



Universidad Nacional  
de **Río Negro**

***Análisis e Interpretación de Datos  
Aplicado al Manejo Orgánico en  
Frambuesos sobre Artrópodos, Patógenos  
y Agentes de Biocontrol, en El Bolsón, Río  
Negro.***

*Trabajo Final de Carrera  
Licenciatura en Sistemas*

***Alumna: Tec. Giuliana Fois***  
[\*fois.giuliana@gmail.com\*](mailto:fois.giuliana@gmail.com)

***Directora: Dra. Paola Verónica Britos***  
[\*pbritos@unrn.edu.ar\*](mailto:pbritos@unrn.edu.ar)

***Co-Directora: Lic. Paola Andrea Pizzingrilli***  
[\*ppizzingrilli@unrn.edu.ar\*](mailto:ppizzingrilli@unrn.edu.ar)

*Viedma, Río Negro, Argentina  
Diciembre 2023*

## ***Dedicatoria***

*A mi hijo Benjamín que me enseñó el verdadero valor de la vida.  
A mi compañero de vida Tomás, por el apoyo incondicional para  
lograr mis metas.*

## Resumen

Este trabajo final de carrera evalúa el efecto de diferentes prácticas de manejo sobre la biodiversidad de artrópodos y la presencia de agentes patógenos y de biocontrol en cultivos orgánicos de frambuesa (*Rubus idaeus*) en chacras de El Bolsón, Río Negro.

A pesar de las ventajas ambientales y productivas de la región, la falta de información sobre el manejo de la salud de los cultivos afecta la producción orgánica de frutas finas.

Se propone aplicar la ciencia de datos para analizar el manejo orgánico de frambuesas, centrándose en artrópodos, patógenos y agentes de biocontrol. La colaboración con biólogos/as del Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural y docentes de la Universidad Nacional de Río Negro respaldará el estudio.

La metodología ASUM (Analytics Solutions Unified Method for Data Mining) se elige para desarrollar el proyecto de ciencia de datos, siguiendo cinco fases: Analizar, Diseñar, Configurar y Construir, Implementar y/o Desplegar, y Operar y Optimizar.

Los datos se obtienen de un proyecto de investigación dirigido por la Dra. María Verónica Chillo. Se utilizan herramientas como Orange para el análisis, y se estudian atributos relevantes como el manejo del suelo, el riego, el desmalezado y la presencia de artrópodos.

La aplicación de la ciencia de datos en la industria agropecuaria desempeña un papel fundamental en la transformación y mejora de las prácticas agrícolas. Al aprovechar las herramientas analíticas y predictivas de la ciencia de datos, los agricultores pueden tomar decisiones más informadas y precisas. El análisis de grandes conjuntos de datos provenientes de diversas fuentes, permite identificar patrones, tendencias y correlaciones que de otra manera podrían pasar desapercibidos. Esta información facilita la optimización de los procesos agrícolas, desde la gestión del riego y la fertilización hasta la prevención de enfermedades y plagas. Además, la ciencia de datos contribuye a la creación de modelos de predicción que ayudan a anticipar y mejorar la planificación de cultivos y maximizar la eficiencia en la producción. En última instancia, la aplicación de la ciencia de datos en la industria agropecuaria no solo impulsa la productividad y la rentabilidad, sino que también fomenta prácticas más sostenibles al minimizar el uso de insumos y optimizar el rendimiento de los cultivos.

**Palabras Clave:** *Frambuesos, Frutas Finas, Ciencia de Datos, Cultivo, Agroecológico.*

## **Agradecimientos**

A la Dra. Paola V. Britos por dirigir mi trabajo final de carrera con tanta dedicación, a la Lic. Paola Pizzingrilli por su acompañamiento durante el proceso y a mi grupo de investigación *I*<sup>3</sup>.

A mis hermanos Fiore y Gian que fueron mis primeros amigos, confidentes y cómplices.

A mi mamá Ana y mi papá Aldo por los valores que me inculcaron.

A mi abuela Erica, que tiene el orgullo de ver su primera nieta licenciada.

A mis amigas, amigos y toda mi familia por compartir la alegría de mis logros.

A los profesores y compañeros de cursada por las incontables horas compartidas.

A Rubén, mi tío del corazón, con quien tuve el placer de aprender, trabajar y tengo muchos lindos recuerdos de la infancia y se que estaría orgulloso de mi. Un beso al cielo.

**Cumple sus sueños quien persiste.**

# Índice

<b>Dedicatoria.....</b>	<b>2</b>
<b>Resumen.....</b>	<b>3</b>
<b>Agradecimientos.....</b>	<b>4</b>
<b>Índice.....</b>	<b>5</b>
<b>1. Introducción.....</b>	<b>6</b>
1.1. Contexto del Trabajo Final de Carrera.....	6
1.2. Contenido del Trabajo Final de Carrera.....	7
<b>2. Estado de la Cuestión.....</b>	<b>9</b>
2.1. Ciencia de Datos.....	9
2.2. Cultivos de Frambuesos.....	10
<b>3. Metodología Aplicada para Desarrollar Proyectos de Ciencia de Datos.....</b>	<b>12</b>
3.1. Metodología ASUM.....	12
3.1.1 Características de la Metodología.....	12
3.1.2 Analizar.....	14
3.1.3 Diseñar.....	15
3.1.4 Configurar y Construir.....	16
3.1.5 Implementar y/o Desplegar.....	18
3.1.6 Operar y Optimizar.....	20
<b>4. Implementación.....</b>	<b>22</b>
4.1. Herramienta Utilizada.....	22
4.2. Estudio de Atributos.....	23
<b>5. Resultados y Análisis.....</b>	<b>25</b>
5.1. Modelo de Datos.....	25
5.2. Descripción de los Resultados.....	26
<b>6. Conclusiones.....</b>	<b>35</b>
<b>7. Referencias Bibliográficas.....</b>	<b>35</b>
<b>Anexo 1 : Glosario.....</b>	<b>38</b>
<b>Anexo 2 : Muestreo de Tabla Utilizada.....</b>	<b>43</b>

## 1. Introducción

En este capítulo se presenta el contexto del trabajo final de carrera (sección 1.1) y se describe el contenido del trabajo final de carrera (sección 1.2).

### 1.1. Contexto del Trabajo Final de Carrera

La producción de fruta fina en la Comarca Andina del Paralelo 42° [1], que incluye las ciudades de El Bolsón, El Hoyo, El Maitén, Lago Puelo y Epuypén, representa el 60% de los productores a nivel nacional. La sanidad de los cultivos es clave para la producción y las condiciones ambientales en la región, como los bajos niveles de humedad y la condición regional de zona libre de mosca<sup>1</sup> de los frutos, permiten disminuir e incluso no utilizar herbicidas y plaguicidas. Además, el uso de materia orgánica como modo de fertilización, favorece las condiciones del suelo. Por lo expuesto anteriormente, podemos concluir que las condiciones de la región facilitan el manejo orgánico de las plantaciones.

A pesar de estas ventajas, la producción de fruta fina de manera orgánica se ve afectada por falta de información sobre cómo gestionar la salud de los cultivos. Esto repercute tanto en la producción actual como en las transiciones hacia prácticas agroecológicas en el futuro.

Los datos utilizados para este trabajo son tomados del Proyecto de Investigación de la Universidad Nacional de Río Negro (PI 40-B-700 2018) “Comparación del efecto de diferentes prácticas de manejo orgánico en frambuesa sobre la biodiversidad de artrópodos, patógenos y agentes de biocontrol en El Bolsón (Río Negro)” dirigido por la Dra. en Ciencias Biológicas María Verónica Chillo.

Como solución para mejorar el rendimiento y la comprensión de los datos, se propone la aplicación de la ciencia de datos. Este enfoque implica la transformación y procesamiento de los datos recopilados, la evaluación de los métodos de extracción de información, y la provisión de apoyo en la toma de decisiones para el área en estudio.

Mediante la aplicación de la ciencia de datos nuestro objetivo es analizar el manejo orgánico en frambuesos, centrándonos en los artrópodos, patógenos y agentes de biocontrol en El Bolsón, Río Negro. Buscamos comprender cómo varía la proporción de gremios tróficos y las posibles plagas en función de las prácticas de manejo de cada chacra y en distintas estaciones del año. El propósito final es proporcionar información precisa a los productores de fruta fina en la Comarca Andina del Paralelo 42° para el manejo efectivo de cultivos orgánicos de frambuesos.

---

<sup>1</sup> La **zona libre de mosca de los frutos** se refiere a áreas geográficas sin presencia detectada de la mosca de los frutos, una plaga que afecta diversos cultivos. Estas zonas se establecen mediante medidas de control y monitoreo para prevenir la propagación de la plaga. Mantener estas áreas libres es esencial para proteger la agricultura, facilitar el comercio de productos agrícolas y garantizar la calidad de los cultivos.

Dentro de los objetivos más específicos se incluye:

- Estructurar los datos de manera precisa, clara y concreta.
- Centralizar la información para facilitar el acceso y la gestión.
- Agrupar datos por áreas y/o sectores específicos para un análisis detallado.
- Maximizar el potencial analítico mediante la explotación eficiente de la información.
- Incrementar la calidad de los informes generados a partir de los datos recopilados.
- Identificar la diversidad de artrópodos asociados a los cultivos de frambuesa.
- Identificar la presencia de agentes fúngicos patógenos y de biocontrol en las líneas de producción de frambuesas.
- Predecir eventos futuros mediante el análisis de patrones en los datos disponibles.
- Evaluar las relaciones entre las prácticas de manejo, las condiciones microambientales, la diversidad de artrópodos y la presencia de agentes patógenos y de biocontrol.
- Generar información útil específicamente orientada al manejo orgánico y ecológico de cultivos de frambuesas.

Para llevar a cabo este estudio, se establecerá una colaboración estrecha con biólogos/as del Instituto de Investigaciones en Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD), así como con docentes de la Universidad Nacional de Río Negro.

## **1.2. Contenido del Trabajo Final de Carrera**

El trabajo se estructura en seis capítulos: “Introducción”, “Estado de la Cuestión”, “Metodología Aplicada para Desarrollar Proyectos de Ciencia de Datos”, “Implementación”, “Resultados y Análisis” y “Conclusiones” a los que se agrega un anexo con información complementaria.

En el capítulo uno “Introducción” se presenta el contexto del trabajo final de carrera y se describe el contenido del trabajo final de carrera.

En el capítulo dos “Estado de la Cuestión” se introduce el concepto de Ciencia de Datos, y se presentan los cultivos de Frambuesas.

En el capítulo tres “Metodología Aplicada para Desarrollar Proyectos de Ciencia de Datos” se expone la metodología ASUM aplicada para desarrollar proyectos de ciencia de datos y se describen las características de dicha metodología.

En el capítulo cuatro “Implementación”, se detallan las herramientas utilizadas y se describe el estudio de atributos.

En el capítulo cinco “Resultados y Análisis” se muestran los casos de éxito y se detallan los puntos relevantes de la investigación.

En el capítulo seis “Conclusiones” se presentan los aportes de este trabajo final de carrera y se señalan futuras líneas de investigación relacionadas con problemas abiertos identificados durante el desarrollo de las investigaciones que se considera de interés continuar.

Finalmente, se incluye una sección con las Referencias Bibliográficas, donde se listan las fuentes consultadas.

En el Anexo 1, se muestra el glosario para entender la terminología de esta investigación.

En el Anexo 2 , se ejemplifican algunos registros de la tabla utilizada en el estudio para ayudar a la comprensión de los datos.



## 2. Estado de la Cuestión

En este capítulo se desarrolla el estado de la cuestión de los dos aspectos vinculados con el objetivo de este trabajo de final de carrera: la Ciencia de Datos y las particularidades del cultivo de frambuesos.

### 2.1. Ciencia de Datos

Durante los últimos años, tanto las personas como los sensores y dispositivos de comunicación utilizados, han ido generando grandes volúmenes de datos. Un estudio realizado por EMC [2] indica que en los últimos siete años hemos sido capaces de generar más información que en toda la historia de la humanidad, y además, la duplicamos cada dos años. A partir de esta comprobación, surge el término de *Datos Masivos* (o *Big Data*).

En la actualidad, podemos encontrar múltiples definiciones para el término de *Datos Masivos*, pero básicamente podemos describirlo como un gran conjunto de datos que supera la capacidad del hardware para su almacenamiento, y el software convencional de gestión y análisis. Esto se fundamenta en que tanto el software como el hardware tradicional son insuficientes no solo por el volumen de información que pueden almacenar, sino también por el contenido heterogéneo y la velocidad con la que se generan.

A dicho término se lo asocia con las 3V: Volumen (cantidad), Velocidad (de creación y utilización) y Variedad (fuentes de datos estructurados, no estructurados o híbridos), tal como lo menciona Joyanes [3] sobre una publicación de Gartner realizada en 2001. Además, en el último período, han aparecido otras dos “V”: el Valor y la Veracidad. El primer término hace referencia a los beneficios asociados al análisis de los datos que logran reducción de costos, eficiencia operacional y mejoras del negocio entre otras. El segundo, hace referencia a que se debe tratar de procesar el gran volumen de datos con el fin de obtener información útil y confiable para tomar decisiones basadas en datos exactos.

La *Ciencia de Datos* (o *Data Science*) surge a partir de la insuficiencia de métodos y análisis clásicos para procesar los grandes volúmenes de datos. Si bien el término mencionado es relativamente nuevo, los orígenes remontan a la década de 1960 donde Turkey [4] [5] ya se preocupaba por los grandes conjuntos de datos y proponía poner énfasis en el análisis de datos para sugerir hipótesis a probar. Por otra parte, la ciencia de datos se establece como multidisciplinar ya que involucra la estadística, matemática e ingeniería de datos entre otros campos, para desarrollar procesos, técnicas y sistemas que extraigan conocimiento de grandes volúmenes de fuentes de información diversas y complejas.

Fayad et al. [6] en 1996 definieron el término *Descubrimiento de Conocimiento en Bases de Datos* (o *Knowledge Discovery in Databases, KDD*) como un proceso que

identifica patrones válidos, novedosos, potencialmente útiles y comprensibles de una manera no trivial.

La *Minería de Datos* (o *Data Mining*) es una parte central dentro del proceso de descubrimiento en bases de datos, lo cual requiere que se ejecuten pasos que definan procesos y tecnologías que transformen en conocimiento la información. A partir de ello, surge la *explotación de información*, una subdisciplina que aporta a la inteligencia de negocio las herramientas para la transformación de información en conocimiento y define la búsqueda de patrones interesantes y de regularidades importantes en grandes masas de información [7], para eso, define un grupo de tareas relacionadas lógicamente que, a partir de un conjunto de información con un cierto grado de valor para la organización, se ejecuta para lograr otro, con un grado de valor mayor que el inicial [8].

Britos [9] propone como procesos de explotación de la información lo siguiente: “*descubrimiento de reglas de comportamiento, descubrimiento de grupos, descubrimiento de atributos significativos (atributos importantes para el entorno de negocio que se aplica), descubrimiento de reglas de pertenencia a grupos y ponderación de reglas de comportamiento o de pertenencia a grupos*”. Cada proceso tiene asociadas técnicas de minería de datos para descubrir patrones a través de su aplicación.

Con el propósito de aplicar de manera organizada y sistemática diversas técnicas de minería de datos, se emplean metodologías específicas. En este contexto, se busca identificar patrones y características relacionados con el cultivo de fruta fina orgánica, así como comprender cómo varían los gremios tróficos en función del manejo de las chacras.

Para dar respuestas a lo expuesto, este trabajo propone investigar, desarrollar y aplicar técnicas de ciencia de datos que permitan descubrir patrones de la dinámica de los cultivos, para luego, de manera simple y comprensible al usuario, visualizar los descubrimientos, para permitir llevar a cabo el proceso de toma de decisiones frente a casos de incertidumbre de manera más efectiva comparada con la habitual.

## **2.2. Cultivos de Frambuesos**

El frambueso (*Rubus idaeus*) [10] es un arbusto de clima templado con el mayor precio unitario en el mercado fresco. Es nativo de Europa y el norte de Asia y da un fruto (la frambuesa) de color rojo y sabor agridulce. Este arbusto llega a tener dos floraciones por temporada dependiendo la variedad. La producción mundial de frambuesa fresca y el consumo mundial están concentrados en el hemisferio norte. Este panorama favorece la venta de frambuesa en contraestación con un importante diferencial de precio para la fruta fresca exportada desde los países del hemisferio sur. El 90% de la producción mundial de frambuesa [10], al ser una fruta altamente perecedera, se congela con destino industrial, principalmente para la elaboración de jugos concentrados.

Argentina, adquirió competitividad en los últimos años como producto congelado; aunque su consumo local es a través de dulces, licores y conservas o fruta fresca. Su fruto es de los de mayor precio unitario en el mercado fresco y con alta demanda por parte de la agroindustria.

Sólo el 14% de la producción nacional de frambuesa se exporta, casi exclusivamente como producto congelado.

El arbusto [11], que alcanza una altura de aproximadamente 40-60 centímetros, está caracterizado por numerosas raíces delgadas y superficiales que se originan desde la corona, presenta también un tallo subterráneo corto, que genera estolones. Estos últimos tienen una duración de dos años; durante el primer año, se dedican al desarrollo vegetativo y exhiben una epidermis gris amarillenta con una pilosidad amarillo-dorada y nudosidades débiles. En el segundo año, estos tallos adquieren una epidermis gris cubierta de espinas. Florecen, dan frutos y, finalmente, se secan después de la maduración de sus frutos, siendo reemplazados por nuevos brotes.

En cuanto a los requisitos agroclimáticos y edáficos, la frambuesa alcanza su desarrollo óptimo en regiones de clima templado húmedo. En términos térmicos, la frambuesa muestra tolerancia a bajas temperaturas y veranos frescos, siendo de 14 a 19°C las temperaturas ideales para un desarrollo fisiológico y productivo [12]. La temperatura base mínima para su crecimiento se sitúa en 10°C, mientras que la temperatura máxima oscila entre 28 y 32°C. En lo que respecta a las condiciones de suelo ideales [13], esta especie demanda suelos profundos, fértiles y ligeramente ácidos, con un contenido mínimo de materia orgánica del 3% y un pH del suelo que oscila entre 6 y 7.

### **3. Metodología Aplicada para Desarrollar Proyectos de Ciencia de Datos**

Como se ha definido en la sección anterior, en el marco de la ciencia de datos se procede a analizar la aplicación de metodologías ágiles existentes que se utilizan en el desarrollo de este tipo de proyectos. En este trabajo se describe la metodología ASUM (sección 3.1.), la que ha sido evaluada en trabajos anteriores [14], y ha determinado ser considerada la que mejor se ajusta a la gestión de dichos proyectos.

#### **3.1. Metodología ASUM**

En esta sección, siguiendo publicaciones anteriores [15], se detallan características de la metodología (sección 3.1.1.) y se describen brevemente las cinco fases que la componen: Analizar (3.1.2.), Diseñar (sección 3.1.3.), Configurar y Construir (sección 3.1.4.), Implementar y/o Desplegar (sección 3.1.5.), Operar y Optimizar (sección 3.1.6.).

##### **3.1.1 Características de la Metodología**

La metodología desarrollada por IBM Analytics ASUM-DM (del acrónimo en inglés 'Analytics Solutions Unified Method for Data Mining' o Método Unificado de Soluciones de Análisis para la Minería de Datos) es un proceso iterativo para implementar un proyecto de minería de datos o análisis predictivo.

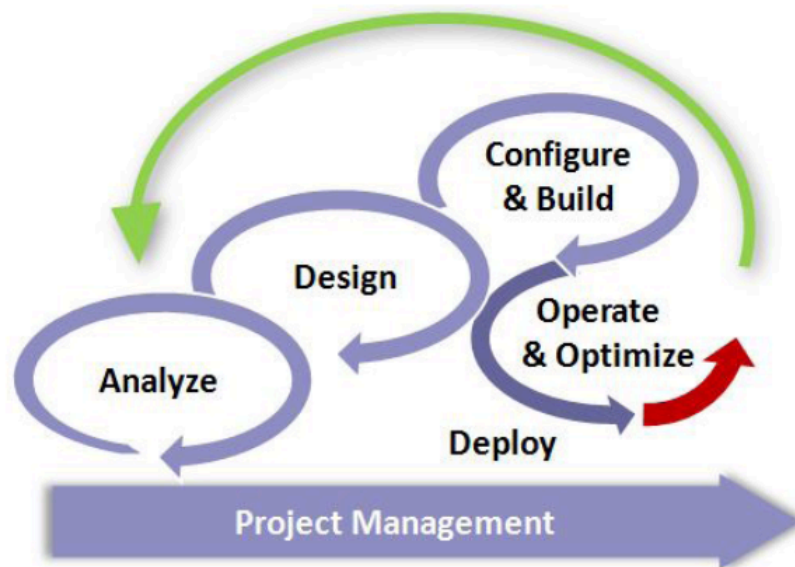
Este método está basado en la metodología CRISP-DM (acrónimo en inglés de 'Cross Industry Standard Process for Data Mining' o Proceso Estándar para Minería de Datos Independiente de la Industria) y fue creado para acelerar los tiempos, valorar y reducir el riesgo y establecer enfoques y procesos consistentes que aumentan la eficiencia del proyecto. Contiene pasos estructurados, actividades de desarrollo, roles y responsabilidades, plantillas y pautas.

Los beneficios que trae su implementación son, en primer lugar, minimizar el riesgo combinando la experiencia del mundo real con prácticas comprobadas de la industria, para lograr implementaciones exitosas y gestión del riesgo. La validación formal sigue las pruebas estándar de la industria y pasos de validación utilizando el modelo V (que demuestra las relaciones entre cada fase del ciclo de vida de desarrollo y su fase asociada de prueba). En segundo lugar, la escalabilidad que permite replicar y elevar el éxito utilizando un enfoque y lenguaje común en diferentes proyectos. Por último, al ser exhaustivo ofrece un valor constante y duradero, ya que comienza en las etapas iniciales de un proyecto y continúa en función de una coherente gestión de proyectos para coordinar acciones, establecer y gestionar las expectativas y comunicar los resultados en cada fase.

ASUM-DM sigue cinco fases completamente definidas donde cada una es supervisada por un flujo de gestión de proyectos que garantiza la comunicación y

colaboración consistente y coordinada. Dichas fases están organizadas de forma jerárquica en tareas que van de niveles generales a específicas (fase, actividad, tarea, subtarea, paso).

En su nivel general, el proceso está organizado en fases tal como se puede observar en la figura 3.1.



**Figura 3.1. Diagrama de las cinco fases de la metodología ASUM-DM. Tomado de IBM - (2015) “Analytics Solution Unified Method (ASUM)”**

ASUM utiliza un híbrido de implementación ágil y principios tradicionales para lograr sus objetivos de solución y proporcionar un resultado óptimo. Estos principios incluyen:

- El proyecto es evaluado para la aplicación de principios ágiles.
- El alcance del proyecto reúne los requisitos comerciales iniciales.
- Tanto el personal de negocios como el de Tecnología de la Información forman parte integral del equipo de implementación del proyecto.
- Los requisitos se aclaran y ajustan a través de un número de sprints iterativos de prototipos.
- Según el número y la prioridad de los requisitos, el cronograma y recursos disponibles, se adopta una implementación por etapas para alcanzar los objetivos.
- Los resultados de la creación de prototipos se comparan con los requisitos totales para evaluar logros y determinar nuevas iteraciones.
- El desarrollo iterativo e incremental se utiliza para finalizar la configuración y construcción.
- Después de pruebas adecuadas realizadas a lo largo del ciclo de vida del proyecto, se pone en marcha la primera etapa de la solución.

- Las etapas restantes del proyecto siguen el mismo camino de sprints de prototipos iterativos e incrementales desarrollados como la primera etapa.
- Los resultados de las etapas implementadas son monitoreados continuamente en la fase "Operar y optimizar" y si se producen desviaciones, entonces podrían tener lugar nuevas rondas de desarrollo.

### 3.1.2 Analizar

La primera fase se denomina *Analizar* (o *Analyze*) y define qué necesita lograr la solución, tanto en términos de características y atributos no funcionales (como el rendimiento y la usabilidad) y obtener acuerdo entre todas las partes sobre estos requisitos. Esta fase incluye las actividades y tareas que se detallan en la tabla 3.1.

**Tabla 3.1. Actividades y tareas de la fase "Analizar" de la metodología ASUM-DM.**

Actividad	Tarea
Prepararse para la implementación	Enlace con ventas y revisión de detalles del proyecto
	Identificar recursos
Realizar evaluación de preparación	Evaluar la preparación del cliente para la implementación
Realizar el lanzamiento del proyecto	Preparar la plataforma de lanzamiento del proyecto
	Orientar los recursos del proyecto
	Ejecutar el lanzamiento del proyecto
Comprender el negocio	Determinar los objetivos
	Evaluar la situación
	Determinar los objetivos de minería de datos
	Crear un plan de Proyecto
	Crear un informe de comprensión empresarial
Comprender los datos	Recopilar datos iniciales
	Describir los datos
	Explorar los datos

	Verificar la calidad de los datos
	Crear un informe de compresión de los datos

La primera actividad correspondiente a “*Prepararse para la implementación*” tiene como objetivo revisar los detalles del proyecto y las expectativas del cliente aclarando cuáles son esas expectativas, los supuestos clave, los puntos de influencia, las personalidades y cualquier cosa que pueda ser valiosa durante el ciclo de vida del proyecto. Además, se identifican los recursos adecuados para el proyecto.

La segunda actividad “*Realizar evaluación de la preparación*” implica evaluar qué tan listo está el cliente para comenzar con el proyecto, para lo cual se requiere analizar los diferentes factores de implementación y administrativos, factores de construcción y ubicación, factores de tecnología, fuentes de datos y factores de volúmenes que podrían afectar la implementación del proyecto e identificar soluciones para mitigar el impacto de esos factores.

La tercera actividad “*Realizar el lanzamiento del proyecto*” implica preparar una plataforma para presentar el equipo del proyecto al cliente y orientarlo con los objetivos y expectativas del proyecto, el alcance, la metodología, los plazos, las limitaciones y los riesgos, y los próximos pasos a seguir. Por otra parte se debe discutir y alinear con los recursos del proyecto sobre lo que debe cubrir durante la sesión de inicio para que los miembros del equipo estén preparados y listos para el evento. Luego, realizar una sesión de inicio y presentar formalmente el equipo al cliente y cubrir temas como los objetivos y expectativas del proyecto, el alcance, la metodología, los plazos, las limitaciones y los riesgos, y los próximos pasos a seguir.

La cuarta actividad “*Comprender el negocio*” implica comprender los objetivos y requisitos del proyecto desde una perspectiva comercial, luego convertir este conocimiento en una definición de problema de minería de datos y un plan preliminar diseñado para lograr los objetivos.

La quinta actividad “*Comprender los datos*” implica observar más de cerca los datos disponibles para la minería. Este paso es crítico para evitar problemas inesperados durante la siguiente actividad, la preparación de datos, que generalmente es la parte más larga de un proyecto. La comprensión de los datos implica acceder a los datos y explorarlos. Esto le permite determinar la calidad de los datos y describir los resultados de estos pasos en la documentación del proyecto.

### 3.1.3 Diseñar

La segunda fase se denomina *Diseñar (o Design)* y tiene como objetivo definir todos los componentes de la solución y sus dependencias, identificar recursos e

instalar un entorno de desarrollo. Los sprints iterativos de prototipos se utilizan cuando corresponde aclarar requisitos. Esta fase incluye las actividades y tareas que se detallan en la tabla 3.2.

**Tabla 3.2. Actividades y tareas de la fase “Diseñar” de la metodología ASUM-DM.**

Actividad	Tarea
Diseñar y validar la infraestructura	Diseñar la infraestructura técnica
	Diseñar la infraestructura de seguridad

La actividad de esta fase “*Diseñar y validar la infraestructura*” tiene por objetivo diseñar la arquitectura de entornos y las estrategias de autenticación y autorización.

### 3.1.4 Configurar y Construir

La tercera fase se denomina *Configurar y Construir* (o *Configure & Build*) y tiene como objetivo configurar, construir e integrar componentes basados en un enfoque iterativo e incremental. Utiliza pruebas de entorno múltiple y plan de validación basado en el modelo V. Esta fase incluye las actividades y tareas que se detallan en la tabla 3.3.

**Tabla 3.3. Actividades y tareas de la fase “Configurar y Construir” de la metodología ASUM-DM.**

Actividad	Tarea
Configurar entornos	Configurar los entornos en el sitio
	Configurar los entornos en la nube
Preparar los datos	Seleccionar los datos
	Limpiar los datos
	Construir los datos
	Integrar los datos
	Formatear los datos
	Crear informe de preparación de los datos
Construir el modelo	Seleccionar técnicas de modelado
	Generar diseño de prueba



	Construir el modelo
	Evaluar el modelo
Evaluar el modelo	Evaluar resultados
	Proceso de revisión
	Determinar los próximos pasos
Realizar transferencia de conocimiento analítico	Orientar y transferir conocimiento al equipo analítico del proyecto del cliente
Definir enfoque de implementación	Crear plan de implementación
Diseño de estrategia de prueba operacional	Identificar y acordar planes de prueba
	Crear planes de prueba
Validar y probar en entorno de control de calidad	Crear archivos de datos de control de calidad
	Asegurarse de que el entorno de control de calidad esté listo
	Migrar/Restaurar modelo analítico en QA
	Sistema de conducta y pruebas de rendimiento
	Realizar prueba de aceptación del usuario

La primera actividad “*Configurar entornos*” implica configurar los entornos de análisis, control de calidad y producción según el diseño y los requisitos en el sitio o en la nube. También incluye eliminar lo que no sea aplicable.

La segunda actividad “*Preparar los datos*” es uno de los aspectos más importantes y a menudo lentos de la minería de datos. Se estima que la preparación de datos generalmente toma del 50 al 70% del tiempo y esfuerzo de un proyecto. Depende en gran medida de las actividades de "comprensión de datos" y "comprensión de negocios", por lo que dedicar la energía adecuada a estas actividades anteriores puede minimizar esta sobrecarga, pero aún debe gastar una buena cantidad de esfuerzo preparando y empaquetando los datos para la minería. Dependiendo de la organización y sus objetivos, la preparación de datos generalmente involucra lo siguiente:

- Fusionar conjuntos de datos y / o registros ,
- Seleccionar un subconjunto de datos de muestra ,
- Agregar registros,

- Obtener nuevos atributos,
- Ordenar los datos para modelar,
- Eliminar o reemplazar espacios en blanco o faltantes valores,
- Dividir los datos en conjuntos de entrenamiento y prueba.

La tercera actividad “*Construir el modelo*” generalmente se realiza en múltiples iteraciones, donde se ejecutan varios modelos utilizando los parámetros predeterminados y luego se ajustan los parámetros o se vuelve a la actividad “*Preparar datos*” para las manipulaciones requeridas por su modelo de elección. Es raro que la pregunta de minería de datos de una organización sea respondida satisfactoriamente con un solo modelo y una sola ejecución.

La cuarta actividad “*Evaluar el modelo*” implica utilizar los criterios de éxito de la organización.

La quinta actividad “*Realizar transferencia de conocimiento analítico*” implica orientar y educar al equipo del cliente sobre el proceso de minería de datos / análisis predictivo que se ha configurado.

La sexta actividad “*Definir enfoque de implementación*” implica determinar y describir cómo la solución se trasladará a todos los usuarios.

La séptima actividad “*Diseño de estrategia de prueba operacional*” implica discutir y acordar con el equipo del proyecto la estrategia de prueba para el flujo operativo del proyecto y cómo se realizarán y ejecutarán las pruebas de rendimiento, sistema y se crearán planes de pruebas que se actualizarán a medida que avance el proyecto.

La octava actividad “*Validar y probar en el entorno de control de calidad*” implica asegurarse de que se hayan tomado todos los pasos correctos hasta ahora, probar la solución en un entorno de control de calidad y tomar una decisión de implementación de producción basada en pasos validados y pruebas exitosas.

### 3.1.5 Implementar y/o Desplegar

La cuarta fase se denomina *Implementar y/o Desplegar (o Deploy)* y tiene como objetivo crear un plan para ejecutar y mantener la solución, incluyendo un planificador de soporte, migrar al entorno de producción, configurar según sea necesario y comunicar la implementación a los usuarios. Esta fase incluye las actividades y tareas que se detallan en la tabla 3.4.

**Tabla 3.4. Actividades y tareas de la fase “Implementar y/o Desplegar” de la metodología ASUM-DM.**

Actividad	Tarea
Realizar transferencia de conocimiento operacional	Orientar y transferir conocimiento al equipo operativo del proyecto del cliente
Prepararse para el mantenimiento	Establecer cronograma para el soporte

continuo	de guardia
	Programar actividades de mantenimiento para el entorno de producción
	Programar actividades de monitoreo para el entorno de producción
Implementar la solución	Crear archivos de datos de producción
	Crear y realizar pruebas de preparación operacional
	Migrar/Restaurar el modelo de control de calidad a producción
Lanzamiento	Ir en vivo
	Implementar plan de lanzamiento de comunicación
	Lanzamiento de revisión
Prepararse para el cierre del proyecto	Crear un informe final
	Seguimiento de las acciones de la reunión de revisión
	Realizar el evento del equipo de cierre de proyecto
	Realizar una revisión interna del proyecto y las lecciones

La primera actividad “*Realizar transferencia de conocimiento operacional*” implica Orientar y educar al cliente sobre el aspecto no analítico de la solución para que la solución pueda ejecutarse de manera efectiva y eficiente.

La segunda actividad “*Prepararse para el mantenimiento continuo*” implica asegurarse de que todas las funciones y actividades de soporte estén establecidas antes de implementar la solución.

La tercera actividad “*Implementar la solución*” implica mover la solución al entorno de producción según el Plan de Implementación y validar que el entorno de producción esté configurado correctamente.

La cuarta actividad “*Lanzamiento*” implica poner en marcha la solución y comunicar a la comunidad de usuarios finales y las partes interesadas que la solución está en vivo y revisar el lanzamiento para recopilar las lecciones aprendidas y los éxitos.

La quinta actividad “*Prepararse para el cierre del proyecto*” implica preparar y ejecutar tareas para cerrar el proyecto.

### 3.1.6 Operar y Optimizar

La quinta fase se denomina *Operar y Optimizar (o Operate & Optimize)* y tiene como objetivo incluir las tareas de mantenimiento y puntos de control después del desarrollo que faciliten el éxito de la solución y lo preserven. Esta fase incluye las actividades y tareas que se detallan en la tabla 3.5.

**Tabla 3.5. Actividades y tareas de la fase “Operar y Optimizar” de la metodología ASUM-DM.**

Actividad	Tarea
Modelo de monitor	Monitorear continuamente la precisión del modelo y la actualización del modelo
Operar optimizar y mejorar el sistema	Programar y operar el sistema
	Monitorear los procesos de producción
	Gestionar procesos de mantenimiento
	Gestionar defectos
	Administrar mejoras
	Optimizar el sistema
	Mantener la documentación de la aplicación
	Gestionar la competencia
Comunidad de usuarios de soporte	Administrar operaciones de la mesa de ayuda
	Administrar problemas
	Administrar recursos de la mesa de ayuda
	Gestionar la comunicación
Administrar la infraestructura	Administrar la configuración del sistema
	Gestionar la capacidad
	Gestionar la disponibilidad del sistema

	Administrar lanzamientos
	Administrar seguridad
Programa de ciclo de vida del sistema de gobierno	Verificar la realización de los beneficios
	Revisar rendimiento del sistema
	Revisar el plan del ciclo de vida del sistema
	Gestionar la calidad

La primera actividad "*Modelo de monitor*" implica monitorear continuamente los resultados de los modelos implementados para garantizar su precisión y que aún satisfagan las metas de minería de datos de la organización y los objetivos comerciales.

La segunda actividad "*Operar optimizar y mejorar el sistema*" implica realizar actividades posteriores al lanzamiento para mantener el sistema funcionando correctamente.

La tercera actividad "*Comunidad de usuarios de soporte*" implica realizar actividades posteriores al lanzamiento para apoyar a la comunidad de usuarios finales.

La cuarta actividad "*Administrar la infraestructura*" implica realizar actividades posteriores al lanzamiento para administrar y mantener la infraestructura.

La quinta actividad "*Programa de ciclo de vida del sistema de gobierno*" implica realizar actividades posteriores al lanzamiento para gestionar el ciclo de vida de la solución.

## 4. Implementación

En este capítulo se explicará la herramienta utilizada (sección 4.1) y el estudio de atributos (sección 4.2).

Para llevar a cabo este estudio se realizó la implementación de la metodología ASUM-DM. En la Tabla 1 se define cual es la relación entre la etapa de la metodología y las secciones del presente trabajo.

*Tabla 1: Relación Fases de la metodología vs. Trabajo Final de Carrera*

Fases de la Metodología ASUM-DM	Relación con el Trabajo Final de Carrera
Analizar	Objetivos (sección 1.1) Estudio de Atributos (sección 4.2)
Diseñar	Herramientas Utilizadas (sección 4.1) Estudio de Atributos (sección 4.2)
Configurar y Construir	Modelo de Datos (sección 5.1)
Implementar y/o Desplegar	Descripción de los resultados (sección 5.2)
Operar y Optimizar	Conclusiones (sección 6)

### 4.1. Herramienta Utilizada

Orange [16] es una herramienta integral de software de código abierto, desarrollada por la Facultad de Informática de la Universidad de Ljubljana, en Eslovenia, diseñada para la minería de datos de análisis predictivo y visualización de datos, que proporciona a los usuarios una plataforma flexible y amigable para explorar, analizar y comprender conjuntos de datos de manera efectiva. Su versatilidad la convierte en una herramienta valiosa en diversas áreas y su interfaz gráfica intuitiva la hace accesible para usuarios con diferentes niveles de experiencia en análisis de datos.

Entre las principales funciones y aplicaciones de Orange, en su versión 3.35.0, se destacan el análisis predictivo mediante la construcción de modelos y la predicción de resultados, basándose en conjuntos de datos existentes; la preparación y limpieza de datos mediante técnicas de preprocesamiento, como la imputación de valores faltantes, la normalización y la transformación de datos; la visualización de datos que permite explorar y representar los datos de manera gráfica, lo que facilita la comprensión de patrones y relaciones en los conjuntos de datos; análisis estadísticos que ayudan a obtener información significativa sobre la distribución y características de los datos; aprendizaje de máquina que incorpora algoritmos de aprendizaje que permiten construir modelos predictivos y clasificar los datos de manera automatizada; visualización de modelos que facilita la interpretación de modelos generados mediante la representación visual de estructuras y relaciones, lo que ayuda a comprender cómo se toman decisiones en un modelo específico; y la

interactividad y experimentación que permite a los usuarios experimentar con diferentes enfoques y algoritmos de manera interactiva, facilitando el proceso de prueba y error en la construcción de modelos.

## 4.2. Estudio de Atributos

Los datos se tomaron del estudio realizado en el PI 40-B-700 2018 “Comparación del efecto de diferentes prácticas de manejo orgánico en frambuesa sobre la biodiversidad de artrópodos, patógenos y agentes de biocontrol en El Bolsón (Río Negro)” dirigido por la Dra. en Ciencias Biológicas María Verónica Chillo.

En dicho estudio se tomaron muestras en tres momentos del año: invierno (agosto 2019), primavera (noviembre 2019) y verano (febrero 2020), en tres chacras con cultivos que son sometidos a manejo orgánico de frambuesos, en la zona “camino de los nogales”, una zona productiva en El Bolsón, provincia de Río Negro. Las chacras elegidas se seleccionaron por su ubicación próxima, para que las condiciones del suelo y clima sean de similares características. Allí se recabaron datos del manejo de los cultivos, la forma de fertilización, el riego, el método de desmalezamiento y el control de plagas.

Los datos obtenidos estaban separados en dos planillas de cálculos (Diversidades y SpsxTraits), se fusionaron (Anexo 2) a través de datos en común. No hubo inconsistencias, como valores nulos, duplicados o incorrectos. Las celdas vacías se dejaron vacías, la información se mantuvo completa. La Lic. Pizzingrilli evaluó la consistencia de los datos y resultó coherente, y se obtuvo una tabla final que consta de 167 registros, con los siguientes campos (Tabla 2) en los que se puede visualizar 54 registros de la chacra 1, 56 de la chacra 2 y 54 de la chacra 3.

*Tabla 2: Campos Utilizados*

Campos Utilizados				
Fecha	Acari.sp1	Dermaptera	Carabidae.sp3	Lepidoptera.sp1
Chacra	Acari.sp2	Collembola	Carabidae.sp5	Lepidoptera.sp2
Muestra	Camponotus	Hemiptera.sp1	Carabidae.sp6	Shannon
Riego	Camponotus.sp2	Aphidae.sp1	Carabidae.sp7	Simpson
Desmalezado	Apis.mel	Aphidae.sp2	Carabidae.sp4	InvSimpson
Control_plagas	Bombus.sp	Aphidae.sp3	Carabidae.sp8	Riqueza
Fertilizacion	Ichneumonidae.sp1	Hemiptera.sp2	Carabidae.sp9	q0
HumedadSuelo	Ichneumonidae.sp2	Pentatomidae	Carabidae.sp10	q1
D.s	Tenthredinidae	Nabidae	Coleoptera.sp1	q2
Cicadellidae	Sphecidae	Iulida	Tenebrionidae	FRic
Ly.sp1	Cynipoidae.sp1	Polydesmia	Scarabidae	FEve

Ly.sp2	Cynipoidae.sp2	Lithobiomorpha	Cerambicidae	FDiv
Araneomorph.sp3	Vespula.ger	Curculionidae.sp1	Elateridae.sp1	FDis
Araneomorph.sp4	Avispa.papel	Curculionidae.sp2	Elateridae.sp2	RaoQ
Araneomorph.sp7	Braconidae	Elmidae	Staphylinidae	CWM.Dieta
Araneomorph.tigre	Proctotrupoidea	Carabidae.sp1	Piesmidae	CWM.SE
Araneomorph.blanca	Forficula.auri	Carabidae.sp2	Cicimidae	CWM.Size

Las chacras tienen tres tipos de manejos de riego, por goteo, por aspersión y goteo y aspersión. Además, los datos cuentan con información sobre diferentes tipos de abono, uno con abono de corral estepa, otro con guanito lombricompuesto, y otro con abono de corral propio. Las chacras tienen tres medios de desmalezamiento, manual o mecánico, y se puede hacer o no control de plagas.

Por otro lado se puede ver que son familias de insectos y cada chacra tiene parecidos pero cambia la abundancia y riqueza.



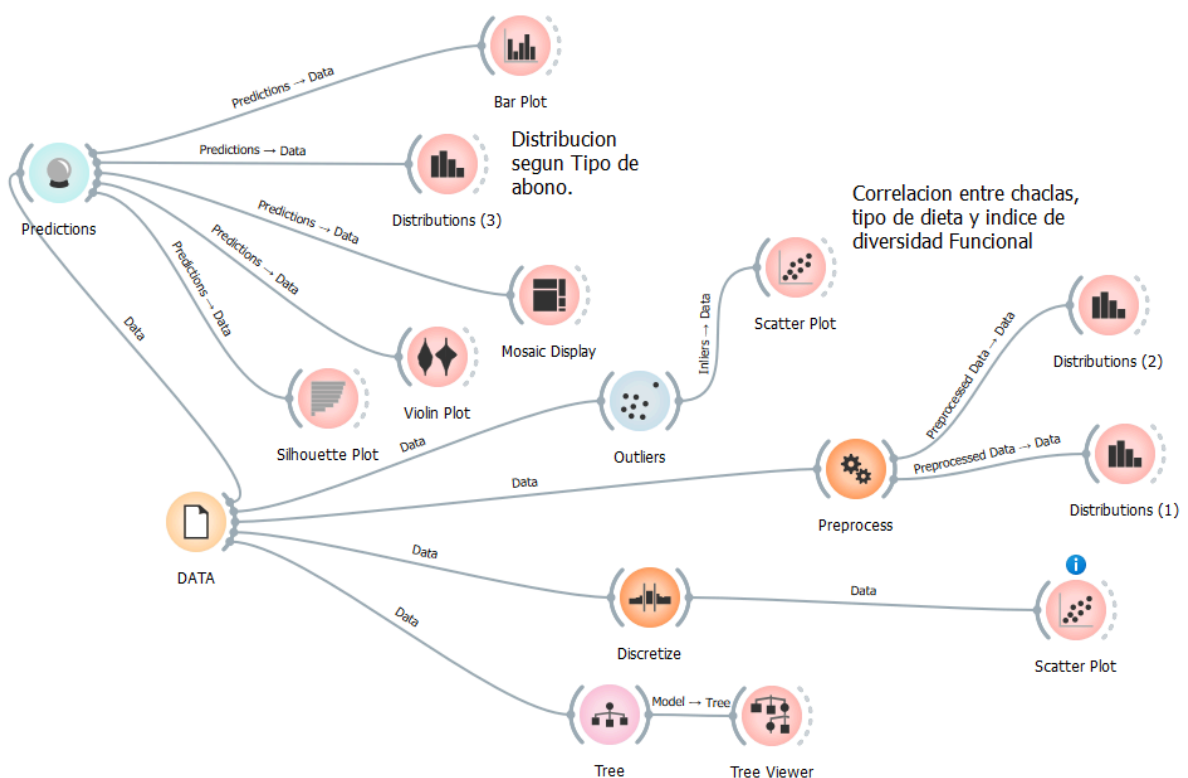
## 5. Resultados y Análisis

En esta sección, se presenta de manera exhaustiva el modelo de datos construido, ofreciendo una visión detallada de su estructura. Además, se proporcionan en detalle los resultados derivados del estudio de datos realizado, ofreciendo una perspectiva comprehensiva de las observaciones y patrones identificados. Se destaca la amplitud de información que se puede analizar a partir de estos resultados, lo que brinda una base sólida para interpretaciones más profundas y para la toma de decisiones informadas en el contexto del estudio.

### 5.1. Modelo de Datos

Después de completar el análisis de los campos (señalados en Tabla 1), se procedió a llevar a cabo la construcción de un modelo de datos, cuya representación se exhibe detalladamente en la figura siguiente. Este modelo, elaborado a partir de los datos recopilados durante el estudio de atributos, proporciona una estructura visual que facilita la comprensión y visualización de las relaciones y componentes identificados en el análisis.

El target definido fue el tipo de dieta, declarado en la columna con el valor CWM.Dieta.



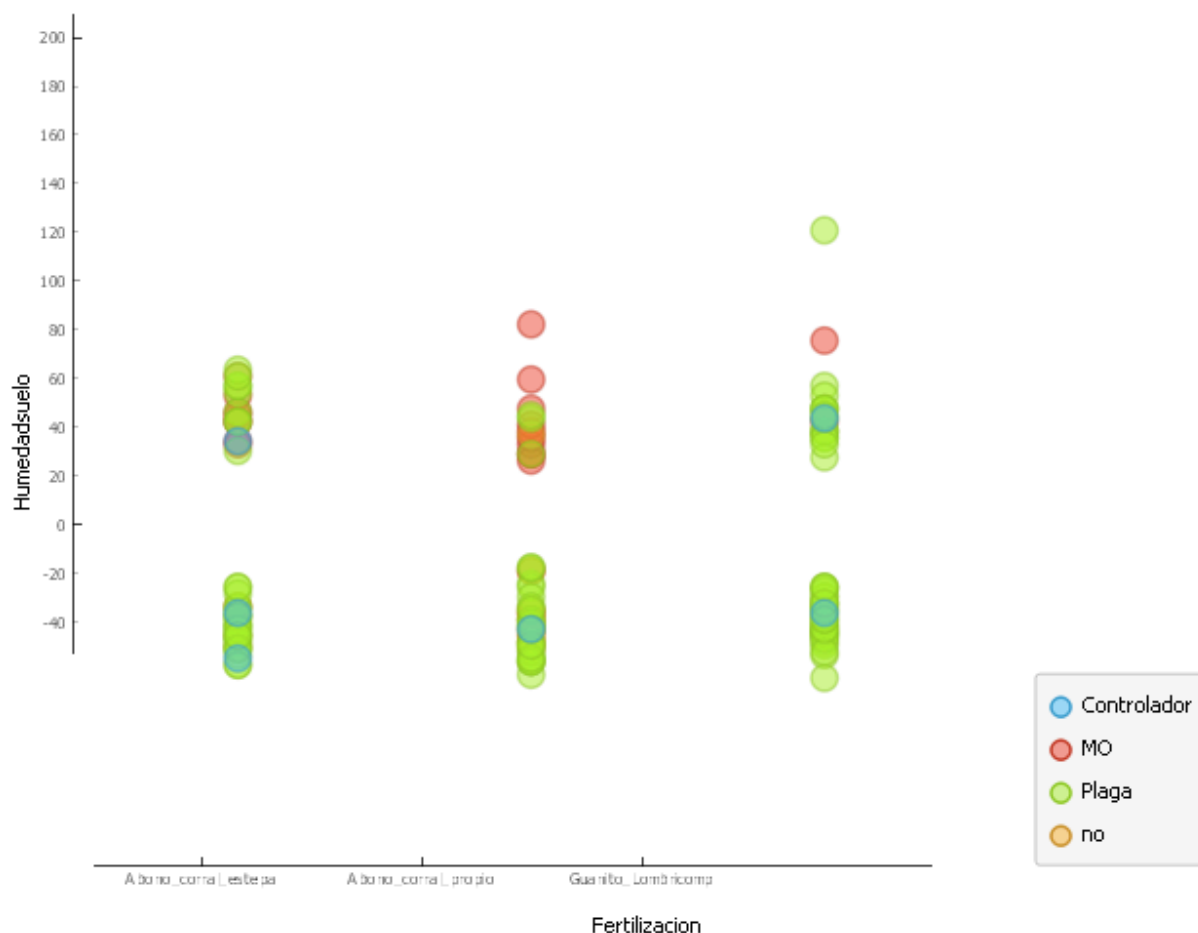
**Figura 5.1 Modelo de Datos**

**Fuente: Elaboración propia**

## 5.2. Descripción de los Resultados

Se procede a realizar una exploración detallada de los hallazgos obtenidos en el análisis. Al observar el gráfico de dispersión presentado en la Figura 5.2, el cual resulta especialmente útil cuando se desea visualizar la relación entre dos variables numéricas, se evidencia una representación visual en un plano cartesiano. En este gráfico, cada punto representa una observación y su posición en los ejes x e y corresponde a los valores de dos variables específicas. Este enfoque facilita la identificación de patrones, tendencias y relaciones entre variables.

Al examinar el gráfico, se nota que en suelos con menor contenido de humedad y donde la fertilización utilizada es el abono de corral estepa, se registra una mayor presencia de agentes de biocontrol (controladores). Estos resultados sugieren una posible correlación entre las condiciones del suelo, el tipo de fertilizante aplicado y la presencia de agentes de biocontrol, lo cual puede tener implicaciones significativas en términos de manejo agrícola y sostenibilidad.

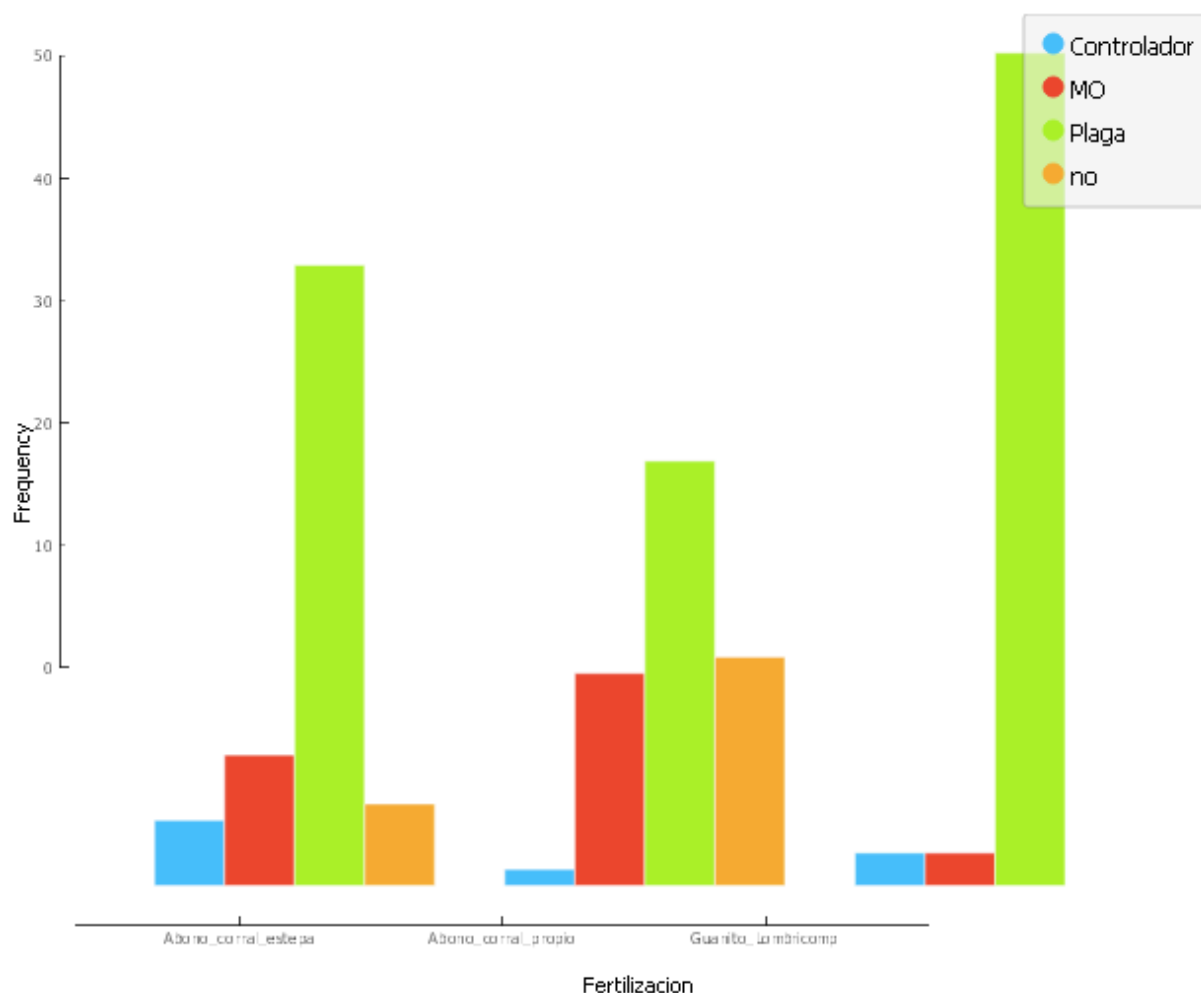


**Figura 5.2 Scatter Plot (Diagrama de Dispersión) según tipo de fertilización y humedad del suelo.**

**Fuente: Elaboración propia.**

Posteriormente, en el gráfico de barras, o "bar plot," presentado en la Figura 5.3, donde se representan datos categóricos mediante barras rectangulares y cada barra representa una categoría específica, observamos que la longitud de cada barra es proporcional a la cantidad o frecuencia de la variable que representa. En este contexto, se destaca que el suelo tratado con abono de corral propio exhibe una mayor diversidad de artrópodos y una menor cantidad de plagas.

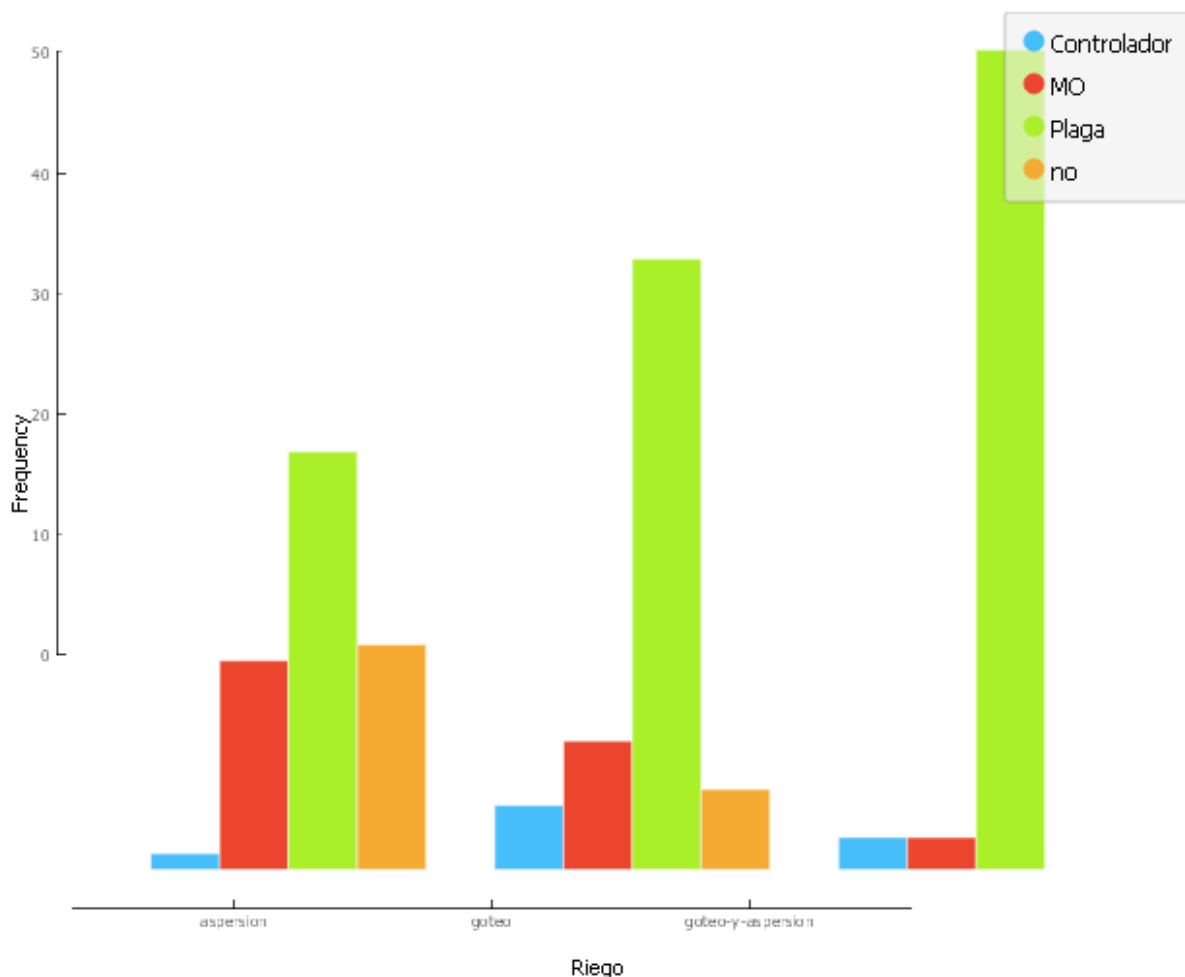
Notablemente, el abono de corral estepa se distingue por tener la mayor presencia de agentes controladores. Estos resultados, visualizados de manera clara en el gráfico de barras, sugieren que la elección del tipo de abono puede influir significativamente en la diversidad de la población de artrópodos y en la capacidad de controlar plagas en el entorno agrícola. Este hallazgo puede ser de gran importancia para la toma de decisiones en prácticas agrícolas y estrategias de manejo integrado de plagas.



**Figura 5.3 Bar Plot (Diagrama de Barras) según tipo de fertilización**  
**Fuente: Elaboración propia.**

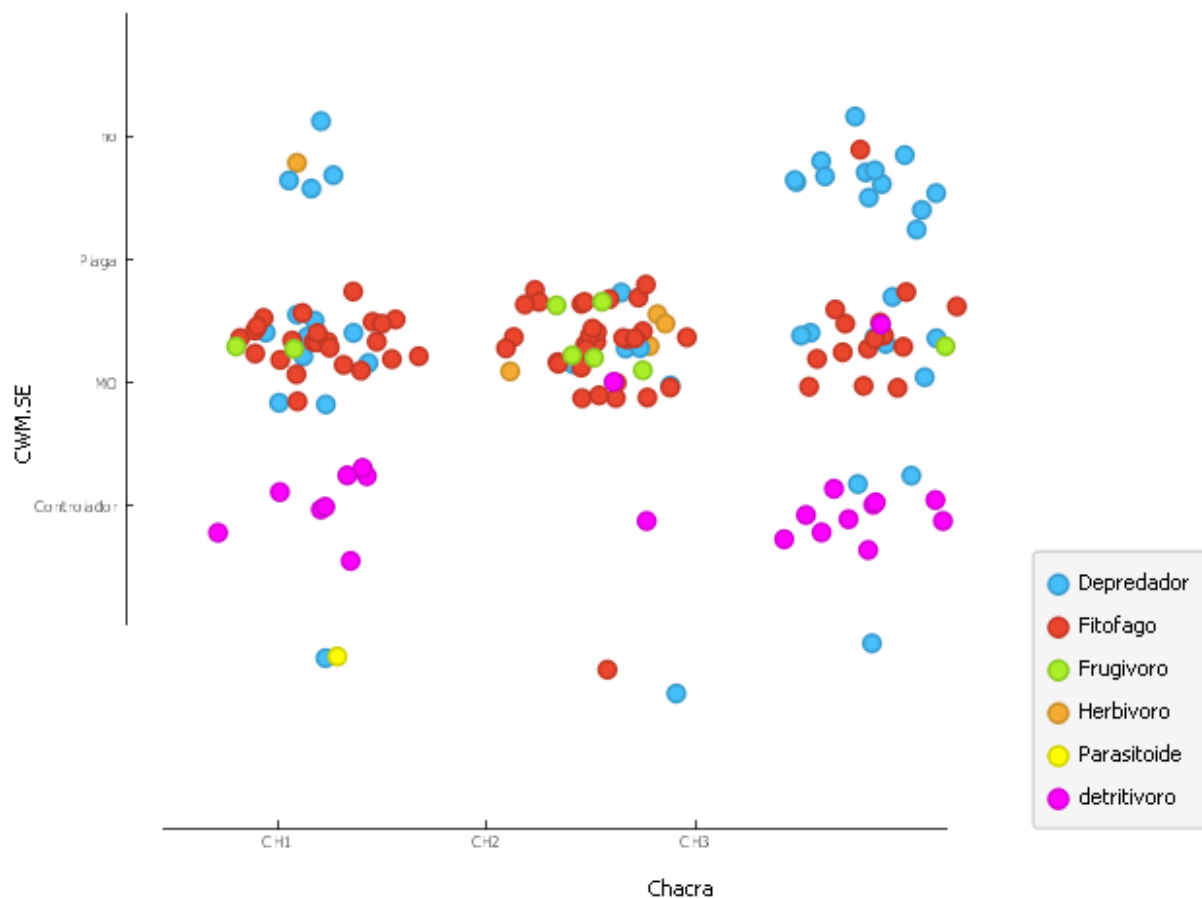
Utilizando el mismo método de visualización de resultados, en el gráfico de barras de la Figura 5.4, categorizado según el tipo de riego, se observa que el riego por aspersión demuestra ser más propicio, con una mayor diversidad de artrópodos y una menor cantidad de plagas. Por otro lado, el riego por goteo muestra favorecer la presencia de controladores, a pesar de que presenta una menor diversidad de artrópodos.

Este gráfico de barras, que representa datos categóricos mediante barras rectangulares, donde la longitud de cada barra es proporcional a la cantidad o frecuencia de la variable que representa, destaca las diferencias significativas en la dinámica de artrópodos y plagas en función del método de riego utilizado. Estos resultados ofrecen información valiosa para la toma de decisiones en la gestión agrícola, especialmente en la elección del tipo de riego para promover la salud del ecosistema y el rendimiento de los cultivos.



**Figura 5.4 Bar Plot (Diagrama de Barras) según tipo de Riego**  
**Fuente: Elaboración propia.**

En la Figura 5.5, al evaluar el manejo de los cultivos en distintas chacras, se observa que la chacra 2 resulta menos favorable, ya que presenta una mayor concentración de fitófagos y herbívoros. Contrariamente, la chacra 3 exhibe una mayor concentración de artrópodos beneficiosos, tales como depredadores, parasitoides y detritívoros. Estos hallazgos sugieren que existe una variabilidad significativa en la composición de la comunidad de artrópodos beneficiosos y perjudiciales entre las chacras evaluadas, lo que puede tener implicaciones importantes para la salud y el rendimiento de los cultivos en cada una de ellas.

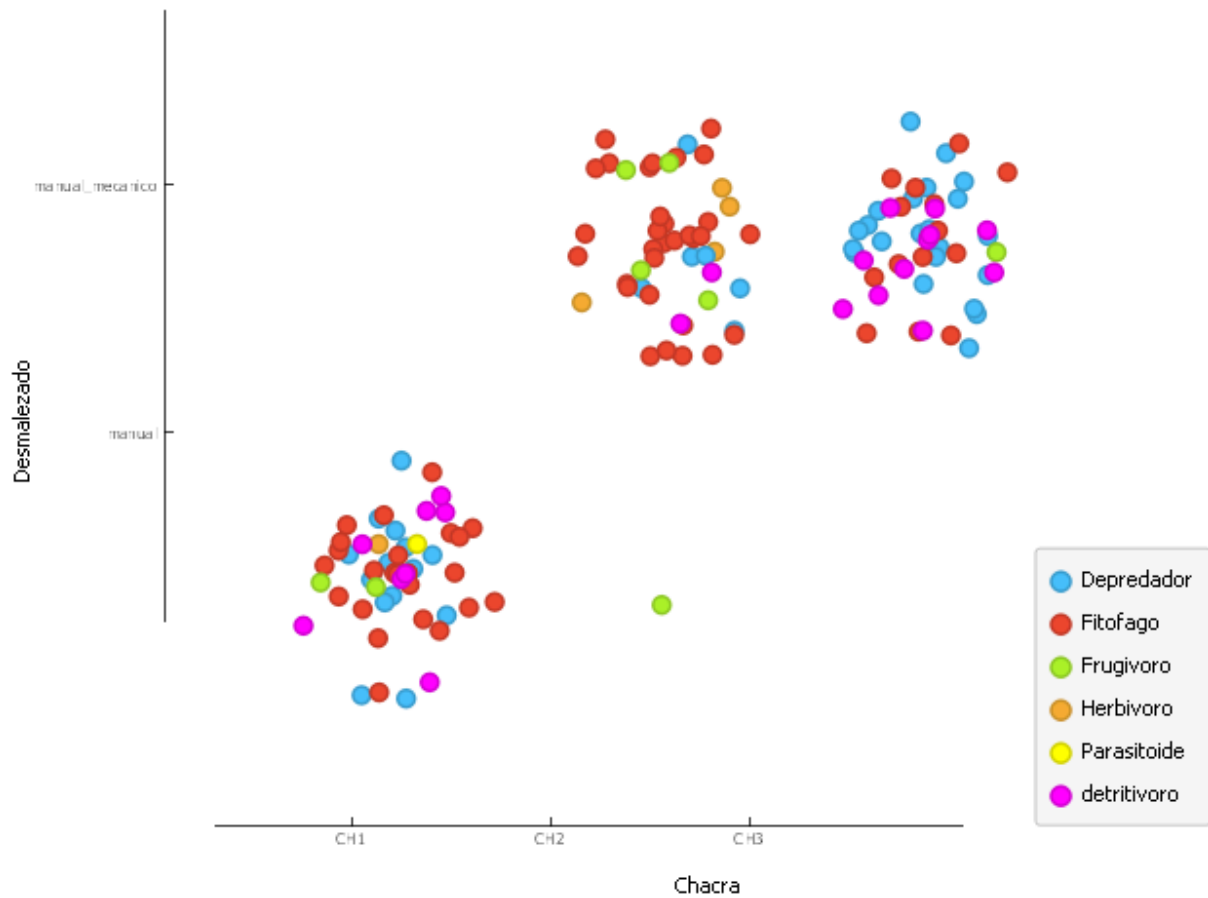


**Figura 5.5 Scatter Plot (Diagrama de Dispersión) según las chacras y el tipo de dieta.**

**Fuente: Elaboración propia.**

Si examinamos el desmalezado a través del diagrama de dispersión de la Figura 5.6, se evidencia que el desmalezado manual resulta más beneficioso, ya que se asocia con una mayor diversidad y, al mismo tiempo, con un menor agrupamiento de artrópodos dañinos como fitófagos, frugívoros y herbívoros. Estos resultados indican que la práctica de desmalezado manual puede contribuir a fomentar un entorno más equilibrado, con una variedad más amplia de artrópodos y una reducción en la concentración de aquellos que pueden ser perjudiciales para los

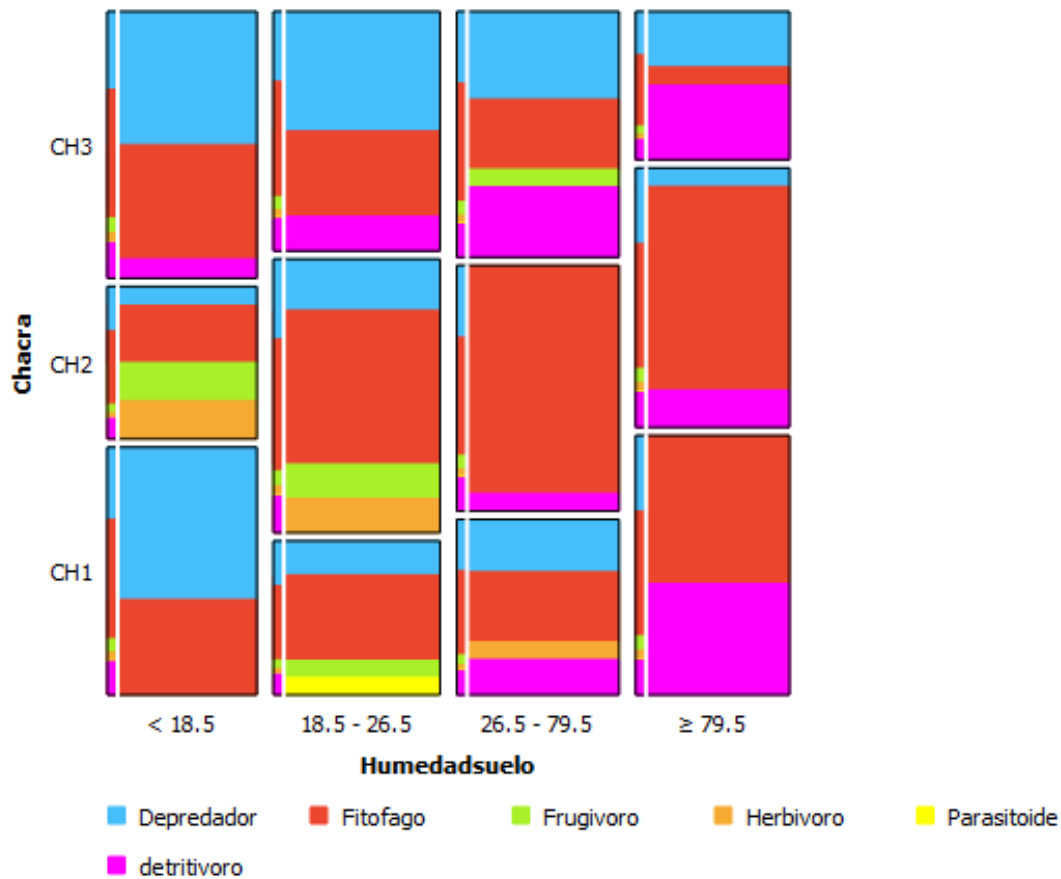
cultivos. Este hallazgo puede tener importantes implicaciones para las estrategias de manejo agrícola y la promoción de la biodiversidad en los campos cultivados.



**Figura 5.6 Scatter Plot (Diagrama de Dispersión) según las chacras y el tipo de desmalezado.**

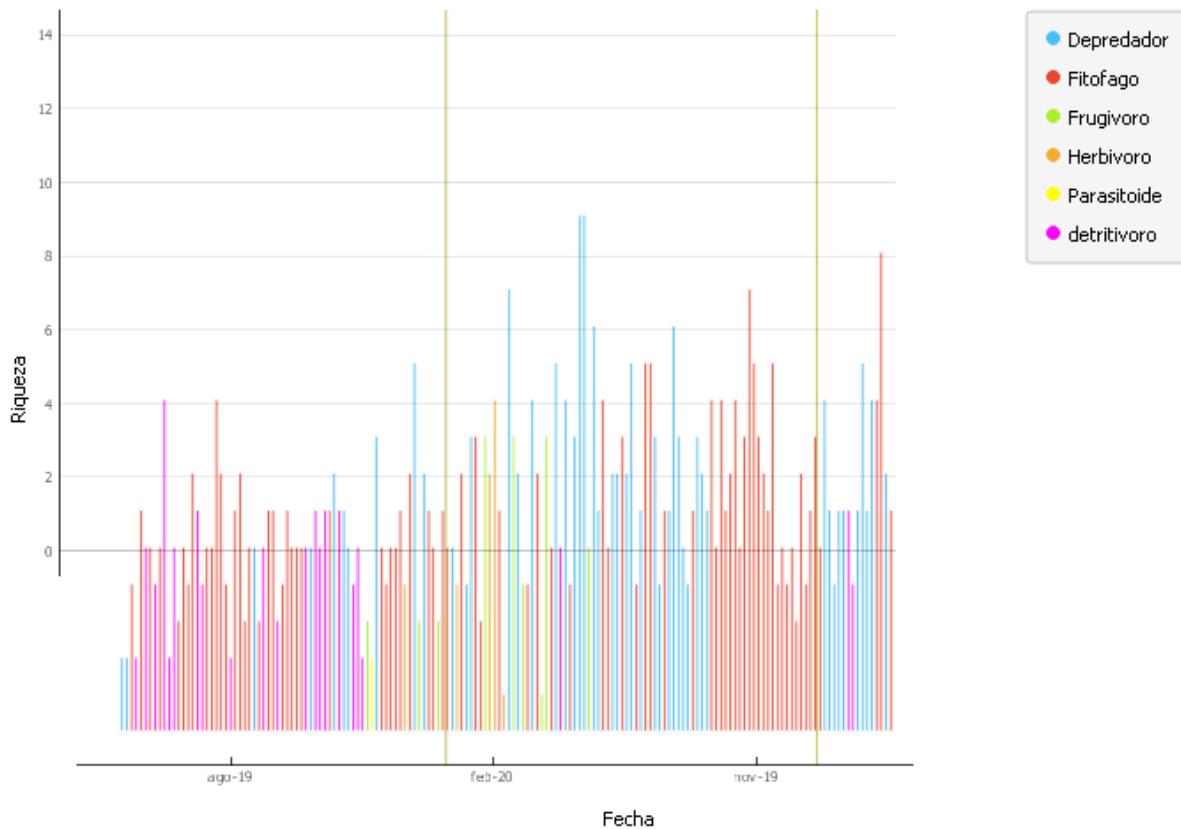
**Fuente: Elaboración propia.**

Con respecto a la humedad, la observación del mosaico en la Figura 5.7 revela que a medida que disminuye la humedad, se encuentran más depredadores. Por otro lado, cuando la humedad supera el 80%, se observa un aumento en la presencia de detritívoros, los cuales son beneficiosos. Sin embargo, esta condición también se asocia con una concentración más abundante de fitófagos, lo que no compensa los beneficios anteriores. En conclusión, los suelos con menor humedad parecen ser más favorables, ya que promueven una mayor presencia de depredadores sin el inconveniente de una proliferación significativa de fitófagos, lo que podría tener implicaciones positivas para la salud y el equilibrio del ecosistema agrícola.



**Figura 5.7 Mosaic Display (Visualización de mosaico) según las chacras y la humedad del suelo.**  
**Fuente: Elaboración propia.**

Se aprecia en la figura 5.8 que durante el verano, específicamente en la muestra de febrero de 2020, se concentra la mayor riqueza de especies. Este hallazgo sugiere que las condiciones estacionales, específicamente en el período estival, pueden influir significativamente en la diversidad y presencia de especies en el ecosistema evaluado. La observación de una mayor riqueza de especies durante el verano puede tener implicaciones importantes para la comprensión de los patrones de biodiversidad y la gestión de ecosistemas en el contexto temporal.



**Figura 5.8 Riqueza de gremios tróficos según estaciones del año.**  
**Fuente: Elaboración propia.**

En la Figura 5.9, se presenta un modelo de aprendizaje automático conocido como árbol de decisión, ampliamente utilizado en el campo de la inteligencia artificial para clasificar ejemplos o predecir valores. Este árbol de decisión específico se ha analizado detalladamente y se han revelado patrones significativos en relación con la clasificación de observaciones como plagas o depredadores.

Para esta visualización de árbol definí el target el campo CWM.SE, dado que la variable CWM.Dieta no arrojaba resultados.

En el nivel 1 de este árbol, se distingue la clasificación como plaga del atributo de índice de diversidad funcional "CWM.SE". Al adentrarse en el nivel 2 y considerar la variable dieta (CWM.Dieta), se destaca que si la dieta es depredadora, la probabilidad de que el organismo sea catalogado como plaga alcanza el 47%. Además, se observa una combinación de condiciones en el nivel 2, donde la diversidad, evaluada mediante el índice inverso de Simpson y registrando valores inferiores a 3.57, y la morfoespecie identificada como "*Araneomorph.tigre*", dan como resultado una probabilidad del 100% de ser clasificado como controlador. Sin embargo, si el índice de Simpson es mayor a 3.57 hay un 54.8% de ser considerado plaga.

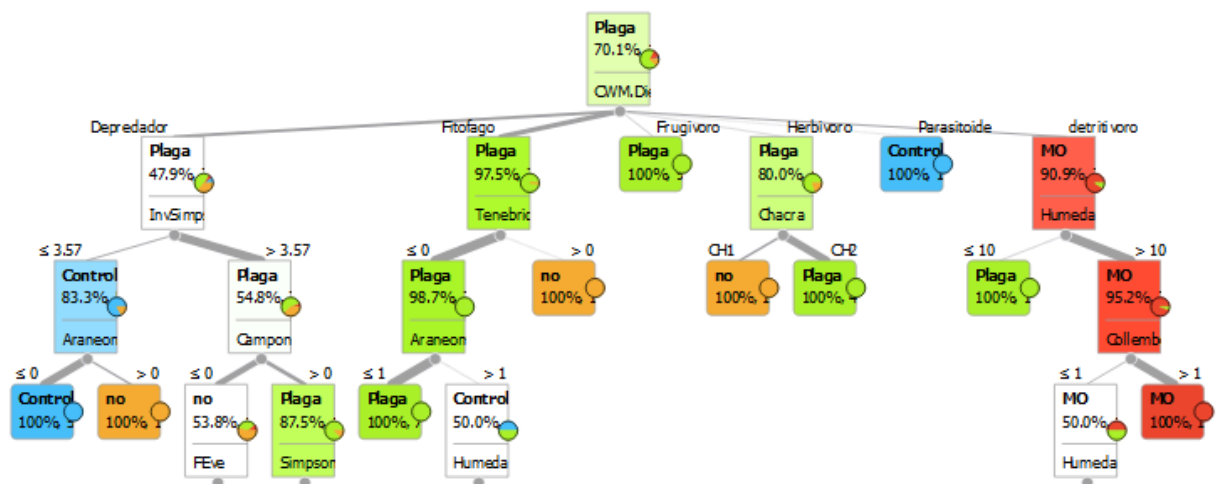


En el caso que la dieta sea Fitofago, la probabilidad de plaga asciende al 97.5%, en cuanto a la diversidad, si es Tenebrionidae con un valor inferior a 0 y la morfoespecie es "Araneomorph.sp3" con valores menores a 1 la probabilidad de plaga asciende a 100%, si es mayor a 1 hay un 50% que sea controlador.

Los Frugívoros son 100% considerados plagas, y por el contrario los parasitoides 100% controladores.

Si la dieta es herbívoro, la probabilidad de plaga es del 80% en el caso de la Chacra dos. Si la dieta es detritívoro, y la humedad del suelo es inferior a 10 hay 100% probabilidad que sea plaga.

Este análisis proporciona información crucial sobre los factores determinantes en la clasificación de organismos como plagas o depredadores en el contexto del estudio. Específicamente, resalta la relevancia crítica de la dieta, la diversidad y la morfoespecie en la asignación de categorías. Estos hallazgos no solo contribuyen a la comprensión de las dinámicas ecológicas, sino que también ofrecen valiosas perspectivas para el desarrollo de estrategias efectivas en el manejo y control en el ámbito de estudio.



**Figura 5.9** Árbol de visualización.  
**Fuente:** Elaboración propia.

Los resultados de este estudio indican que el uso de abono propio está asociado con una mayor diversidad, equitatividad y heterogeneidad funcional en la comunidad. Esto posiblemente se deba a que la composición de este abono es más variada en comparación con los productos industriales. Es crucial destacar que el tipo de fertilizante orgánico, como el abono de corral, es ampliamente utilizado en la región, al igual que los bioestimulantes que implican la adición de micronutrientes principalmente.

Es importante señalar que estos insumos externos, aunque beneficiosos para la producción, también añaden un costo monetario significativo, especialmente en función de la extensión del cultivo. En este contexto, el hecho de que el abono de corral contribuya no solo a la fertilidad del suelo sino también a la diversidad y composición funcional de la comunidad biológica puede considerarse como un punto a favor. Esto sugiere que la adopción de prácticas que utilizan abonos orgánicos propios podría ser una estrategia positiva para fomentar transiciones agroecológicas, promoviendo prácticas más sostenibles y resilientes en la agricultura.

## **6. Conclusiones**

Luego del Estudio realizado de Ciencia de Datos bajo un enfoque ágil, los resultados revelan la importancia crucial de incorporar la Ciencia de Datos y otras prácticas de la Tecnología de la Información en diversas ramas de la industria agropecuaria, para un abordaje multidisciplinar. El análisis detallado de aspectos clave relacionados con la gestión agrícola, como el tipo de suelo, fertilizantes, métodos de riego y condiciones estacionales, destaca la relevancia de estas herramientas en la toma de decisiones informadas. La influencia del abono de corral estepa y los métodos de riego subraya la necesidad de estrategias tecnológicas para favorecer la salud del ecosistema agrícola. Además, los hallazgos indican que la aplicación de tecnologías informáticas, como el aprendizaje automático, ofrece valiosas perspectivas para clasificar organismos, resaltando la importancia de la dieta y diversidad en esta tarea. La combinación de prácticas sostenibles, como el desmalezado manual y el uso de abono propio, contribuye positivamente a la ampliación de hábitats y la diversidad de artrópodos, a lo que hay que agregar la relevancia de la informática para gestionar la complejidad de las interacciones en el ecosistema agrícola.

En futuras líneas de investigación, se sugiere una evaluación a largo plazo de prácticas agrícolas específicas, el desarrollo de estrategias específicas de Manejo Integrado de Plagas y la expansión de estudios a diferentes regiones geográficas, mediante herramientas informáticas para lograr avances en prácticas agrícolas más sostenibles y resilientes.

## 7. Referencias Bibliográficas

- [1] - Chillo, M. (2018), Descripción técnica PI UNRN 2018. Comparación del efecto de diferentes prácticas de manejo orgánico en frambueso sobre la biodiversidad de artrópodos, patógenos y agentes de biocontrol en El Bolsón, Río Negro. Directora María Verónica Chillo. Universidad Nacional de Río Negro. Sede Andina. Argentina.
- [2] - EMC – “The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things” – (2014). Publicada en: <https://bit.ly/46SWVrb>. Último acceso Junio 2023.
- [3] - Joyanes Aguilar, Luis, (2013). “Big Data: Análisis de grandes volúmenes de datos en organizaciones”. AlfaOmega, Grupo Editor, S.A de C.V., Mexico. ISBN: 978-607-707-689-6
- [4] - Tukey, J.W., (1962). The future of data analysis. Ann. Math. Stat. 33 (1), 1e67.
- [5] - Tukey, J.W., (1977). Exploratory Data Analysis. Addison-Wesley.
- [6] - Fayad, U. M., Piatetsky-Shapiro, G., Smyth, P., Uhturudsamy, R. (1996). Advances in Knowledge Discovery and Data Mining. AAAI Press.
- [7] - Martins, S. (2014). Derivación del Proceso de Explotación de Información Desde el Modelado del Negocio. Revista Latinoamericana de Ingeniería de Software, 2(1): 53-76, ISSN 2314-2642.
- [8] - García-Martínez, R., Britos, P., Rodríguez, D. (2013). Information Mining Processes Based on Intelligent Systems. Lecture Notes on Artificial Intelligence, 7906: 402-410. ISBN 978-3- 642- 38576-6.
- [9] - Britos, P. (2008). Procesos de explotación de información basados en sistemas inteligentes. (Tesis doctoral). Universidad Nacional de La Plata. Argentina.
- [10] - Bruzone, I. (2005). Dirección Nacional de Alimentos, Dirección de Industria Alimentaria. <https://bit.ly/3T3YmQc>. Último acceso Noviembre 2023
- [11] - InfoAgro - “El cultivo de la frambuesa”. Publicada en: <https://bit.ly/41isEAC> Último acceso Diciembre 2023.
- [12] - Morales, C., Pedreros, L., Riquelme, J., Uribe, H., Hirzel, J., Abarca, P., Rayentué, A. (eds.) 2017. Manual de manejo agronómico del frambueso. Boletín INIA N° 007. 114 p. Instituto de Investigaciones Agropecuarias INIA, Santiago de Chile, Chile
- [13] - CIREN. 2016. Modelo de adaptación al cambio climático por medio de la zonificación de aptitud productiva de especies hortofrutícolas priorizadas en la Región del Biobío.
- [14] - Fois, G; Agüero, G.; Britos, P. (2020). Evaluación comparativa de las metodologías Team Data Science Process TDSP y Analytics Solutions Unified Method for Data Mining ASUM-DM desde la perspectiva de la ciencia de datos. En Investigación Formativa en Ingeniería. 4ta. Edición. ISBN 978-958-52333-5-5, DOI <http://doig.org/10.5824/zenodo4031253>. Medellín. Antioquia. Colombia.
- [15] - IBM - (2015) “Analytics Solution Unified Method (ASUM)”. Publicada en: <https://bit.ly/47QMuvC>. Último acceso Noviembre 2023.

[16] - Orange. Publicada en: <https://orangedatamining.com/> Último acceso  
Noviembre 2023.

## Anexo 1 : Glosario

<i>Término</i>	<i>Descripción</i>
A	Agosto
A19	Agosto 2019
Abono_corral_estepa	Tipo de fertilización
Abono_corral_propio	Tipo de fertilización
Acari	Código de morfoespecie identificada en la muestra
Acari.sp1	Código de morfoespecie identificada en la muestra
Acari.sp2	Código de morfoespecie identificada en la muestra
Aphidae	Código de morfoespecie identificada en la muestra
Aphidae.sp1	Código de morfoespecie identificada en la muestra
Aphidae.sp2	Código de morfoespecie identificada en la muestra
Aphidae.sp3	Código de morfoespecie identificada en la muestra
Apidae	Código de morfoespecie identificada en la muestra
Apis.mel	Código de morfoespecie identificada en la muestra
Araneomorph.blanca	Código de morfoespecie identificada en la muestra
Araneomorph.sp3	Código de morfoespecie identificada en la muestra
Araneomorph.sp4	Código de morfoespecie identificada en la muestra
Araneomorph.sp7	Código de morfoespecie identificada en la muestra
Araneomorph.tigre	Código de morfoespecie identificada en la muestra
Araneomorphae	Código de morfoespecie identificada en la muestra
Avispa.papel	Código de morfoespecie identificada en la muestra
Bombus.sp	Código de morfoespecie identificada en la muestra
Braconidae	Código de morfoespecie identificada en la muestra
Bt_Tracer	Tipo de control de plagas
Camponotus	Código de morfoespecie identificada en la muestra

Carabidae	Código de morfoespecie identificada en la muestra
Cerambycidae	Familia taxonómica
CH1	Chacra 1
CH2	Chacra 2
CH3	Chacra 3
Chelicerata	Orden taxonómico
Cicadellidae	Familia taxonómica
Cicimidae	Familia taxonómica
Cicimidae (FCH1M11)	Código de morfoespecie identificada en la muestra
Coleoptera	Orden taxonómico
Collembola	Orden taxonómico
Controlador	Rol de la morfoespecie en el ecosistema
Curculinoidea	Familia taxonómica
CWM.Dieta	Índice de diversidad funcional
CWM.SE	Índice de diversidad funcional
CWM.Size	Índice de diversidad funcional
Cynipodae	Familia taxonómica
D.s	Código de morfoespecie identificada en la muestra
Depredador	Rol de la morfoespecie en el ecosistema
Dermaptera	Orden taxonómico
Detritívoro	Dieta de la morfoespecie
Diplopoda	Orden taxonómico
Díptero	Orden taxonómico
Drosophilidae	Familia taxonómica
Elateridae	Familia taxonómica
Elmidae	Familia taxonómica
F20	Febrero 2020

FDis	Índice de diversidad funcional
FDiv	Índice de diversidad funcional
FEve	Índice de diversidad funcional
Fitofago	Dieta de la morfoespecie
Folívoro	Dieta de la morfoespecie
Forficula.auri	Código de morfoespecie identificada en la muestra
Forficulidae	Familia taxonómica
FRic	Índice de diversidad funcional
Frugívoro	Dieta de la morfoespecie
G	Tipo de riego (goteo)
Guanito_Lombricomp	Tipo de fertilizante
GyA	Tipo de riego (goteo y aspersión)
Hemiptera	Orden taxonómico
Herbívoro	Dieta de la morfoespecie
Heteroptera	Orden taxonómico
Homoptera	Orden taxonómico
Hymenoptera	Orden taxonómico
Ichnumonidae	Familia taxonómica
InvSimpson	Índice de diversidad
Isotomidae	Familia taxonómica
Iulida	Suprafamilia taxonómica
Lepidoptera	Orden taxonómico
Lithobiomorpha	Familia taxonómica
Ly.sp1	Código de morfoespecie identificada en la muestra
Ly.sp2	Código de morfoespecie identificada en la muestra
Lycosidae	Familia taxonómica
M1	Muestra 1



M10	Muestra 10
M11	Muestra 11
M12	Muestra 12
M13	Muestra 13
M14	Muestra 14
M15	Muestra 15
M16	Muestra 16
M17	Muestra 17
M18	Muestra 18
M19	Muestra 19
M2	Muestra 2
M20	Muestra 20
M3	Muestra 3
M4	Muestra 4
M5	Muestra 5
M6	Muestra 6
M7	Muestra 7
M8	Muestra 8
M9	Muestra 9
MO	Rol de la morfoespecie en el ecosistema
Myriapoda	Familia taxonómica
N19	Noviembre 2019
NA	Sin datos
Nabidae	Familia taxonómica
Nectar	Dieta de la morfoespecie
nn (fitogafo)	Código de morfoespecie identificada en la muestra
Pentanomidae	Familia taxonómica

Pentatomidae	Código de morfoespecie identificada en la muestra
Piesmidae	Familia taxonómica
Plaga	Rol de la morfoespecie en el ecosistema
Polifago	Dieta de la morfoespecie
Polinizador	Rol de la morfoespecie en el ecosistema
Polydesmia	Suprafamilia taxonómica
Proctotrupoidea	Familia taxonómica
$q_0$	Índice de diversidad
$q_1$	Índice de diversidad
$q_2$	Índice de diversidad
RaoQ	Índice de diversidad funcional
Riqueza	Índice de diversidad
Scarabaide	Familia taxonómica
Scarabidae	Familia taxonómica
Shannon	Índice de diversidad
Simpson	Índice de diversidad
sp.	morfoespecie
Sphecidae	Familia taxonómica
Staphylinidae	Familia taxonómica
Tenebrionidae	Familia taxonómica
Tenthredinidae	Familia taxonómica
Thripidae	Familia taxonómica
Thysanoptera	Familia taxonómica
Vespidae	Familia taxonómica
Vespula.ger	Código de morfoespecie identificada en la muestra

## Anexo 2 : Muestreo de Tabla Utilizada

En la siguiente tabla se ejemplifica un registro para dar a conocer los valores de los campos, de la tabla en estudio. En la primera columna se encuentra el nombre del campo y en la segunda el valor del primer registro de la tabla.

Si desea visualizar los datos en profundidad, puede verlos a través del siguiente [enlace](#).

<b>Fecha</b>	Ago-19
<b>Chacra</b>	CH1
<b>Muestra</b>	M1
<b>Riego</b>	Goteo
<b>Desmalezado</b>	Manual
<b>Control_Plagas</b>	Bt_Tracer
<b>Fertilizacion</b>	Abono_corral_estepa
<b>HumedadSuelo</b>	78
<b>D.s</b>	0
<b>Cicadellidae</b>	1
<b>Ly.sp1</b>	50
<b>Ly.sp2</b>	0
<b>Araneomorph.sp3</b>	0
<b>Araneomorph.sp4</b>	0
<b>Araneomorph.sp7</b>	0
<b>Araneomorph.tigre</b>	0
<b>Araneomorph.blanca</b>	0
<b>Acari.sp1</b>	0
<b>Acari.sp2</b>	0
<b>Camponotus</b>	0
<b>Camponotus.sp2</b>	0

<b>Apis.mel</b>	0
<b>Bombus.sp</b>	0
<b>Ichnumonidae.sp1</b>	0
<b>Ichnumonidae.sp2</b>	0
<b>Tenthredinidae</b>	0
<b>Sphecidae</b>	0
<b>Cynipoidae.sp1</b>	0
<b>Cynipoidae.sp2</b>	0
<b>Vespula.ger</b>	0
<b>Avispa.papel</b>	0
<b>Braconidae</b>	0
<b>Proctotrupoidea</b>	0
<b>Forficula.auri</b>	0
<b>Dermaptera</b>	0
<b>Collembola</b>	0
<b>Hemiptera.sp1</b>	0
<b>Aphidae.sp1</b>	0
<b>Aphidae.sp2</b>	0
<b>Aphidae.sp3</b>	0
<b>Hemiptera.sp2</b>	0
<b>Pentatomidae</b>	0
<b>Nabidae</b>	0
<b>Iulida</b>	0
<b>Polydesmia</b>	0
<b>Lithobiomorpha</b>	0
<b>Curculionidae.sp1</b>	0
<b>Curculionidae.sp2</b>	0

<b>Elmidae</b>	0
<b>Carabidae.sp1</b>	0
<b>Carabidae.sp2</b>	0
<b>Carabidae.sp3</b>	0
<b>Carabidae.sp5</b>	0
<b>Carabidae.sp6</b>	0
<b>Carabidae.sp7</b>	0
<b>Carabidae.sp4</b>	0
<b>Carabidae.sp8</b>	0
<b>Carabidae.sp9</b>	0
<b>Carabidae.sp10</b>	0
<b>Coleoptera.sp1</b>	0
<b>Tenebrionidae</b>	0
<b>Scarabidae</b>	0
<b>Cerambycidae</b>	0
<b>Elateridae.sp1</b>	0
<b>Elateridae.sp2</b>	0
<b>Staphylinidae</b>	0
<b>Piesmidae</b>	0
<b>Cicimidae</b>	0
<b>Lepidoptera.sp1</b>	0
<b>Lepidoptera.sp2</b>	0
<b>Shannon</b>	0.1
<b>Simpson</b>	0.04
<b>InvSimpson</b>	1.04
<b>Riqueza</b>	2
<b>q0</b>	2

<b>q1</b>	1.1
<b>q2</b>	1.04
<b>FRic</b>	NA
<b>FEve</b>	Na
<b>FDiv</b>	NA
<b>FDis</b>	0.04
<b>RaoQ</b>	0.02
<b>CWM.Dieta</b>	Depredador
<b>CWM.SE</b>	Controlador
<b>CWM.Size</b>	mediano