



UNIVERSIDAD TÉCNICA DEL NORTE

FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS

**CARRERA DE INGENIERÍA EN ELECTRÓNICA Y REDES DE
COMUNICACIÓN**

ARTÍCULO CIENTÍFICO

TEMA:

**ANÁLISIS DE DATOS BASADO EN TÉCNICAS DE BIG DATA Y DATA MINING
PARA CULTIVOS DE HORTALIZAS EN EL INVERNADERO DE LA GRANJA LA
PRADERA DE LA UNIVERSIDAD TÉCNICA DEL NORTE**

AUTOR: KARINA LISETH PONCE GUEVARA

DIRECTOR: MSC. EDGAR ALBERTO MAYA OLALLA

Ibarra-Ecuador

2017

Análisis de datos basado en técnicas de Big Data y Data mining para cultivos de hortalizas en el invernadero de la granja La Pradera de la Universidad Técnica del Norte

Autores – Karina Liseth PONCE GUEVARA, Ing Edgar Alberto MAYA OLALLA, MSc.

Facultad de Ingeniería en Ciencias Aplicadas, Universidad Técnica del Norte, Avenida 17 de Julio 5-21 y José María Córdova, Ibarra, Imbabura

klponceg@utn.edu.ec, eamaya@utn.edu.ec

Resumen. *El presente proyecto es una propuesta del uso de técnicas de Big Data y Data Mining (minería de datos) aplicados a cultivos de hortalizas en el invernadero de la granja "La Pradera", con el objetivo de analizar los factores que influyen en el crecimiento de los cultivos, y determinar un modelo predictivo de la humedad del suelo.*

Dentro de un invernadero, las variables que inciden en el crecimiento de los cultivos son: Humedad relativa, humedad del suelo, temperatura ambiental, y niveles de iluminación y CO₂. Estos parámetros son esenciales para la fotosíntesis, es decir, durante los procesos donde las plantas adquieren la mayoría de nutrientes, y por tanto, con un buen control de dichos parámetros las plantas podrían crecer más sanas y producir mejores frutos. El proceso de análisis de los factores en un contexto de minería de datos requiere diseñar un sistema de análisis y establecer una variable objetivo a ser predicha por el sistema. En este caso, con el fin de optimizar el gasto de recurso hídrico, se ha escogido como variable objetivo la humedad del suelo.

El sistema de análisis propuesto es desarrollado en una interfaz de usuario implementada en Java y NetBeans IDE 8.2, y consta principalmente de dos etapas: Una de ellas es la clasificación a través del algoritmo C4.5, el cual emplea un árbol de decisión basado en la entropía de los datos, y permite visualizar los resultados de manera gráfica. La segunda etapa principal es la predicción, en la cual, a partir de la clasificación obtenida en la etapa anterior, se predice la variable objetivo con base en un nuevo conjunto de datos. En otras palabras, la interfaz construye un modelo predictivo para determinar el comportamiento de la humedad de suelo.

Palabras Claves

Agricultura de precisión, Big Data, minería de datos, KDD, open data, limpieza de datos, bodega de datos, C4.5.

Abstract. *This work outlines the use of Big Data and Data Mining techniques on vegetable crops data from the greenhouse of the farm "The Pradera", which is aimed at*

analyzing the factors that influence the growth of the crops, and determine a predictive model of soil moisture.

Within a greenhouse, the variables that affect crop growth are: relative humidity, soil moisture, ambient temperature, and levels of illumination and CO₂. These parameters are essential for photosynthesis, i.e. during processes where plants acquire the most nutrients, and therefore, if performing a good control on these parameters, plants might grow healthier and produce better fruits. The process of analysis of such factors in a data mining context requires designing an analysis system and establishing an objective variable to be predicted by the system. In this case, in order to optimize water resource expenditure, soil moisture has been chosen as the target variable.

The proposed analysis system is developed in a user interface implemented in Java and NetBeans IDE 8.2, and consists mainly of two stages. One of them is the classification through algorithm C4.5, which uses a decision tree based on the data entropy, and allows to visualize the results graphically. The second main stage is the prediction, in which, from the classification results obtained in the previous stage, the target variable is predicted from information of a new set of data. In other words, the interface builds a predictive model to determine the behavior of soil moisture.

Keywords

Precision agriculture, Big Data, data mining, KDD, open data, data cleaning, data warehouse, C4.5

1. Introducción

Big Data es un concepto que ha tomado fuerza en los últimos años, debido principalmente a la gran cantidad de información que se genera en diversos contextos, entre ellos: Economía, educación, medio ambiente, redes de sensores y redes móviles. Por tanto, se puede decir que se ha convertido en un área multidisciplinaria. En efecto, existen diversas fuentes generando un flujo de información, el cual, a través

de un procesado permite el análisis de diferentes variables (atributos o características) y provee elementos para realizar una toma de decisiones inteligente.

Particularmente, en la agricultura se tiene varios factores y variables que producen una cantidad inimaginable de información. A través de herramientas que usan internet de las cosas (IoT - *Internet of Things*) e internet de todo (IoE - *Internet of Everything*), es decir, haciendo uso de sistemas embebidos con redes de sensores inalámbricos, puede realizarse la adquisición de datos masivos y, adicionalmente, usando herramientas de análisis de datos puede lograrse una gestión agrícola sustentable que, en cierta medida, genere impactos favorables al medio ambiente debido a que una adecuada toma de decisiones permitiría optimizar los recursos naturales

2. Materiales y Métodos.

2.1 Invernadero

Un invernadero es una construcción de madera, hierro u otro material, que tradicionalmente está cubierto por cristales, aunque existen modelos básicos cubiertos por plástico. En general, su estructura está provista de calefacción y está iluminada artificialmente, y por tanto en su interior es factible cultivar diferentes hortalizas, flores u otras plantas fuera de su estación. Los materiales usados para recubrirlos, así como los sistemas de control de los factores ambientales son de gran de variedad. [1]

2.2 Factores o variables a medir

Humedad del suelo.

Este factor se refiere a la cantidad de agua por volumen de tierra que existe en el terreno de un cultivo. El buen manejo de la humedad del suelo permite mejorar la producción de las plantas. Su medida es gravimétrica, y se da entre 0.1 y 0.3 bares de presión. Se relaciona con la capacidad de las raíces de las plantas para realizar la absorción de nutrientes del suelo. [2]

Humedad Relativa.

Las plantas tienen que transpirar agua para poder transportar nutrientes y regular su crecimiento, este factor depende de la transpiración y de la temperatura que posea el invernadero. El porcentaje de humedad relativa en el cual las plantas tienen un correcto desarrollo es del 55% al 70%. [3]

Temperatura.

Para sus procesos de crecimiento y correcto desarrollo, las plantas necesitan de una temperatura adecuada, de no ser así, estos procesos se detienen. Cuando este factor desciende a cero grados o menos, las plantas pueden sufrir daños severos en sus tejidos, así como suele suceder cuando se encuentran

al en el aire libre durante las heladas nocturnas. En general, el efecto favorable que produce el invernadero sobre el desarrollo de las raíces y del cultivo es mantener la adecuada temperatura tanto del aire como del suelo. [4]

Luminosidad

Esencialmente toda la luz visible es capaz de promover la fotosíntesis, pero las regiones de 400 a 500 y de 600 a 700 nm son las más eficaces. La cantidad de iluminación necesaria para que los cultivos dentro de un invernadero sobrevivan está entre los 10000 a 40000 lux. [3]

Dióxido de Carbono.

Este gas carbónico es de suma importancia en el ciclo de vida de los cultivos, es un material indispensable para la fotosíntesis y la clorofila de las plantas. Combinado con agua y energía luminosa, el CO₂ se emplea para que las plantas puedan producir carbohidratos y oxígeno, además, este factor está presente en la actividad estomática. La concentración de CO₂ en la atmósfera debe estar entre los 100 ppm (0,2 g/m³) y los 2000 ppm (4g/m³). [5].

2.3 Big Data.

Big Data es un término que se refiere a la información que excede la capacidad de procesamiento de los sistemas convencionales de bases de datos, o también que dicha cantidad sea tan grande, constantemente cambiante, o simplemente que no encaje en la estructura de las bases de datos. Dentro de dicha de información, se encuentra patrones e información oculta valiosos para la toma de decisiones. Actualmente, el hardware básico, las arquitecturas de cloud y el software libre implica casi que directamente el uso de Big Data para obtener resultados que no provee el hardware. [6]

2.4 Proceso KDD.

Este término se originó con investigaciones en el campo de la inteligencia artificial, este proceso involucra algunas etapas en el análisis de datos: Selección, procesamiento, transformación (en caso de ser necesaria), la realización de minería de datos (Data Mining) para extraer patrones y relaciones, y por último interpretación y evaluación de las estructuras descubiertas. En la figura 1 se muestra una ilustración del proceso KDD y sus etapas. [7]

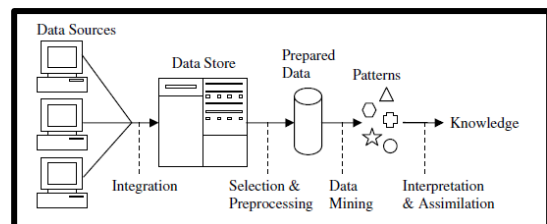


Figura 1. Proceso KDD.

Fuente: [8]

2.5 Minería de datos.

Minería de datos o data mining viene a ser el proceso de la extracción de información implícita y potencialmente útil de los datos. “Data minig es el proceso de descubrir patrones interesantes y conocimiento de una gran cantidad de datos” [9], donde la idea es construir programas computacionales que examinen cuidadosamente las bases de datos, en busca de aspectos similares o patrones para hacer predicciones exactas en los datos futuros. Muchos de los valores pueden ser banales y sin interés, mientras que otros falsos o inconsistentes, algunas partes serán ilegibles y otras se perderán en el proceso.

3. Desarrollo del software.

Realizar esta interfaz de análisis de datos involucra diferentes procesos, uno de ellos es la clasificación de los datos a través de una variable objetivo y el uso de un algoritmo de clasificación, se pretende proponer un modelo predictivo con los datos recolectados.

Los datos han sido obtenidos por medio de repositorios en línea que son confiables y poseen las variables necesarias, tales como: humedad del suelo, humedad relativa, temperatura ambiental, nivel de iluminación y Co2, las cuales interfieren en el buen desarrollo de los cultivos. El desarrollo de la solución posee varias etapas, entre estas: Selección de datos (búsqueda del archivo), pre procesamiento (selección de la variable objetivo), algoritmo de clasificación (árbol de clasificación C4.5), y el algoritmo de predicción.

3.1 Herramienta débilmente acoplada.

Herramienta	Característica
Débilmente acoplada	Las técnicas y algoritmos se encuentran fuera del Sistema Gestor de Base de Datos (SGBD).
Medianamente Acoplada	Ciertas funciones y tareas forman parte del SGBD.
Fuertemente acoplada	Todas las funciones, tareas, algoritmos se encuentran en el SGBD, con operaciones primitivas.

Tabla 1. Análisis comparativo de los sensores de temperatura y humedad relativa.

Fuente: Desarrollo del proyecto

En la Figura 2 se muestra la arquitectura de una herramienta débilmente acoplada, en esta la interfaz gráfica de usuario y los algoritmos de minería de datos se encuentran desarrollados de manera que sean fáciles de usar, y que permita realizar el modelo predictivo de la variable objetivo (humedad del suelo) con base en los valores de las otras

variables ambientales que inciden en el desarrollo de los cultivos del invernadero.

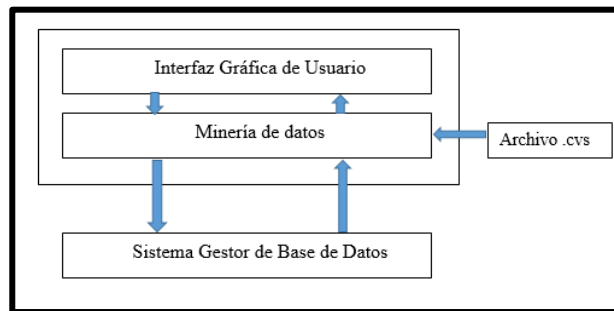


Figura 2 Diagrama de la arquitectura de una herramienta débilmente acoplada.

Fuente: Propia.

3.2 Diagramas de caso de uso del software.

En la interfaz desarrollada, existe un caso de uso general, donde se muestra el proceso que un usuario debería tomar para tener la visualización del árbol de decisión y las reglas de clasificación, que vienen a ser el resultado del análisis de los datos. Otra manera de usarlo es generando la predicción en base a las reglas calculadas por el proceso anterior.

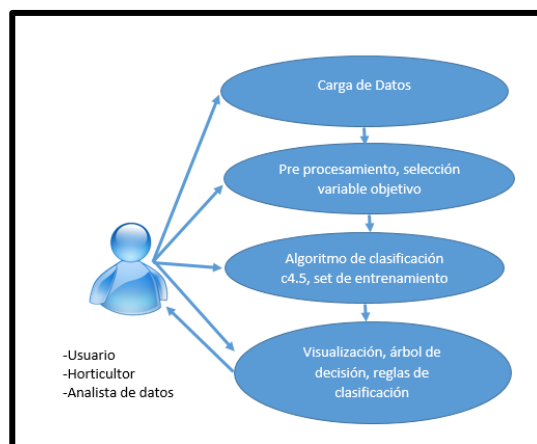


Figura 3 Caso de uso general de la interfaz desarrollada.

Fuente: Propia.

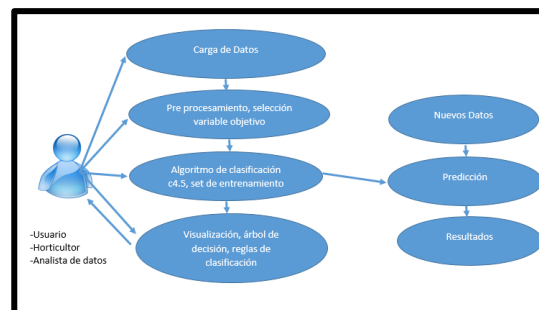


Figura 4 Caso de uso predictivo de la interfaz desarrollada.

Fuente: Propia.

3.3 Módulos de la interfaz

A continuación, se describe el funcionamiento de la interfaz de manera modular, explicando cada etapa y las funciones del usuario. En la Figura 5 se describe cada uno de los módulos que conforman la interfaz.

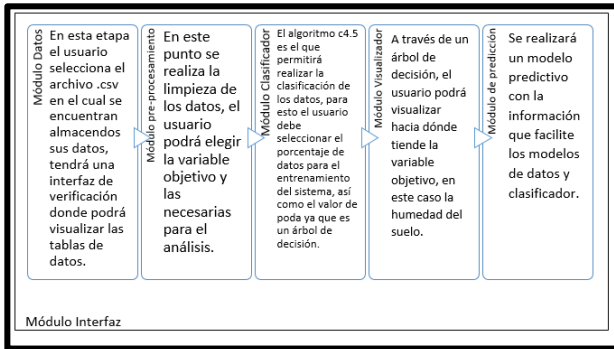


Figura 5 Funcionamiento de la interfaz en estructura de módulos
Fuente: Propia.

3.4 Paquetes usados en el desarrollo del software

Este software ha sido desarrollado con el uso de varios paquetes y clases, las cuales se describen a continuación en la Figura 6

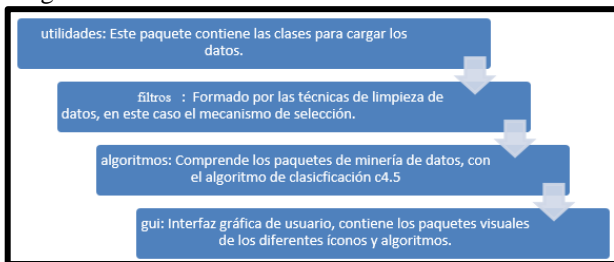


Figura 6 Paquetes y clases usados en la programación de la interfaz
Fuente: Propia.

4. Pruebas de Funcionamiento.

El presente proyecto se ha implementado de tal forma que sea de fácil uso para el usuario, a continuación, se muestra cada una de las partes de las que está conformado y cómo estas funcionan. La interfaz se encuentra conformado por dos partes que se muestran en forma de panel de selección, posee un panel de selección con dos pestañas, la primera ha sido denominada inicio posee las opciones para abrir, crear o guardar un proyecto, además de tener un botón de ayuda. Las figuras 7, 8, 9 y 10 muestran la interfaz desarrollada, indicando las partes que esta posee.

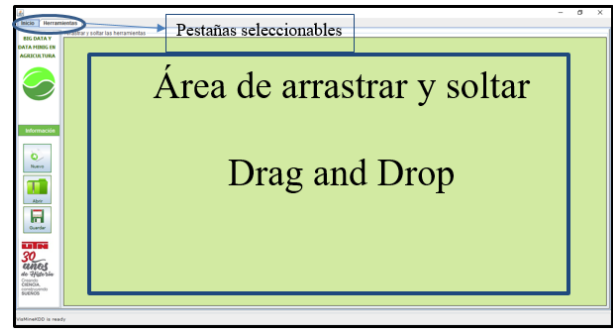


Figura 7 Pantalla de inicio de la interfaz gráfica de usuario.
Fuente: Propia.

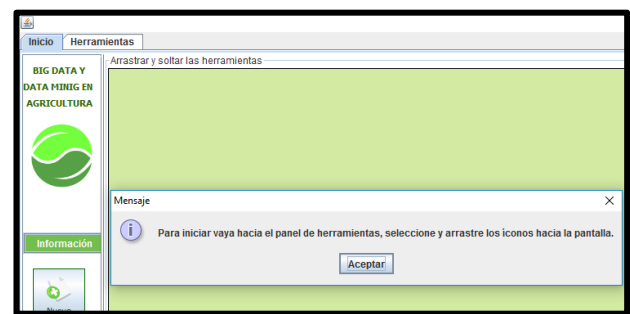


Figura 8 Texto mostrado por el botón informativo de la pantalla de inicio.
Fuente: Propia

Esta pestaña de inicio posee seis partes, un área informativa que muestra el motivo del sistema, a continuación, un botón informativo que muestra el siguiente paso a seguir, los tres botones son para empezar un nuevo proyecto, abrir uno ya realizado o guardar lo que se ha venido trabajando, por último, se tiene el logo representativo de la institución.



Figura 9 Descripción de los íconos formados por la pestaña Inicio
Fuente: Propia.

La segunda pestaña llamada herramientas posee todos los procesos que se usan en el análisis: la selección del archivo de datos, el pre procesamiento con la elección de la variable objetivo, el algoritmo de clasificación de datos c4.5, y la visualización que se realiza a través de un árbol de decisión, además de una opción de predicción. En el centro se encuentra el canvas, es decir, el área donde el usuario podrá configurar los parámetros de cada uno de los procesos y conectarlos entre sí para su funcionamiento.



Figura 10 Descripción de los íconos formados en la pestaña Herramientas.
Fuente: Propia.

Datos

En esta sección el usuario puede seleccionar un archivo de extensión .csv (valores separados por comas) alojado en su computador, dependiendo del caracter por el que se encuentren separados los datos sea una “,” o “;”. Este proceso permite visualizar la información en una tabla, y de esta manera se pueda cerciorar que el archivo es el correcto. Se utilizó la librería *JavaCSV*, la cual permite la lectura y escritura de archivos de este tipo. En la figura 11 aprecia la carga del archivo .csv para el análisis de datos.

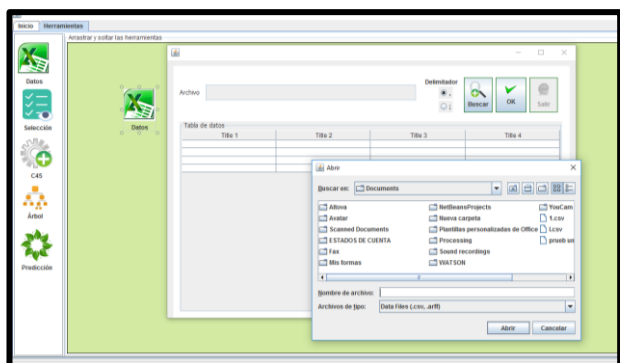


Figura 11 Funcionamiento módulo datos de interfaz de análisis de datos.
Fuente: Propia.

Selección

Como su nombre lo indica, este módulo permite al usuario seleccionar la variable objetivo o target (humedad del suelo), así como también las variables con las que se desea trabajar. Este proceso almacena los datos seleccionados en una nueva tabla y la muestra al usuario, también etiqueta a la variable objetivo para que el siguiente módulo la tome en cuenta. La Figura 12 muestra el funcionamiento del módulo selección.

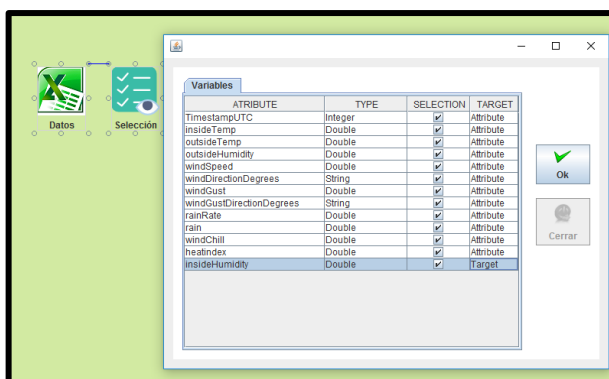


Figura 12 Funcionamiento módulo selección en la interfaz de análisis de datos.
Fuente: Propia.

Clasificación

En este módulo se desarrolló el algoritmo de clasificación, el cual forma las reglas para el árbol de decisión C4.5, para esto el usuario debe dar parámetros al sistema para su correcto funcionamiento, esto dependerá del punto de vista del analista. El set de entrenamiento es el conjunto de datos con el que se va a construir las reglas de clasificación, las filas por nodo, son el número de filas de datos que algoritmo analizará por cada nodo que ser forme, y el porcentaje de límite se refiere a cuan frondoso se quiere visualizar el árbol, este se puede ir cambiando de acuerdo al nivel de entendimiento del usuario. En la Figura 13 se aprecia al módulo clasificación funcionando.

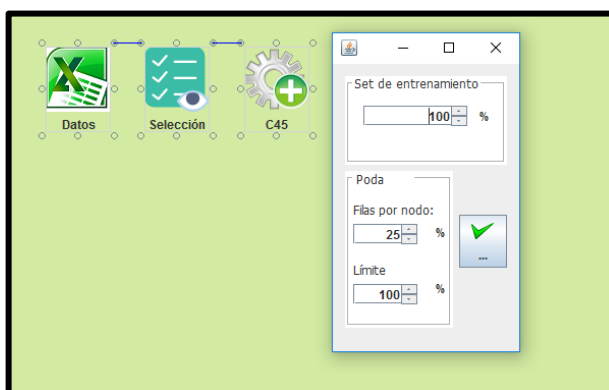


Figura 13 Funcionamiento módulo clasificación en la interfaz de análisis de datos.
Fuente: Propia.

Visualización

Para visualizar los resultados del algoritmo de clasificación se ha utilizado el WekaTreeViewer, tomado de la herramienta weka. Con base en estos datos y las reglas que se generaron en el algoritmo C4.5, el árbol se grafica, sin necesidad de pedir otro parámetro adicional. La Figura 14 muestra un esquema de conexión para llegar a este módulo, en la Figura 15 se puede observar un árbol generado por la herramienta, el cual ha resultado particularmente muy frondoso debido a la naturaleza numérica de los datos y la cantidad de los mismos. Las reglas que han sido utilizadas en este proceso de clasificación, se muestran la pestaña Rules, la Figura 16 indica aquellas que han sido particularmente calculadas para este proceso y el porcentaje de confianza que genera.

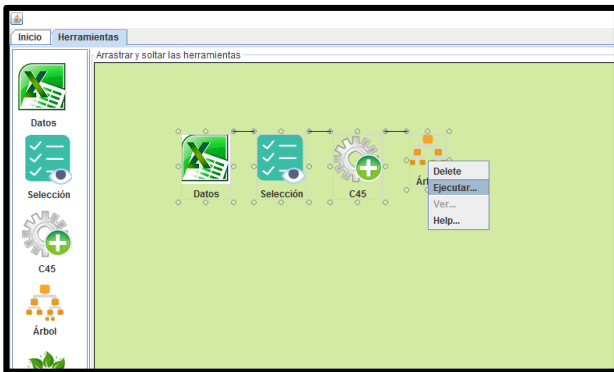


Figura 14 Funcionamiento módulo visualización en la interfaz de análisis de datos.

Fuente: Propia.

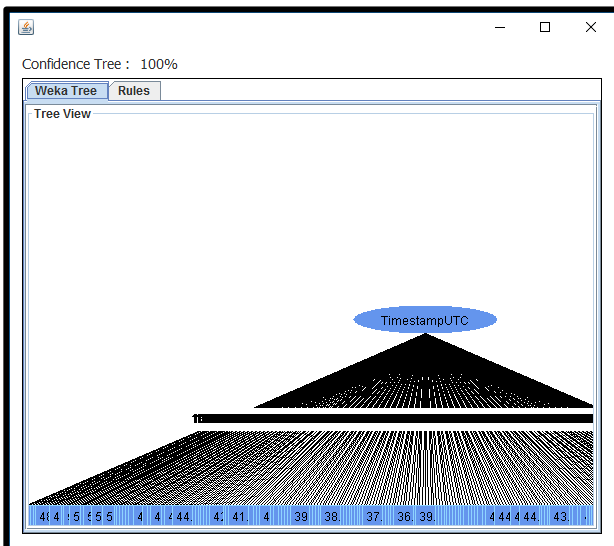


Figura 15 Visualización del árbol de decisión en la interfaz de análisis de datos.

Fuente: Propia.

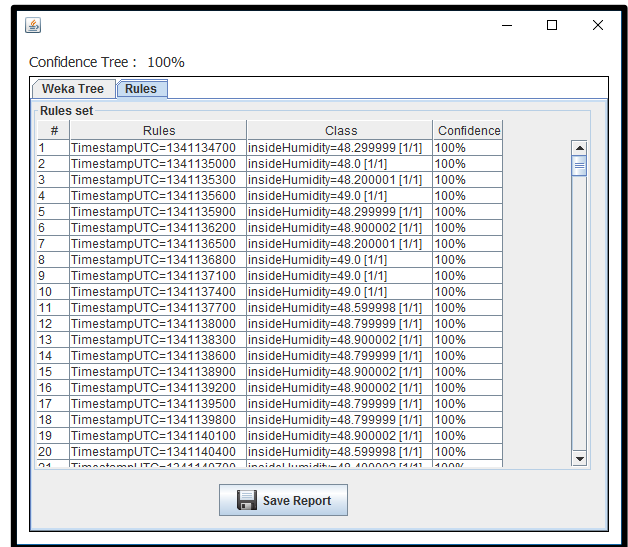


Figura 16 Visualización de las reglas de clasificación en la interfaz de análisis de datos.

Fuente: Propia.

Predicción

El algoritmo de predicción trabaja con base en las reglas formadas por el módulo clasificación C4.5. El set de datos con el que se realizaron las etapas anteriores funciona ahora como el conjunto de información de entrenamiento, ahora para realizar predicción se necesita de un archivo con datos nuevos, el cual desconoce la variable objetivo (humedad del suelo). Por lo tanto, los parámetros que se deben enviar a esta herramienta, mediante la conexión entre estas (esto se realiza manualmente en la interfaz, conectándolos en un punto, el proceso se muestra en las figuras 17 y 18) son la clasificación realizada y el nuevo archivo. Con la información ya mencionada, se realiza el proceso de predicción, y dando como resultado los valores que tomará la variable objetivo.

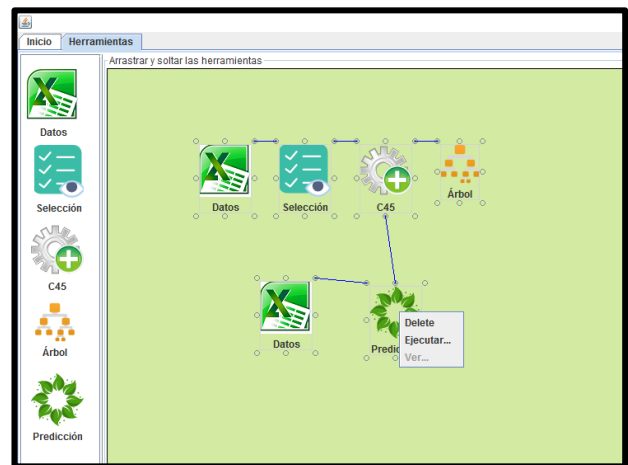


Figura 17 Funcionamiento del módulo predicción en la interfaz de análisis de datos.

Fuente: Propia.

Time	inside	outsid	windS	wind	windR	rain	rainC	heatIn	inside	inside
1341	72.5	61.286	89.3	0	0	0	0	61.286	61.286	48.348.29
1341	72.5	61.178	88.8	0	0	0	0	61.178	61.178	48.48.0
1341	72.5	61.322	88.8	0	0	0	0	61.322	61.322	48.248.20
1341	72.5	61.034	92	0	0	0	0	61.034	61.034	49.49.0
1341	72.5	61.052	90.6	0	0	0	0	61.052	61.052	48.348.29
1341	72.5	60.8	92.7	0.39	0	1315.0	0	60.8	60.8	48.948.90
1341	72.5	61.034	91.1	0.29	0	190.0	0	61.034	61.034	48.248.20
1341	72.5	61.034	91.7	0	0	0	0	61.034	61.034	49.49.0
1341	72.5	60.836	91.1	0	0	0	0	60.836	60.836	49.49.0
1341	72.5	60.98	90	0	0	0	0	60.98	60.98	49.49.0
1341	72.5	61.034	91.5	0	0	0	0	61.034	61.034	48.648.59
1341	72.41	61.25	91.8	0.384	0	2315.0	0	61.25	61.25	48.848.79
1341	72.32	61.304	93.3	0.195	0	2338.0	0	61.304	61.304	48.948.90
1341	72.32	61.664	93.6	0.9110	0	3135.0	0	61.664	61.664	48.848.79
1341	72.32	61.934	94.2	0.670	0	3338.0	0	61.934	61.934	48.948.90
1341	72.32	62.204	91.7	0	0	0	0	62.204	62.204	48.948.90
1341	72.32	62.456	91	0	0	0	0	62.456	62.456	48.848.79
1341	72.302	62.762	90.5	0	0	0	0	62.762	62.762	48.848.79
1341	72.212	63.088	91	0	0	0	0	63.088	63.088	48.948.90
1341	72.14	63.392	90.9	0.2351	0	10.0	0	63.392	63.392	48.648.59
1341	72.14	63.878	91.2	0.154	0	0.45	0	63.878	63.878	48.448.40
1341	72.032	64.418	89.6	0	0	0	0	64.418	64.418	49.49.0
1341	71.96	64.85	88.6	0	0	0	0	64.85	64.85	48.48.0
1341	71.978	65.354	87.6	0	0	0	0	65.354	65.354	48.48.0
1341	72.14	65.804	87.4	0	0	0	0	65.804	65.804	48.48.0
1341	72.14	66.254	86.2	0	0	0	0	66.254	66.254	48.248.20

Figura 18 Visualización de la predicción de la variable target (Humedad del suelo)

Fuente: Propia.

5. Resultados.

Prueba

Resultados

Data Set for Sustainability Clasificación

En este caso, se observó que las reglas de clasificación apuntaban a la variable TimeStamp como la de mayor peso, de esta dependía directamente la humedad.

Data Set for Sustainability Predicción

Al realizar el algoritmo de predicción, se comprobó que el modelo si funciona, pues la variable Humedad del Suelo fue predicha con un alto grado de confianza.

Datos reales Clasificación

Al poseer datos reales el algoritmo determinó que la humedad del suelo depende directamente de la variable iluminación, cuando esta no posee valor, la segunda con mayor peso es la humedad relativa. Por ende, en base a estos dos factores se forman las reglas de clasificación

Datos reales Predicción

La variable Humedad del Suelo fue predicha con un alto grado de confianza, ya que los datos obtenidos por el proceso de predicción se acercan a los reales, con alta precisión.

Tabla2.Resultados del proyecto

Fuente: Desarrollo del proyecto

6. Conclusiones.

Después de realizar una investigación acerca de los factores que inciden en el crecimiento de los cultivos dentro de un invernadero, se determinó que los factores más importantes son: Humedad del suelo, humedad relativa, temperatura ambiental, nivel de iluminación y CO₂, dado que éstos influyen directamente en el proceso de fotosíntesis de las plantas y, estableciendo valores correctos, logran una mejor captación de nutrientes y obtienen mejores frutos.

Las herramientas y técnicas de *Big Data* y, específicamente, de *data mining* son fundamentales para realizar procesos de analítica de datos. Si bien estas dos áreas surgieron como soporte para la toma de decisiones en economía y negocios, hoy en día son de uso transversal en diversos escenarios y se enfocan al descubrimiento de patrones dentro de una montaña de datos. A través de una revisión de la documentación empleada para esta investigación, se pudo encontrar que estas herramientas y técnicas de analítica de datos no se rigen, particularmente, por un estándar, sino que representan un conjunto de algoritmos que permiten realizar modelos descriptivos sobre un conjunto de datos con el fin de clasificar y/o predecir información.

En las primeras etapas del desarrollo de este proyecto, se realizó una búsqueda de bases de datos que contengan diversas mediciones representando variables capturadas en un invernadero real. En este sentido, se determinó usar la base de datos *Environmental data (indoor and outdoor)* del repositorio *UMass Trace Repository*, debido a que se ajustó adecuadamente a los requerimientos del proyecto, es decir, que contiene un conjunto de muestras suficiente y posee las variables con los factores más importantes.

Como resultado significativo de este proyecto, se encontró que el diseño de un software de *data mining* con una interfaz de uso intuitivo en un *framework* de *drag and drop* es una alternativa adecuada para procesar variables de invernadero. Se comprobó que puede realizarse el procesamiento de los datos conformando de forma secuencial las etapas del proceso de descubrimiento de conocimiento en base de datos (KDD) a través de la unión de objetos que representan módulos de programación. Específicamente, el entorno de desarrollo NetBean IDE 8.2, que trabaja con el lenguaje de programación java, comprobó ser un software adecuado y de precisión para el desarrollo de las técnicas y algoritmos de la analítica de datos, y que también permitió implementar una interfaz amigable con el usuario.

Después del desarrollo de un software para el análisis de los parámetros ambientales que inciden en el crecimiento de cultivos en invernaderos y con el fin de comprobar su correcto funcionamiento, se realizó diversas pruebas con diferentes fuentes de información, siendo una de ellas la encontrada en el *UMass Trace Repository*, dando como resultado que la humedad interior se puede predecir con base en las demás variables con una precisión, significativamente buena. Particularmente, se logró determinar que dicha variable se encuentra ligada a la variable *TimeStampUTM*, la cual se refiere al tiempo en que ha sido tomada la muestra.

Las pruebas comprobaron la usabilidad y confiabilidad del software.

Para comprobar el funcionamiento del sistema con datos reales, se realizó pruebas con información obtenida del invernadero de la granja “La pradera”, a través de una red de sensores inalámbricos instalada en el lugar. Al igual que en los experimentos con los datos de prueba, se encontró nuevamente que la humedad del suelo depende del tiempo de toma de la muestra. Adicionalmente, usando la herramienta de selección se determinó las cinco variables más importantes y, en ese caso, se obtuvo como resultado que el nivel de iluminación (denominado luz) es el factor más importante del cual depende la humedad del suelo. Además, con base en este factor se calcularon las reglas de clasificación, de forma que cuando éste tiene un valor de 0, la variable que considera el sistema es la humedad relativa, y se comprobó experimentalmente que así se genera mayor conocimiento a través de la exploración de datos.

Las áreas de *Big Data* y *Data Mining* son relativamente emergentes y se encuentran en constante desarrollo, y, particularmente, su aplicación en el sector agrícola es un tema amplio y diverso que busca, entre otros aspectos, optimizar los recursos. En efecto, la agricultura de precisión está enfocada al uso de herramientas tecnológicas para hacer eficientes en el uso y administración de recursos. Dicho esto, las técnicas de minería de datos representan una buena alternativa para explorar la información de las variables relacionadas con agricultura y soportar la toma de decisiones inteligentes.

Referencias Bibliográficas.

- 1] A. Alpi y F. Tognoni, Cultivo en invernadero, Madrid: Ediciones Mundi-Prensa, 1999.
- 2] J. J. Ibáñez, «Madrid,» Junio 2006. [En línea]. Available: <http://www.madrimasd.org/blogs/universo/2006/06/26/33002>.
- 3] N. Iglesias, «INTA,» 2009. [En línea]. Available: http://inta.gob.ar/sites/default/files/script-tmp-inta_produccion-de-hortalizas-bajo-cubierta_2006.pdf.
- 4] O. Barrios, Construcción de un Invernadero, Santiago, 2004.
- 5] D. Marlow, «Hortalizas,» Marzo 2011. [En línea]. Available: <http://www.hortalizas.com/horticultura-prottegida/invernadero/aporte-de-co2-en-un-invernadero/>.
- 6] E. H. J. Z. M. L. M. Dumbill, M. Slocum, A. Croll and C. Hill, Big Data Now-2012 Edition, Sebastopol, CA: O'Reilly Media, 2012.
- 7] D. Hand, H. Mannila y P. Smyth, Principles of Data Mining, Massachusetts: MIT, 2001.
- 8] M. Bramer, Principles of Data Mining, Springer, 2007.
- 9] J. Han, M. Kamber y J. Pei, DATA MINING Concepts and Techniques, Morgab Kaufmann Publishers, 2012.
- 10] J. Erazo y C. Hervas, «Revista Científica Maskana,» 20 Octubre 2014. [En línea]. Available: <http://www.ucuenca.edu.ec/ojs/index.php/maskana/article/view/532>.
- 11] Secretaría de Gestión de Riesgos, «Secretaría de Gestión de Riesgos | Ecuador,» 2015. [En línea]. Available: <http://www.gestionderiesgos.gob.ec/>.
- 12] Consejo de Federaciones de Bomberos de Argentina, «Consejo Nacional Bomberos Argentina,» 2006. [En línea]. Available: http://scoutsanpatricio.com.ar/libros/bomberos_03.pdf.
- 13] OMS, GUÍAS PARA LA CALIDAD DEL AIRE, Lima: Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente, 2004.
- 14] J. González, «Tesis Electrónicas de la Universidad Austral de Chile,» 2014. [En línea]. Available: <http://cybertesis.uach.cl/tesis/uach/2014/bmfcig643m/doc/bmfcig643m.pdf>.
- 15] Bomberos de Granollers, «Bombers de Granollers Visita d'escoles al parc,» Mayo 2011. [En línea]. Available: http://bombersgranollers.org/documentacio/incendio_forestal/CIF01_texto&imágenes_CAST_Intro.pdf.
- 16] O. Rosas, J. Flores, M. Meléndez y R. Cabrera, «EVALUACION DEL COMPORTAMIENTO DEL FUEGO BAJO DIFERENTES CONDICIONES AMBIENTALES EN EL ESTADO DE CHIHUAHUA A PARTIR DE LA APLICACIÓN DE QUEMAS PRESCRITAS,» *Scientia CUCBA*, p. 7, 2006.
- 17] R. F. Martínez, J. O. Meré, F. J. de Pisón Ascacibar, A. G. Marcos y F. A. Elías, Redes Inalámbricas de sensores: teoría y aplicación práctica, Publicaciones Universidad de la Rioja, 2009.

- CCM Benchmark Group, «CCM - Online Community,» Junio 2014. [En línea]. Available: <http://es.ccm.net/faq/9457-benchmark-definicion>.
- 18] ISO/IEC, «ISO/IEC 27005,» 30 Junio 2008. [En línea]. Available: http://www.pqm-online.com/assets/files/lib/std/iso_iec_27005-2008.pdf.
- 19] iso 27000, «Gestión de Riesgos tecnológicos basada en ISO 27005 para la continuidad de negocio,» 2011. [En línea]. Available: [file:///C:/Users/HENRY/Downloads/Dialnet-GestionDeRiesgosTecnologicosBasadaEnISO31000EISO27-4797252%20\(3\).pdf](file:///C:/Users/HENRY/Downloads/Dialnet-GestionDeRiesgosTecnologicosBasadaEnISO31000EISO27-4797252%20(3).pdf). [Último acceso: 1 Abril 2016].
- 20] T. R, «Arquitecturas de integración del proceso de descubrimiento de conocimiento con sistemas de gestión de bases de datos: un estado del arte.,» *Revista de Ingeniería y Competitividad*, 2012.
- 21]

Sobre los Autores.



Karina L. PONCE GUEVARA. Nació en Ibarra el 07 de mayo de 1993. Realizó sus estudios primarios en la Escuela de niñas “Sarance” Los estudios secundarios los realizó en el Instituto Superior Tecnológico “República del Ecuador” donde finalizó en el año 2010, obteniendo el título de Bachiller

en Ciencias Especialización Físico Matemático. Actualmente, está realizando su proceso de titulación en Ingeniería en Electrónica y Redes de Comunicación, Universidad Técnica del Norte – Ecuador.



Edgar A. MAYA OLALLA. Nació en Ibarra – Ecuador el 22 de abril del año 1980. Ingeniero en Sistemas Computacionales en la Universidad Técnica del Norte en el año 2006. Magíster en Redes de Comunicaciones en la Pontificia Universidad Católica del Ecuador en el año 2014.

Actualmente es docente de la Carrera de Ingeniería en Electrónica y Redes de Comunicación de la Universidad Técnica del Norte.



Data analysis based on Big Data and Data Mining techniques for vegetable crops in the “La Pradera” green farm.

Autores – Karina Liseth PONCE GUEVARA, Ing Edgar Alberto MAYA OLALLA, MSc.

Facultad de Ingeniería en Ciencias Aplicadas, Universidad Técnica del Norte, Avenida 17 de Julio 5-21 y José María Córdova, Ibarra, Imbabura

klponceg@utn.edu.ec, eamaya@utn.edu.ec

Abstract. *This work outlines the use of Big Data and Data Mining techniques on vegetable crops data from the greenhouse of the farm "The Pradera", which is aimed at analyzing the factors that influence the growth of the crops, and determine a predictive model of soil moisture.*

Within a greenhouse, the variables that affect crop growth are: relative humidity, soil moisture, ambient temperature, and levels of illumination and CO₂. These parameters are essential for photosynthesis, i.e. during processes where plants acquire the most nutrients, and therefore, if performing a good control on these parameters, plants might grow healthier and produce better fruits. The process of analysis of such factors in a data mining context requires designing an analysis system and establishing an objective variable to be predicted by the system. In this case, in order to optimize water resource expenditure, soil moisture has been chosen as the target variable.

The proposed analysis system is developed in a user interface implemented in Java and NetBeans IDE 8.2, and consists mainly of two stages. One of them is the classification through algorithm C4.5, which uses a decision tree based on the data entropy, and allows to visualize the results graphically. The second main stage is the prediction, in which, from the classification results obtained in the previous stage, the target variable is predicted from information of a new set of data. In other words, the interface builds a predictive model to determine the behavior of soil moisture.

Keywords

Precision agriculture, Big Data, data mining, KDD, open data, data cleaning, data warehouse, C4.5

1. Introducción

Big Data is a concept that has gained momentum in recent years, mainly due to the large amount of information generated in various contexts, including: Economy, education, environment, sensor networks and mobile

networks. Therefore, it can be said that it has become a multidisciplinary area. In fact, there are several sources generating a flow of information, which, through a process allows the analysis of different variables (attributes or characteristics) and provides elements to make intelligent decision making.

In particular, agriculture has several factors and variables that produce an unimaginable amount of information. Through tools that use internet of things (IoT - *Internet of Things*) and internet all (SUI - *Internet of Everything*), ie, using embedded systems with wireless sensor networks, acquiring massive data can be performed. And, additionally, using data analysis tools can achieve a sustainable agricultural management that, to a certain extent, generates favorable impacts to the environment because an adequate decision making would optimize the natural resources

2. Materials and methods

2.1 Green Famr

A greenhouse is a construction of wood, iron or other material, which is traditionally covered by crystals, although there are basic models covered by plastic. In general, its structure is provided with heating and is illuminated artificially, and therefore inside it is feasible to grow different vegetables, flowers or other plants outside its season. The materials used to coat them as well as environmental factors control systems are of great variety. [1]

2.2 Factors or variables to be measured

Soil moisture.

This factor refers to the amount of water per volume of land that exists on the land of a crop. The good management of the humidity of the soil allows to improve the production of the plants. Its measurement is gravimetric, and gives between 0.1 and 0.3 bar of pressure. It relates to the ability

of the roots of plants to perform the absorption of nutrients from the soil. [2]

RH.

Plants have to transpire water in order to transport nutrients and regulate their growth, this factor depends on the transpiration and temperature that the greenhouse has. The percentage of relative humidity in which the plants have a correct development is of 55% to 70%. [3]

Temperature.

For their processes of growth and correct development, plants need a suitable temperature, otherwise, these processes will stop. When this factor drops to zero degrees or less, the silver can suffer severe damage to their tissues, as it usually happens when they are in the open air during the night frosts. In general, the favorable effect of the greenhouse on the development of roots and cultivation is to maintain the adequate temperature of both air and soil. [4]

Brightness

Essentially all visible light is capable of promoting photosynthesis, but the 400 to 500 and 600 to 700 nm regions are the most effective. The amount of illumination required for crops within a greenhouse to survive is between 10,000 and 40,000 lux. [3]

Carbon dioxide.

This carbonic gas is extremely important in the life cycle of crops, it is an indispensable material for photosynthesis and chlorophyll of plants. Combined with water and light energy, CO₂ is used so that plants can produce carbohydrates and oxygen, in addition, this factor is present in the stomatal activity. The CO₂ concentration in the atmosphere should be between 100 ppm (0.2 g / m³) and 2000 ppm (4g / m³). [5].

2.3 Big Data.

Big Data is a term that refers to information that exceeds the processing capacity of conventional database systems, or that such a quantity is so large, constantly changing, or simply that it does not fit into the structure of the databases. data. Within such information, hidden patterns and information are valuable for decision making. Today, basic hardware, cloud architectures, and free software almost directly involve using Big Data to get results that are not provided by hardware. [6]

2.4 KDD Process.

This term originated with research in the field of artificial intelligence, this process involves some stages in the analysis of data: Selection, processing, transformation (if necessary), the performance of data mining (Data Mining)

to extract Patterns and relationships, and finally interpretation and evaluation of discovered structures. Figure 1 shows an illustration of the KDD process and its stages. [7]

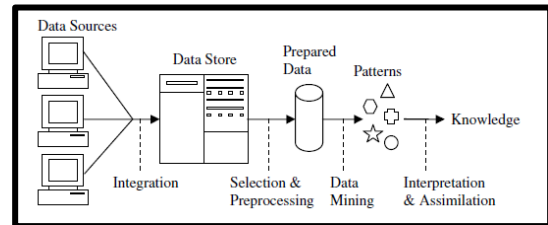


Figure 1. Process KDD.

Source: [8]

2.5 Data mining

Data mining or data mining is the process of extracting implicit and potentially useful information from the data. "Data minig is the process of discovering interesting patterns and knowledge of a large amount of data" [9], where the idea is to build computer programs that carefully examine databases, looking for similar aspects or patterns to make exact predictions in Future data. Many of the values may be banal and uninteresting, while others may be false or inconsistent, some parts will be illegible and others will be lost in the process.

3. Software development.

To perform this data analysis interface involves different processes, one of them is the classification of the data through a target variable and the use of a classification algorithm, we intend to propose a predictive model with the data collected.

The data have been obtained through online repositories that are reliable and have the necessary variables such as: soil moisture, relative humidity, ambient temperature, lighting level and Co₂, which interfere with the good development of the crops. The development of the solution has several stages, among them: Data selection (file search), preprocessing (target variable selection), classification algorithm (classification tree C4.5), and prediction algorithm.

3.1 Herramienta débilmente acoplada.

Tool	caracter ística
Weakly coupled	Algorithms are outside the System Manager Database (DBMS).

Moderately Engaged	Certain functions and tasks are part of the DBMS.
Strongly coupled	All functions, tasks, algorithms are in the DBMS, with primitive operations.

Table 1. Comparative analysis of temperature and relative humidity sensors.

Source: Project Development

Figure 2 shows the architecture of a weakly coupled tool, in this the graphical user interface and the data mining algorithms are developed in a way that is easy to use, and allows to realize the predictive model of the objective variable (Soil moisture) based on the values of the other environmental variables that affect the development of the greenhouse crops.

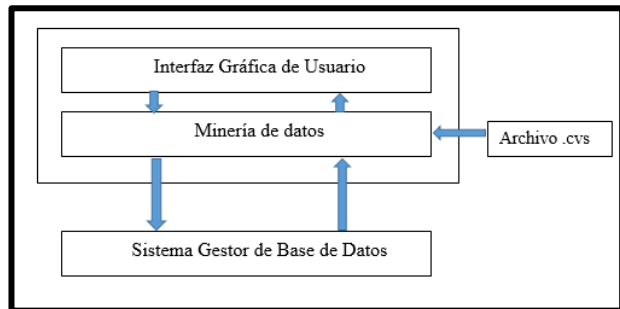


Figure 2 architecture diagram of a weakly coupled tool.

Source: Project development

3.2 Software use case diagrams

In the developed interface, there is a general use case, which shows the process that a user should take to have the visualization of the decision tree and the classification rules, which come to be the result of the analysis of the data. Another way to use it is to generate the prediction based on the rules calculated by the previous process.

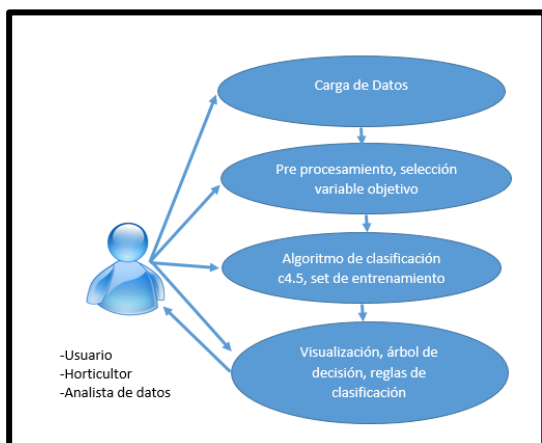


Figure 3 Case general purpose interface developed.

Source: project development

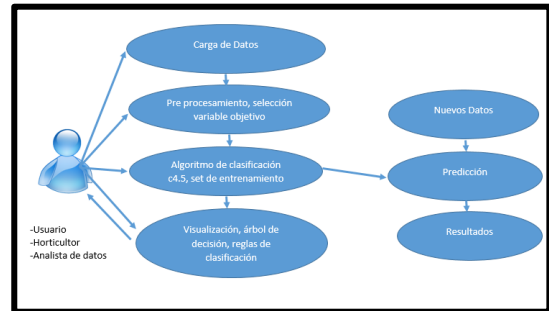


Figure 4. Case predictive use interface developed.

Source: project development

3.3 Paquetes usados en el desarrollo del software

This software has been developed with the use of several packages and classes, which are described below in Figure 6.

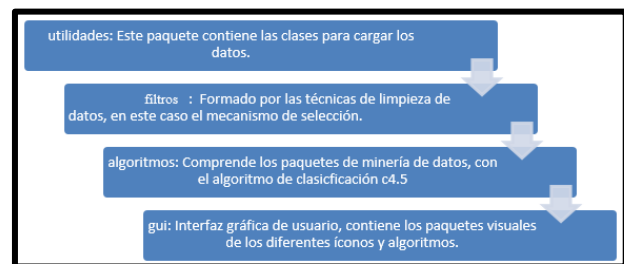


Figure 6 . Packages and classes used in programming interface

Fuente: Propia.

4. Funcionalidad test.

The present project has been implemented in such a way that it is easy to use for the user, then it shows each of the parts of which it is formed and how they work. The interface consists of two parts that are shown in the form of selection panel, has a selection panel with two tabs, the first has been called start has the options to open, create or save a project, in addition to having a button Of help. Figures 7, 8, 9 and 10 show the developed interface, indicating the parts that it has.

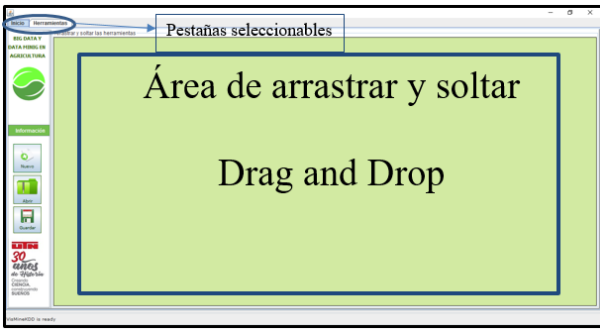


Figure 7 home screen graphical user interface.
Source: Project development

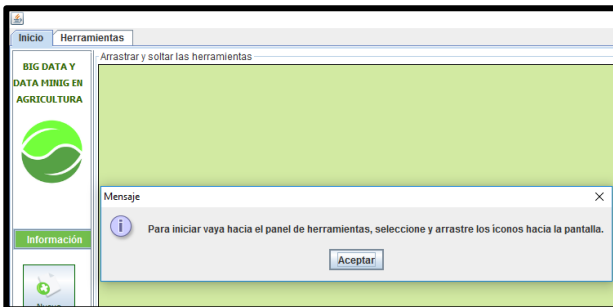


Figure 8 text displayed by the information button on the home screen..
Source: Project development

This home tab has six parts, an information area that shows the reason for the system, then an information button that shows the next step, the three buttons are to start a new project, open an already done or save what Has been working, finally, have the logo representative of the institution.



Figure 9 Description of icons formed by the Home tab
Source: Project development

The second tab named tools has all the processes that are used in the analysis: data file selection, preprocessing with the choice of the target variable, the data classification algorithm c4.5, and the visualization that is performed Through a decision tree, in addition to a prediction option. In the center is the canvas, that is, the area where the user can configure the parameters of each of the processes and connect them to each other for their operation.



Figure 10 Description of icons formed in the Tools tab.
Source: Project development

Data

In this section the user can select a .csv extension file (comma separated values) hosted in his computer, depending on the character by which the data are separated, a ", " or ";". This process allows you to view the information in a table, and in this way you can make sure that the file is correct. The library was then used *JavaCSV*, Which allows the reading and writing of files of this type. Figure 11 shows the loading of the .csv file for data analysis.

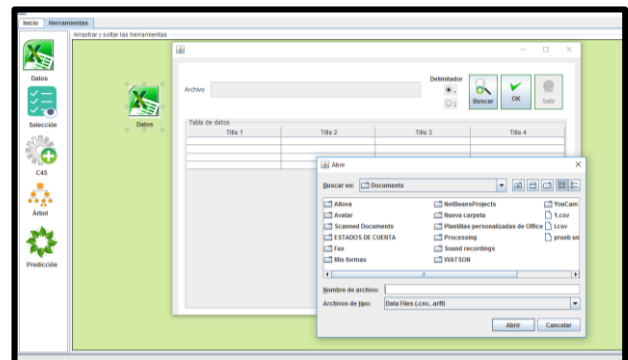


Figure 11 Operation module data interface for data analysis.
Source: Project development

Selection

As its name implies, this module allows the user to select the target variable (soil moisture) as well as the variables with which it is desired to work. This process stores the selected data in a new table and shows it to the user, it also tags the target variable for the next module to take into account. Figure 12 shows the operation of the selection module.

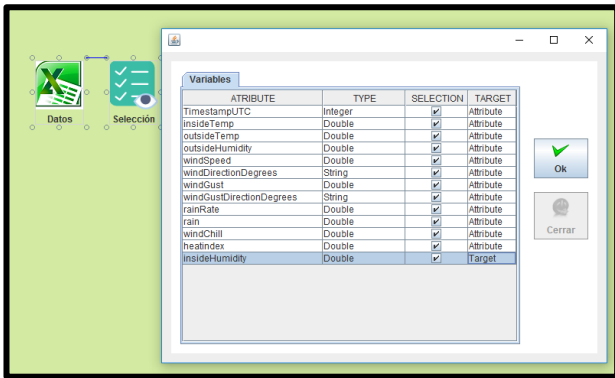


Figure 12 Operation module selection interface data analysis. Source: Project development

Classification

In this module the classification algorithm was developed, which forms the rules for the decision tree C4.5, for this the user must give parameters to the system for its correct operation, this will depend on the analyst's point of view. The training set is the set of data with which to build the classification rules, the rows per node, are the number of rows of data that algorithm will analyze for each node to be formed, and the percentage of limit refers To how leafy it is wanted to visualize the tree, this one can be changed according to the level of understanding of the user. Figure 13 shows the working classification module.

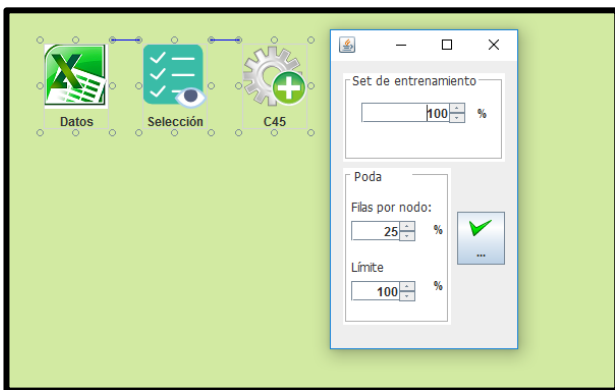


Figure 13 Operation classification module interface data analysis. Source: Project development

Visualization

To view the results of the classification algorithm we used the WekaTreeViewer, taken from the weka tool. Based on these data and the rules that were generated in the algorithm C4.5, the tree is graphical, without having to ask for another additional parameter. Figure 14 shows a connection scheme to reach this module, in Figure 15 a tree generated by the tool can be observed, which has been particularly very leafy due to the numerical nature of the data and the quantity thereof. The rules that have been used in this classification process are shown in the Rules tab, Figure 16 indicates those that have been particularly calculated for this process and the percentage of confidence it generates.

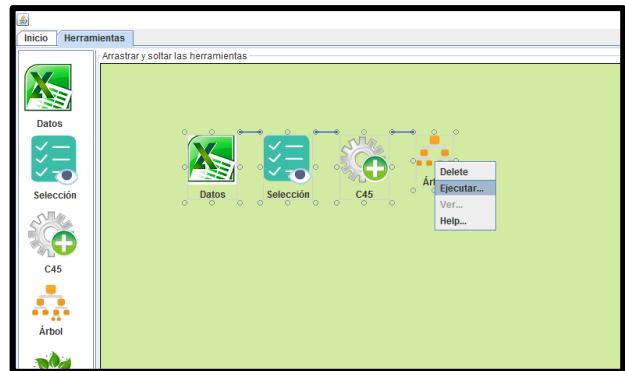


Figure 14 Operation module interface display data analysis.de datos. Source: Project development

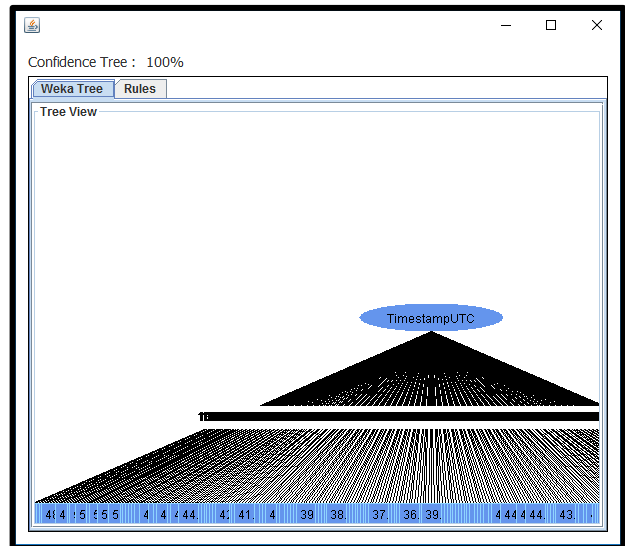


Figure 15 Viewing decision tree analysis interface. Source: Project development

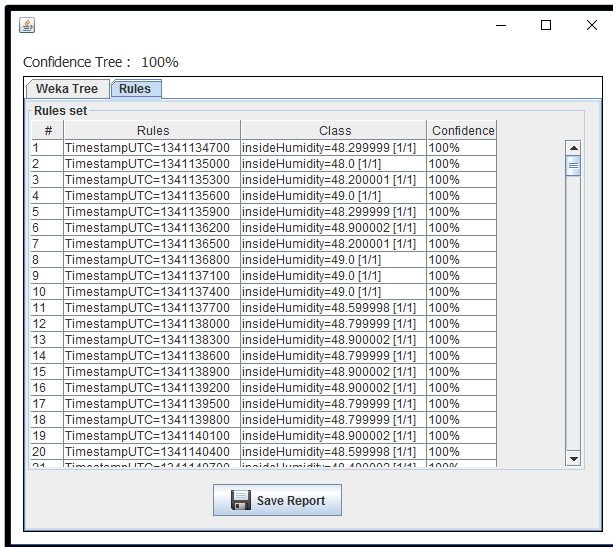


Figure 16 display the classification rules in the data analysis interface.
Source: Project development

Prediction

The prediction algorithm works based on the rules formed by the classification module C4.5. The set of data with which the previous steps were performed now works as the training information set, now to make a prediction, a file with new data is needed, which does not know the objective variable (soil moisture). Therefore, the parameters to be sent to this tool, through the connection between these (this is done manually in the interface, connecting them in a point, the process is shown in Figures 17 and 18) are the classification made and the New file. With the information already mentioned, the prediction process is performed, and resulting in the values that the target variable will take.

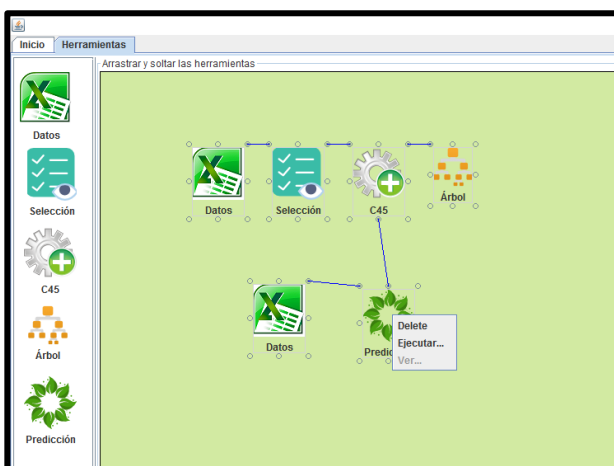


Figure 17 operation prediction module interface data analysis
Source: Project development

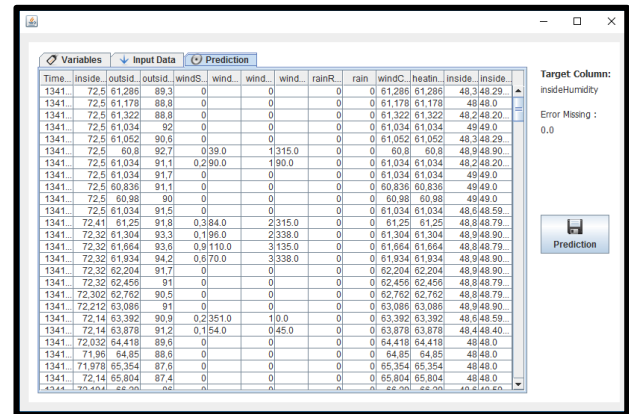


Figure 18 Viewing predicting the target variable (soil moisture)
Source: Project development

5. Resultados.

Proof	Results
Sorting Data Set for Sustainability ng	In this case, it was observed that the rules or classification aimed at the TimeStamp variable as the heaviest, this depended directly moisture.
Data Set for Sustainability prediction ng	When performing prediction algorithm ng, it was found that the model if it works, then the variable Soil Moisture was predicted with a high degree of confidence.
Classification ng actual data	Possessing actual data the algorithm determined that soil moisture directly dependent variable illumination, when this has no value, the second most weight is relative humidity. Therefore, based on these two factors, the classification rules are formed Soil Moisture variable was predicted with a high degree of confidence, since the data obtained by the prediction process ng approach the real, with high precision.
Actual data prediction ng	

Table2. Results of the project

Source: Project Development

6. Conclusions.

After conducting research about the factors that affect the growth of crops in a greenhouse, it was determined that the most important factors are: soil moisture, relative humidity, ambient temperature, illumination level and CO₂ since These directly influence the process of photosynthesis of plants and, establishing correct values, achieve a better nutrient uptake and obtain better fruits.

The tools and techniques of *Big Data*, and specifically *data mining* processes are critical for data analytics. While these two areas emerged as a support for decision making in economics and business, today they are of transversal use in various scenarios and focus on the discovery of patterns within a mountain of data. Through a review of the documentation used for this research, it could be found that these tools and techniques of data analytics are not governed, particularly, by a standard, but represent a set of algorithms that allow to make descriptive models on a set Of data in order to classify and / or predict information.

In the early stages of the development of this project, a search was made for databases containing various measurements representing variables captured in a real greenhouse. In this regard, it was determined using the database *Environmental data (indoor and outdoor)* repository *UMass Trace Repository*, because properly adjusted to the requirements of the project, that is, containing a sufficient set of samples and has the variables With the most important factors.

As a significant result of this project, it was found that the design of *data mining* software with an intuitive interface in a *framework of drag and drop* is a suitable alternative to process variables emissions. It was found that data processing can be performed by sequentially forming the steps of the knowledge discovery process in database (KDD) through the joining of objects representing programming modules. Specifically, the NetBean IDE 8.2 development environment, which works with the Java programming language, proved to be adequate and accurate software for the development of techniques and algorithms of data analytics, and also allowed the implementation of a user-friendly interface the user.

After the development of a software for the analysis of the environmental parameters that affect the growth of crops in greenhouses and in order to verify their correct functioning, several tests were carried out with different sources of information, one of them being found in the *Trace UMass Repository*, resulting in indoor humidity can be predicted based on other variables with **accuracy**, significantly good. Particularly, it was determined that that variable is linked to the *TimeStampUTM* variable, which refers to the time the sample was taken. The tests tested the usability and reliability of the software.

In order to verify the operation of the system with real data, tests were carried out with information obtained from the greenhouse of the farm "The prairie", through a network of wireless sensors installed in the place. As in the experiments with the test data, it was again found that the

soil moisture depends on the sampling time. Additionally, using the selection tool, the five most important variables were determined and, in that case, the result was that the level of illumination (called light) is the most important factor on which soil moisture depends. In addition, based on this factor the classification rules were calculated, so that when it has a value of 0, the variable that considers the system is the relative humidity, and it was proved experimentally that this way more knowledge is generated through the Data exploration.

Areas of *Big Data* and *Data Mining* are relatively emerging and are in constant development, and particularly its application in agriculture is a broad and diverse subject that seeks, among other things, optimize resources. Indeed, precision agriculture is focused on the use of technological tools to make efficient use and management of resources. That said, data mining techniques represent a good alternative for exploring information on agriculture-related variables and supporting intelligent decision-making.

Bibliographic references.

- 1] A. Alpi y F. Tognoni, Cultivo en invernadero, Madrid: Ediciones Mundi-Prensa, 1999.
- 2] J. J. Ibáñez, «Madrid,» Junio 2006. [En línea]. Available: <http://www.madrimasd.org/blogs/universo/2006/06/26/33002>.
- 3] N. Iglesias, «INTA,» 2009. [En línea]. Available: http://inta.gob.ar/sites/default/files/script-tmp-inta_produccion-de-hortalizas-bajo-cubierta_2006.pdf.
- 4] O. Barrios, Construcción de un Invernadero, Santiago, 2004.
- 5] D. Marlow, «Hortalizas,» Marzo 2011. [En línea]. Available: <http://www.hortalizas.com/horticultura-prottegida/invernadero/aporte-de-co2-en-un-invernadero/>.
- 6] E. H. J. Z. M. L. M. Dumbill, M. Slocum, A. Croll and C. Hill, Big Data Now-2012 Edition, Sebastopol, CA: O'Reilly Media, 2012.
- 7] D. Hand, H. Mannila y P. Smyth, Principles of Data Mining, Massachusetts: MIT, 2001.
- 8] M. Bramer, Principles of Data Mining, Springer, 2007.

- 9] J. Han, M. Kamber y J. Pei, DATA MINING Concepts and Techniques, Morgab Kaufmann Publishers, 2012.
- 10] J. Erazo y C. Hervás, «Revista Científica Maskana,» 20 Octubre 2014. [En línea]. Available: <http://www.ucuenca.edu.ec/ojs/index.php/maskana/article/view/532>.
- 11] Secretaría de Gestión de Riesgos, «Secretaría de Gestión de Riesgos | Ecuador,» 2015. [En línea]. Available: <http://www.gestionderiesgos.gob.ec/>.
- 12] Consejo de Federaciones de Bomberos de Argentina, «Consejo Nacional Bomberos Argentina,» 2006. [En línea]. Available: http://scoutsanpatricio.com.ar/libros/bomberos_03.pdf.
- 13] OMS, GUÍAS PARA LA CALIDAD DEL AIRE, Lima: Centro Panamericano de Ingeniería Sanitaria y Ciencias del Ambiente, 2004.
- 14] J. González, «Tesis Electrónicas de la Universidad Austral de Chile,» 2014. [En línea]. Available: <http://cybertesis.uach.cl/tesis/uach/2014/bmfcig643m/doc/bmfcig643m.pdf>.
- 15] Bomberos de Granollers, «Bombers de Granollers Visita d'escoles al parc,» Mayo 2011. [En línea]. Available: http://bombersgranollers.org/documentacio/incendio_forestal/CIF01_texto&imagenes_CAST_Intro.pdf.
- 16] O. Rosas, J. Flores, M. Meléndez y R. Cabrera, «EVALUACION DEL COMPORTAMIENTO DEL FUEGO BAJO DIFERENTES CONDICIONES AMBIENTALES EN EL ESTADO DE CHIHUAHUA A PARTIR DE LA APLICACIÓN DE QUEMAS PRESCRITAS,» *Scientia CUCBA*, p. 7, 2006.
- 17] R. F. Martínez, J. O. Meré, F. J. de Pisón Ascacíbar, A. G. Marcos y F. A. Elías, *Redes Inalámbricas de sensores: teoría y aplicación práctica*, Publicaciones Universidad de la Rioja, 2009.
- 18] CCM Benchmark Group, «CCM - Online Community,» Junio 2014. [En línea]. Available: <http://es.ccm.net/faq/9457-benchmark-definicion>.
- 19] ISO/IEC, «ISO/IEC 27005,» 30 Junio 2008. [En línea]. Available: http://www.pqm-online.com/assets/files/lib/std/iso_iec_27005-2008.pdf.
- 20] iso 27000, «Gestión de Riesgos tecnológicos basada en ISO 27005 para la continuidad de negocio,» 2011. [En línea]. Available: [file:///C:/Users/HENRY/Downloads/Dialnet-GestionDeRiesgosTecnologicosBasadaEnISO31000EISO27-4797252%20\(3\).pdf](file:///C:/Users/HENRY/Downloads/Dialnet-GestionDeRiesgosTecnologicosBasadaEnISO31000EISO27-4797252%20(3).pdf). [Último acceso: 1 Abril 2016].
- 21] T. R, «Arquitecturas de integración del proceso de descubrimiento de conocimiento con sistemas de gestión de bases de datos: un estado del arte.,» *Revista de Ingeniería y Competitividad*, 2012.

Sobre los Autores.

7.



Karina L. PONCE GUEVARA. She was born in Ibarra on May 7, 1993. She completed her primary studies at the Sarance Girls' School. She studied secondary education at the Higher Technological Institute "Republic of Ecuador" where she finished in 2010, obtaining a Bachelor's Degree in Sciences Mathematical

Physical Specialization. He is currently completing his degree in Engineering in Electronics and Communication Networks, Universidad Técnica del Norte - Ecuador



Edgar A. MAYA OLALLA. He was born in Ibarra, Ecuador, on April 22, 1980. He holds a degree in Computer Systems at Universidad Técnica del Norte in 2006. Master's degree in Communication Networks at the Pontifical Catholic University of Ecuador in the year 2014. He is

currently a lecturer at Career of Engineering in Electronics and Communication Networks of the Universidad Técnica del Norte