoría única	6
Coautores por artículo	3.44
Coautores internacionales %	20.75
<b>Tipo de artículos</b>	
Artículos de revista	99
Review	7

*Basado en Bibliometrix y Scopus.*

### Publicaciones por año

En la Figura 1 se observa la variación por año de las publicaciones en el área de estudio, teniendo 22 publicaciones (2019), 18 publicaciones (2020 y 2022) y son los años con mayor crecimiento en publicaciones. Se evidencia el constante interés y marcada tendencia en esta área de conocimiento dentro de la comunidad científica y académica.



*Figura 1 Publicaciones por año.*

## Citas promedio por año

Con el propósito de examinar la calidad, impacto de la investigación, evolución de la influencia y reconocimiento de las publicaciones a lo largo del tiempo se identifica las citas promedio por año, siendo el 2015 el año con mayor número de citas. Ver Tabla 2.

**Tabla 2: Citas promedio por año**

<b>Año</b>	<b>Promedio de citas por artículo</b>	<b>N</b>	<b>Promedio de citas por año</b>	<b>Citas por año</b>
<b>2013</b>	1	2.00	0.09	11
<b>2014</b>	20	3.00	2.00	10
<b>2015</b>	38.17	6.00	4.24	9
<b>2016</b>	23.67	6.00	2.96	8
<b>2017</b>	26.7	10.00	3.81	7
<b>2018</b>	31.6	10.00	5.27	6
<b>2019</b>	6.82	22.00	1.36	5
<b>2020</b>	7.89	18.00	1.97	4
<b>2021</b>	11.38	8.00	3.79	3
<b>2022</b>	2.72	18.00	1.36	2
<b>2023</b>	1.67	3.00	1.67	1

*Basado en Bibliometrix*

## Revistas más relevantes

La evolución de la influencia y reconocimiento de las publicaciones en las principales revistas científicas a lo largo del tiempo se describe en la Figura 2. Se destaca la revista *Computer and Electronics in Agriculture* que se enfoca en la solución de problemas en las áreas de agricultura, agronomía, horticultura desde el enfoque de hardware y software [16], esta revista cuenta con un factor de impacto de cuartil Q1, de acuerdo con indicador SCImago [17]. En Scopus se registra desde el año 1985 hasta la actualidad [18]. En este estudio se identifican 6 publicaciones resultantes de la ecuación de búsqueda, registrados desde el año 2014 al 2022. Hasta la fecha existen 6763 publicaciones registradas en Scopus de esta revista.

Seguidamente se encuentra como segunda revista con mayores publicaciones en el área, la revista *International Journal of Innovative Technology and Exploring Engineering*, enfocada en soluciones de ingeniería y tecnología [19], cuenta con un factor de impacto de cuartil Q4 [19] de acuerdo con el indicador SCImago, en esta investigación se identifican 4 artículos publicados en esta revista. Mientras que, en Scopus se registra la primera publicación de esta revista desde el año 2018, y, hasta la actualidad cuenta con 8594 publicaciones registradas, debido a su periodicidad mensual y de acceso libre.

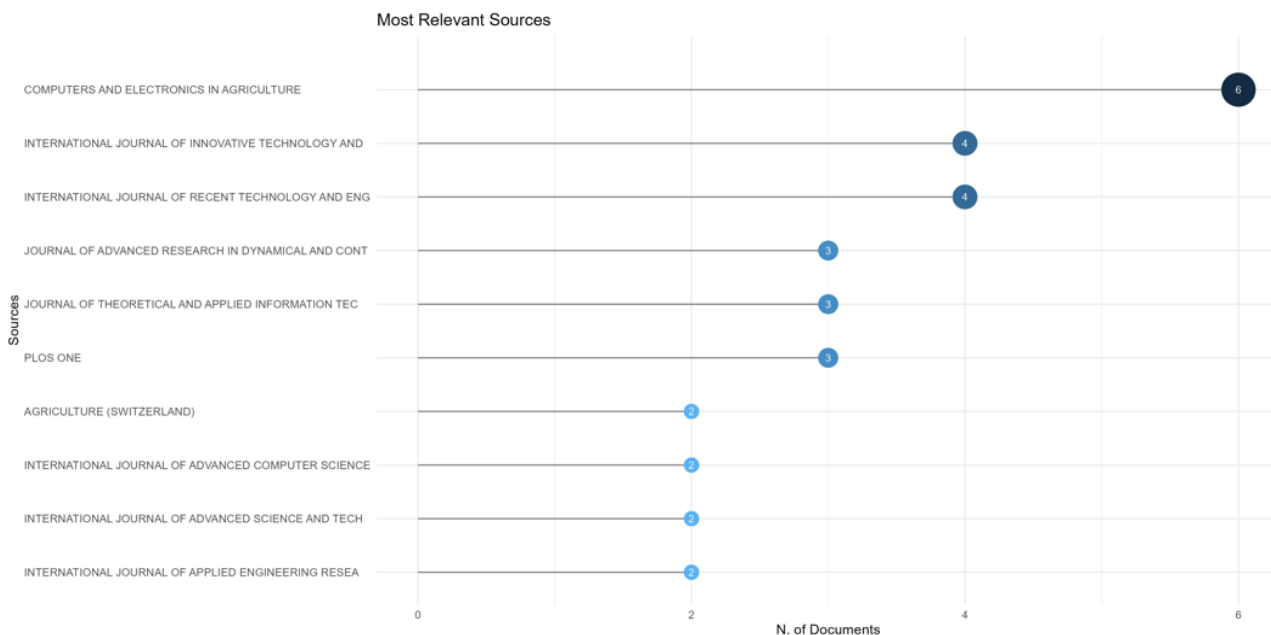


Figura 2: Revistas más relevantes Basado en Bibliometrix

### Revistas más citadas

A continuación, se presenta la evolución de la influencia y el reconocimiento de los trabajos en dichas revistas altamente citadas. Ver Figura 3 y Tabla 3.

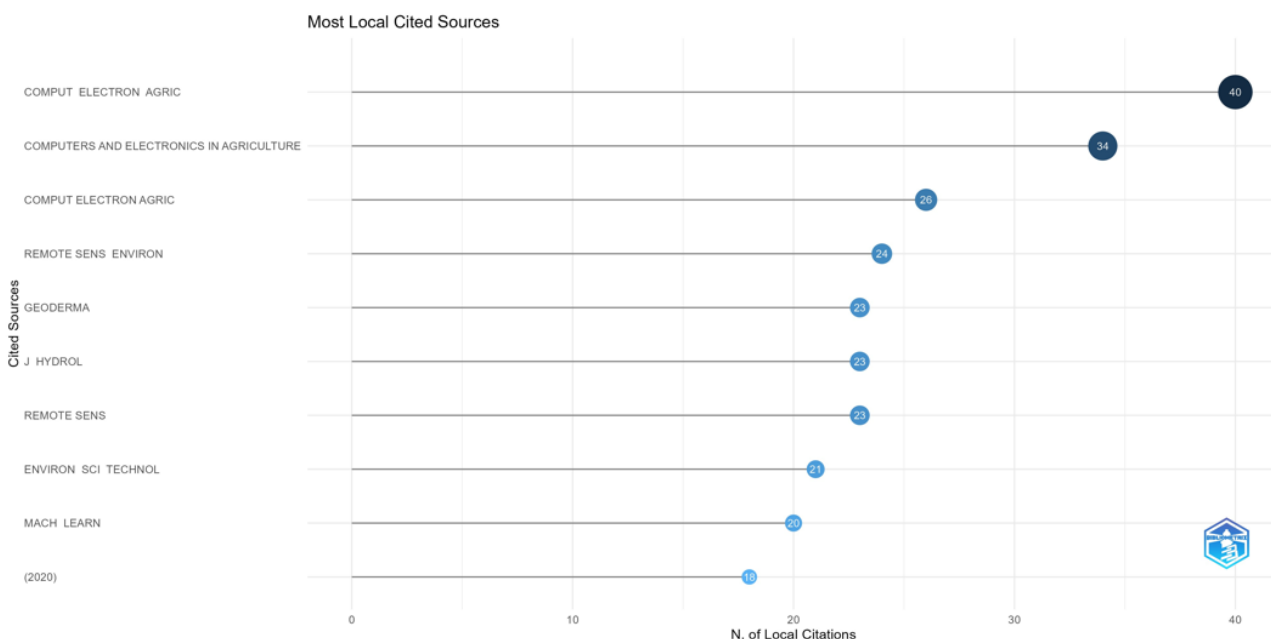


Figura 3: Revistas más citadas. Basado en Bibliometrix.

Para evaluar el impacto y la productividad se analizaron métricas como el h-index o también llamado Índice de Hirsch, que es la métrica más utilizada, se enfoca en la producción e impacto del investigador o investigadores [20]. Por otra parte, el Índice g-index es una mejora del índice h, se enfoca en los artículos altamente citados [20], mientras que el Índice m-index busca el equilibrio entre la productividad y el impacto.

Tabla 3: Análisis de indicadores por revistas.

Revistas	h_index	g_index	m_index	T o t a l Citas	NP	PY_start
Computers and electronics agriculture	5	6	0.5	124	6	2014
Plos one	3	3	0.333	139	3	2015
International journal of applied engineering research	2	2	0.2	10	2	2014
International journal of engineering and advanced technology	2	2	0.4	5	2	2019
International journal of innovative technology and exploring engineering	2	4	0.4	16	4	2019
Journal of big data	2	2	0.286	137	2	2017
Remote sensing	2	2	0.286	91	2	2017
Accounting	1	1	0.25	2	1	2020
A g r i c u l t u r e (switzerland)	1	2	0.5	4	2	2022
Applied soft computing	1	1	0.333	6	1	2021

Basado en Bibliometrix

### Autores más relevantes

La relevancia y la contribución de estos autores destacados a la comunidad académica y científica se analiza en la Tabla 4, la misma que presenta como autor más citado a Geeta M con 3 publicaciones en esta área de estudio.

Tabla 4: Autores más relevantes

Autores	Publicaciones	Publicaciones fraccionadas
Geetha Mcs	3	2.00
Elango Nm	2	1.00
Meganathan S	2	0.53
Aamir M	1	0.17
Abdelmajid B	1	0.25
Abel thangaraja G	1	0.50
Abuamra I	1	0.33
Abushawish Hf	1	0.33
Adil N	1	0.33
Adinarayana J	1	0.13

Basado en Bibliometrix

## Países de producción

La contribución de los autores destacados de diferentes países a la comunidad académica y científica se puede ver la Figura 4, se evidencia que la India es el país con mayor cantidad de publicaciones científicas 89, seguido por China con 18 publicaciones, mientras que Brasil 16 publicaciones, lo que permite visualizar el interés de este tipo de investigaciones.

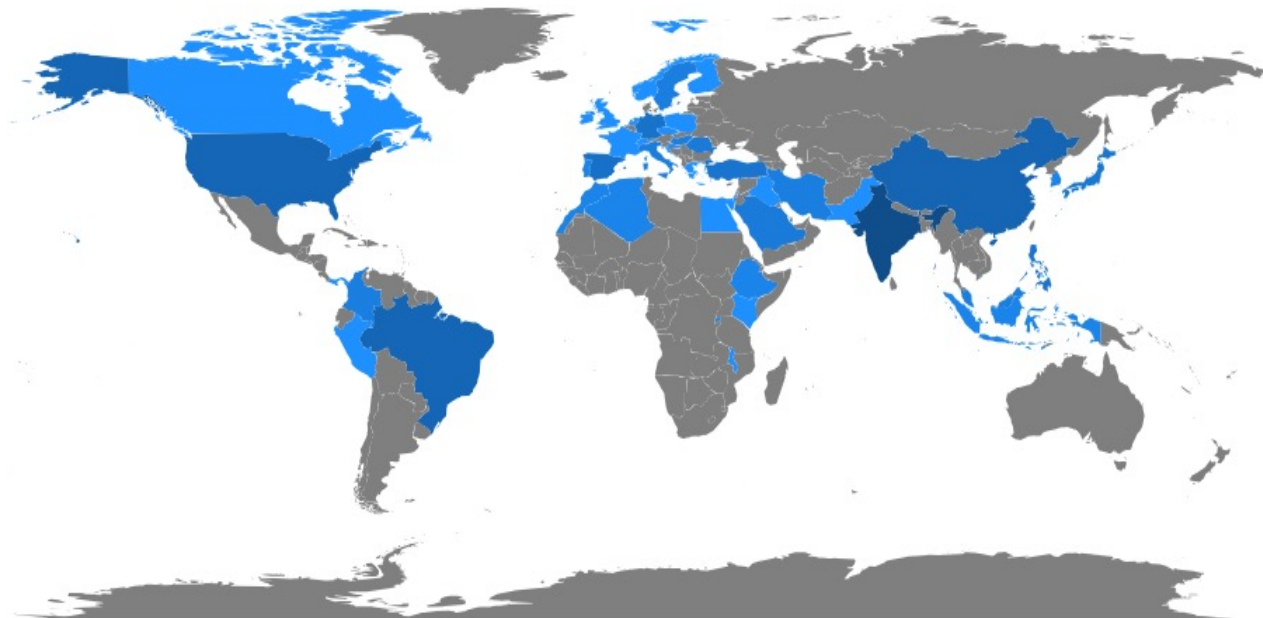


Figura 4: Producción científica por países

Basado en Bibliometrix

Región	Nro. Publicaciones
India	89
China	18
Brasil	16
EE.UU	16
España	8
Alemania	7
Filipinas	6
Turquía	6
Japón	5
Rumania	4

Basado en Bibliometrix



País	Año	Nro. Artículos	País	Año	Nro. Artículos	País	Año	Nro. Artículos
Brasil	2013	2	China	2017	2	E E UU.	2015	3
	2014	2		2018	8		2016	3
	2015	2		2019	8		2017	3
	2016	4		2020	9		2018	3
	2017	4		2021	10		2019	3
	2018	8		2022	18		2020	12
	2019	15		2023	18		2021	12
	2020	16		2014	6		2022	16
	2021	16		2015	10		2023	16
	2022	16		2016	10			
2023	16	2017	17					
España	2018	2	India	2018	24			
	2019	2		2019	48			
	2020	5		2020	57			
	2021	8		2021	63			
	2022	8		2022	87			
	2023	8		2023	89			

Basado en bibliometrix

## Análisis sistemático

En esta sección se realiza el estado del arte que contiene la clasificación de las diversas técnicas que se aplican en este sector, desde el área de conocimiento de minería de datos, así como las variables de interés y los enfoques de estudio.

## Taxonomía

Entre los principales enfoques de interés identificados se encuentran:

### Análisis del agua

#### Manejo inteligente de agua

Uno de los enfoques investigativos de interés es el Enrutamiento inteligente para manejo del agua con la propuesta de un Terrain based Routing, utilizando reglas difusas, se compara "Region Based Routing" y "Equalized Cluster Head Election Routing Protocol" [6]. En [21], se aborda el estudio de patrones de resistencia de sequía. Por otra parte, se realizan estudios de la calidad del agua (análisis hidrológico) en cuanto a la concentración de contaminantes nitrogenados [22].

### Patrones de lluvia

La investigación de [23] se enfoca en la predicción de la lluvia por medio de la recopilación de información de sensores remotos para el análisis de precipitación en zonas agrícolas, datos de cultivos y precipitación. En esta área se destaca [24], analiza técnicas estadísticas y de minería de datos con el objetivo de identificar predicciones precisas de lluvia. El conjunto de datos abarca un período de 31 años y se basa en observaciones realizadas en el área de Langat, lo que permite un análisis a largo plazo en esta región.

En este mismo contexto, [25] estudia la predicción de lluvias con Deep Convolutional Neural Network (DCNN), mientras que, en [26] analiza desde deep learning el nivel de agua de los cultivos.

### **Predicción de sequías**

[27] Random Subspace, Random Tree y Random Forest, Bagging, para predecir futuras sequías. Adicionalmente en [22] se presenta el estudio sobre predecir los niveles de nitrato, mientras que [28] aplica redes neuronales (ANN) para la predicción temprana de sequía utilizando datos de series temporales.

### **Crecimiento de cultivos**

El área de agricultura analiza el crecimiento de cultivos como arroz [29] [30], [31], [13], [32], [33] [34], así como alfalfa, algodón, Brassica napus, cacahuetes, col, caña de azúcar, hongos (comestibles o no comestibles), hojas de cebada, maíz, papas irlandesas, pepinos, trigo, soja, leguminosas utilizadas como abono verde, manzana: Amasya, Starking, Granny Smith, Pink Lady, Golden Delicious y Arapkız, Dendrobium officinale, que en la sección variables se detalla con mayor profundidad.

### **Análisis del suelo**

- En [35] estudia el empleo del índice de vegetación espectroscopía, la variabilidad de la biomasa de cuatro cultivos importantes (arroz, alfalfa, algodón, maíz) [30], así como estudios para predecir las emisiones CO<sub>2</sub> en el suelo en los cultivos de caña de azúcar [36]. En la investigación de [37] se enfoca en el estudio de los humedales y su monitorización. En la investigación de [38] analiza la predicción de la humedad del suelo de datos de series de tiempo, la diversidad de tipos de estudio de tipos de suelo con el apoyo de datos como imágenes satelitales analizando (i) monitoreo forestal, (ii) monitoreo de inundaciones, (iii) monitoreo macroeconómico / urbano [29].
- Mientras que existen estudios como [39] técnicas "omics" y minería de datos para el estudio de los metales pesados en el suelo agrícola.
- Segmentación del suelo [40] en las que dependiendo la zona de interés el suelo es segmentado, para este estudio se aplica el Coefficient, Jaccard Index, Sensitivity, Specificity, Precision.
- Monitoreo de deslizamientos de tierra [41] que se apoya con redes de sensores inalámbricos WSN.
- Predicción de futuras cosechas como papas irlandesas y maíz, analizando la lluvia y la temperatura como predictores [42].

### **Abono**

- Abono: el estudio descomposición de semivida de leguminosas utilizadas como abono verde es abordado por [43] con el enfoque de agricultura orgánica.
- Elección de fertilizantes y pesticidas [44].

## Análisis de plantas

### Enfermedades

- Hojas de cebada: En [45] estudia la técnica de factorización de matrices de tiempo lineal llamada "Simplex Volume Maximization" para determinar 3 tipos de enfermedades Pyrenophora teres, Puccinia hordei y Blumeria graminis hordei.
- Hojas de cacahuates: donde se identifica la temperatura mínima entre los 18 a 20°C, humedad de las hojas 7 a 10 horas, alta humedad 75% o mayor, aplica clasificación de Bayes ingenuo con distribución gaussiana, reglas de asociación, minería de regresión multivariable para identificar la relación del cultivo con el clima, el ambiente [46].
- Enfermedad cultivo en [9].

## Agricultura de precisión

La agricultura de precisión se enfoca en el estudio de cómo mejorar la productividad de los cultivos [47], entre los mayores exponentes se identifican:

- Cosecha robótica: Clasificador de píxeles SVM, aplicación de distancia euclidiana, segmentación transformación de watershed en cultivo de pepinos [48].
- Agricultura inteligente - smart farming: Estimación del riego de cultivo [49].
- Imagen de Teledetección Hiperespectral (HRS) para monitoreo de cultivo [50].
- Fenotipificación: Termografía para determinar el estrés biótico y abiótico como medios de afectación a las mediciones térmicas[31].

## Análisis de datos agrícolas para maximizar la producción de cultivos

### Maximización de producción de cultivo:

- Entre las técnicas para el analizar los datos agrícolas y encontrar parámetros óptimos que maximicen la producción de cultivos, en [7] estudia Partitioning Around Medoids (PAM), Clustering Large Applications (CLARA), Density-Based Spatial Clustering of Applications with Noise (DBSCAN) y Regresión Lineal Múltiple (RLM) para la maximización de cultivos en big data, considerando como variables: (i) condiciones del ambiente, (ii) variabilidad del suelo, (iii) niveles de insumos , (iv) precios de los productos agrícolas. Random Forest (RF) en el desarrollo y rendimiento de la caña de azúcar es estudiado en [51] identificando que la arcilla, cantidad de materia orgánica, intercambio catiónico como variables, inciden en el crecimiento.
- Adicionalmente, existen estudios donde se detecta el crecimiento de cultivos como es el caso de la col [26], que se basa en el estudio de análisis catastrales y de imágenes satelitales.

### Rendimiento de cultivo

- Mientras que [52] analiza el contexto general del aporte de la minería de datos en la optimización de cultivos en agricultura, [53] analiza el rendimiento de cultivos Fuzzy C Mean , Perceptrón Multicapa (Multilayer Perceptron – MLP). Así como [54] se apoya con minería de datos para el estudio, [55] se enfoca en la toma de decisiones [56], Predicción plan de cultivo para los agricultores [33], [57] [58]. En cuanto a [59], analiza el rendimiento de cultivo enfocándose en deep learning con el método QRECF-DFFMPC para refinar las precisiones.

- La evaluación de las relaciones: (i) cambios climáticos, (ii) variabilidad del rendimiento del cultivo a escala local, donde se identifica que, los factores climáticos afectan al desarrollo del cultivo, son abordadas en [13] con las técnicas Conditional Inference Forest y agrupamiento (clustering).
- Recomendaciones de cultivo [60] basándose en Navegación de usuarios, comportamiento de la comunidad de usuarios, enlaces en páginas web, o en su defecto, soluciones web basadas en Marketplace para este sector [61].
- Gestión de residuos de cultivos en el carbono orgánico del suelo (SOC) y los rendimientos de los cultivos [62].
- Control de plagas: en [63] se analiza la clasificación de plagas agrícolas.

## Otros

En la investigación de [64] se enfoca en el estudio de la construcción de un mapa de susceptibilidad a inundaciones, combinando técnicas de regresión logística (LR), Random Forest (RF) y Máquinas de Soporte Vectorial (SVM), evaluando (AUC), para el modelo "fuzzy WofE-SVM, del cual fuzzy WofE-SVM presentó un mejor rendimiento predictivo AUC 0.9865.

Software de agricultura inteligente [44], Soporte línea call center [65], Planificación agrícola del rendimiento de los cultivos [66], así como análisis de comercio agrícola internacional con datos comerciales [66], detección de fraudes en la industria agrícola [67], Cadena de valor [68] y otros.

## Tipos de cultivos

En este apartado se identifica los tipos de cultivos analizados en la producción científica identificada. Ver Tabla 5.

**Tabla 5: Tipos de cultivo**

<b>Cultivo</b>	<b>Autor</b>
Arroz	[30], [31], [13], [32], [33] Predecir el rendimiento del cultivo de arroz [29]
Alfalfa	[30]
Algodón	[30]
Brassica napus	[69]
Cacahuetes	[46] enfermedad
Col	Estudio del crecimiento [26]
Caña de azúcar	predecir las emisiones de CO2[51] [36] Recomendación de cultivo [60]
Hongos (comestibles o no comestibles)	Clasificación de tipos de hongos[70]
Hojas de cebada	Pyrenophora teres, Puccinia hordei y Blumeria graminis hordei [45]
Maíz	[30] Predicción de cosecha [42] índices de precios agrícolas [71]
Papas irlandesas	Predicción de cosecha [42]

Pepinos	[48]
Trigo	enfermedad (brotes de roya amarilla) [9]
Soja	índices de precios agrícolas [71]
Leguminosas utilizadas como abono verde	Descomposición de semivida de leguminosas utilizadas como abono verde [43]
Manzana: Amasya, Starking, Granny Smith, Pink Lady, Golden Delicious y Arapkiz	Predicción de las características de la manzana[41]
Dendrobium officinale	
Atributos de contratos de leasing	Otros análisis de fraude en empresas agrícolas [67]
Parámetros de cultivo y crecimiento de la vegetación	Monitoreo de crecimiento de cultivo, teledetección espectral [50]

### Variables de estudio y enfoque

A continuación, se describe las principales variables que identificadas en los 83 artículos científicos seleccionados. Ver Tabla 6.

**Tabla 6: Variables de estudio y enfoque**

Variable	Autor	Enfoque
Condiciones del ambiente	[7], [32], [46]	Análisis de crecimiento de cultivo
Condiciones bioclimáticas / climáticas	[9], [72] [73]	Predicción de emergencia agrícola
Condiciones topográficas	[9]	Predicción de emergencia agrícola, Inseguridad alimentaria
Condiciones edafológicas	[9]	Predicción de emergencia agrícola
Variabilidad del suelo	[7], [74]	Análisis de datos agrícolas Clasificación de datos agrícolas
Niveles de insumos	[7]	Análisis de datos agrícolas
Precios de los productos agrícolas	[7]	Análisis de datos agrícolas
Carbono orgánico del suelo (SOC)	[62]	Gestión de residuos de cultivos
Rendimientos de los cultivos	[62]	Gestión de residuos de cultivos
Agua	[6]	Enrutamiento inteligente
Datos meteorológicos	[33]	
Datos agronómicos	[33]	Inseguridad alimentaria
Situación socioeconómica	[72]	
Producción	[72]	
Comercio agrícola	[72]	

Lluvia		Datos de precipitación para estudio de eficiencia agrícola[23] Predicción de cosecha [42]
Cobertura de cultivos	[36]	
Laboreo mínimo	[36]	
Humedad del suelo	[36], [38] Crecimiento col [75]	Predecir las emisiones de CO2 del suelo
Temperatura del suelo (Ts)	[36], Crecimiento col [75]	
Precipitación, pH	[36]	
Carbono orgánico	[36]	
Nivel concentración de Nitrato	[22]	
Carga	[22]	
Tipo de cuenca	[22]	Calidad de Agua
Uso de tierra	[22]	
Rendimiento	[22]	
Evapotranspiración	[49]	
Temperatura	[49], Predicción de cosecha [42]	
Transferencia de masa	[49]	Predicción del nivel de humedad
Radiación	[49]	
Variables meteorológicas principales	[49]	
Tiempos de descomposición de semivida de leguminosas	[76]	Descomposición abono verde
Ancho de la cavidad del tallo	[41]	
Profundidad de la cavidad del tallo	[41]	Predicción de las propiedades de la manzana
Ancho del cuenco del ojo	[41]	
Profundidad del cuenco del ojo	[41]	
Forma	[77]	
Color	[77]	Identificación varios tipos de trigo
textura	[77]	
Crecimiento de las plantas	[73]	
Datos agrícolas no detallados	[56]	

Datos de llamadas de consulta en el Kisan área de centro De Atención Telefónica [65]	Predicción de llamadas telefónicas a línea de agricultura centro de atención telefónica
Documentos de plataforma tecnológicas [57]	Documentos tecnológicos de plataformas agricultura sostenible
Navegación de usuarios [60]	
Comportamiento de la comunidad de usuarios [60]	Recomendación cultivo [60]
Enlaces en páginas web [60]	
Datos de micronutrientes (Fe, Mn, Zn, Cu), crecimiento agrícola [75]	Crecimiento de cultivo
Velocidad del viento [75]	
Duración de la luz solar [75]	Crecimiento col
Radiación [75]	

### Técnicas de minería de datos

Seguidamente se presenta la contribución de las técnicas de minería de datos identificadas en este estudio. Ver Tabla 7.

**Tabla 7: Técnicas de minería de datos**

<b>Técnicas de minería de datos</b>	<b>Autor y enfoque</b>
Regresión Logística (LR)	[64], [78] Predicción de emisiones de CO2 [36]
Support Vector Machine (SVM).	[64], imágenes [48], [74] y variante H-SVM, [79], [78]
Decision Tree	[74], [33], [27], [78] [37] Predicción de emisiones de CO2 [36] Descomposición de abono verde [43] Toma de decisiones datos agrícolas [56] Aprendizaje de cultivo <i>Dendrobium officinale</i> [73]
Random Forest (RF)	[64], [33], [51], [37] Predicción de cosecha [42] Toma de decisiones datos agrícolas [56]
Gradient Boosted Trees	Toma de decisiones datos agrícolas [56]
Support Vector Regressor-	Predicción de cosecha [42] Predicción llamadas línea soporte agricultura [65]
Combina con el enfoque de imagen hiperespectral	Teledetección hiperespectral para monitoreo de cultivos [45], [50]
Conditional Inference Forest y	[32]

Agrupamiento (clustering).	[32], [3]. Predicción de lluvia [24] Agglomerative Hierarchical Clustering
Region based routing	[6]
Equalized cluster head election routing protocol	[6]
Terrain based Routing using Fuzzy rules for precision agriculture"	[6], [53]
Segmentación	[48], [40] Segmento de tiempo [73]
Clasificación de Bayes ingenuo con distribución gaussiana	[46]
Reglas de asociación	[46], [3] Descomposición de abono verde y Redes de reglas de asociación [43] Recomendación cultivo caña de azúcar [60]
Minería de regresión multivariable)	[46]
Ensemble learning	[9]
Red neuronal recurrente espacio-temporal (STRNN)	[9]
Deep learning	Toma de decisiones de datos agrícolas [56] Rendimiento de cultivo - QRECF-DFFMPC [59] Estudio del crecimiento de la col [26]
Sistemas de recomendación	Recomendación cultivo caña de azúcar [60]
K-NN	[74], [37], [78] , [3], Calidad de agua [22] Predicción del nivel de humedad [38]
Naive bayes ( nb)	[74], [78] Mejorar la calidad de los cultivos [58] clasificación de hongos comestibles o no comestibles [70] Rendimiento del cultivo de arroz [29]
Bayes net	clasificación de hongos comestibles o no comestibles [70] [80]
ZeroR	clasificación de hongos comestibles o no comestibles [70]
K-means clustering	[33]
Bagged-C4.5	[79]
Regresión lineal	[72], Calidad de agua [22]
ARIMA (Autoregressive Integrated Moving Average)	[81]
Bagging (BG)	[27] Predicción de emisiones de CO2 [36]
Random Subspace (RSS)	[27]



Random Tree (RT)	[27]
Standardized Precipitation Index (SPI-3 and SPI-12)	Calidad de agua. REP Tree [22]
Bayesian Models	[27]
Redes Neuronales Artificiales (ANN)	Predicción de emisiones de CO2 [36] [37], [3] Humedad del suelo [38] Rendimiento de cultivo [55], [55] Predicción de lluvias/sequias [28]
Multilayer Perceptron Design (MLP)	[53], [3], Calidad de agua [22] Predicción llamadas línea soporte agricultura [65]
Random Forest (RF)	[3]Calidad de agua [22] Planificación agrícola [66]
K-Means,	[3], Predicción de lluvia K-Means Partitional Clustering, [24]
Minería espacial	[23]
Clasificación (general)	[3]
Regresión Polinómica	Predicción de cosecha [42]
Visualización	[3]Identificación de trigo [77]
Self-Organising Map	Predicción de lluvia [24] Análisis de fraudes empresas agrícolas [67]
Descomposición de Tucker	[71]
Find Law	Predicción características de manzana [76]
Minería de Datos Difusos, NS-2	Deslizamiento de tierra [41] Predicción del nivel del agua [26]
Deep Convolutional Neural Network (DCNN)	Predicción de lluvia [25]
Improved Particle Swarm Optimization (IPSO)	Predicción de lluvia [25]
A priori	Rendimiento del cultivo de arroz [29]
Generalized Linear Model	Toma de decisiones de datos agrícolas [56]
Long Short-Term Memory Networks, Gated Recurrent	Predicción llamadas línea soporte agricultura [65]
CART	Planificación Comercio agrícola internacional [66]
BERT (bidirectional encoder representations from transformers)	Planificación Comercio agrícola internacional [66]
Web scraping	Clasificación de documentos para promoción de agricultura orgánica [57]
Minería de datos web	
Teledetección Hiperespectral (HRS)	Monitoreo de cultivo [50]

J48	Rendimiento de cultivo de arroz [29]
Series temporales	[34]
Map reduce	Clasificación de plagas de cultivos [63]
C5.0	Clasificación de plagas de cultivos [63]
Exploratory Factor Analysis (EFA)	[69]
Kaiser Criterion	[69]
Cattell's Scree test	[69]
Monte Carlo Parallel Analysis	[69]
Principal Component Analysis (PCA) Total Cumulative Variance Explained	[69], [37]

## Conclusiones

El estudio ha permitido destacar los hallazgos y contribuciones principales derivados de su investigación. Se ha evidenciado el valor y la relevancia de la minería de datos en el ámbito de la agricultura, mostrando cómo estas técnicas pueden mejorar la productividad y eficiencia en el negocio agrícola. Además, proporcionan información valiosa para la toma de decisiones y la gestión de recursos.

Asimismo, se ha subrayado la importancia de los análisis bibliométricos para identificar las tendencias y actores clave en este campo de estudio, lo que orienta futuras investigaciones.

El uso de la minería de datos y técnicas de análisis en la agricultura ofrece información valiosa para la toma de decisiones y la optimización de recursos. Los estudios revisados han demostrado que la aplicación de estas técnicas ha llevado a mejoras significativas en la predicción de cosechas, la gestión del agua y la detección de cambios climáticos, contribuyendo a una agricultura más sostenible y productiva.

La combinación de datos de satélite, información catastral y técnicas de aprendizaje ha demostrado ser efectiva para predecir el crecimiento de cultivos y el nivel de agua en la agricultura. Estos enfoques han brindado resultados prometedores y pueden ser herramientas valiosas para la gestión eficiente de recursos en el sector agrícola.

## Referencias

1. Choudhury, A., Biswas, A., Prateek, M., Chakrabarti, A.: *Agricultural Informatics: Automation Using the IoT and Machine Learning*. Wiley (2021). <https://doi.org/10.1002/9781119769231>.
2. Castillejo, P., Johansen, G., Cürüklü, B., Bilbao-Arechabala, S., Fresco, R., Martínez-Rodríguez, B., Pomante, L., Rusu, C., Martínez-Ortega, J.-F., Centofanti, C., Hakojärvi, M., Santic, M., Häggman, J.: *Aggregate Farming in the Cloud: The AFarCloud ECSEL project*. *Microprocess. Microsyst.* 78, (2020). <https://doi.org/10.1016/j.micpro.2020.103218>.
3. Kavitha, G., Elango, N.M.: *An overview of data mining techniques and its applications*. *Int. J. Civ. Eng. Technol.* 8, 1013–1020 (2017).

4. Kommineni, M., Perla, S., Yedla, D.B.: A survey of using data mining techniques for soil fertility. *Int. J. Eng. Technol.* 7, 917–918 (2018).
5. Belinda, M.J.C.M., Umamaheswari, R., David, S.A.: Study of high yielding crops cultivation in India using data mining techniques. *Int. J. Eng. Technol.* 7, 121–124 (2018).
6. Pandiyaraju, V., Logambigai, R., Ganapathy, S., Kannan, A.: An Energy Efficient Routing Algorithm for WSNs Using Intelligent Fuzzy Rules in Precision Agriculture. *Wirel. Pers. Commun.* 112, 243–259 (2020). <https://doi.org/10.1007/s11277-020-07024-8>.
7. Majumdar, J., Naraseeyappa, S., Ankalaki, S.: Analysis of agriculture data using data mining techniques: application of big data. *J. Big Data.* 4, (2017). <https://doi.org/10.1186/s40537-017-0077-4>.
8. Vandna, Bansal, K.L.: Data mining techniques for increasing smart farming in agrarian sector. *Int. J. Sci. Technol. Res.* 9, 3555–3562 (2020).
9. Xu, W., Wang, Q., Chen, R.: Spatio-temporal prediction of crop disease severity for agricultural emergency management based on recurrent neural networks. *Geoinformatica.* 22, 363–381 (2018). <https://doi.org/10.1007/s10707-017-0314-1>.
10. de Lima, M.D., Barbosa, R.: Methods of authentication of food grown in organic and conventional systems using chemometrics and data mining algorithms: A review. *Food Anal. Methods.* 12, 887–901 (2019). <https://doi.org/10.1007/s12161-018-01413-3>.
11. Organización de las Naciones Unidas para la Alimentación y la Agricultura: Ecuador en una mirada | FAO en Ecuador | Organización de las Naciones Unidas para la Alimentación y la Agricultura, <https://www.fao.org/ecuador/fao-en-ecuador/ecuador-en-una-mirada/es/>, last accessed 2022/10/02.
12. Cutamora, J.C., Padua, R.: An Economic Rationalization Framework for Higher Education. *Recoletos Multidiscip. Res. J.* 8, 43–65 (2020). <https://doi.org/10.32871/rmrj2008.01.04>.
13. Ait Issad, H., Aoudjit, R., Rodrigues, J.J.P.C.: A comprehensive review of Data Mining techniques in smart agriculture. *Eng. Agric. Environ. Food.* 12, 511–525 (2019). <https://doi.org/10.1016/j.eaef.2019.11.003>.
14. Sunhare, P., Chowdhary, R.R., Chattopadhyay, M.K.: Internet of things and data mining: An application oriented survey. *J. King Saud Univ. - Comput. Inf. Sci.* 34, 3569–3590 (2022). <https://doi.org/10.1016/j.jksuci.2020.07.002>.
15. Aria, M., Cuccurullo, C.: bibliometrix: An R-tool for comprehensive science mapping analysis. *J. Informetr.* 11, 959–975 (2017). <https://doi.org/10.1016/J.JOI.2017.08.007>.
16. Computers and Electronics in Agriculture | Journal | ScienceDirect.com by Elsevier, <https://www.sciencedirect.com/journal/computers-and-electronics-in-agriculture>, last accessed 2023/09/09.

17. Computers and Electronics in Agriculture, <https://www.scimagojr.com/journalsearch.php?q=30441&tip=sid&clean=0>, last accessed 2023/09/09.
18. Scopus - Computers and Electronics in Agriculture | Signed in, <https://www.scopus.com/sourceid/30441>, last accessed 2023/09/09.
19. International Journal of Innovative Technology and Exploring Engineering, <https://www.scimagojr.com/journalsearch.php?q=21100889409&tip=sid&clean=0>, last accessed 2023/09/11.
20. Harzing, A.W., Alakangas, S.: LibGuides: Impact Metrics: Author Impact: h-index, g-index... *Scientometrics*. 106, 787–804 (2016). <https://doi.org/10.1007/S11192-015-1798-9>.
21. Nasrnia, F., Ashktorab, N.: Sustainable livelihood framework-based assessment of drought resilience patterns of rural households of Bakhtegan basin, Iran. *Ecol. Indic.* 128, (2021). <https://doi.org/10.1016/j.ecolind.2021.107817>.
22. Li, S., Bhattarai, R., Cooke, R.A., Verma, S., Huang, X., Markus, M., Christianson, L.: Relative performance of different data mining techniques for nitrate concentration and load estimation in different type of watersheds. *Environ. Pollut.* 263, (2020). <https://doi.org/10.1016/j.envpol.2020.114618>.
23. Kumar, Y.J.N., Kanth, T.V.R.: GIS-MAP based spatial analysis of rainfall data of Andhra Pradesh and telangana states using R. *Int. J. Electr. Comput. Eng.* 7, 460–468 (2017). <https://doi.org/10.11591/ijece.v7i1.pp460-468>.
24. Othman, Z.A., Ismail, N., Hamdan, A.R., Sammour, M.A.: Klang vally rainfall forecasting model using time series data mining technique. *J. Theor. Appl. Inf. Technol.* 92, 372–379 (2016).
25. Refonaa, J., Lakshmi, M.: Accurate prediction of the rainfall using convolutional neural network and parameters optimization using improved particle swarm optimization. *J. Adv. Res. Dyn. Control Syst.* 11, 318–328 (2019).
26. Hsiou, D.-C., Huang, F., Tey, F.J., Wu, T.-Y., Lee, Y.-C.: An Automated Crop Growth Detection Method Using Satellite Imagery Data. *Agric.* 12, (2022). <https://doi.org/10.3390/agriculture12040504>.
27. Mohammed, S., Elbeltagi, A., Bashir, B., Alsafadi, K., Alsilibe, F., Alsalman, A., Zeraatpisheh, M., Széles, A., Harsányi, E.: A comparative analysis of data mining techniques for agricultural and hydrological drought prediction in the eastern Mediterranean. *Comput. Electron. Agric.* 197, (2022). <https://doi.org/10.1016/j.compag.2022.106925>.
28. Sheik Mohideen Shah, S., Meganathan, S., Kamali, A.: Soft computing research for weather prediction using multilayer architecture. *Int. J. Eng. Adv. Technol.* 8, 3779–3783 (2019). <https://doi.org/10.35940/ijeat.F9390.088619>.
29. Vinoth, B., Elango, N.M.: An effective data mining techniques based optimal paddy yield cultivation: a rational approach. *Paddy Water Environ.* 19, 331–343 (2021). <https://doi.org/10.1007/s10333-021-00845-8>.

30. Marshall, M., Thenkabail, P.: Advantage of hyperspectral EO-1 Hyperion over multispectral IKONOS, GeoEye-1, WorldView-2, Landsat ETM+, and MODIS vegetation indices in crop biomass estimation. *ISPRS J. Photogramm. Remote Sens.* 108, 205–218 (2015). <https://doi.org/10.1016/j.isprsjprs.2015.08.001>.
31. Pineda, M., Barón, M., Pérez-Bueno, M.-L.: Thermal imaging for plant stress detection and phenotyping. *Remote Sens.* 13, 1–21 (2021). <https://doi.org/10.3390/rs13010068>.
32. Delerce, S., Dorado, H., Grillon, A., Rebolledo, M.C., Prager, S.D., Patiño, V.H., Varón, G.G., Jiménez, D.: Assessing weather-yield relationships in rice at local scale using data mining approaches. *PLoS One.* 11, (2016). <https://doi.org/10.1371/journal.pone.0161620>.
33. Arumugam, A.: A predictive modeling approach for improving paddy crop productivity using data mining techniques. *Turkish J. Electr. Eng. Comput. Sci.* 25, 4777–4787 (2017). <https://doi.org/10.3906/elk-1612-361>.
34. Ahmed, G.N., Kamalakannan, S., Kavitha, P.: A Machine Learning Approach for Stochastic Pattern Analysis for the Measurement of Time-Series Datasets. *Instrum. Mes. Metrol.* 21, 199–205 (2022). <https://doi.org/10.18280/i2m.210505>.
35. Gholizadeh, A., Carmon, N., Klement, A., Ben-Dor, E., Borůvka, L.: Agricultural soil spectral response and properties assessment: Effects of measurement protocol and data mining technique. *Remote Sens.* 9, (2017). <https://doi.org/10.3390/rs9101078>.
36. Farhate, C.V. V, De Souza, Z.M., De Medeiros Oliveira, S.R., Tavares, R.L.M., Carvalho, J.L.N.: Use of data mining techniques to classify soil CO<sub>2</sub> emission induced by crop management in sugarcane field. *PLoS One.* 13, (2018). <https://doi.org/10.1371/journal.pone.0193537>.
37. Radhika, A., Masood, M.S.: Effective dimensionality reduction by using soft computing method in data mining techniques. *Soft Comput.* 25, 4643–4651 (2021). <https://doi.org/10.1007/s00500-020-05474-7>.
38. Kiruthika, V.G., Arutchudar, V., Senthil Kumar, P.: Highest humidity prediction using data mining techniques. *Int. J. Appl. Eng. Res.* 9, 3259–3264 (2014).
39. Uchimiya, M., Bannon, D., Nakanishi, H., McBride, M.B., Williams, M.A., Yoshihara, T.: Chemical Speciation, Plant Uptake, and Toxicity of Heavy Metals in Agricultural Soils. *J. Agric. Food Chem.* 68, 12856–12869 (2020). <https://doi.org/10.1021/acs.jafc.0c00183>.
40. Prathik, A., Anuradha, J., Uma, K.: A novel algorithm for soil image segmentation using color and region based system. *Int. J. Innov. Technol. Explor. Eng.* 8, 3544–3550 (2019). <https://doi.org/10.35940/ijitee.J9762.0881019>.
41. Sumathi, M.S., Anitha, G.S.: Energy efficient wireless sensor network with efficient data handling for real time landslide monitoring system using fuzzy data mining technique. *Int. J. Mob. Netw. Des. Innov.* 8, 179–193 (2018). <https://doi.org/10.1504/IJMNDI.2018.093701>.

42. Kuradusenge, M., Hitimana, E., Hanyurwimfura, D., Rukundo, P., Mtonga, K., Mukasine, A., Uwitonze, C., Ngabonziza, J., Uwamahoro, A.: Crop Yield Prediction Using Machine Learning Models: Case of Irish Potato and Maize. *Agric.* 13, (2023). <https://doi.org/10.3390/agriculture13010225>.
43. Calçada, D.B., Rezende, S.O., Teodoro, M.S.: Analysis of green manure decomposition parameters in northeast Brazil using association rule networks. *Comput. Electron. Agric.* 159, 34–41 (2019). <https://doi.org/10.1016/j.compag.2019.02.013>.
44. Chandak, P.P., Agrawal, A.J.: Smart farming system using data mining. *Int. J. Appl. Eng. Res.* 12, 2788–2791 (2017).
45. Wahabzada, M., Mahlein, A.-K., Bauckhage, C., Steiner, U., Oerke, E.-C., Kersting, K.: Metro maps of plant disease dynamics-automated mining of differences using hyperspectral images. *PLoS One.* 10, (2015). <https://doi.org/10.1371/journal.pone.0116902>.
46. Tripathy, A.K., Adinarayana, J., Vijayalakshmi, K., Merchant, S.N., Desai, U.B., Ninomiya, S., Hirafuji, M., Kiura, T.: Knowledge discovery and Leaf Spot dynamics of groundnut crop through wireless sensor network and data mining techniques. *Comput. Electron. Agric.* 107, 104–114 (2014). <https://doi.org/10.1016/j.compag.2014.05.009>.
47. Krishna Priya, C.B., Venkateswari, S.: Analysis of different clustering algorithms on management zones in precision agriculture. *J. Adv. Res. Dyn. Control Syst.* 11, 489–494 (2019).
48. Fernández, R., Montes, H., Surdilovic, J., Surdilovic, D., Gonzalez-De-Santos, P., Armada, M.: Automatic detection of field-grown cucumbers for robotic harvesting. *IEEE Access.* 6, 35512–35526 (2018). <https://doi.org/10.1109/ACCESS.2018.2851376>.
49. Martín, J., Sáez, J.A., Corchado, E.: On the suitability of stacking-based ensembles in smart agriculture for evapotranspiration prediction. *Appl. Soft Comput.* 108, (2021). <https://doi.org/10.1016/j.asoc.2021.107509>.
50. Yu, H., Kong, B., Hou, Y., Xu, X., Chen, T., Liu, X.: A critical review on applications of hyperspectral remote sensing in crop monitoring. *Exp. Agric.* 58, (2022). <https://doi.org/10.1017/S0014479722000278>.
51. Sanches, G.M., Graziano Magalhães, P.S., Junqueira Franco, H.C.: Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. *Geoderma.* 334, 90–98 (2019). <https://doi.org/10.1016/j.geoderma.2018.07.051>.
52. Saritha, S., Abel Thangaraja, G.: CROP YIELD PREDICTION IN BIG DATA USING MARGALEF KERNEL PERCEPTRON BASED WINNOWER BOOST CLASSIFIER. *J. Theor. Appl. Inf. Technol.* 101, 2091–2107 (2023).
53. Geetha, M.C.S., Elizabeth Shanthi, I.: Forecasting the crop yield production in trichy district using fuzzy C-Means Algorithm and Multilayer Perceptron (MLP). *Int. J. Knowl. Syst. Sci.* 11, 83–98 (2020). <https://doi.org/10.4018/IJKSS.2020070105>.
54. Maury, R.K., Yadav, S.K., Sharma, T.K.: Estimation of major agricultural crop with effective yield prediction using data mining. *Int. J. Innov. Technol. Explor. Eng.* 8, 170–174 (2019).

55. Chiche, A.: Hybrid decision support system framework for crop yield prediction and recommendation. *Int. J. Comput.* 18, 181–190 (2019).
56. Adil, N., Dewangan, S., Sharma, K.: Efficient classification and regression techniques to predict crop yield. *Int. J. Sci. Technol. Res.* 8, 378–382 (2019).
57. Blazquez, D., Domenech, J., Garcia-Alvarez-Coque, J.-M.: Assessing technology platforms for sustainability with web data mining techniques. *Sustain.* 10, (2018). <https://doi.org/10.3390/su10124497>.
58. Patil, N.N., Saiyyad, M.A.M.: Machine learning technique for crop recommendation in agriculture sector. *Int. J. Eng. Adv. Technol.* 9, 1359–1363 (2019). <https://doi.org/10.35940/ijeat.A1171.109119>.
59. Sivanantham, V., Sangeetha, V., Alnuaim, A.A., Hatamleh, W.A., Anilkumar, C., Hatamleh, A.A., Sweidan, D.: Quantile correlative deep feedforward multilayer perceptron for crop yield prediction. *Comput. Electr. Eng.* 98, (2022). <https://doi.org/10.1016/j.compeleceng.2022.107696>.
60. de Barros, F.M.M., Oliveira, S.R.M., de Oliveira, L.H.M.: Development and validation of a recommender system for technological information on sugarcane [Desenvolvimento e validação de um sistema de recomendação de informações tecnológicas sobre cana-de-açúcar]. *Bragantia.* 72, 387–395 (2013). <https://doi.org/10.1590/brag.2013.049>.
61. Kaviarasan, S., Vanitha, M.: E-farming management system using data mining techniques. *Int. J. Intell. Unmanned Syst.* 10, 257–266 (2022). <https://doi.org/10.1108/IJIUS-05-2020-0018>.
62. Warren Raffa, D., Bogdanski, A., Tittonell, P.: How does crop residue removal affect soil organic carbon and yield? A hierarchical analysis of management and environmental factors. *Biomass and Bioenergy.* 81, 345–355 (2015). <https://doi.org/10.1016/j.biombioe.2015.07.022>.
63. Revathy, R., Balamurali, S., Lawrance, R.: Classifying agricultural crop pest data using hadoop MapReduce based C5.0 algorithm. *J. Cyber Secur. Mobil.* 8, 393–408 (2019). <https://doi.org/10.13052/jcsm2245-1439.835>.
64. Hong, H., Tsangaratos, P., Ilia, I., Liu, J., Zhu, A.-X., Chen, W.: Application of fuzzy weight of evidence and data mining techniques in construction of flood susceptibility map of Poyang County, China. *Sci. Total Environ.* 625, 575–588 (2018). <https://doi.org/10.1016/j.scitotenv.2017.12.256>.
65. Godara, S., Toshniwal, D.: Deep Learning-based query-count forecasting using farmers' helpline data. *Comput. Electron. Agric.* 196, (2022). <https://doi.org/10.1016/j.compag.2022.106875>.
66. Suarez, A.J.B., Singh, B., Almkhtar, F.H., Kler, R., Vyas, S., Kaliyaperumal, K.: Identifying Smart Strategies for Effective Agriculture Solution Using Data Mining Techniques. *J. Food Qual.* 2022, (2022). <https://doi.org/10.1155/2022/6600049>.
67. Bach, M.P., VlahoviÄ‡, N., Pivar, J.: Fraud Prevention in the Leasing Industry Using the Kohonen Self-Organising Maps. *Organizacija.* 53, 128–145 (2020). <https://doi.org/10.2478/orga-2020-0009>.

68. Ngabalin, A.M.: An investigation on value chain cooperation attributes in fisheries micro-enterprises. *Accounting*. 6, 301–306 (2020). <https://doi.org/10.5267/j.ac.2020.2.005>.
69. Iantovics, L.B., Rotar, C., Morar, F.: Survey on establishing the optimal number of factors in exploratory factor analysis applied to data mining. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 9, (2019). <https://doi.org/10.1002/widm.1294>.
70. Beniwal, S., Das, B.: Mushroom classification using data mining techniques. *Int. J. Pharma Bio Sci.* 6, B1170–B1176 (2015).
71. Correa, F.E., Oliveira, M.D.B., Gama, J., Corrêa, P.L.P., Rady, J.: Analyzing the behavior dynamics of grain price indexes using Tucker tensor decomposition and spatio-temporal trajectories. *Comput. Electron. Agric.* 120, 72–78 (2016). <https://doi.org/10.1016/j.compag.2015.11.011>.
72. Shortridge, J.E., Falconi, S.M., Zaitchik, B.F., Guikema, S.D.: Climate, agriculture, and hunger: statistical prediction of undernourishment using nonlinear regression and data-mining techniques. *J. Appl. Stat.* 42, 2367–2390 (2015). <https://doi.org/10.1080/02664763.2015.1032216>.
73. Sun, L., Zheng, Z., Zhu, J.: Mining spatio-temporal knowledge of climate for dendrobium officinale in greenhouse cultivation. *Recent Adv. Electr. Electron. Eng.* 11, 160–166 (2017). <https://doi.org/10.2174/2352096510666170921162448>.
74. Shastry, K.A., Sanjay, H.A., Deexith, G.: Quadratic-radial-basis-function-kernel for classifying multi-class agricultural datasets with continuous attributes. *Appl. Soft Comput. J.* 58, 65–74 (2017). <https://doi.org/10.1016/j.asoc.2017.04.049>.
75. Rajagopal, M., Ponnuchamy, M., Kapoor, A.: Water management for irrigation scheduling by computing evapotranspiration using ANFIS modelling. *Desalin. Water Treat.* 251, 123–133 (2022). <https://doi.org/10.5004/dwt.2022.28290>.
76. Demir, B., Gurbuz, F., Eski, I., Kus, Z.A., Yilmaz, K.U., Ercisli, S.: Possible Use of Data Mining for Analysis and Prediction of Apple Physical Properties [Die Anwendung statistischer Methoden (Data-Mining) zur Analyse und Prognose physikalischer Eigenschaften bei Apfel]. *Erwerbs-Obstbau.* 60, 1–7 (2018). <https://doi.org/10.1007/s10341-017-0330-1>.
77. Ullah, A., Mohd Nawari, N., Arifianto, A., Ahmed, I., Aamir, M., Khan, S.N.: Real-time wheat classification system for selective herbicides using broad wheat estimation in deep neural network. *Int. J. Adv. Sci. Eng. Inf. Technol.* 9, 153–158 (2019). <https://doi.org/10.18517/ijaseit.9.1.5031>.
78. Rahman, N.A.B.A., Tan, K.L., Lim, C.K.: Supervised and unsupervised learning in data mining for employment prediction of fresh graduate students. *J. Telecommun. Electron. Comput. Eng.* 9, 155–161 (2017).
79. Mousavizadegan, M., Mohabatkar, H.: An evaluation on different machine learning algorithms for classification and prediction of antifungal peptides. *Med. Chem. (Los Angeles)*. 12, 795–800 (2016). <https://doi.org/10.2174/1573406412666160229150823>.



80. Krause, P.J., Bokinala, V.: A tutorial on data mining for Bayesian networks, with a specific focus on IoT for agriculture. *Internet of Things (Netherlands)*. 22, (2023). <https://doi.org/10.1016/j.iot.2023.100738>.
81. Shivhare, N., Rahul, A.K., Dwivedi, S.B., Dikshit, P.K.S.: ARIMA based daily weather forecasting tool: A case study for Varanasi. *Mausam*. 70, 133–140 (2019).