

Algoritmos de minería de datos para el análisis del riesgo de desastre fitosanitario en el municipio Nueva Paz.



Data mining algorithms for the analysis of phytosanitary disaster risk in the municipality of Nueva Paz.

Ing. Adianys Gil Montero ¹, Dr.C. Yasser Vázquez Alfonso ² & Dr.C. Efraín Velasteguí López³

Recibido: 10-02-2019 / Revisado: 15-02-2019 / Aceptado: 04-03-2019 / Publicado: 04-04-2019

Abstract.

DOI: <https://doi.org/10.33262/cienciadigital.v9i2.310>

Disaster is a function of the risk process and result of the combination of threats, conditions of vulnerability and insufficient capacity or measures to reduce the negative and potential consequences of the risk. Faced with these emergencies, studies of danger, vulnerability and risks of health disasters arise that today play a very important role for all sectors, both in the international and national community. This subject of considerable interest is the objective of this research that allows the management of risk indicators of phytosanitary disasters (epiphytic) in the municipality of Nueva Paz. A study was carried out on the different dimensions of disaster risk, defining the technologies and methodologies to be used in conjunction with mathematical and computational algorithms using Data Mining techniques implemented through mathematical algorithms: Apriori, J48, BsgNet, Ridor, Multilayer Perceptron and SimpleKMeans to analyze the information and thus discover trends and behaviors of the data and at the same time facilitate the process of decision making by the specialists of the Nueva Paz municipality.

Keywords: phytosanitary Disaster, danger, vulnerability, risk, administration, epiphytic, mathematical algorithms.

¹Universidad Agraria de la Habana, Facultad de Ciencias Técnicas, Mayabeque, Cuba, adianys@unah.edu.cu

²Universidad de la Habana, Facultad de Turismo, La Habana, Cuba, yasser@ftur.uh.cu

³Revista Ciencia Digital, Ambato, Ecuador, luisefrainvelastegui@cienciadigital.org

Resumen.

El desastre es función del proceso de riesgo y resultado de la combinación de amenazas, condiciones de vulnerabilidad e insuficiente capacidad o medidas para reducir las consecuencias negativas y potenciales del riesgo. Ante estas emergencias surgen estudios de peligro, vulnerabilidad y riesgos de desastres sanitarios que hoy en día juegan un rol muy importante para todos los sectores, tanto de la comunidad internacional como nacional. Esta temática de considerable interés constituye el objetivo de esta investigación que permite la gestión de los indicadores de riesgo de desastres fitosanitarios (epifitias) en el municipio Nueva Paz. Se realizó un estudio sobre las diferentes dimensiones que contemplan los algoritmos de minería de datos para el análisis del riesgo de desastre, definiéndose las tecnologías y metodologías a utilizar en conjunto con los algoritmos matemáticos y computacionales mediante técnicas de Minería de Datos implementadas a través de algoritmos matemáticos: Apriori, J48, BsyNet, Ridor, Perceptrón Multicapa y SimpleKMeans para analizar la información y así descubrir tendencias y comportamientos de los datos y a su vez facilitar el proceso de toma de decisiones por parte de los especialistas del municipio Nueva Paz.

Palabras claves: Desastre fitosanitario, peligro, vulnerabilidad, riesgo, gestión, epifitias, algoritmos matemáticos.

Introducción

La Organización de las Naciones Unidas para la Agricultura y la Alimentación o FAO por sus siglas en inglés, plantea que la comunidad internacional asiste hoy a un incremento sostenido de situaciones de desastres. Entre ellos los de tipo sanitario, con un elevado costo económico, social y ambiental que ponen en riesgo la sanidad humana, animal y vegetal sin descontar las afectaciones a la producción sostenible de alimentos, amenazas que se espera aumenten ante el cambio climático global (FAO, 2005).

Dentro de la clasificación de los desastres sanitarios están: los que atacan al hombre o epidemias, los que atacan los animales o epizootias y los que atacan las plantas o fitosanitarios. Estos últimos son llamados Epifitias, que pueden ser causados por organismos o agentes nocivos que constituyen plagas, enfermedades o malezas que en determinado nivel poblacional producen daño económico a las plantas (Machado, 2007).

El creciente impacto en la economía mundial de enfermedades y plagas transfronterizas exóticas ha colocado a los países ante el reto de fortalecer sus capacidades defensivas para disminuir la vulnerabilidad institucional, como ha sido reconocido por organismos internacionales como la Organización Mundial de la Sanidad Animal (OIE), la Organización Mundial de la Salud (OMS) y la FAO (Percedo, 2006).

Internacionalmente destacan otras instituciones como la Organización de las Naciones

Unidas para la Cultura y las Artes (UNESCO por sus siglas en inglés) que están relacionadas con esta temática, realizan investigaciones y llevan a cabo proyectos con la finalidad de fortalecer la capacidad de los países (en los campos del uso de la información técnica y científica y el desarrollo de metodologías participativas) para la producción de información relacionada con la zonificación de amenazas y riesgos, haciendo uso de los Sistemas de Información Geográfica (SIG) que debe permitir mejorar la toma de decisiones para la reducción de los desastres. El “Regional Action Program for Central America” (RAPCA, por sus siglas en inglés) está incluido en el Programa “Capacity Building for Natural Disaster Reduction” (CBNDR por sus siglas en inglés), ambos financiados por el Gobierno de Holanda y administrados por la UNESCO. En América, de manera general, la creciente necesidad de contar con una herramienta que facilite la gestión del riesgo ha traído consigo la creación de diversos softwares que valgan como índice de comparación y admitan un mejor alcance de la información entre políticos, científicos y defensores del ambiente y de igual modo facilite la toma de decisiones por parte de los directivos de los órganos de gobierno (SENASICA, 2013).

En Cuba las acciones para el enfrentamiento a desastres sanitarios por enfermedades y plagas exóticas y cuarentenadas (aisladas o apartadas) en animales y plantas, incluso por epidemias en la población, se ejecutan según la Directiva No. 1 para la Planificación, Organización y Preparación del País para Situaciones de Desastres, del Consejo de Defensa Nacional (CDN) (CDN, 2005).

A nivel nacional se ha reforzado la investigación e implementación de herramientas que contribuyan en esta temática. Se desarrolló REDesastres, la primera red del sector agropecuario nacional en este campo, la cual brinda soporte al trabajo del Centro de Capacitación para la Reducción de Desastres Sanitarios en Animales y Plantas (CEDESAP), adscrito al Centro Nacional de Sanidad Agropecuaria (CENSA) y auspiciado por el Ministerio de Educación Superior (MES) y el Estado Mayor Nacional de la Defensa Civil (EMNDC) (Percedo, 2008).

La idea de obtener información a partir de un conjunto de datos no es nueva. Desde hace siglos, la gente ha tratado de entender cómo extraer datos valiosos o útiles a partir de una colección de datos y luego cómo hacer uso de la información obtenida para lograr un beneficio o un objetivo útil siendo este el propósito de la minería de datos (Martínez, 2005).

La disciplina denominada Minería de Datos estudia métodos y algoritmos que permiten la extracción automática de información sintetizada que permite caracterizar las relaciones escondidas en la gran cantidad de datos; también se pretende que la información obtenida posea capacidad predictiva, facilitando el análisis de los datos de forma eficiente.

El departamento de Medio Ambiente del CITMA del municipio Nueva Paz no se encuentra ajeno a la necesidad de mantener su información recopilada bien tratada y analizada al incorporar avances tecnológicos en cuanto a automatización de procesos a través de la Minería de Datos, ya que puede reportar grandes beneficios a las organización en cuanto en

los EPVR sobre todo al Estado Mayor Nacional de la Defensa Civil al explicar problemáticas aparentemente aleatorias y abrir nuevos horizontes a la hora de la toma de decisiones, además de garantizar el cumplimiento de estos estudios para la gestión de los indicadores de riesgos de desastres fitosanitarios y recoger esta información por cada consejo popular, actividad que hasta hoy se realizan manualmente en el departamento.

En la actualidad el CITMA del municipio Nueva Paz exige a cada consejo popular entregar los resultados de los EPVR. Los indicadores del riesgo de desastre fitosanitario recogidos de estos estudios son gestionados por los especialistas en el CITMA municipal manualmente apoyándose principalmente en el uso de hojas de cálculo (Microsoft Excel) para almacenar los datos, que no están protegidas por un control de accesibilidad con privilegios, dando cabida a que cualquier persona pueda ver los datos que necesite y los que no, teniendo la posibilidad de modificarlos sin autorización. La interpretación de los datos se realiza de acuerdo a las experiencias de los especialistas y la manera en que estos los entienden, lo que ralentiza el procesamiento de la información. Las valoraciones son en la mayoría de las ocasiones acertadas, pero pueden estar sujetas a errores subjetivos. Sucede además que en ocasiones se torna difícil acceder a los datos cuando se precisa de alguno en específico, pues están contenidos en diferentes formatos.

Metodología

En la presente investigación se utilizó la información proveniente de los EPVR (Estudios de Peligro Vulnerabilidad y Riesgo) desarrolladas durante un año (2017) para la gestión de los indicadores de riesgos de desastres fitosanitarios exigidos a cada consejo popular en el departamento de Medio Ambiente del CITMA municipal de Nueva Paz, situado al extremo este sudeste de la provincia Mayabeque, en los 22o46' Lat. N y los 81o45' Long. W, limitando al Norte con el municipio de Madruga; al Este con el municipio Unión de Reyes, perteneciente a la provincia de Matanzas; al Sur con las aguas de la Ensenada de la Broa, en el Golfo de Batabanó y al Oeste con el municipio de San Nicolás.

Para llevar a cabo el análisis de los EPVR (especialistas de la dirección provincial de veterinarios, de la agricultura, de la gestión del riesgo, de la defensa civil, de aprovechamiento hidráulico, de salud pública, de higiene y epidemiología, entre otros) propuestos por el CITMA municipal de Nueva paz se tuvo en cuenta el “Índice General de Riesgo Biológico”, que se calcula a partir de los siguientes componentes:

Peligro: Constituidos por las plagas, entendiéndose como tales, cualquiera de las especies, razas, líneas o biotipos vegetales, animales o agentes patógenos dañinos para las plantas o productos vegetales, que causan perjuicio o que interfieren de cualquier otra forma en su producción, elaboración, almacenamiento, transporte o comercialización

Vulnerabilidad de los OPB: Dada por las brechas sanitarias detectadas en la protección fitosanitaria de los objetivos con riesgo biológico.

Se define como objetivo con peligro biológico (OPB) el lugar o instalación donde concurren de forma eventual o permanente, fuentes primarias o secundarias de plagas cuyo desarrollo puede dar lugar a un desastre, y las posibles vías de dispersión directa o indirecta de estas hacia la población vegetal potencialmente expuesta.

Vulnerabilidad de la población vegetal: Dada por los factores que la hacen más propensa al impacto de un brote de plaga, por ser los elementos expuestos a ese peligro.

Al evaluar el peligro biológico se tienen en cuenta el factor país (FP), el factor local (FL) y se incluyen los siguientes indicadores:

- La identificación de las plagas que puedan dar lugar a una emergencia capaz de desencadenar un desastre fitosanitario.
- La clasificación de esas plagas según la probabilidad de ocurrencia de una situación de desastre debida a su introducción y/o diseminación en el territorio objeto de estudio (nación, provincia o municipio).
- La identificación y caracterización de los escenarios de peligro y la evaluación de la probabilidad de su ocurrencia a través de métodos cualitativos y/o cuantitativos.

Factor País: Vínculos comerciales, medios de transporte, mercancías y personas, así como cercanía geográfica a países afectados, entre otros. (Alto, Medio y Bajo)

Factor Local: Evaluación del escenario de peligro según tipos de OPB existentes, relacionados con las características geográficas y las actividades económicas que en ellos se realizan (Alto, Medio o Bajo).

Para este análisis del peligro de desastre fitosanitario se utilizaron las planillas y tablas propuestas por la dirección del CITMA municipal de Nueva Paz que facilitaron el desempeño durante todo el proceso, así como las formulas diseñadas al respecto.

También aportaron información estadística a este estudio, el Circuito de Guarda Bosques de Nueva Paz, la dirección de la Empresa Agropecuaria Nueva Paz y los planes Municipales y Provinciales de producción facilitados por los organismos de la agricultura de dichas instancias.

Este estudio se llevó a cabo para 10 grupos de cultivos teniendo en cuenta 27 y 16 plagas reglamentadas y no reglamentadas respectivamente ajustadas a las especies establecidas en Lista Oficial de Plagas Reglamentadas y Plagas no Reglamentadas de la República de Cuba vigente (2007).

La evaluación del peligro biológico se realiza individualmente para cada cultivo o especie vegetal, y sus resultados se expresan cualitativamente en términos de probabilidades (Alta,

Media o Baja) o de manera cuantitativa si la información acopiada permite responder a las preguntas sobre dónde, cuándo y cómo pudiera ocurrir el evento adverso (brote de una plaga). A partir de las planillas para registrar los resultados de los EPVR orientados a la población vegetal correspondientes al CITMA del municipio Nueva Paz se procede a realizar el análisis de estos riesgos de desastres fitosanitarios a través de la aplicación los algoritmos de minería de datos correspondientes a clasificar, agrupar y asociar tales como; Perceptrón Multicapa, SimpleKMeans, J48, A priori, BayesNet y Ridor (implementados en Weka) con el objetivo de predecir de manera automatizada tendencias y comportamientos de la misma y responder a una mejor toma de decisiones por parte de los especialistas en cuestión.

Resultados

El desarrollo de esta investigación cuenta con un sistema informático que parte de una base de datos para almacenar toda la información necesaria la cual contiene de forma digitalizada las planillas para registrar los resultados de los EPVR orientados a la población vegetal del municipio Nueva paz. Utiliza, los algoritmos de minería de datos Perceptrón Multicapa, SimpleKMeans, J48, A priori, BayesNet y Ridor (implementados en Weka) para el análisis de la información referida al peligro, con el objetivo de predecir de manera automatizada tendencias y comportamientos de la misma.

Las ventajas que aporta la implementación de este sistema según los autores de esta investigación son:

- Facilita la recogida y validación de la información relacionada con el análisis del riesgo de desastre y mejora de la calidad del dato, al eliminar el tratamiento manual de la información.
- Permite mantener una alta disponibilidad de la información, homogénea y actualizada, lo que ahorra tiempo y recursos a la hora de analizar las métricas del negocio.
- Posibilita el control y comunicación: La información agregada es traducida a lenguaje de negocio y se hace visible y accesible a todos los miembros de la organización.
- Obviamente, cada usuario tendrá acceso y visibilidad únicamente a la información que necesita.
- Facilita la toma de decisiones para análisis del riesgo de desastre fitosanitario en el municipio nueva paz.

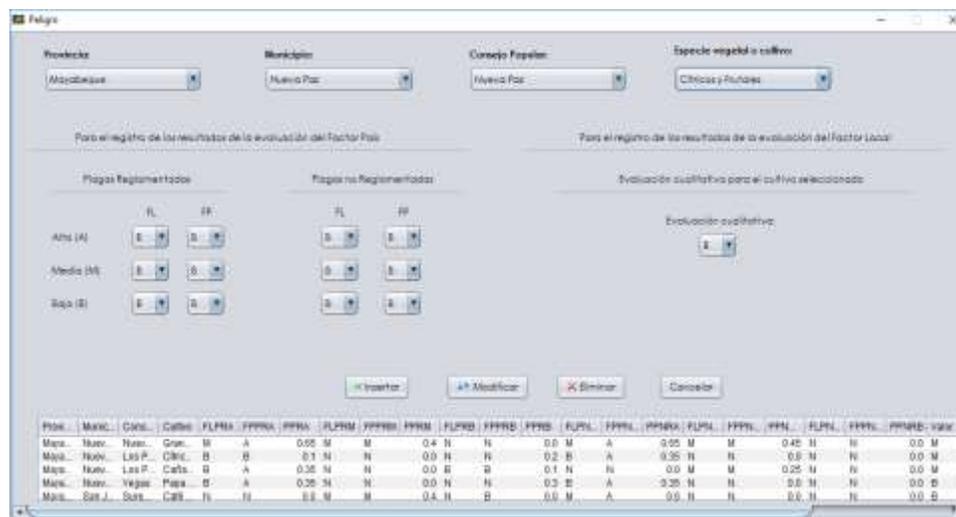
A continuación, se muestran las interfaces de inicio al sistema (figura 1 y 2), en las mismas se muestran imágenes sobre la información proveniente de los EPVR obtenida a partir de planillas y tablas propuestas por la dirección del CITMA municipal de Nueva Paz y según el “Índice General de Riesgo Biológico”, que se calcula a partir de los componentes antes mencionados.

Figura 1: Interfaz de la página inicio al sistema.



Fuente: Elaboración propia

Figura 2: Interfaz de la página inicio al sistema.

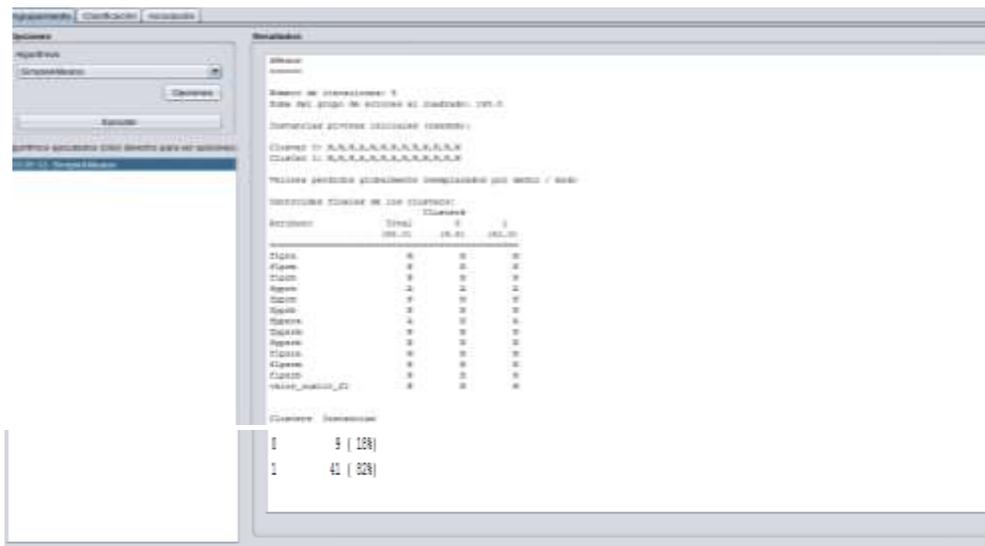


Fuente: Elaboración propia

La figura 3 muestra la salida del algoritmo Simple Kmeans, proporcionando al especialista conocer los parámetros que se tienen en cuenta durante el proceso de agrupamiento en este caso el conjunto de instancias que no son más que los indicadores antes planteados que se tienen en cuenta EPVR para analizar el riesgo de desastres fitosanitario para estos 10 plantas con sus respectivas plagas reglamentadas o no reglamentadas.

De ahí se procede a calcular la distancia euclidiana de esta instancia y a partir de los clustering seleccionados se procede a calcular sus centroides esta acción es repetida hasta que se encuentre la mayor de las similitudes entre las instancias y se dé como conclusión los grupos correspondientes.

Figura 3: Interfaz de la página Modelos de Agrupamiento (Método SimpleKMeans).



Fuente: Elaboración propia

La figura 4 muestra la salida del algoritmo Simple Kmeans, proporcionando la asignación de instancias por grupos, pero en este caso visualizando graficas de pastel.

Figura 4: Interfaz de la página Modelos de Agrupamiento.



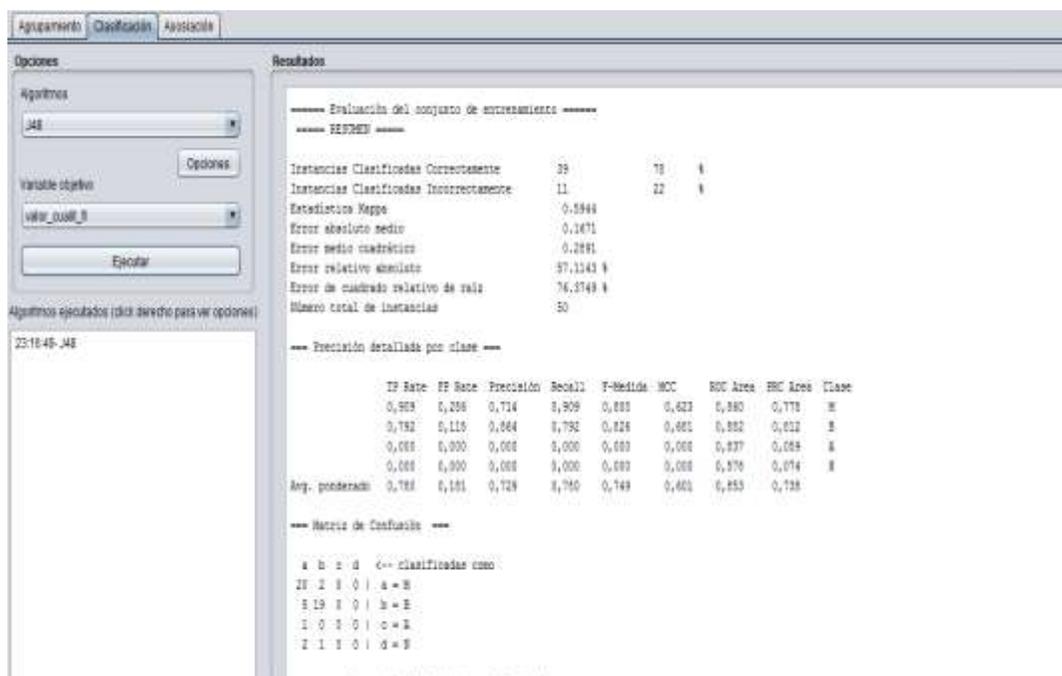
Fuente: Elaboración propia

La figura 5 muestra la salida de uno de los algoritmos utilizados dentro de los modelos de Clasificación para esta investigación en este caso el método J48 por ser entre los modelos antes mencionados el mejor clasificador para los datos analizados, proporcionando al especialista conocer los parámetros que se tienen en cuenta durante el proceso de clasificación como son los casos correctamente clasificados, el índice estadístico Kappa que refleja la coorcondancia de dichos indicadores que se interpreta que hay un acuerdo máximo entre todos(1) si no (0 que no hay).

La matriz de Confusión facilitara al especialista comprobar o por lo menos obtener una probabilidad del conjunto de instancias clasificadas según los indicadores de riesgo fitosanitario por cada una de las 10 plantas reglamentadas o no teniendo en cuenta el factos pais y el factor local todas estas clasificadas en alta media y baja según los formularios, tablas y cálculos realizados para los EPVR.

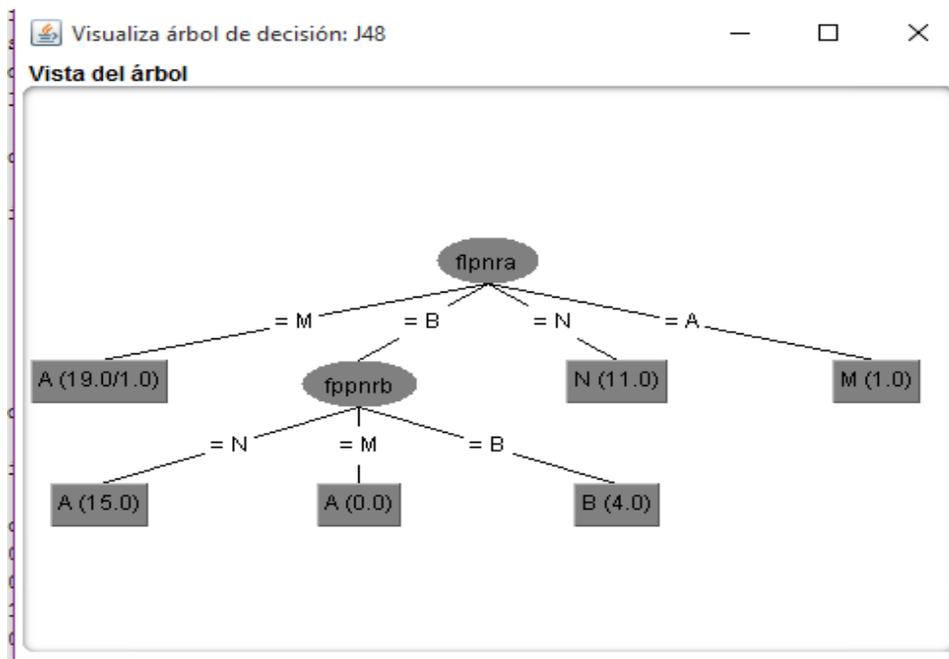
La figura 6 muestra además para este algoritmo su clasificador creando un árbol de decisión para predecir la clasificación. Los árboles de decisiones a veces también se llaman árboles de clasificación cuando se utilizan para clasificar valores nominales objetivos, o árboles de regresión cuando se utilizan para predecir un valor numérico.

Figura 5: Interfaz de la página Modelos de Clasificación (Método J48).



Fuente: Elaboración propia

Figura 6: Interfaz de la página Modelos de Clasificación (Vista del Árbol de decisión).



Fuente: Elaboración propia.

La figura numero 7 muestra la salida para el modelo de Asociación a partir de una sus aplicaciones a través de las reglas de asociación teniendo en cuenta el soporte para la toma de decisiones siendo este el principal objetivo de la investigación, facilitar el análisis del riesgo de desastre fitosanitario de los especialistas en el tema dentro del municipio Nueva Paz.

Lo que se brinda a los especialista con la salida de este algoritmo según los datos de entradas es encontrar asociaciones o correlaciones sobre un conjunto de datos ,en este caso entre los indicadores analizados. Dicho de otra manera que condiciones deben ocurrir para que se produzca cierta condición. Es decir que es lo que ocurre tantas veces antes de, por ejemplo el método Apriori extrae un conjunto de reglas de los datos y destaca aquellas reglas con un mayor contenido de información.

El proceso de evaluación de las reglas de asociación en minería de datos se realiza de acuerdo a: La cobertura (soporte): Es el número de instancias para las cuales ella predice correctamente (soporte). La precisión (confianza): Es el número de instancias que predice correctamente, expresado como una proporción de todas las instancias a las que se aplica.

Figura 7: Interfaz de la página Modelos de Asociación (Método Apriori).

The screenshot shows a software interface for the Apriori algorithm. On the left, there are configuration options for 'Apriori', including a dropdown menu set to 'Apriori', a 'Criterios' button, a dropdown menu set to 'valor_qualit_2', and a 'Ejecutar' button. Below these are 'Algoritmos aplicados' with 'Apriori' selected. The main area displays 'Resultados' for 'Relacion Base: Orogrencia'. It shows 'Reg. Contados: 10' and 'Reg. Asociados: 13'. A table lists 13 associated rules with columns: Item, Type, Num, Item, Reg., Support, Unique, and Dist. Below the table, there are 'Apriori' parameters: 'Mínimo soporte: 0,50 (50 porcentajes)', 'Mínimo soporte-confianza: 0,5', and 'Máximo de reglas asociadas: 1'. There are also four lines of text indicating the number of generated rules for different support and confidence thresholds.

Item	Type	Num	Item	Reg.	Support	Unique	Dist
0 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
1 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
2 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
4 Siglos	Reg 1004	10	0,7	0,8	1,2	2,8	0
5 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
6 Siglos	Reg 1004	10	0,7	0,8	1,2	2,8	0
7 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
8 Siglos	Reg 1004	10	0,7	0,8	1,2	2,8	0
9 Siglos	Reg 1004	10	0,7	0,8	1,2	2,8	0
10 Siglos	Reg 1004	10	0,7	0,8	1,2	2,8	0
11 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
12 Siglos	Reg 1004	10	0,7	0,8	0,7	0,8	0
13 valor_qualit_2	Reg 1004	10	0,7	0,8	1,2	2,8	0

Fuente: Elaboración propia

Conclusiones

En la investigación se arriba a las siguientes conclusiones:

- Consultaron los referentes teóricos sobre la investigación a realizar, enfocándose en la teoría referente a los sistemas informáticos que sustenta la aplicación de los algoritmos de Minería de Datos en el riesgo fitosanitario.
- Se determinaron los componentes que se relacionan con los riesgos de desastres fitosanitarios.
- Se implementaron los algoritmos de Minería de Datos en una aplicación informática para el análisis de los riesgos de desastres fitosanitarios.

Referencia bibliográfica

- AMA (2012). Propuesta de metodología para la realización de los estudios de peligro, vulnerabilidad y riesgos de desastres fitosanitarios en la agricultura, 28 de junio, La Habana, Cuba.: Agencia de Medio Ambiente (AMA), p.80.
- CDN (2005). Directiva No. 1 del Vicepresidente del Consejo de Defensa Nacional para la Planificación, Organización y Preparación del País para Situaciones de Desastres., 1ro de junio, Ciudad de La Habana, p.30.
- FAO (2005). El cambio climático puede incrementar el número de víctimas del hambre. Comité de Seguridad. Alimentaria Mundial, 26 de mayo, Roma, Italia, s.l.: s.n.

- MACHADO (2007). Manual para la preparación comunitaria en situaciones de Desastres., Rodas: Ediciones Damují, Cuba.p.170..
- MARTÍNEZ, B. B.(2005). Minería de Datos. Tesis en opción al título de Licenciado en Ciencias de la Computación. Facultad de Ciencias de la Computación. Universidad Autónoma de Puebla.México. 67p.
- PERCEDO (2006). ESTRATEGIA PARA LA REDUCCIÓN DE DESASTRES SANITARIOS EN LA POBLACIÓN ANIMAL EN CUBA. EL CASO DE LA INFLUENZA AVIAR., Mayabeque, Cuba: Rev. Salud Anim. Vol. 28 No. 3: 174-181.
- PERCEDO (2008). CEDESAP y REDesastres. Una contribución a la preparación y gestión intersectorial y multidisciplinaria, San José de las Lajas, La Habana, Cuba.: Veterinaria Organización.
- SENASICA (2013). La Sanidad e Inocuidad en México. Disponible en <https://www.gob.mx/senasica/>, Consultado el 20 de enero 2019.

Para citar el artículo indexado.

Gil A, Vázquez Y. & Velasteguí E. (2019) Algoritmos de minería de datos para el análisis del riesgo de desastre fitosanitario en el municipio Nueva Paz. *Revista electrónica Ciencia Digital* 3(2), 306-318. Recuperado desde:

<http://cienciadigital.org/revistacienciadigital2/index.php/CienciaDigital/article/view/310/904>



El artículo que se publica es de exclusiva responsabilidad de los autores y no necesariamente reflejan el pensamiento de la **Revista Ciencia Digital**.

El artículo queda en propiedad de la revista y, por tanto, su publicación parcial y/o total en otro medio tiene que ser autorizado por el director de la **Revista Ciencia Digital**.

