



Ingeniería en Informática

Proyecto de Trabajo de Fin de Carrera

**“Aplicación de técnicas de minería de datos a
un repositorio de variables fitofenológicas de
cultivos cítricos”**

Alumnos:

Martín Darío Ehman (65347)

&

Gabriel Adrián Surraco (65326)

Director: Ing. Karina Eckert

Primera Instancia de Presentación

Año: 2016

ÍNDICE

1. TÍTULO DEL TRABAJO DE INVESTIGACIÓN	3
2. PALABRAS CLAVES.....	3
3. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN	3
4. OBJETIVOS.....	4
4.1. GENERAL.....	4
4.2. ESPECÍFICOS.....	4
5. ALCANCE	5
6. MARCO TEÓRICO CONCEPTUAL.....	5
6.1. CONCEPTOS REFERIDOS A TICS	5
6.1.1. <i>Minería de datos</i>	5
6.1.1.1 Agrupamiento	7
6.1.1.2 Asociación	7
6.1.1.3 Clasificación.....	8
6.1.1.4 Regresión.....	8
6.1.2. <i>Metodologías para la minería de datos</i>	8
6.1.2.1 La metodología CRISP-DM.....	9
6.1.3. <i>Herramientas utilizadas en minería de datos:</i>	9
6.1.3.1 R-Project	9
6.1.3.2 Weka.....	10
6.2. CONCEPTOS REFERIDOS AL CULTIVO DE LOS CÍTRICOS.....	10
6.2.1. <i>Fitofenología</i>	10
6.2.2. <i>Triángulo de enfermedades en las plantas</i>	10
6.2.3. <i>Diaphorina Citri</i>	11
6.2.4. <i>Minador de la hoja de los cítricos</i>	12
6.2.5. <i>Moscas de los frutos</i>	12
6.3. TRABAJOS RELACIONADOS	13
6.3.1. <i>Knowledge Discovery and Data Mining to Identify Agricultural Patterns.</i>	13
6.3.2. <i>Knowledge Discovery on Agricultural Dataset Using Association Rule Mining.</i>	14
6.3.3. <i>Modelado para la predicción de enfermedades en cultivos de alto valor comercial.</i>	14
7. JUSTIFICACIÓN	14
8. METODOLOGÍA.....	14
9. CRONOGRAMA DE ACTIVIDADES.....	16
TABLA 1: CRONOGRAMA DE ACTIVIDADES.....	17
10. BIBLIOGRAFÍA	18

1. TÍTULO DEL TRABAJO DE INVESTIGACIÓN

Aplicación de técnicas de minería de datos a un repositorio de variables fitofenológicas de cultivos cítricos.

2. PALABRAS CLAVES

Minería de datos. CRISP-DM. Fitofenología de cultivos cítricos.

3. PLANTEAMIENTO DEL PROBLEMA DE INVESTIGACIÓN

El estado de la sanidad de los cultivos es un factor primordial para el progreso económico y social. El acelerado incremento de las poblaciones obliga a mantener una alta eficiencia y productividad de las agro industrias para obtener productos de excelente calidad que satisfaga a los consumidores.

Durante el crecimiento de una planta se tienen en cuenta tres ejes fundamentales conocidos como el triángulo de la enfermedad. Este triángulo está compuesto por factores que son imprescindibles monitorear en un cultivo: el ambiente (el estado meteorológico, así como también las condiciones del suelo y de la comunidad biótica que conviven con las plantas del cultivo en ese mismo ambiente), los patógenos (plagas y enfermedades) y el hospedero (cultivos). Es importante entonces el conocimiento y manejo de estos factores para estar bien informados del entorno en el que crecen los cultivos y de esta manera lograr una mejor fuente de información para la toma de decisiones.

En la actualidad la tecnología ha evolucionado lo suficiente como para facilitar muchas tareas por medio de sensores que permiten la recolección de datos y su posterior procesamiento. De esta manera se minimizan los costos y tiempos de trabajo sobre las plantaciones. En el ámbito de la agricultura en los últimos años la cantidad de información se ha incrementado exponencialmente gracias a innovadoras técnicas de monitoreo. A medida que esta cantidad de información aumenta, los métodos tradicionales de almacenamiento, procesamiento y análisis de datos se vuelven insuficientes. [1]

En Argentina, en un proyecto conjunto, el INTA (Instituto Nacional de Tecnología Agropecuaria), la Asociación de Citricultores de Concordia, el ACDI (Asociación Cultural para el Desarrollo Integral), una ONG con capacidades en informática y comunicaciones; y contando con el apoyo financiero del BID (Banco Interamericano de Desarrollo), se desarrolló una herramienta para el manejo integrado del cultivo cítrico llamada FruTIC. Este sistema permite la captura de datos meteorológicos, fenológicos (etapas de crecimiento de los cultivos), información fitosanitaria e información del suelo de las distintas zonas que conforman la región citrícola del río Uruguay¹. Esta infraestructura tecnológica permite capturar, transferir, procesar, generar y divulgar en tiempo operativo información de gran utilidad para productores, investigadores y técnicos. [2]

La región citrícola del río Uruguay está dedicada principalmente a la producción de mandarinas y naranjas destinadas al mercado fresco; con un total de 55.839 hectáreas plantadas con cítricos, es la de mayor extensión de la República Argentina, siendo la exportación un objetivo principal en las empresas que buscan mayor rentabilidad. Las PyMEs argentinas productoras de frutos cítricos enfrentan importantes problemas de competitividad. La competencia internacional se incrementa continuamente, así como los requerimientos de precio, calidad y certificación. Son crecientes también las restricciones vinculadas con aspectos sanitarios y el nivel de los residuos de agroquímicos en los alimentos. Estas exigencias constituyen verdaderas trabas para el comercio externo y aumentarán en el futuro. [3]

¹Sureste de Corrientes, Noreste y Este de Entre Ríos.

Si bien el INTA posee registros fenológicos de las principales combinaciones cítricas comerciales, solo en años recientes se ha implementado una metodología más detallada y precisa para su registro.

Actualmente la información capturada por el sistema FruTIC se utiliza a través de una interfaz Web, esto permite a productores y asesores técnicos tener información climática, de la floración y brotación de los cultivos, y presencia de plagas en tiempo real para mejorar la calidad de las frutas y la eficiencia del productor. Los productores incorporados en este sistema pueden saber cuándo comienza a desarrollarse algún patógeno en el ambiente, de esta manera pueden aplicar controles oportunos y así evitar la propagación de las plagas o patógenos a tiempo. Además, el productor puede aplicar controles puntuales y no a todo el lote reduciendo costos agropecuarios. Este sistema permite tomar decisiones a partir de si la plaga se encuentra en el ambiente y si este ambiente es adecuado para que se desarrolle la plaga, en base a esto puede decidirse si aplicar el agroquímico o no. Este tipo de proceso demuestra que aún no es necesario entrar en genómica para mejorar la calidad de los frutos y/o disminuir costos de producción.

Si bien los datos presentados por la interfaz web son en tiempo real, no se tiene en cuenta los datos históricos. Este sistema presenta cierta debilidad para la toma de decisiones ya que no brinda una información precisa sobre qué relaciones existen entre las variables presentes, y si es que la hay realmente, el grado de influencia entre una y otra; los productores y técnicos se basan en cuestiones empíricas. La ausencia de dicho análisis dificulta o impiden responder preguntas como: ¿Qué relación existe entre las infestaciones de moscas de los frutos y la temperatura?, ¿Qué relación existe entre el crecimiento de los frutos y la temperatura y/o la humedad del suelo?, ¿Cuál es el nivel de influencia entre las variables presentes? o inquietudes como el estado de los cultivos luego de temporadas de heladas y duración de las mismas.

El presente trabajo de investigación tiene como propósito estudiar dicha información histórica mediante la aplicación de técnicas de minería de datos sobre los repositorios para encontrar correlaciones entre variables fitofenológicas, descubrir patrones de comportamiento a través del tiempo y establecer modelos que representen dichas relaciones; y de esta forma tener un respaldo para la toma de decisiones. Estas decisiones influyen fuertemente sobre el tipo de acciones y el momento del ciclo de los cultivos cítricos en el que se deben llevar a cabo: fumigaciones, riegos, fertilizaciones, podas, adición de nutrientes en el suelo, con el fin de optimizar el uso de estos recursos, entre otras.

4. OBJETIVOS

4.1. General

Determinar las características que influyen en el desarrollo de los cultivos cítricos a través de las variables fitofenológicas y del triángulo de las enfermedades que son almacenadas en el sistema FruTIC, aplicando técnicas de minería de datos.

4.2. Específicos

- Realizar una revisión teórica sobre los principales tópicos acerca del manejo integrado de cultivos y las técnicas de minería de datos más utilizadas.
- Seleccionar y pre-procesar los datos con los que se van a trabajar durante la investigación.
- Aplicar las técnicas de minería de datos que se adapten a las características de los datos pre-procesados del FruTIC.

- Evaluar y validar los modelos de minería de datos aplicados de acuerdo a los métodos de validación estudiados. Elaborar reportes y visualizaciones.

5. ALCANCE

Los datos que se utilizarán fueron aportados por el INTA ubicado en Concordia, Entre Ríos. Estos repositorios están almacenados en la plataforma del FruTIC, una herramienta que se brinda al productor o técnico que contiene información actualizada sobre los cultivos. Estos datos son capturados a partir de monitoreos que realizan los técnicos en quintas oficiales y sensores meteorológicos, con el fin de facilitar el manejo de información sobre el triángulo de las adversidades, y de esta manera tener una fuente para la toma de decisiones técnicas.

Con respecto a los datos, se cuenta con la información meteorológica, la cual contiene las siguientes variables: temperatura mínima, máxima y media; precipitación acumulada; humedad mínima, máxima y media; velocidad media y máxima del viento a 2 metros de altura; radiación global media y máxima; y humedad de la hoja. Las variables meteorológicas son tomadas cada una hora por una estación automática.

Además, se utilizará información fenológica, la cual contiene datos de dos lotes de naranjas (042-INTA-Salustiana, Don Tito Valencia Late) y tres lotes de mandarinas (058-INTA-NovaR, Lote II Nova, Don Tito Nova) con los siguientes atributos: lote, número de planta, altura de la observación, orientación de observación sobre la planta, etapa del brote y etapa de floración. También se cuenta con información de plagas presentes, con las siguientes variables: lote, fecha del registro, número de planta, método (monitoreo de Diaphorina / monitoreo de minador), valor según punto cardinal (número de ramas con presencia de plagas. Del 0 al 3). Conjuntamente se dispone de información sobre el estado del cultivo, con las siguientes variables: lote, fecha de registro, número de planta, estado de la misma. Además, se tiene datos del fruto con las siguientes variables: lote correspondiente, fecha de registro, número de planta, punto cardinal de donde se tomó la muestra, color y calibre del fruto. Dicha información comprende registros desde el año 2006 al 2015 para los lotes Don Tito Nova y Don Tito Valencia Late, y desde el año 2010 al 2015 para los demás lotes mencionados.

Si bien la investigación comprende el ámbito de las granjas propias del INTA Concordia, se pretende que la misma sea de utilidad también para productores particulares de la región.

6. MARCO TEÓRICO CONCEPTUAL

Dado que el trabajo se centra en tareas de minería de datos y una plataforma de información fitofenológica, haremos una reseña de los principales aportes en estas temáticas agrupando en dos categorías, conceptos referidos a TICs y por otra parte conceptos referidos a temas agrícolas.

6.1. Conceptos referidos a TICs

6.1.1. Minería de datos

“Se define minería de datos al proceso de descubrir patrones interesantes y novedosos, así como también la obtención de modelos descriptivos, predictivos y comprensibles a partir de datos a gran escala”. [4]

La minería de datos es el resultado de los avances en las tecnologías de la información en los últimos tiempos. Las funcionalidades en el manejo de datos fueron evolucionando en diferentes aspectos: la cantidad de información almacenada, las velocidades de intercambio de datos, la disponibilidad y forma de acceso a Internet en el mundo, la forma de coleccionar datos y el análisis de los mismos.

Se han estado almacenando enormes volúmenes de información en bases de datos y *data warehouses* en donde los análisis de datos tradicionales sobre los mismos se volvieron insuficientes. Actualmente existen nuevos retos como trabajar con datos no estructurados, incompletos, de diferentes tipos, con ruido y no representativos. [5]

La minería de datos no es tomada como una disciplina apartada de las demás, sino que reúne técnicas interdisciplinarias tomadas de la estadística, inteligencia artificial y tecnologías de bases de datos. La minería de datos también es conocida como el Descubrimiento de Conocimiento en Bases de Datos, o en sus siglas en inglés, KDD (*Knowledge Discovery in Databases*), y se refiere a un proceso completo que parte de la obtención de los datos y pre-procesamiento de los mismos antes de sumergirse en ellos para descubrir patrones interesantes y finalizando en la presentación de los resultados obtenidos en la extracción de conocimiento. Si bien ciertos autores toman a la minería de datos como una etapa dentro del proceso KDD, puntualmente donde se aplican las técnicas de aprendizaje automático, otros autores definen a la minería de datos como el proceso KDD completo, en esta investigación adoptaremos la última definición. [6]

Además de la cantidad de datos en la actualidad, otra característica importante es la variedad en cuanto a tipos de datos. Las fuentes de datos pueden incluir bases de datos tradicionales, *data warehouses*, la Web, datos multimedia, datos provenientes de *streaming* y datos provenientes de sistemas geográficos. [6]

Uno de los aspectos iniciales con los que se debe lidiar es el pre-procesamiento de los datos debido a que pueden coexistir varias fuentes de datos heterogéneas, por ello estos datos deben ser obtenidos, limpiados, formateados e integrados. Luego estos datos pueden ser almacenados en un sistema de bases de datos para luego ser analizados con métodos analíticos. Si bien el término minería de datos da una noción de métodos analíticos, la realidad es que la mayoría del trabajo está en la parte de la preparación de los datos. [7]

Los datos pueden ser de diferentes formatos o tipos. Pueden ser de tipo numérico, categóricos, texto, espaciales, entre otros. La implementación de distintas técnicas de minería de datos tiene en cuenta los tipos de datos utilizados, incluso si existen datos faltantes o no. [8]

Según el objetivo que se busca en un proyecto de minería de datos puede categorizarse cinco tipos de tareas: Análisis de datos exploratorio, modelado descriptivos, modelado predictivos, descubrimiento de patrones y reglas, recuperación por contenido. [8]

En los análisis de datos exploratorios el objetivo es explorar los datos sin tener una idea clara de lo que se está buscando. Se suelen utilizar técnicas interactivas y visuales. A medida que la cantidad de dimensiones crece, se hace más difícil poder visualizar los datos. Cuando la cantidad de atributos o dimensiones es mayor a 3 o 4 se suelen utilizar técnicas de proyección. [8]

La utilización de modelos descriptivos busca describir todo el conjunto de datos o especificar el proceso que los genera. Se utilizan estimaciones de densidad, agrupamientos, segmentación y modelos que describen relaciones entre atributos denominados modelos de dependencia. Son muy utilizados en segmentación de mercado estableciendo grupos de clientes con perfiles similares y así poder generar campañas de marketing personalizadas a cada grupo. [8]

Se utilizan modelos predictivos, clasificación y regresión, para poder predecir valores de alguna variable teniendo en cuenta todas las otras variables. La principal diferencia entre una regresión y clasificación es el tipo de dato a predecir, numérica y categórica respectivamente. La diferencia clave entre un modelo descriptivo y un predictivo es que en el predictivo se tiene como objetivo predecir una única variable, en cambio en el modelo descriptivo ninguna variable es central en el modelo. [8]

Las tres tareas presentadas anteriormente se centran en la construcción de un modelo que generalice los datos, en cambio según el objetivo del proyecto puede que lo que

se busque sea detectar patrones. La detección de patrones puede utilizarse por ejemplo para descubrir fraude en transacciones de tarjetas de crédito, o para descubrir combinaciones de ítems frecuentes en las transacciones de supermercados a través del uso de reglas de asociación. [8]

En la recuperación por contenido el usuario tiene un patrón de interés y quiere encontrar patrones similares en los datos. Por lo general se utiliza para bases de datos de textos e imágenes. Un ejemplo claro es el buscador de Google, a partir de una palabra o conjuntos de palabras encuentra las páginas más apropiadas para la búsqueda, o bien, a partir de una imagen puede buscarse otras imágenes similares a la especificada. [8]

Entre las técnicas de aprendizaje automático más utilizadas en la minería de datos podemos nombrar a las regresiones lineales y múltiples, clasificadores probabilísticos, árboles de decisión, algoritmos de agrupamiento, redes neuronales, reglas de asociación y algoritmos genéticos. La determinación de la técnica y el algoritmo a utilizar va a depender de varios factores entre ellos los tipos de datos, los objetivos del analista de datos, la velocidad y precisión del modelo que se quiera obtener.

Las técnicas de minería de datos también pueden ser clasificadas según su tarea de descubrimiento: Agrupamiento, Asociación, Clasificación, Regresión.

6.1.1.1 Agrupamiento

Las técnicas de agrupamiento es el proceso de dividir un conjunto de datos u observaciones en subconjuntos. Cada subconjunto es un *cluster* o grupo tal que los objetos en un grupo son similares entre sí, pero son diferentes a los objetos de los otros grupos. En este contexto, diferentes métodos de agrupamiento pueden generar diferentes grupos en los mismos conjuntos de datos. Esta técnica de minería de datos es útil ya que puede conducir al descubrimiento de grupos previamente desconocidos dentro de los datos. [6]

El análisis de *clustering* ha sido ampliamente utilizado en muchas aplicaciones, tales como la inteligencia de negocio, reconocimiento de patrones en imágenes, búsquedas Web, biología y seguridad. En la inteligencia de negocio el agrupamiento puede utilizarse para organizar un número grande de clientes en grupos, donde cada cliente dentro de un grupo comparte características similares. Esto facilita el desarrollo de estrategias de negocio para mejorar la relación con los clientes, por ejemplo, campañas de marketing personalizadas para cada grupo de clientes. El *clustering* también puede ser utilizado para la detección de valores anómalos, donde estos valores pueden ser más interesantes que los casos comunes. Las aplicaciones de detecciones de valores anómalos incluyen la detección de casos como el fraude en tarjetas de crédito y el monitoreo de actividades criminales en comercio electrónico. Por ejemplo, casos que no son tan comunes en transacciones de tarjeta de crédito, tales como compras no frecuentes o compras de elevado valor, estas pueden ser vistas como posibles actividades fraudulentas. [6] Entre las técnicas de agrupamientos de datos podemos encontrar *k-means* y los mapas auto-organizados de Kohonen (SOM).

6.1.1.2 Asociación

Las técnicas de asociación en minería de datos consisten en encontrar relaciones y asociaciones significativas entre los valores de los atributos en un conjunto de datos. Una de las técnicas más utilizadas son las reglas de asociación. Las mismas surgieron a partir de la necesidad de analizar bases de datos de ventas para encontrar relaciones fuertes entre la compra de diferentes productos dentro de las transacciones. Generalmente las bases de datos de ventas consisten en un conjunto de transacciones donde cada transacción contiene productos adquiridos por un cliente. En este contexto, se buscan encontrar reglas de la forma (ítem1, ítem2, ..., ítem n) => (ítem1, ítem2, ..., ítem m). De esta manera se puede decir que la ocurrencia de un evento dado por el antecedente de la regla dispara la ocurrencia de un segundo evento que corresponde al consecuente de la regla. Por ejemplo, si un cliente compra pan y leche, entonces también compra mantequilla. La fuerza de esta regla dentro

de la base de datos está dada por la cantidad de veces que aparece esta combinación de elementos o conjunto de elementos dentro del conjunto de datos. [9]

Uno de los algoritmos más utilizado es el *Apriori* propuesto por Agrawal y Srikant. [10]

6.1.1.3 Clasificación

El problema de clasificación consiste en asignar una observación a cierta categoría o grupo teniendo en cuenta los valores de los atributos. La clasificación encuentra características comunes entre las diferentes instancias y los clasifica en diferentes clases o grupos. Para ello crea un modelo a partir de los datos existentes teniendo en cuenta los atributos de cada instancia y el valor de la clase a la que pertenece. A partir de este modelo lo que se busca es asignar una instancia con clase desconocida a un grupo o clase en particular. [7] Entre las técnicas de clasificación podemos encontrar clasificadores Bayesianos, Redes Neuronales, kNN, entre otras.

6.1.1.4 Regresión

El análisis de regresión consiste en estimar relaciones entre una variable dependiente y una o varias variables independientes, a éstas últimas también se las denomina variables predictoras. La regresión ayuda a entender como varían los valores de la variable dependiente al ir modificando los valores de las variables independientes.

A diferencia de la clasificación, que busca predecir valores discretos o categorías, en la regresión se busca predecir valores continuos teniendo en cuenta algún modelo o función. Dentro de estas técnicas podemos encontrar las regresiones simples y múltiples, también pueden utilizarse redes neuronales para la predicción de datos continuos. [6]

6.1.2. Metodologías para la minería de datos

Para la realización de proyectos de minería de datos es apropiado seguir una metodología para una mejor organización y ejecución de las tareas, desarrollando la misma en diferentes etapas secuenciales y/o paralelas. En la actualidad existen varias metodologías estandarizadas que proveen las etapas y actividades generales que se necesitan desarrollar en un plan de minería de datos. Las mismas pueden modificarse teniendo en cuenta las necesidades del proyecto que se lleve a cabo. Las más utilizadas son el proceso KDD (*Knowledge Discovery in Databases*), SEMMA (*Sample, Explore, Modify, Model and Assess*), CRISP-DM (*Cross Industry Standard Process for Data Mining*) y metodologías propias del investigador.

El modelo KDD fue el primero en surgir en la comunidad científica, en el año 1996, y rige los pasos generales del proyecto teniendo en cuenta que la minería de datos en sí es un la tarea dentro del proyecto en donde se realiza la extracción de los patrones, aunque hoy en día los términos KDD y minería de datos se utilizan indistintamente haciendo referencia al proceso completo de descubrimiento de patrones (desde la obtención de datos, limpieza del conjunto de datos, extracción del conocimiento y evaluación de resultados).

Con los avances en la minería de datos surgieron otras metodologías: SEMMA, Catalyst y CRISP-DM. Según una encuesta en la comunidad KDnuggets publicado en el 2007 la metodología más utilizada dio como resultado a CRISP-DM. [11]

La metodología CRISP-DM profundiza en mayor parte las tareas y actividades a realizar durante el proyecto mientras que los demás modelos proveen etapas más generalizadas a ejecutar.

Por tal motivo en este trabajo de investigación se eligió la misma realizando algunas modificaciones necesarias.

6.1.2.1 La metodología CRISP-DM

El modelo consiste en 6 fases relacionadas entre sí de una forma cíclica (con retroalimentación). [12]

- **Entendimiento del negocio:** Se basa en el estudio de la situación actual, planteamiento de la problemática, establecimiento de los objetivos, comprensión del contexto, desarrollo de un plan del proyecto.
- **Entendimiento de los datos:** Una vez que se establecieron los objetivos del proyecto hay que comprender los datos y los requerimientos de información para llevar a cabo el proyecto. Algunas de las tareas que se encuentran en esta fase son: recolección de los datos, descripción de los mismos, exploración y verificación de la calidad de datos. Se pueden utilizar métodos visuales de estadísticas, análisis de clusters para ver si el conjunto de datos mantiene un patrón de agrupamiento. En esta fase es necesario definir claramente lo que se quiere analizar para poder describir claramente los datos que se va a utilizar en el análisis (aplicación de técnicas de minería).
- **Preparación de los datos:** Una vez que se tienen los datos a utilizar en el proyecto es necesario realizar un pre-procesamiento de los mismos. Esta fase implica la limpieza, transformación y formateo de los datos para luego poder realizar un buen modelado.
- **Modelado:** En esta etapa se seleccionan y se aplican las técnicas de minería de datos sobre los conjuntos de datos. Es frecuente tener que volver a la fase de preparación de datos para revisar aspectos que no se tuvo en cuenta. Por ello CRISP-DM es una metodología cíclica (con retroalimentación).
- **Evaluación:** El modelo o los modelos construidos en la etapa anterior deben ser evaluados en contexto con los objetivos que se plantearon inicialmente. Esto puede llevar a volver a fases anteriores para cubrir requerimientos que no han sido cubiertos (si se encuentra, por ejemplo, una variable que afecta al análisis pero que no se ha tenido en cuenta al definir los datos) o bien por la aparición de nuevas necesidades. Es la fase más crítica porque se hace la interpretación de los resultados por lo tanto es de gran importancia revisar y volver hacia atrás si fuese necesario.
- **Despliegue:** En esta etapa se realiza un plan de despliegue, es decir, detallar los pasos de cómo se integrará los modelos y/o descubrimientos obtenidos para su utilización. Además, se debe prever un plan de monitoreo y mantenimiento para evaluar el desempeño de los modelos a medida que se los va utilizando. Finalmente se presentan los reportes del proyecto y una evaluación general del mismo.

6.1.3. Herramientas utilizadas en minería de datos:

6.1.3.1 R-Project

R es un lenguaje de programación utilizado para computación estadística y visualización de información. Corresponde a un proyecto GNU similar a el lenguaje S que fue desarrollado por *Bell Laboratories* (antes AT&T, ahora llamado *Lucent Technologies*) por John Chambers y otros. [13]

Es utilizado por miles de personas diariamente para realizar análisis estadísticos. Es gratis y de código libre. R provee una amplia variedad de técnicas estadísticas (modelado lineal y no lineal, pruebas estadísticas clásicas, análisis de series de tiempo, clasificación, agrupamiento, ...) y gráficas, y es altamente extensible.

El entorno de R incluye: facilidad en el manejo y almacenamiento de datos, un conjunto de operadores para cálculos sobre vectores y matrices, herramientas integradas para el análisis de datos, facilidades en cuanto a la generación de gráficos, un lenguaje

efectivo que incluye estructuras de bucles, condicionales, funciones definidas por el usuario y métodos para leer y escribir datos.

El lenguaje y entorno R se ha vuelto una de las tecnologías más utilizadas por analistas de datos y científicos, proveyendo el fácil acceso a miles de paquetes que contribuyen los usuarios. [14]

6.1.3.2 Weka

Weka es una colección de algoritmos de aprendizaje automático y herramientas de pre-procesado de datos. Está escrito en Java y provee una API bien documentada la cual permite la integración con aplicaciones propias. Weka es de código libre bajo la licencia GNU GPL, es decir que cualquier aplicación en donde sea utilizado Weka debe ser liberado bajo la licencia GPL. [15]

Weka posee una interfaz gráfica amigable lo cual permite la implementación de algoritmos directamente sobre los datos sin la necesidad de programar. Además, esta interfaz también permite la visualización de los datos. Weka tiene una variada gama de herramientas disponibles partiendo desde el pre-procesamiento de datos, como algoritmos de agrupamiento, regresión, clasificación, reglas de asociación y visualizaciones.

Debido a que Weka está disponible para su libre descarga y ofrece una gran variedad de poderosas características (que por lo general no se encuentran en software comercial), se ha vuelto una de las herramientas de minería de datos más utilizada. Weka también se ha vuelto uno de los medios favoritos dentro del campo de investigación en minería de datos.

6.2. Conceptos referidos al cultivo de los cítricos

6.2.1. Fitofenología

“La fitofenología tiene como finalidad estudiar y describir de manera integral los diferentes eventos fenológicos que se dan en las especies vegetales dentro de ecosistemas naturales o agrícolas, en su interacción con el medio ambiente. En este sentido, la realización de las observaciones fenológicas, consideradas importantes, son la base para la implementación de todo sistema agrícola, permitiendo a los productores agrarios obtengan con su aplicación una mayor eficiencia en la planificación y programación de las diferentes actividades agrícolas conducentes a incrementar la productividad y producción de los cultivos”. [16]

“Una fase fenológica viene a ser el periodo durante el cual aparecen, se transforman, o desaparecen los órganos de las plantas. También pueden entenderse como el tiempo de una manifestación biológica”. [16]

“Una etapa fenológica está delimitada por dos fases fenológicas sucesivas. Dentro de ciertas etapas se presentan periodos críticos, que son el intervalo breve durante el cual la planta presenta la máxima sensibilidad a determinado evento meteorológico, de manera que las oscilaciones en los valores de este evento se reflejan en el rendimiento del cultivo; estos periodos críticos se presentan generalmente poco antes o después de las fases, durante dos o tres semanas. El comienzo y fin de las fases y etapas sirven como medio para juzgar la rapidez del desarrollo de las plantas”. [16]

6.2.2. Triángulo de enfermedades en las plantas

Una enfermedad es un deterioro sostenido o progresivo de las células o tejidos de un organismo que causan anomalías funcionales o estructurales. Algunas enfermedades son superficiales, estropeando solo la apariencia física de la planta infectada, mientras que otras enfermedades pueden tener un mayor número de víctimas al afectar tejidos críticos para el crecimiento del organismo, debilitando o inclusive eliminando a la planta.

La enfermedad ocurre cuando un agente causante de la enfermedad conocido como patógeno se encuentra con el organismo hospedante adecuado, bajo condiciones ambientales favorables para el desarrollo de la enfermedad.

Estos tres elementos, patógeno, hospedante y ambiente forman parte de lo que denominamos *triángulo epidemiológico* o *triángulo de la enfermedad*. El triángulo de la enfermedad es un concepto que ilustra la importancia de los tres elementos. [17]

En primer lugar, un patógeno tiene que estar presente en el ambiente. Los patógenos causantes de las enfermedades de las plantas pueden ser hongos, virus, bacterias y otros microbios. Luego, el patógeno tiene que entrar en contacto con una planta hospedante susceptible. Las interacciones entre el patógeno y el hospedante son por lo general muy específicos, algunos patógenos han evolucionados para atacar solamente a un tipo particular o incluso a especies de organismos. Si no se encuentra un hospedante apropiado, no se presentaría la enfermedad, ya que el patógeno no posee su fuente de alimentación.

Adicionalmente, incluso si una especie es susceptible, no todos los individuos de esa especie tendrán el mismo nivel de susceptibilidad.

Finalmente, deben presentarse las condiciones ambientales adecuadas para que el patógeno cause la enfermedad. Muchos patógenos se desarrollan en árboles bajo condiciones de estrés: inundaciones, sequías, nutrición no balanceada, luz solar inadecuada, heridas, por ejemplo, pueden inducir estrés o lesiones en plantas huéspedes, haciéndolos vulnerables a la enfermedad.

Los productores y científicos deben considerar las variables económicas y de mercado, como también las complejas interacciones entre el patógeno, el hospedante y el ambiente, que determinan la calidad del cultivo. A pesar de que el mercado se preocupa por manejar tendencias en la agricultura, el ambiente sigue siendo la influencia más crucial en la incidencia de la enfermedad de la planta de año a año. El triángulo de la enfermedad es un concepto clásico de patologías en las plantas utilizado para examinar el rol del ambiente en procesos de enfermedad durante la vida de la planta.

El concepto del triángulo de la enfermedad fue formalizado en 1960 por George McNew, un científico del Instituto Boyce Thompson para Investigación Vegetal. McNew sugirió que el triángulo de la enfermedad puede ser utilizado para estudiar la interrelación de varios factores en una epidemia y para entender como las epidemias pueden ser predichas, limitadas o controladas. Fue destinado para utilizarse como una herramienta empírica hasta que madure la investigación para proveer nuevos métodos para predecir y controlar enfermedades. McNew definió ampliamente los parámetros del triángulo de la enfermedad como “la susceptibilidad inherente del hospedante, el potencial inóculo del parásito, y el impacto del ambiente en el parasitismo y la patogénesis”. [17]

6.2.3. Diaphorina Citri

Es considerada una plaga importante en plantaciones cítricas debido a que produce malformaciones de tallos y hojas por la inyección de toxinas, y además puede ser vector de la bacteria *Liberibacter* la cual produce HLB (*Huanglongbing*) o *greening*. La duración del período embrionario varía de 9,7 días a 15°C, a 3,5 días a 28°C. Los huevos son colocados en el extremo de los brotes tiernos, sobre y entre hojas tiernas desplegadas. La oviposición está condicionada a la presencia de brotes tiernos. La hembra alcanza a poner 800 en toda su vida. Posteriormente, nacen las ninfas que se encargan de formar colonias en las ramitas tiernas. Las ninfas secretan una sustancia que se depositan sobre las hojas. [18]

La duración del ciclo biológico (huevo adulto) varía de 14,1 días (a 28°C) a 49,3 días (a 15°C), siendo las temperaturas de 25°C a 28°C las más adecuadas para su desarrollo. El psílido no se desarrolla a temperaturas de 33°C y 10°C. Presenta un pico poblacional al final de la primavera e inicios del verano coincidente con el periodo de brotación de los cítricos. [18]

Los síntomas están relacionados con los frutos y el árbol. Los arboles afectados quedan achaparrados, sin crecimiento y la fruta no tiene valor comercial. En las hojas aparece un amarillamiento irregular a lo largo de las nervaduras que se disemina en los tejidos adyacentes produciendo además un síntoma de moteado. Los frutos en los árboles afectados son más pequeños que los normales y tienen sabor amargo. [19]

La inspección del insecto y el manejo de la enfermedad requieren un programa de manejo integrado de plagas. En el caso de que el vector y la enfermedad estén presentes, es necesario el control. Una población es considerada elevada cuando se observan tres ninfas y cinco adultos por rama. Durante el monitoreo es conveniente observar el incremento de ninfas en los tallos o muestrear los adultos grávidos luego del período invernal.

6.2.4. Minador de la hoja de los cítricos

Según Asplanato [20], el minador de la hoja de los cítricos, *Phyllocnistiscitrella Stainton*, es la plaga foránea de introducción más reciente y es el único lepidóptero considerado actualmente de importancia en el cultivo. La larva de *Phyllocnistiscitrella* realiza galerías en las hojas tiernas, daño que afecta en mayor medida a las plantas más jóvenes. Además, provoca daños indirectos de relevancia, debido a que las heridas causadas son una vía de entrada para la bacteria *Xanthomonas axonopodis* pv. *citri*, organismo causante de la canchrosis en los cítricos, considerada uno de los problemas sanitarios de mayor relevancia en la agricultura. Para la supresión de sus poblaciones son necesarias repetidas intervenciones químicas, lo que genera un aumento en el costo de producción, aparición de resistencia de la plaga, presencia de residuos en los alimentos y medio ambiente, y disminución de enemigos naturales del minador y de otras especies fitófagas.

En la etapa de adulto el minador presenta características morfológicas similares a una pequeña mariposa, cubiertas de un color blanco nacarado. Presenta medidas promedio de 3 mm de largo y de 5 mm de expansión alar. Las alas son escamosas con numerosos pelos marginales largos que poseen varias franjas oscuras transversales y longitudinales. En su extremo apical presentan una mancha negra bien notoria y un fleco de cerdas largas. Las alas posteriores son muy estrechas y con un fleco de color blanco. Posee antenas filiformes largas. Los huevos son colocados individualmente tanto en el haz como en el envés de las hojas. Con respecto a la larva, la misma traspasa la epidermis de la hoja y comienza inmediatamente a formar la mina separando, mediante movimientos de la capsula cefálica, las células por debajo de la cutícula y alimentándose de los jugos producidos. La larva siempre permanece dentro de la galería que construye. [20]

La temperatura óptima para la cópula es de alrededor de 25°C, no logran oviponer tanto a temperaturas menores de 15°C como a mayores de 35°C. La fecundidad es muy variable, claramente dependiente de la temperatura. El insecto presenta un desarrollo ininterrumpido a lo largo del año. El rango de temperaturas entre 18 y 32°C es adecuado para las diferentes fases de desarrollo. La duración del ciclo de vida es muy variable, oscila entre 11 a 52 días dependiendo de la temperatura.

Para el manejo de la plaga es necesario determinar los momentos adecuados de tratamientos químicos. Estos deben ser dirigidos cuando la brotación y la densidad de la plaga justifiquen la aplicación de fumigaciones. Es recomendable que el monitoreo se realice en las etapas iniciales de brotación de la planta y cuando las plagas se encuentren en sus primeras etapas de desarrollo ya que son más sensibles a los productos químicos. Los mayores deterioros de los cultivos se presentan en brotaciones de verano y otoño, por ello debe intensificarse el monitoreo en éstos periodos.

6.2.5. Moscas de los frutos

Una de las principales plagas en el NEA son las moscas de los frutos. Principalmente la mosca del Mediterráneo y, en segundo lugar, pero no menos relevante la mosca

Sudamericana. Los problemas derivados de las moscas no son solo los daños directos sobre los frutos, sino que son plagas cuarentenarias por lo que generan problemas en la comercialización.

La mosca del Mediterráneo (*Ceratitis capitata*) está presente en toda la zona del Litoral, sin embargo, las poblaciones más altas se registran en el noreste de Corrientes y en la provincia de Entre Ríos. Esto es debido a la gran cantidad de hospederos tanto cultivados como silvestres, además del clima adecuado para el desarrollo de estas moscas. Las variedades más afectadas son las naranjas Valencia Late y las mandarinas Satsuma. Por lo general las poblaciones comienzan a incrementar en septiembre-octubre y continúan hasta marzo-abril, disminuyendo en los meses de invierno. [21]

El adulto mide de 4 a 5mm de largo, por 10-15mm de envergadura alar. El tórax es de color negro en la parte superior, con diseños simétricos blancos. Una característica de esta mosca es la de posarse apoyando las alas sobre la superficie. Las hembras adultas alcanzan la madurez sexual en 4-5 días. Luego de aparearse inician la postura, tras un periodo de 4-5 días a una temperatura de 24-27°C, durante los cuales los huevos maduran en su interior. Posteriormente, selecciona el fruto donde va a oviponer, camina sobre él para determinar el mejor lugar para la oviposición, introduce el ovipositor hasta el mesocarpio; moviéndolo, forma una cámara donde coloca los huevos, variando de 4-10 huevos. Después del nacimiento la larva se introduce hasta el endocarpio o pulpa haciendo galerías hasta llegar al centro de la fruta. Cuando termina el periodo larval sale de la fruta y se entierra en el suelo a poca profundidad para transformarse en pupa. Al terminar el periodo pupa emergen como adultos. La temperatura óptima de desarrollo es de 24-27°C. En esas condiciones el ciclo biológico dura alrededor de 17-29 días. [21]

La mosca Sudamericana (*Anastrepha fraterculus*) es una especie de mayor tamaño a la anterior, el adulto mide aproximadamente 6,5mm; es de color castaño claro; posee una mancha amarillo oscuro en forma de S en las alas. El ciclo biológico es similar a la mosca del Mediterráneo, no obstante, este ciclo puede extenderse aún más, dependiendo de las condiciones meteorológicas. [21]

El control químico de las moscas de las frutas se realiza con cebos tóxicos, formados por insecticidas y sustancias proteínicas; estas últimas resultan atractantes para las moscas. Un atrayente muy utilizado es la melaza de la caña de azúcar al 5%. También se utiliza extracto de levadura que se encuentra en el mercado nacional y además otros compuestos importados.

6.3. Trabajos relacionados

6.3.1. Knowledge Discovery and Data Mining to Identify Agricultural Patterns.

En esta investigación los autores proponen la utilización de un proceso de descubrimiento del conocimiento sobre datos provenientes de actividades agrícolas de cultivos de trigo y arroz. Se basan en utilización de diferentes técnicas supervisadas y no supervisadas para encontrar patrones importantes que describan el comportamiento de los datos. Para ello realizan el todo el proceso de descubrimiento y hacen un mayor hincapié en la técnica de minería de datos conocida como agrupamiento. Desarrollan varios modelos haciendo variar los parámetros y así poder evaluar los resultados obtenidos para luego optar por los modelos que tuvieron un mejor desempeño. Los autores realizan las implementaciones correspondientes utilizando Weka como herramienta de minería de datos. [22]

6.3.2. Knowledge Discovery on Agricultural Dataset Using Association Rule Mining.

En el presente trabajo los autores utilizan el proceso de descubrimiento de conocimiento enfocándose en el uso de reglas de asociación sobre datos provenientes de cultivos agrícolas. De esta manera buscan encontrar reglas y relaciones fuertes entre las características de los suelos y los cultivos. Posteriormente realizan una comparación en cuanto al desempeño de dos algoritmos que difieren en la generación de reglas para optar por el que de mejores resultados usando este conjunto de datos. [23]

6.3.3. Modelado para la predicción de enfermedades en cultivos de alto valor comercial.

En esta tesis de grado el autor propone modelos matemáticos como herramienta para predecir enfermedades en los cultivos de arándanos. Los datos colectados están referidos al concepto del triángulo de la enfermedad (datos del patógeno, del ambiente y del hospedante). Se basa en una técnica matemática ampliamente utilizada en la minería de datos y el descubrimiento del conocimiento conocido como regresión logística. De esta manera construye modelos que representen a los datos y con estos modelos poder predecir el comportamiento futuro de las enfermedades y así valerse de las gestiones correspondientes para evitar propagaciones de la plaga tanto a nivel técnico como a nivel económico. [24]

7. JUSTIFICACIÓN

El INTA cuenta con repositorios de datos capturados por una herramienta de manejo integrado de cultivos cítricos, denominado FruTIC.

Actualmente las decisiones tomadas sobre los lotes se realizan de manera empírica teniendo en cuenta los datos en tiempo real del FruTIC. Realizar un estudio de mayor complejidad permitiría potenciar el análisis sobre las relaciones existentes entre las variables y el comportamiento de las mismas. Todo esto es posible gracias a que las técnicas de minería de datos permiten encontrar patrones ocultos en los datos, que con otras técnicas tradicionales no serían posibles. Además, el estudio del comportamiento de las variables del triángulo de la enfermedad dentro de los registros históricos del FruTIC permite un espectro mucho más amplio para la toma de decisiones en lugar de tener solamente una base fundada sobre datos en tiempo real.

Los resultados que se pretenden obtener tales como patrones y correlaciones entre atributos pueden ayudar a una eficiente toma de decisiones para los productores y técnicos, entre estas decisiones se puede encontrar diferentes acciones a realizar para optimizar el cultivo, tales como frecuencia de riegos, cantidad de aplicaciones de pesticidas, cantidad de nutrientes para el suelo, entre otras.

Por otro lado, la presente investigación busca impulsar y promover el análisis de datos provenientes de sectores agrícolas para que los actores involucrados en las actividades tengan una fuente de información idónea para la toma de decisiones, y así poder realizar acciones sobre los cultivos en los momentos más oportunos. Un análisis más exhaustivo de datos también permitiría conocer los factores que limitan la productividad de los cultivos y aumentar el margen de rendimiento. De esta manera, se pueden reducir consumos innecesarios de agua en riego, como así también reducir el uso de plaguicidas, conociendo patrones climáticos favorables, y lograr una producción más ecológica y sustentable.

8. METODOLOGÍA

Para cumplir con los objetivos propuestos se seleccionó como base metodológica la metodología CRISP-DM (*Cross Industry Standard Process for Data Mining*), la cual consiste

en un modelo de procesos jerárquicos. La misma será adaptada en cuanto a tareas adicionales que no refleja la metodología y sean necesarias para la investigación.

Las herramientas a utilizar tanto para los análisis descriptivos y predictivos son R-Project y Weka.

Las etapas contempladas son:

I. Estudio del estado del arte

El estudio del estado del arte se realizará consultando bibliografía existente, sitios en Internet, trabajos de investigación, consultas a expertos en cuestiones de Agricultura y Minería de datos.

Tareas:

- a) Descripción de los conceptos claves de minería de datos y manejo integrado de cultivos cítricos.
- b) Estudio de antecedentes relacionados al tema del proyecto de investigación en cuestión.

II. Comprensión de los datos

Los datos fueron provistos por el INTA en hojas de cálculos de Excel los cuales van a codificarse en formato *csv* (valores separados por comas) para luego ser cargadas al entorno de *RStudio*.

La exploración de datos se realizará en el mismo entorno a través de técnicas estadísticas de descripción de datos y visualización.

- a) Recopilación inicial de los conjuntos de datos.
- b) Descripción de los datos.
- c) Exploración de datos.
- d) Verificación de calidad de datos.

III. Preparación de los datos

La preparación de los datos se realizará en el lenguaje R utilizando como entorno de trabajo el *RStudio*.

Tareas:

- a) Selección de datos.
- b) Limpieza de datos.
- c) Integración de datos.
- d) Formateo de datos.

IV. Modelado

La selección de las técnicas adecuadas dependerá de la estructura de los conjuntos de datos armados (tipos de variables, existencia de variables de respuesta), así como también de los objetivos propuestos.

Tareas:

- a) Selección de la técnica de minería de datos.
- b) Generación de diseño de prueba.
- c) Construcción del modelo.
- d) Evaluación del modelo.

V. Evaluación

Para evaluar los modelos se pretende dividir los datos en conjuntos de entrenamiento y conjuntos de validación.

Se realizará un análisis de sensibilidad variando los parámetros de cada técnica de minería de datos utilizada.

También estos resultados deben tener sentido en el contexto que se quieran aplicar, por ello también se dispone de la ayuda de los expertos en el campo de Agricultura y Minería de datos.

Tareas:

- a) Evaluación de resultados.
- b) Revisión de procesos.

VI. Cierre y conclusiones finales

- a) Planificación de despliegue.
- b) Presentación de resultados.
- c) Documentación.
- d) Revisión del proyecto.
- e) Generación de conclusiones finales.
- f) Identificación de líneas futuras de investigación.

Redacción del informe final

9. CRONOGRAMA DE ACTIVIDADES

El siguiente cronograma está desarrollado teniendo en cuenta los siguientes recursos: dos tesisistas, ambos con disponibilidad de 20 horas semanales. Se estima que el trabajo final de carrera se elaborará en el término de 8 meses a partir de la fecha en que el anteproyecto sea aprobado por el Comité Evaluador. En la siguiente tabla se muestra la distribución de etapas y actividades durante el proyecto.

Tabla 1: Cronograma de actividades

[illegible]

10. BIBLIOGRAFÍA

- [1] M. Castro y M. Domenech, «Agro Big Data: el próximo desafío.,» *AgroBarrow*, nº 55, pp. 23-24, 2014.
- [2] V. Hochmaier, S. Garran, R. Mika, J. Mousques, E. Zaballo, L. Burdyn, A. Taie, A. Freixas, J. Cerrudo y G. Blanco, «FruTIC y MEF: herramientas básicas para poder implementar un manejo fitosanitario integrado en el cultivo cítrico de la región del río Uruguay.,» *XV Jornadas Fitosanitarias*, 2015.
- [3] A. Stablum, S. Franco, S. Ibarrola, S. Milera, S. Garrán, R. Mika y S. Marnetto, «FruTIC: un sistema interactivo que permite un manejo integrado del cultivo cítrico.,» *39JAIIO - CAI*, pp. 696-710, 2010.
- [4] M. Zaki y W. Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press ed., New York, 2014, p. 12.
- [5] H. David, «Data mining: Statistics and More?,» *The American Statistician*, nº 55, pp. 112-118, 1998.
- [6] J. Han, M. Kamber y J. Pei, *Data mining: Concepts and Techniques*, Amsterdam: Elsevier, 2012, pp. 2-8.
- [7] C. Aggarwal, *Data Mining: The Textbook*, New York: Springer, 2015, pp. 1-6.
- [8] H. David, H. Mannila y P. Smyth, *Principles of data mining*, Cambridge: Mass.: MIT Press, 2001.
- [9] A. Rodriguez, J. Martinez, J. Carrasco y J. Ruíz Shulcopler, «Minería de reglas de asociación sobre datos mezclados.,» *Coordinacion de Ciencias Computacionales - INAOE*, 2009.
- [10] R. Agrawal y R. Srikant, «Fast algorithms for mining association rules in large databases.,» *Proceedings of the International Conference on Very Large Databases*, pp. 478-499, 1994.
- [11] J. Moine, A. Haedo y S. Gordillo, «Estudio comparativo de metodologías para minería de datos.,» *XIII Workshop de Investigadores en Ciencias de la Computación*, pp. 278-281, 2011.
- [12] P. Chapman, R. Kerber, C. Julian, K. Thomas, R. Thomas, C. Shearer y R. Wirth, «Keithmccormick.,» [En línea]. Available: <http://keithmccormick.com/wp-content/uploads/CRISP-DM%20No%20Brand.pdf>. [Último acceso: 20 03 2016].
- [13] R-project.org, «R-project.,» [En línea]. Available: <https://www.r-project.org/about.html>. [Último acceso: 20 03 2016].
- [14] G. Daroczi, *Mastering Data Analysis With R*. Birmingham, Birmingham: Packt Publishing Ltd.,

2015.

- [15] Weka, «Weka 3 - Data Mining Wit Open Source Machine Learning Software in Java,» [En línea]. Available: <http://www.cs.waikato.ac.nz/ml/weka/>. [Último acceso: 03 06 2016].
- [16] W. J. Yzarra Tito y F. M. Lopez Ríos, *Manual de observaciones fenológicas*, pp. 9-12.
- [17] K.-B. Scholthof, «The disease triangle: pathogens, the environment and society.,» *Nature Reviews Microbiology*, nº 5, pp. 152-156, 2007.
- [18] C. García Darderes, «Diaphorina citri Kuwayama (Hemíptera: Psyllidae), vector de la bacteria que causa el Huanglongbing (HLB - Greening),» SENASA, Argentina.
- [19] R. Bernal, «Diaprorina Citri (Homóptera: Psyllidae) Nuevo insecto detectado en montes cítricos en el área de Salto, Uruguay,» *Instituto Nacional de investigación Agropecuaria*, nº 25, p. 1, 1991.
- [20] G. Asplanato, «El minador de la hoja de los citricos. Phyllocnistis citrella (Lepidoptera: Gracillariidae): Bioecología y control biológico.,» *Instituto Nacional de Investigación Agropecuaria*, pp. 9-11, 2009.
- [21] M. Ragone y S. Garrán, «Manual para productores de naranja y mandarina de la Región del Rio Uruguay.,» INTA, 1996, p. Capítulo 11.
- [22] K. Kaur y M. Singh, «Knowledge Discovery and Data Mining to Identify Agricultural Patterns,» *International Journal of Engineering Sciences and Research Technology*, vol. 3, nº 3, pp. 1337-1345, 2014.
- [23] F. Khan y S. Divakar, «Knowledge Discovery on Agricultural Dataset Using Association Rule Mining,» *International Journal of Emerging Technology and Advanced Engineering*, vol. 4, nº 2, pp. 925-930, 2014.
- [24] E. Bombelli, *Modelado para la predicción de enfermedades en cultivos de alto valor comercial*, Ciudad Autónoma de Buenos Aires: Editorial de la Universidad Tecnológica Nacional, 2011.