

Optimización del riego en producción de Maíz CL-86 utilizando Minería de Datos y Aprendizaje automático

Wilyereny Lorenzo, George Báez, Margari Pérez, *Willy Maurer*



Instituto Especializado
de Estudios Superiores

LOYOLA

Octubre, 2022

1. Introducción

1.1 Definición del problema



La finca recurre al conocimiento empírico para realizar el riego del cultivo; por lo que la humedad de suelo y otras variables no son tomadas en cuenta.



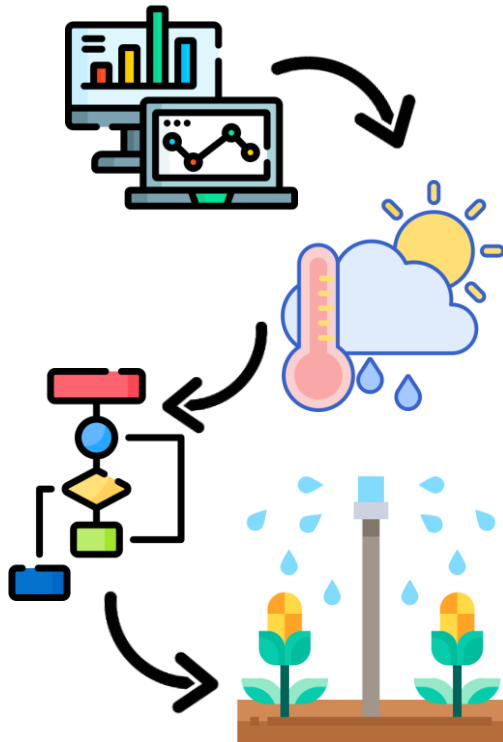
Existe un conjunto de datos que resulta ser ineficiente para un análisis de datos efectivo que permita optimizar el modelo de riego usado en la finca.



No existía una logística adecuada para los equipos y parámetros con los que se capturan los datos.

1. Introducción

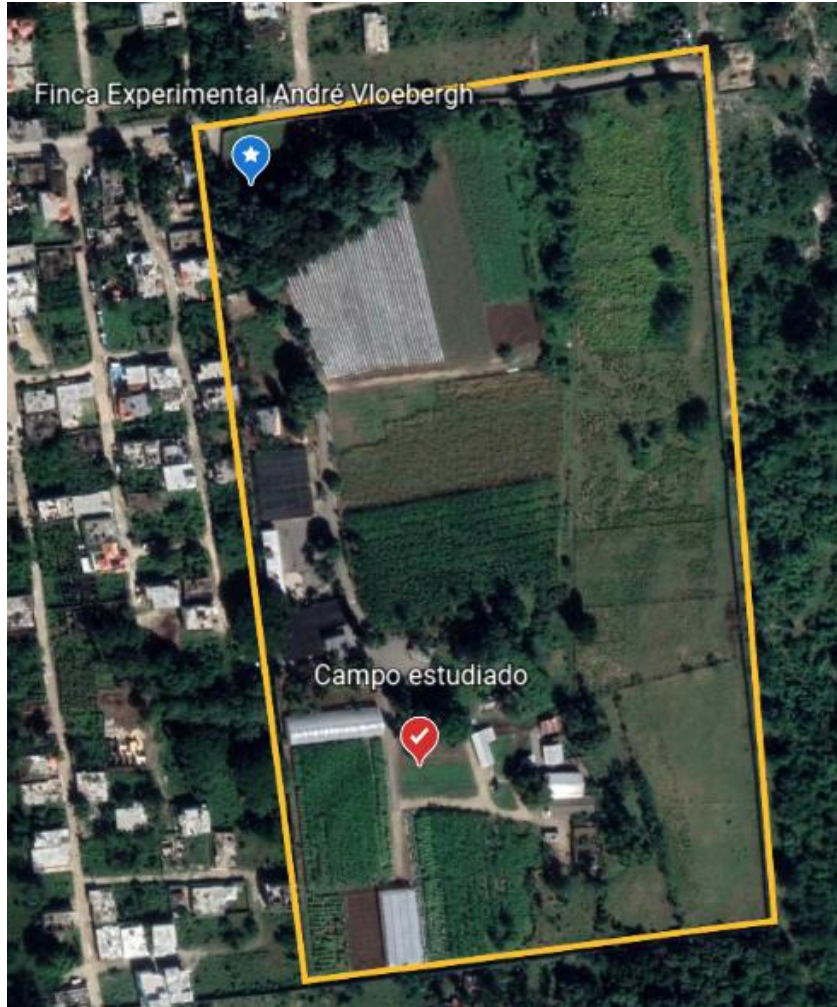
1.3 Objetivo principal



Diseñar un método para la captura y análisis de datos sobre las características edafoclimáticas de una plantación de maíz, para la creación de un algoritmo que permita la implementación de un modelo de riego eficiente en la Finca experimental profesor “André Vloebergh”.

2. Metodología

2.1 Localización del estudio



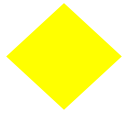
Finca Experimental André Vloebergh



3.4 kilómetros de la planta física del IPL

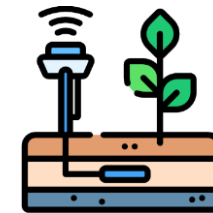


18° 25' N y 70° 07' latitud oeste.



2. Metodología

2.1 Localización del estudio



Instalación de la estación

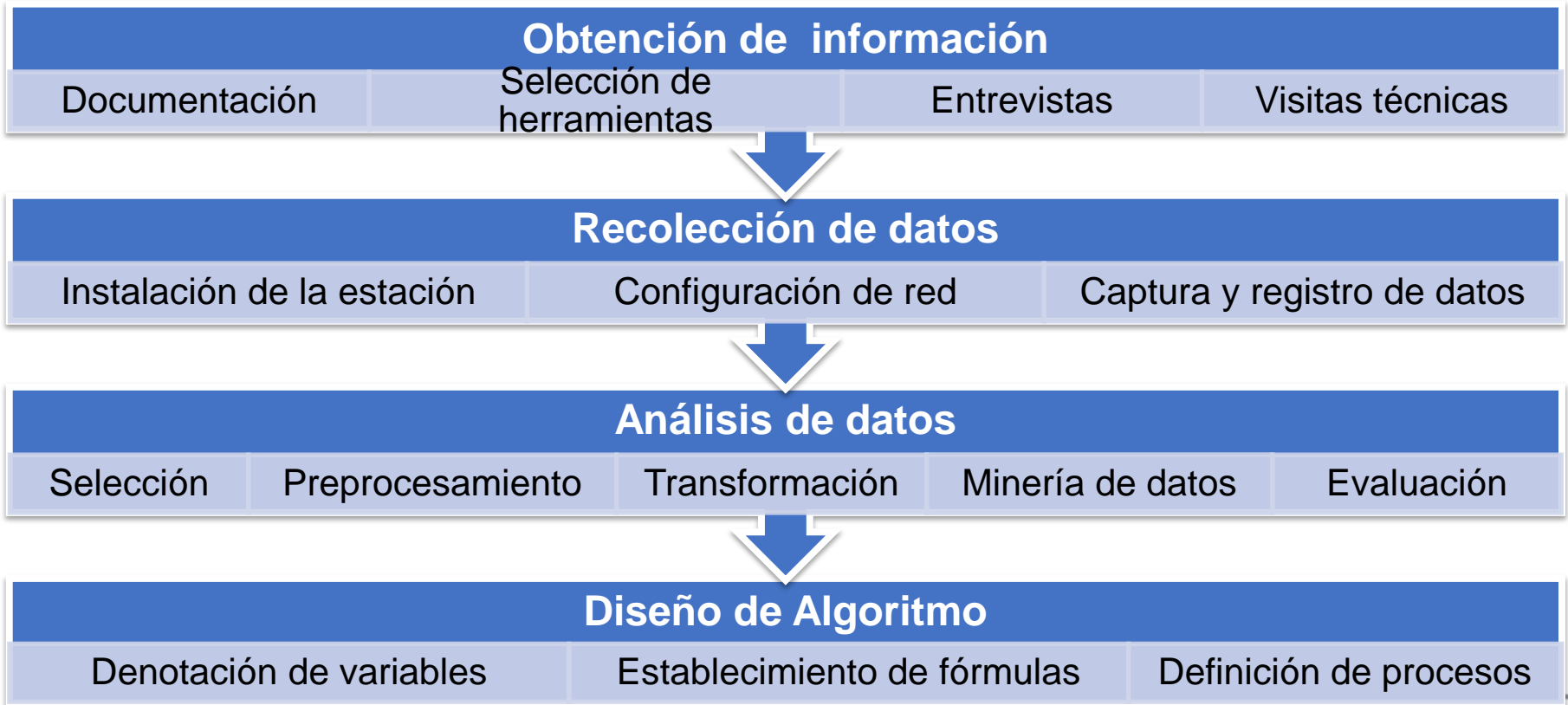


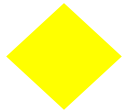
Instalación de sensores



2. Metodología

2.2 Procedimientos





2. Metodología

2.3 Documentación, entrevistas y visitas

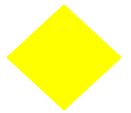


TECNOLOGÍA



AGRONOMIA





2. Metodología

2.4 Selección de Herramientas - Análisis



2. Metodología

2.4 Selección de Herramientas - Análisis



- Visualización y análisis grandes cantidades de datos
- Interfaz gráfica sencilla e intuitiva
- Configuración a través de widgets
- Utilización optima de recursos de memoria y CPU

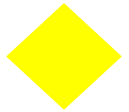
2. Metodología

2.4 Selección de Herramientas – Recolección

Estación Meteorológica Vantage Pro2 Plus



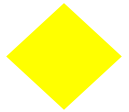
- 40+ Variables
- Acceso remoto
- 10+ Sensores
- Interfaz única
- Respaldo de datos
- Estructura robusta



2. Metodología

2.5 Variables

CULTIVO	CLIMA		
Ancho del tallo	Temperatura ambiente	Índice THW	Radiación Solar Máxima
Alto de la planta	Temperatura ambiente Máxima	Índice THSW	Evapotranspiración
PARÁMETROS TÉCNICOS	Temperatura ambiente Mínima	Índice UV	Temperatura Interior
Recepción	Humedad Relativa	Dosis UV	Humedad Interior
Intervalo	Punto de rocío	Índice UV máximo	Rocío Interior
SUELO	Velocidad viento	Frío grado días	Calor Interior
Humedad suelo 1	Velocidad viento máxima	Calor grado días	Densidad del aire
Humedad suelo 2	Dirección del viento	Presión atmosférica	Muestras de viento
Humedad suelo 3	Dirección del viento predominante	Lluvia	Tasa de Viento
Humedad suelo 4	Corriente de viento	Tasa de Lluvia	Bulbo Húmedo
Temperatura suelo 1	Sensación térmica	Radiación Solar	11 Viento Frío
Temperatura suelo 2	Índice de calor	Energía Solar	
Temperatura suelo 3			
Temperatura suelo 4			



2. Metodología

2.6 Instalación de la estación



2. Metodología

2.6 Instalación de la estación

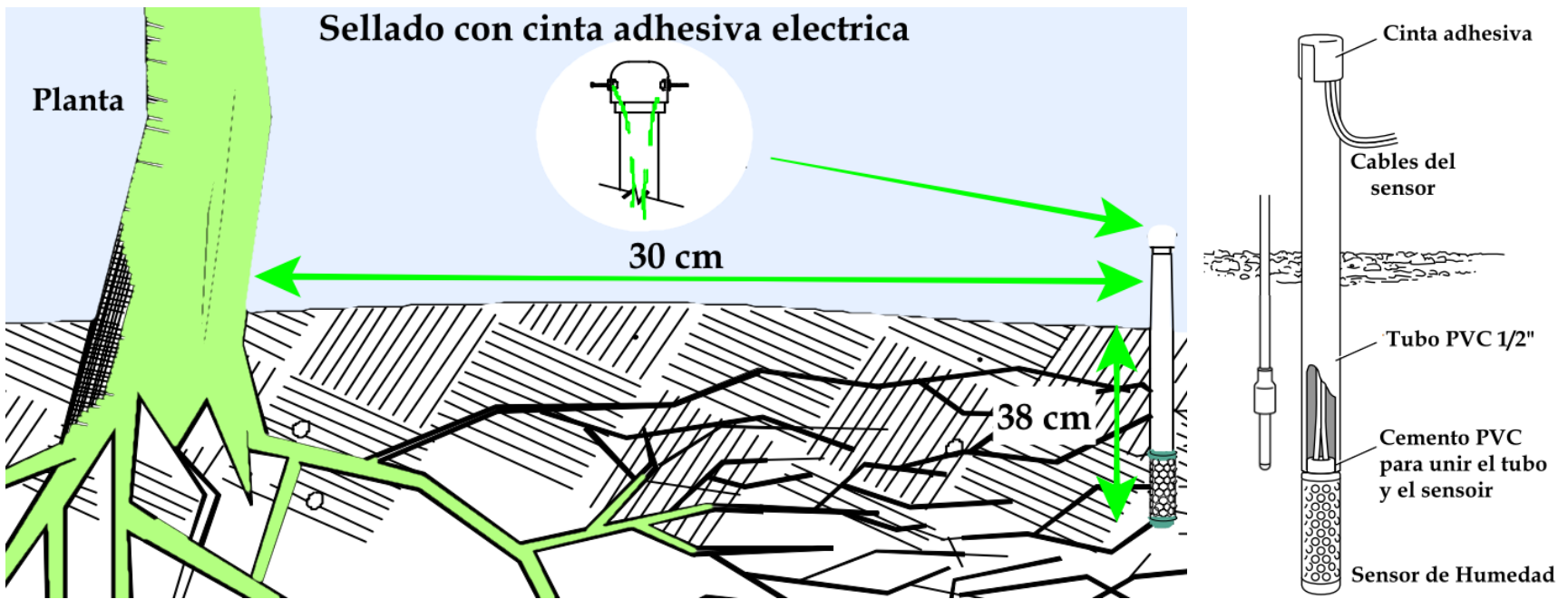


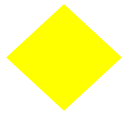
Estación de parámetros de
conjunto de variables (temperatura, humedad y velocidad de viento)



2. Metodología

2.6 Instalación de la estación

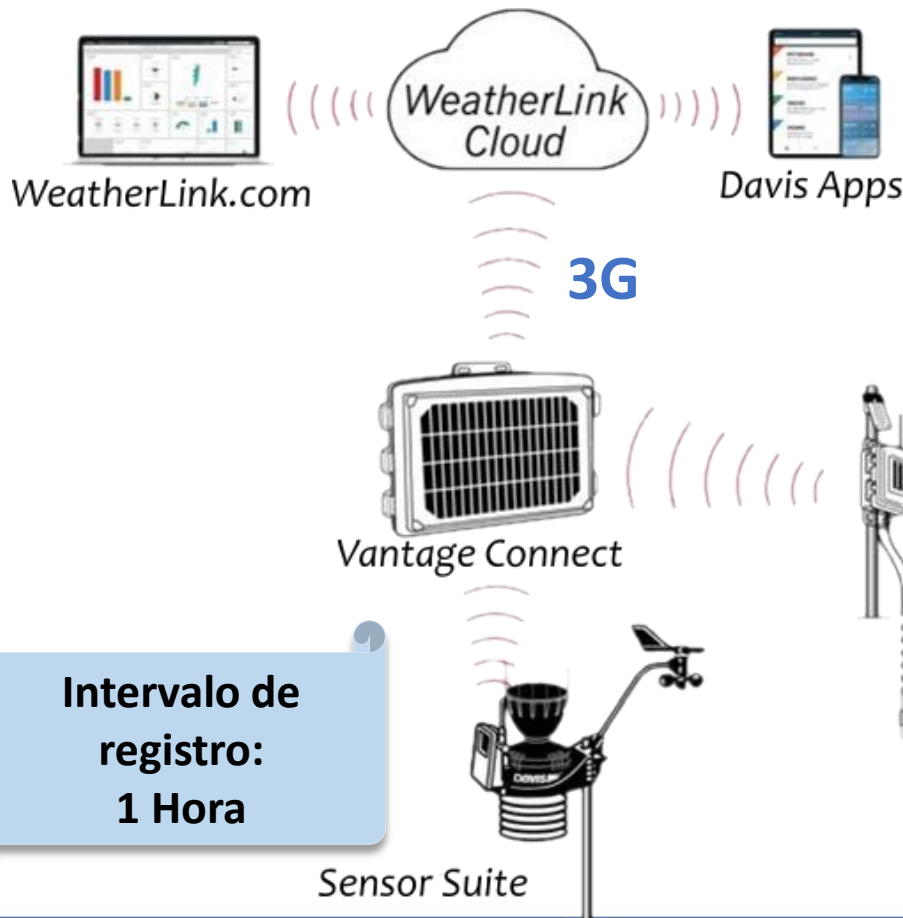




2. Metodología

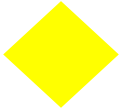
2.7 Configuración de red

Recolección de datos con registro remoto



**Intervalo de registro:
1 Hora**

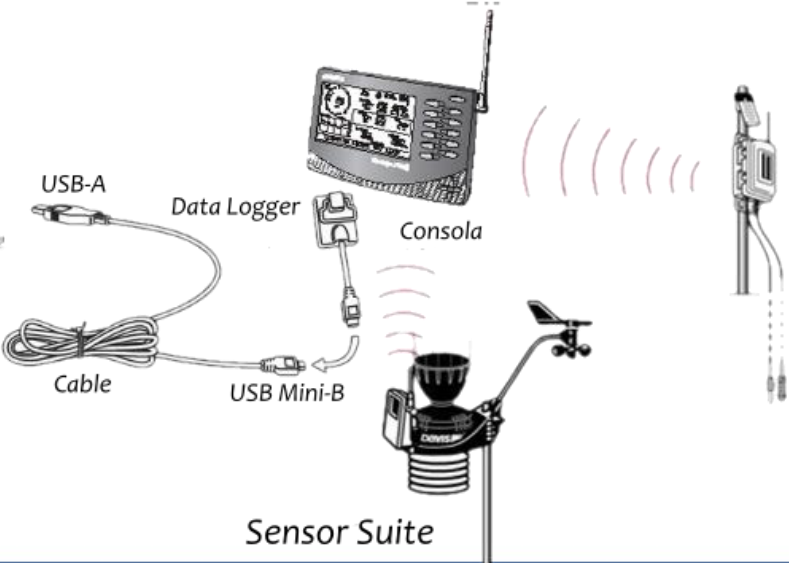
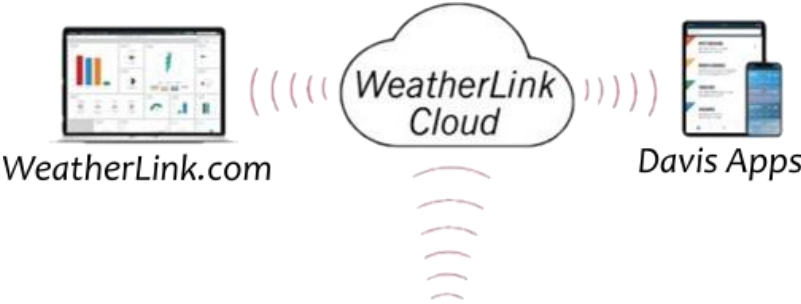
Estación inalámbrica de Temperatura - Humedad de Suelo y Hoja



2. Metodología

2.8 Configuración de red

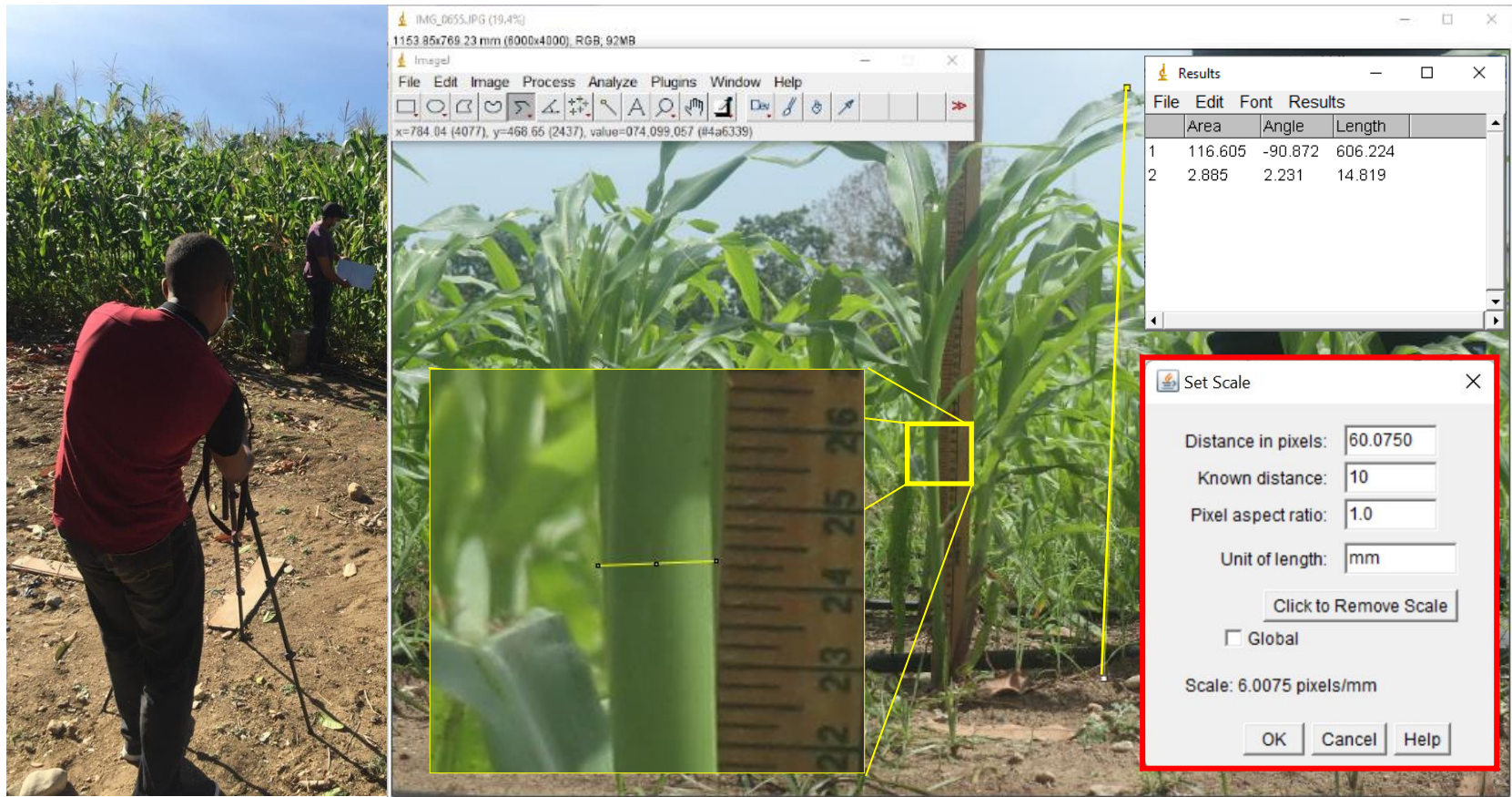
Recolección de datos con registro local



**Intervalo de registro:
1 Minuto**

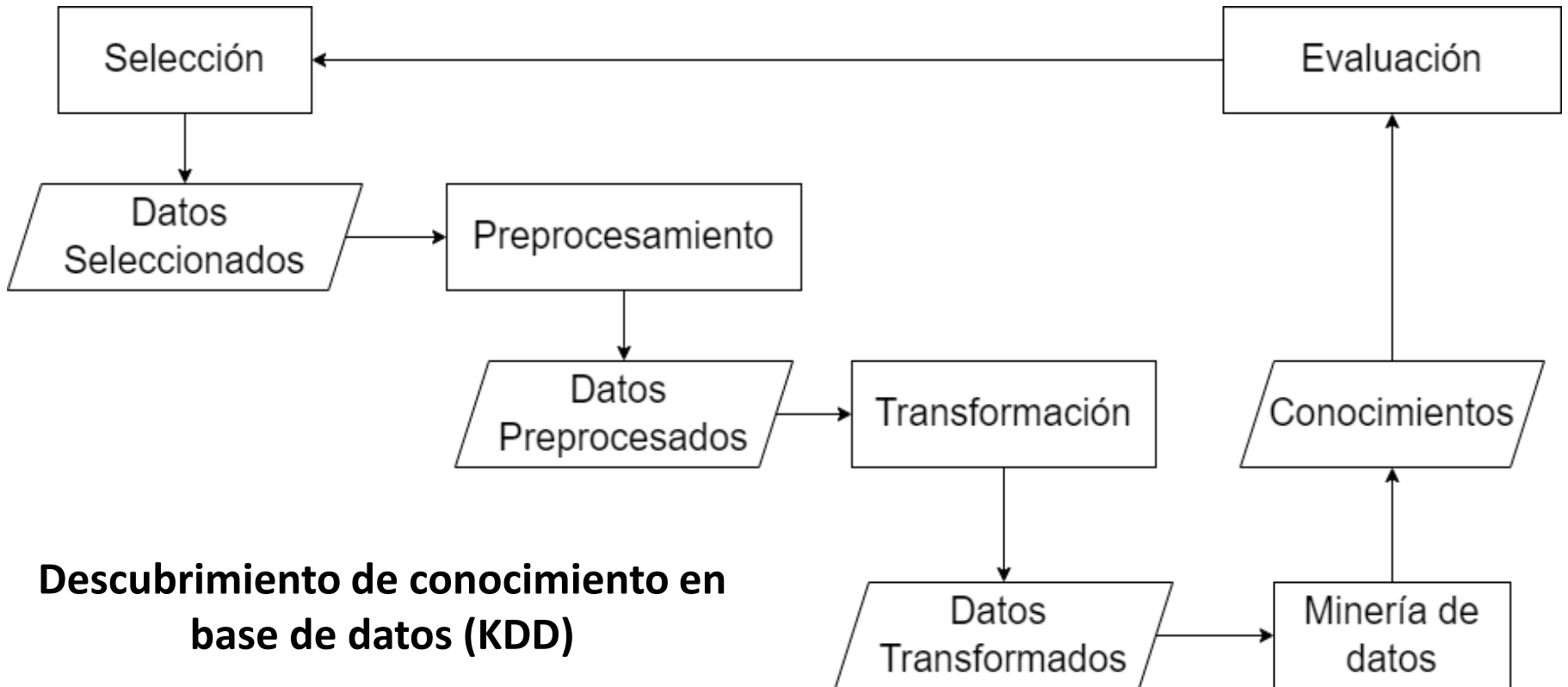
2. Metodología

2.9 Recolección de datos fotogramétricos

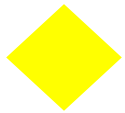


2. Metodología

2.10 Análisis de datos



Descubrimiento de conocimiento en base de datos (KDD)

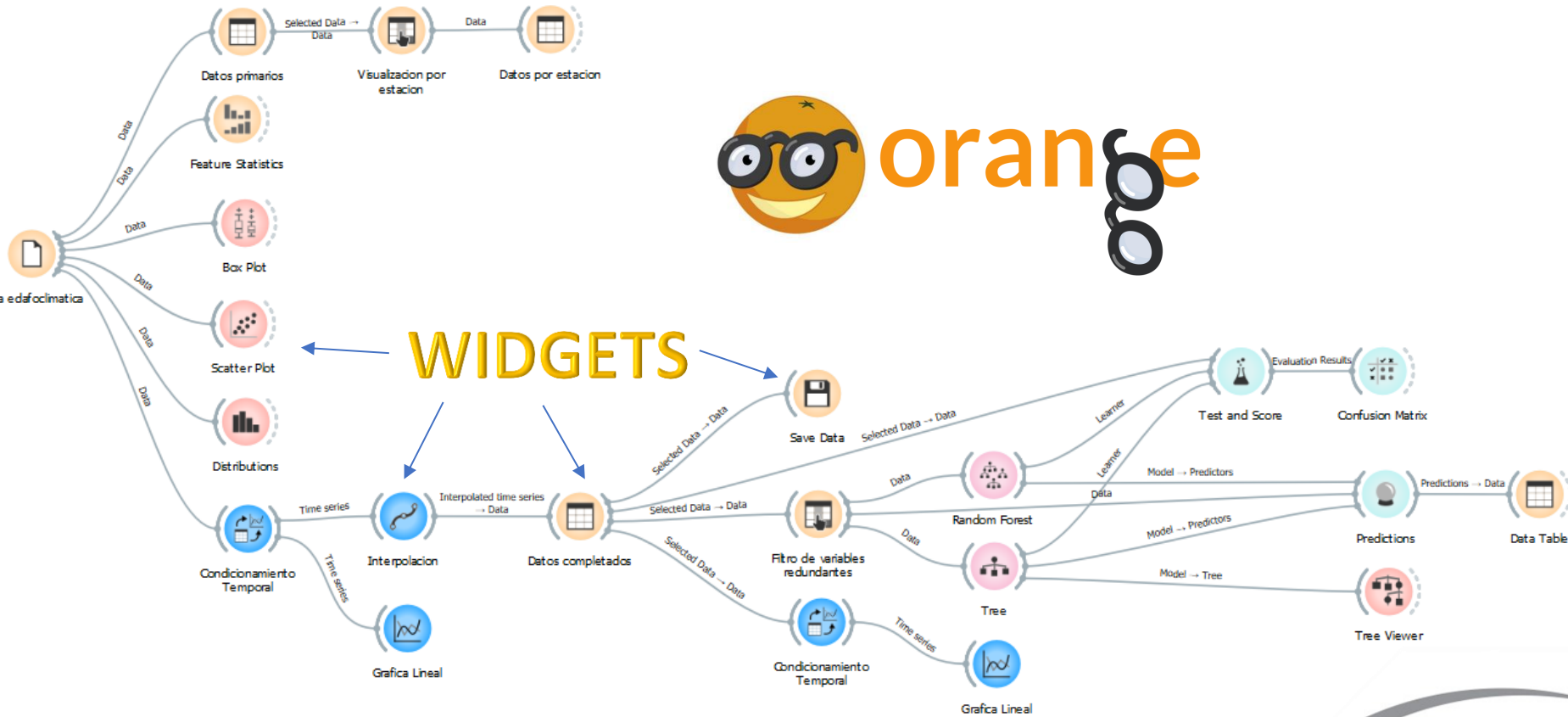


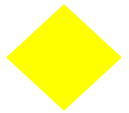
2. Metodología

2.11 Análisis de datos – Flujo de Análisis



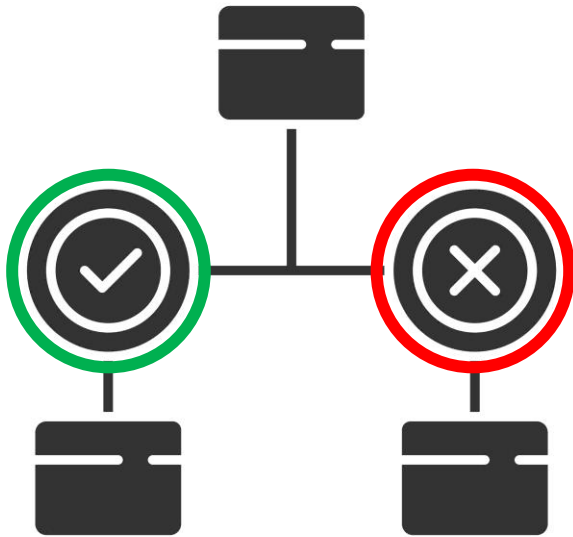
WIDGETS



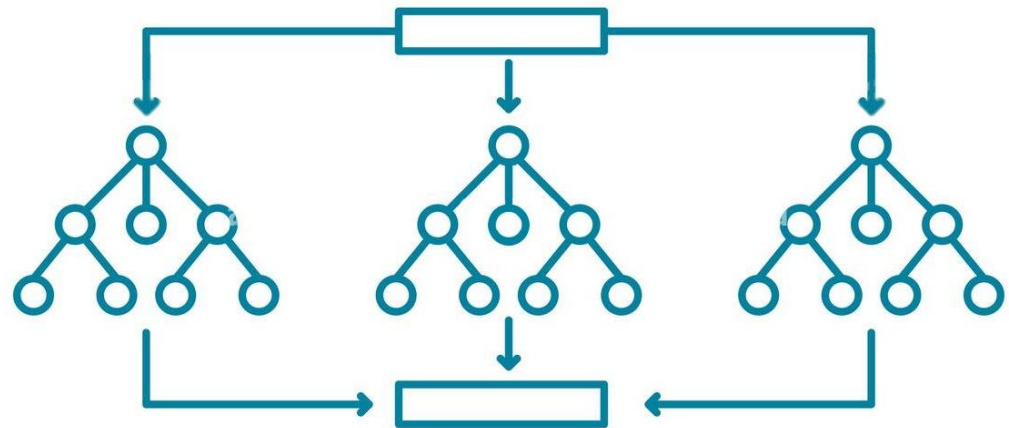


2. Metodología

2.12 Análisis de datos – Algoritmos



Árbol de decisión



Bosque Aleatorio

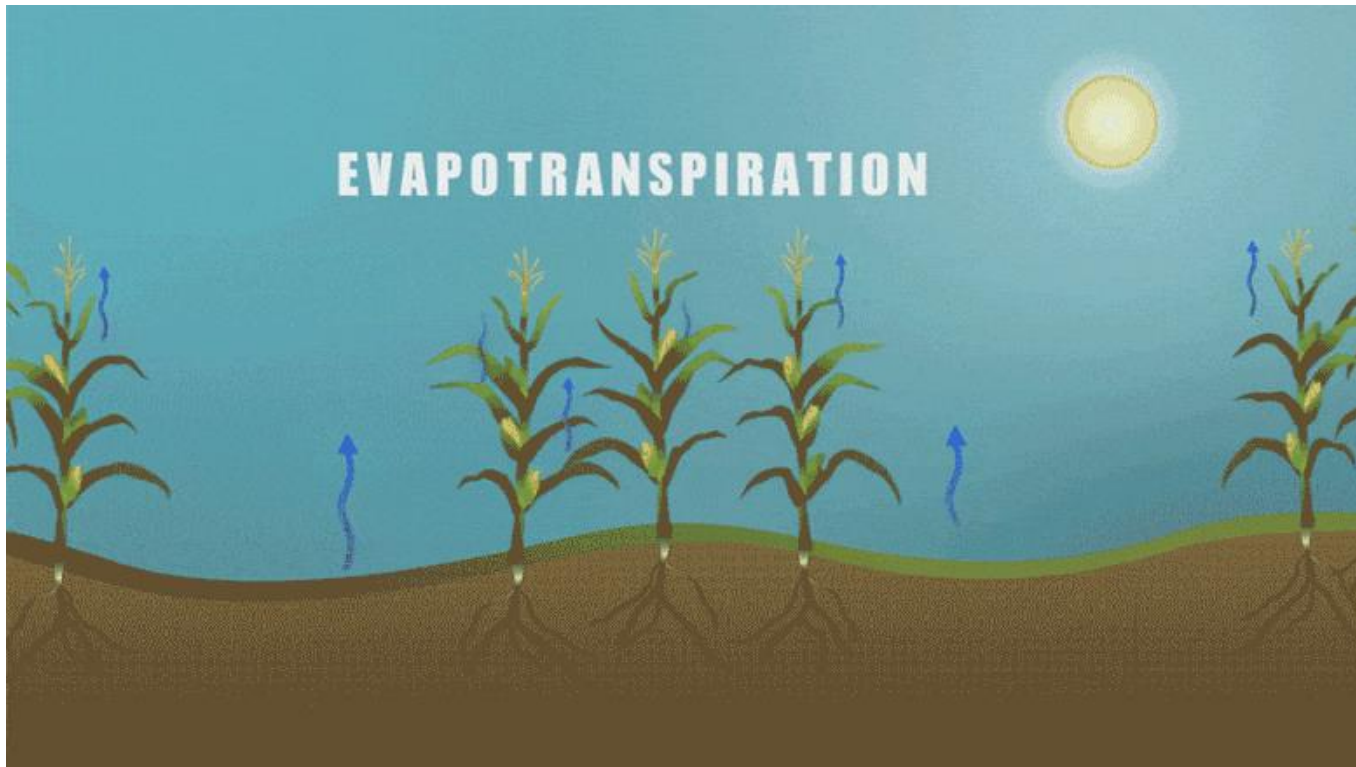
2. Metodología

2.13 Análisis de datos – Variable Objetivo

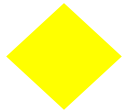


2. Metodología

2.13 Análisis de datos – Variable Objetivo



Proceso de transferencia Hídrica
Precipitación-Evapotranspiración

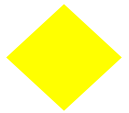


2. Metodología

2.13 Análisis de datos – Variable Objetivo

Cuadro 2.1 Condición de riego según las características de humedad de suelo

Lectura en Centibar (cb)	Condición del suelo	Categoría	Acción	Humedad disponible
0 - 10	Suelo Saturado. Ocurre durante uno o dos días después del riego.	Saturado	No Regar	100%
10 - 20	El suelo está suficientemente húmedo.	Húmedo		75%
30 - 60	Rango habitual para regar (excepto suelos arcillosos pesados).	Seco	Regar	50%
60 - 100	Rango habitual para regar suelos arcillosos pesados	Muy seco		25%
100 - 200	Suelo extremadamente seco	Árido		0%



2. Metodología

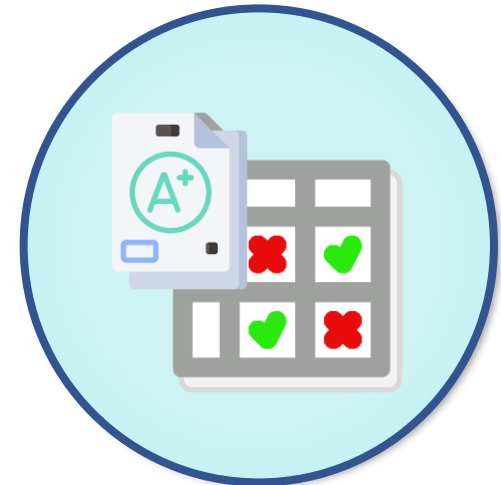
2.14 Análisis de datos – Evaluación



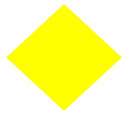
Predicciones



Prueba y calificación



Matriz de confusión

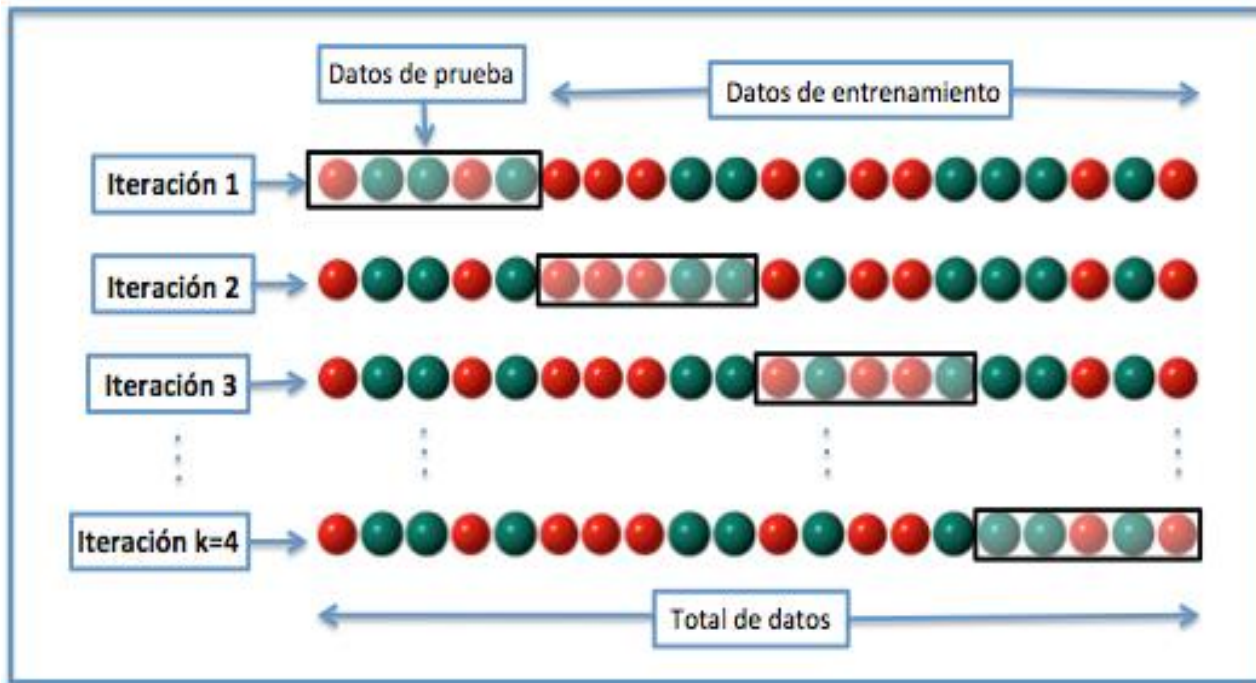


2. Metodología

2.14 Análisis de datos – Evaluación

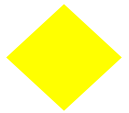


Prueba y calificación



Test and Score - Orange

- Cross validation
 - Number of folds: 10
 - Stratified
- Cross validation by feature
- Random sampling
 - Repeat train/test: 10
 - Training set size: 66 %
 - Stratified
- Leave one out
- Test on train data
- Test on test data



3. Resultados

3.1 Recolección y transformación de datos

87513 registros
47 variables
9.7% datos faltantes



Datos.txt

Datos locales

1223 registros
38 variables
3.4% datos faltantes



Datos remotos

51 registros
4 variables
35.9% datos faltantes

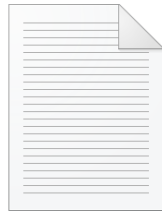


Datos fotogramétricos

3. Resultados

3.1 Recolección y transformación de datos

- 87513 registros
 - 47 variables
- 9.7% datos faltantes



Datos.txt

Datos locales

- 1223 registros
 - 38 variables
- 3.4% datos faltantes



- 43199 registros
- 40 variables
- 0.0% datos faltantes

Datos remotos

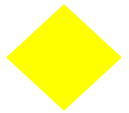
- 51 registros
 - 4 variables
- 35.9% datos faltantes



- 1223 registros
- 30 variables
- 0.0% datos faltantes

- 51 registros
- 3 variables
- 0.0% datos faltantes

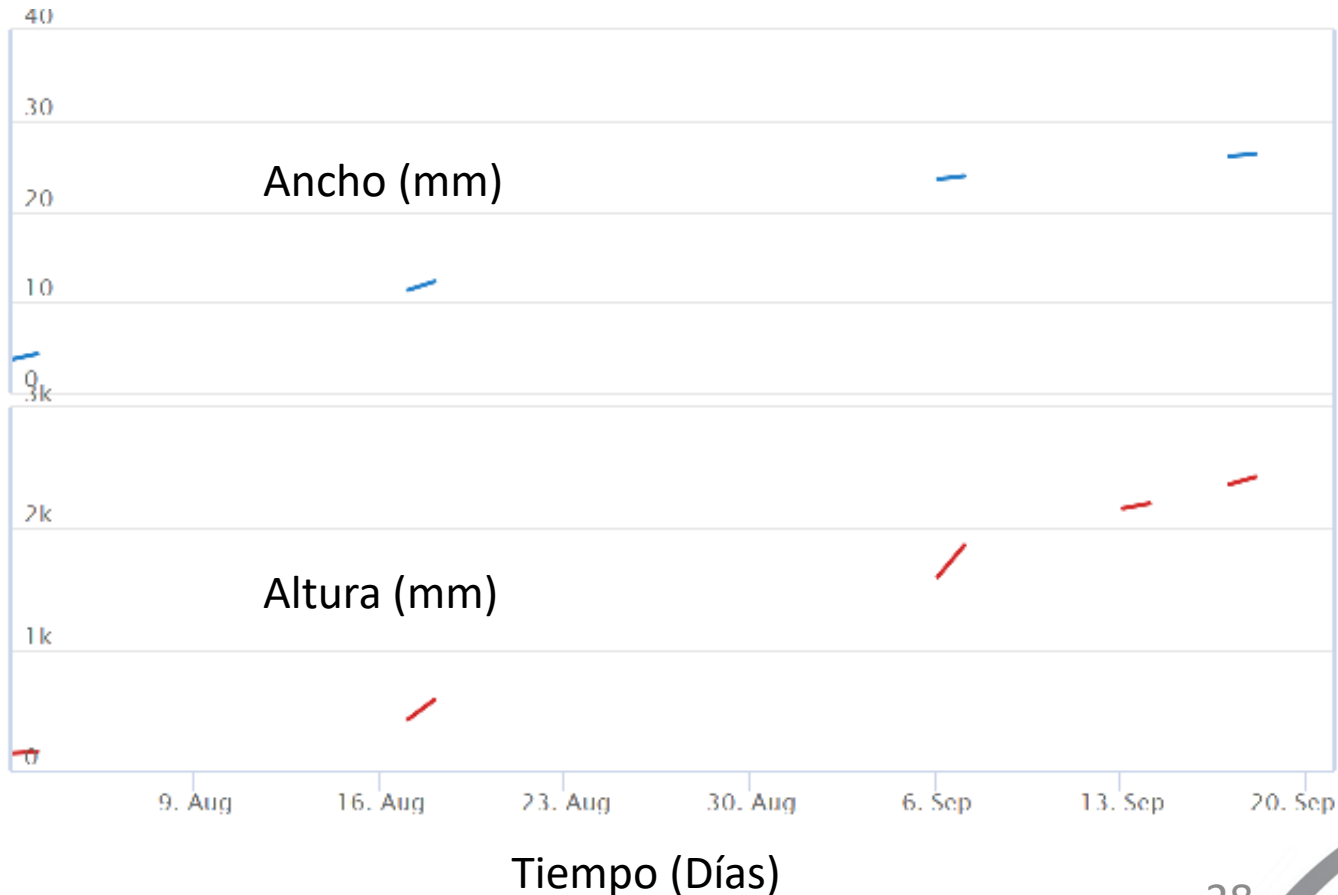
Datos fotogramétricos



3. Resultados

3.2 Análisis de datos – Relación de crecimiento

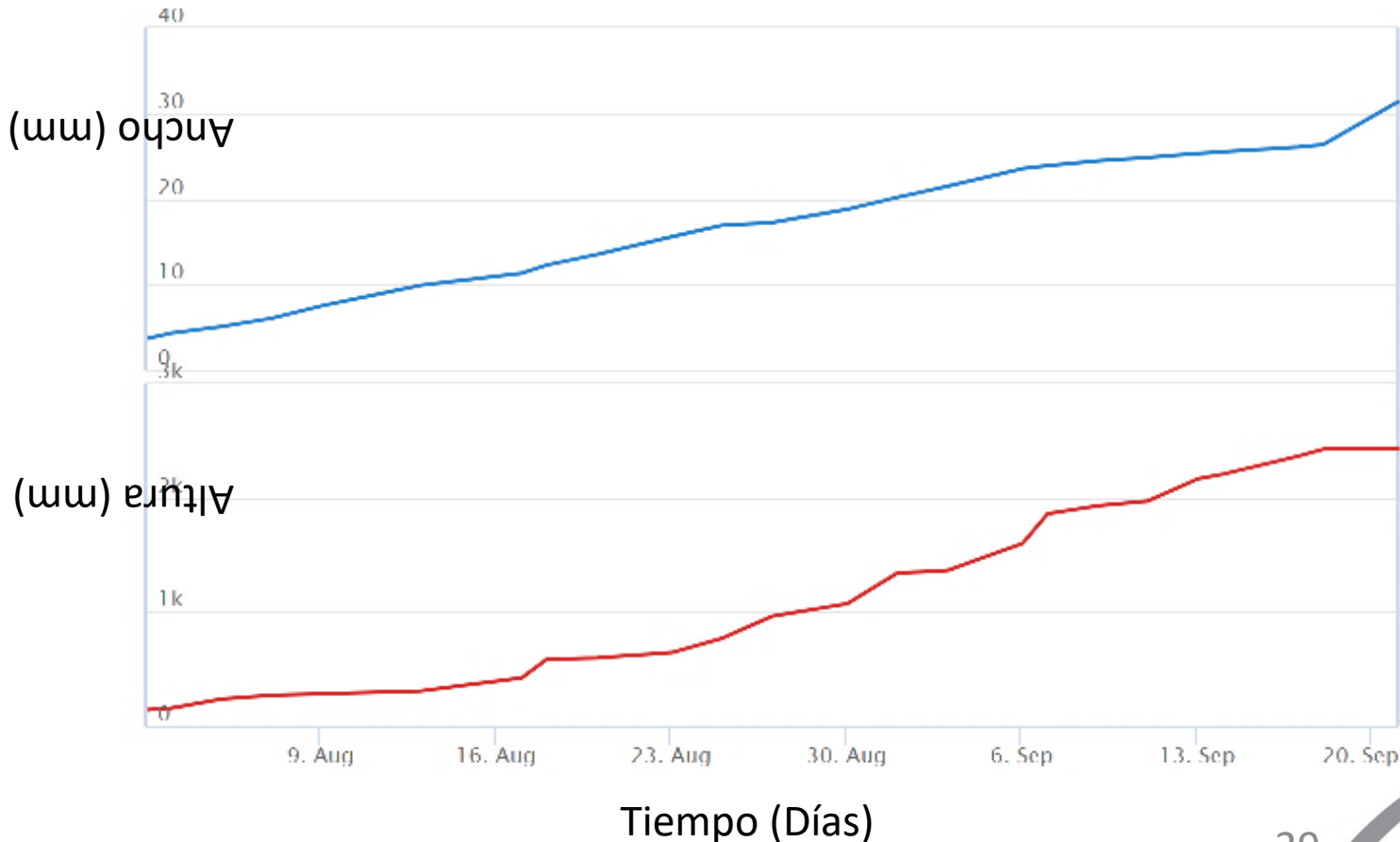
Gráfico 3.1 Altura y ancho del tallo de la planta a través del estudio (No completado)



3. Resultados

3.3 Análisis de datos – Relación de crecimiento

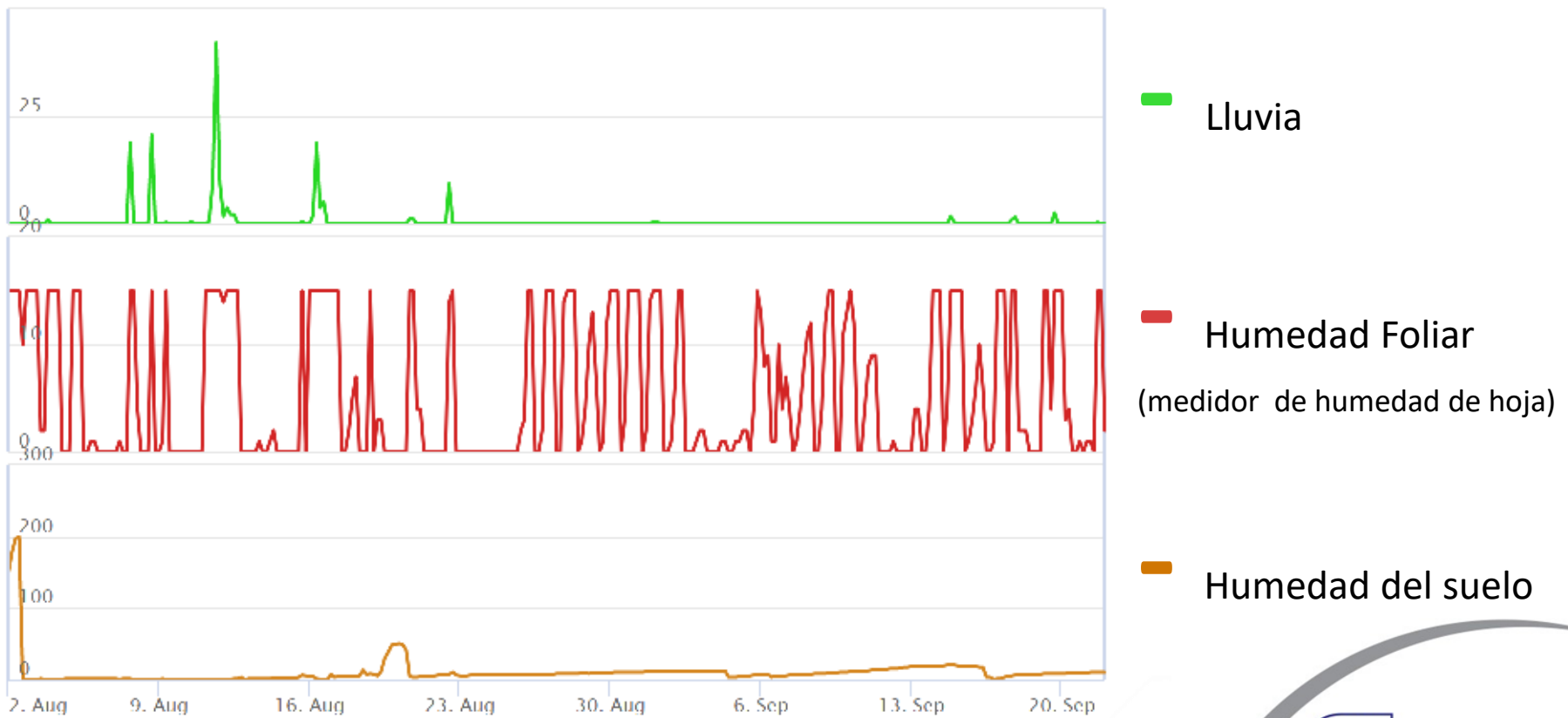
Gráfico 3.2 Altura y ancho del tallo de la planta a través del estudio (Completado)

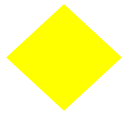


3. Resultados

3.4 Análisis de datos – Relación de Humedad

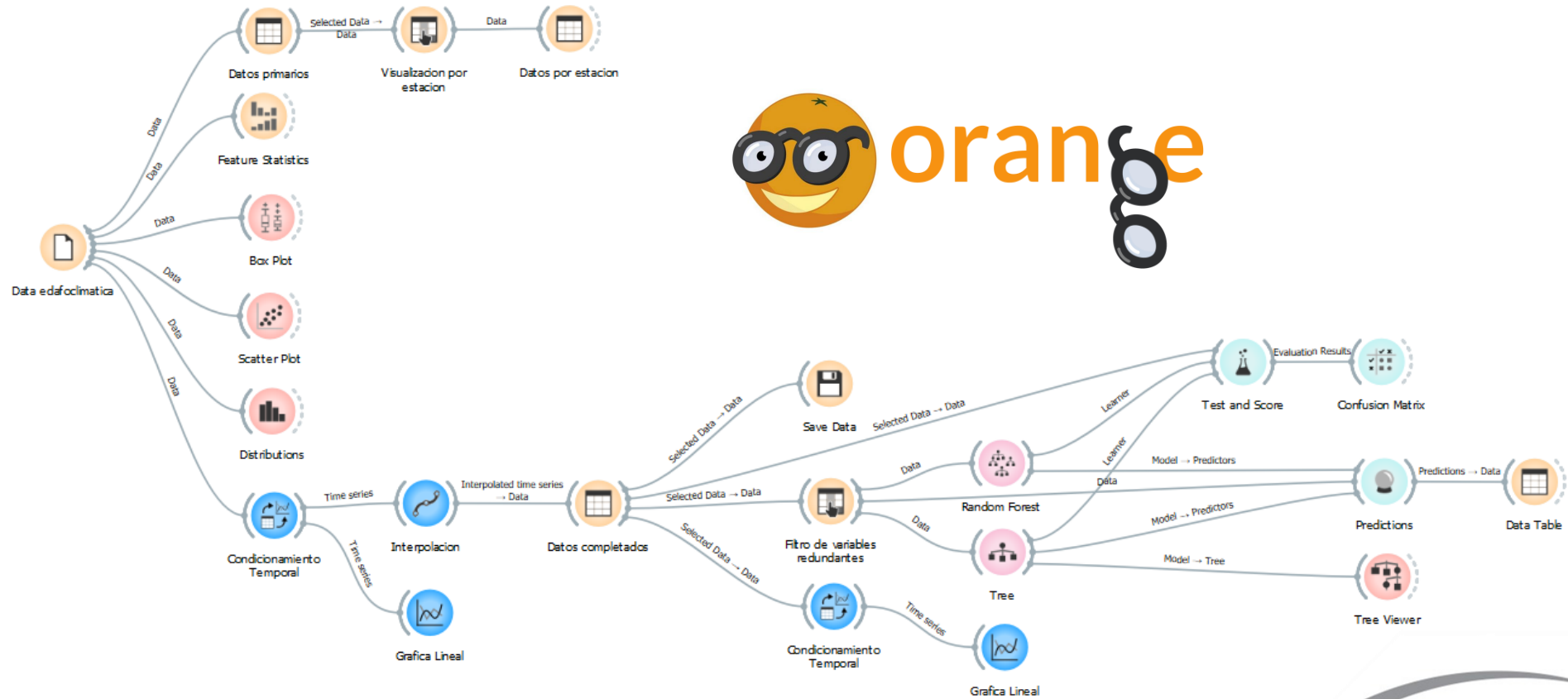
Gráfico 3.3 Comportamiento de la Humedad Foliar y Humedad del suelo respecto a la Lluvia





3. Resultados

3.5 Implementación de modelo de análisis



3. Resultados

3.6 Evaluación – Prueba y calificación



Recomendaciones de riego



Evapotranspiración y necesidad hídrica



Humedad del suelo

9 niveles
Árbol de decisiones
98.7 %
Bosque Aleatorio
99.2 %

15 niveles
Árbol de decisiones
98.7 %
Bosque Aleatorio
99.2 %

14 niveles
Árbol de decisiones
99.9 %
Bosque Aleatorio
99.9 %

3. Resultados

3.6 Evaluación – Matriz de confusión



Recomendaciones de riego

		Predicted		
		N	S	Σ
Actual	N	1192	9	1201
	S	8	14	22
Σ		1200	23	1223



Evapotranspiración y necesidad hídrica

		Predicted		
		N	S	Σ
Actual	N	1153	33	1186
	S	33	4	37
Σ		1186	37	1223



Humedad del suelo

		Predicted			
		Saturado	no regar	regar	Σ
Actual	Saturado	883	0	0	883
	no regar	0	303	0	303
	regar	0	1	36	37
Σ		883	304	36	1223

3. Resultados

3.6 Evaluación – Matriz de confusión



Recomendaciones de riego

		Predicted		Σ
		N	S	
Actual	N	1192	9	1201
	S	8	14	22
Σ		1200	23	1223



Evapotranspiración y necesidad hídrica

		Predicted		Σ
		N	S	
Actual	N	1153	33	1186
	S	33	4	37
Σ		1186	37	1223



Humedad del suelo

		Predicted			Σ
		Saturado	no regar	regar	
Actual	Saturado	883	0	0	883
	no regar	0	303	0	303
	regar	0	1	36	37
Σ		883	304	36	1223

3. Resultados

3.6 Evaluación – Matriz de confusión



Recomendaciones de riego

Predicted

		Predicted		
		N	S	Σ
Actual	N	1192	9	1201
	S	8	14	22
Σ		1200	23	1223



Evapotranspiración y necesidad hídrica

Predicted

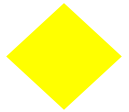
		Predicted		
		N	S	Σ
Actual	N	1153	33	1186
	S	33	4	37
Σ		1186	37	1223



Humedad del suelo

Predicted

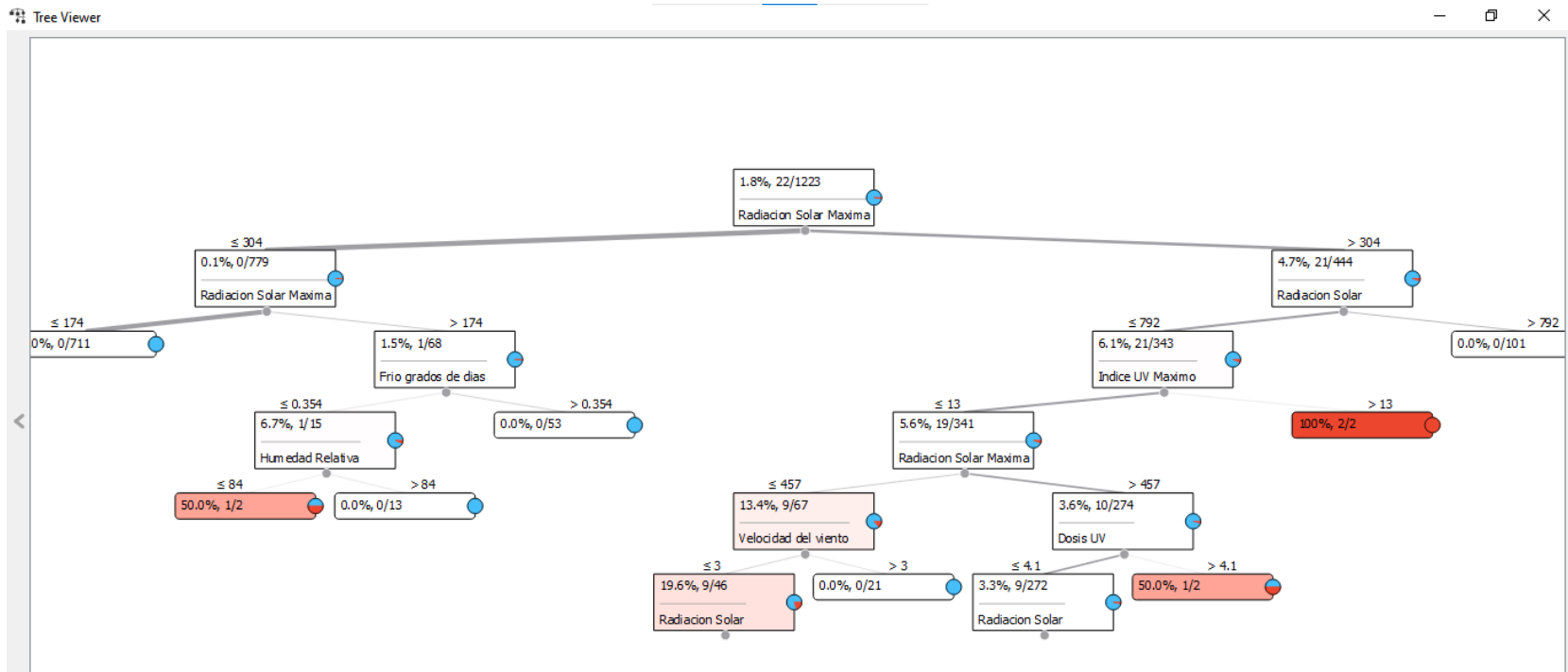
		Predicted			
		Saturado	no regar	regar	Σ
Actual	Saturado	883	0	0	883
	no regar	0	303	0	303
	regar	0	1	36	37
Σ		883	304	36	1223

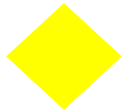


3. Resultados

3.7 Evaluación – Árboles de decisión

Gráfico 3.4 Ejemplo de árbol de decisión obtenido

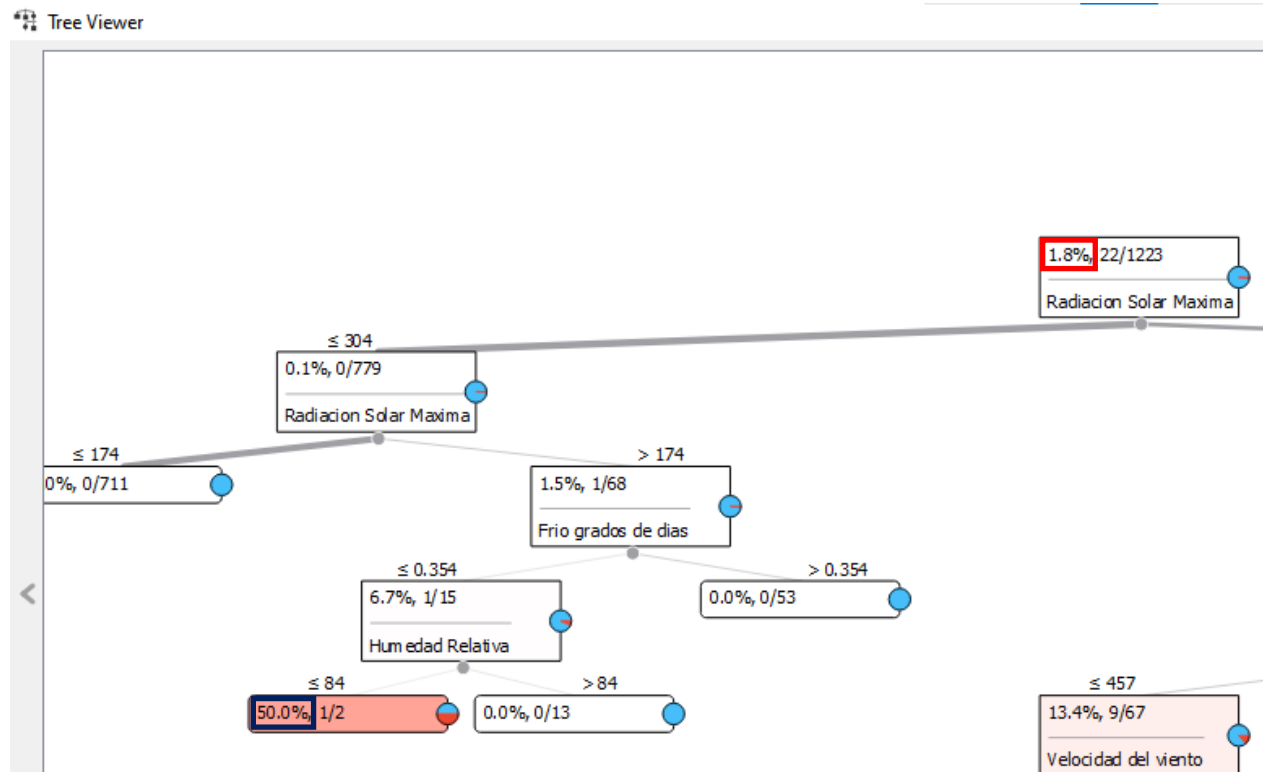




3. Resultados

3.7 Evaluación – Árboles de decisión

Gráfico 3.4 Ejemplo de árbol de decisión obtenido



3. Resultados

3.9 Variables primordiales



Recomendaciones de riego

Precisión: 98.7 %



Evapotranspiración y necesidad hídrica

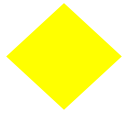
Precisión: 96.9%



Humedad del suelo

Precisión: 99.5 %

- Evapotranspiración
- Radiación Solar
- Humedad Foliar
- Humedad Relativa
- Velocidad del viento
- Humedad del suelo
- Lluvia

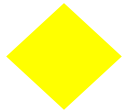


3. Resultados

3.10 Algoritmo

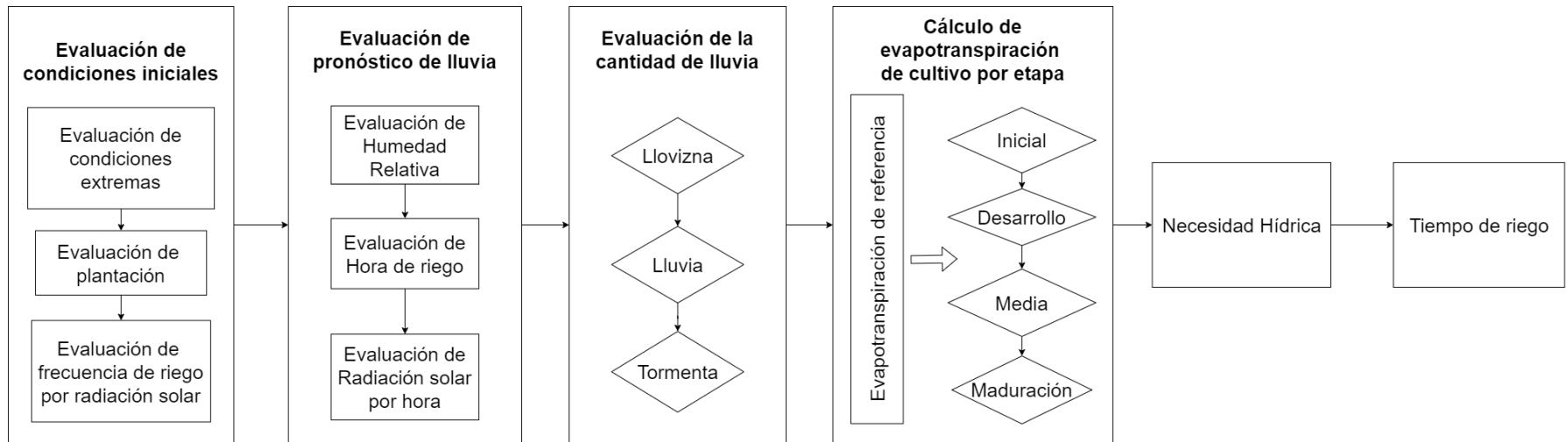


El algoritmo debe decidir el momento y periodo de riego en base a las variables primordiales



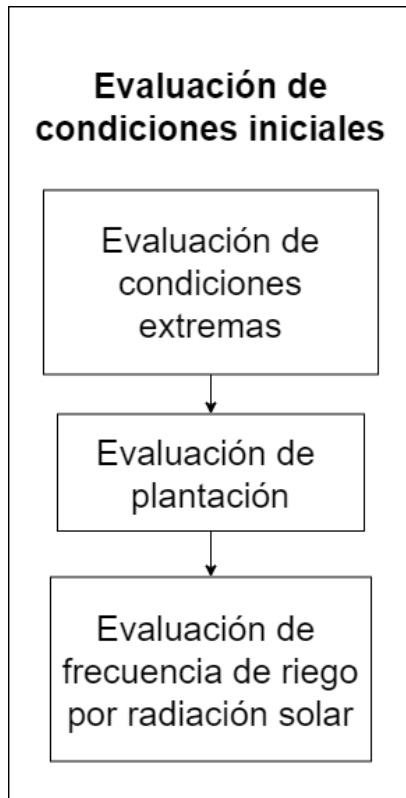
3. Resultados

3.10 Algoritmo



3. Resultados

3.10 Algoritmo

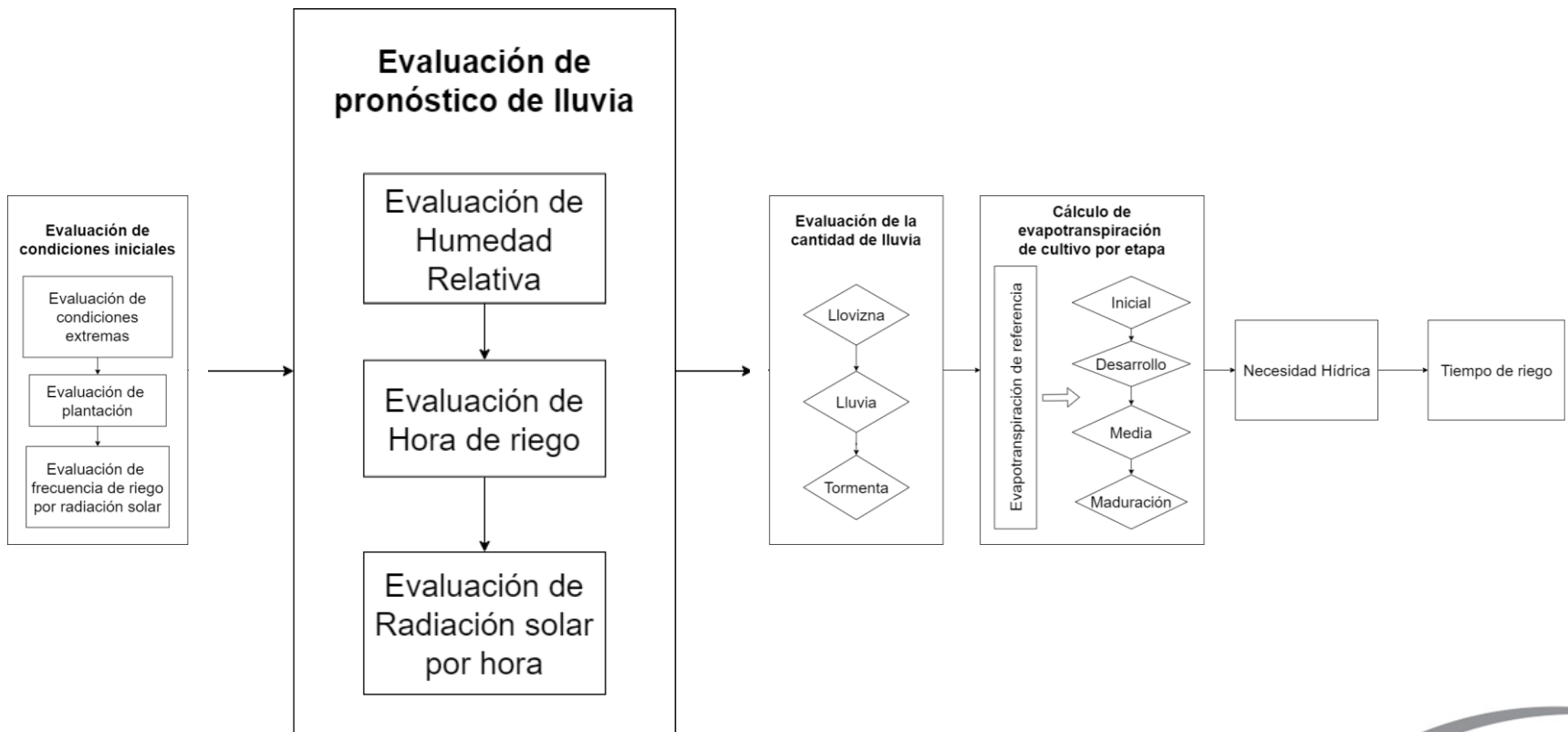


Cuadro 3.2 Condiciones extremas de riego según las características de humedad de suelo

Lectura en Centibar	Condición del suelo	Categoría	Acción	Humedad disponible
0 - 10	Suelo Saturado. Ocurre durante uno o dos días después del riego.	Saturado	No Regar	100%
60 - 100	Rango habitual para regar suelos arcillosos pesados	Muy seco	Regar	25%
100 - 200	Suelo extremadamente seco	Árido		0%

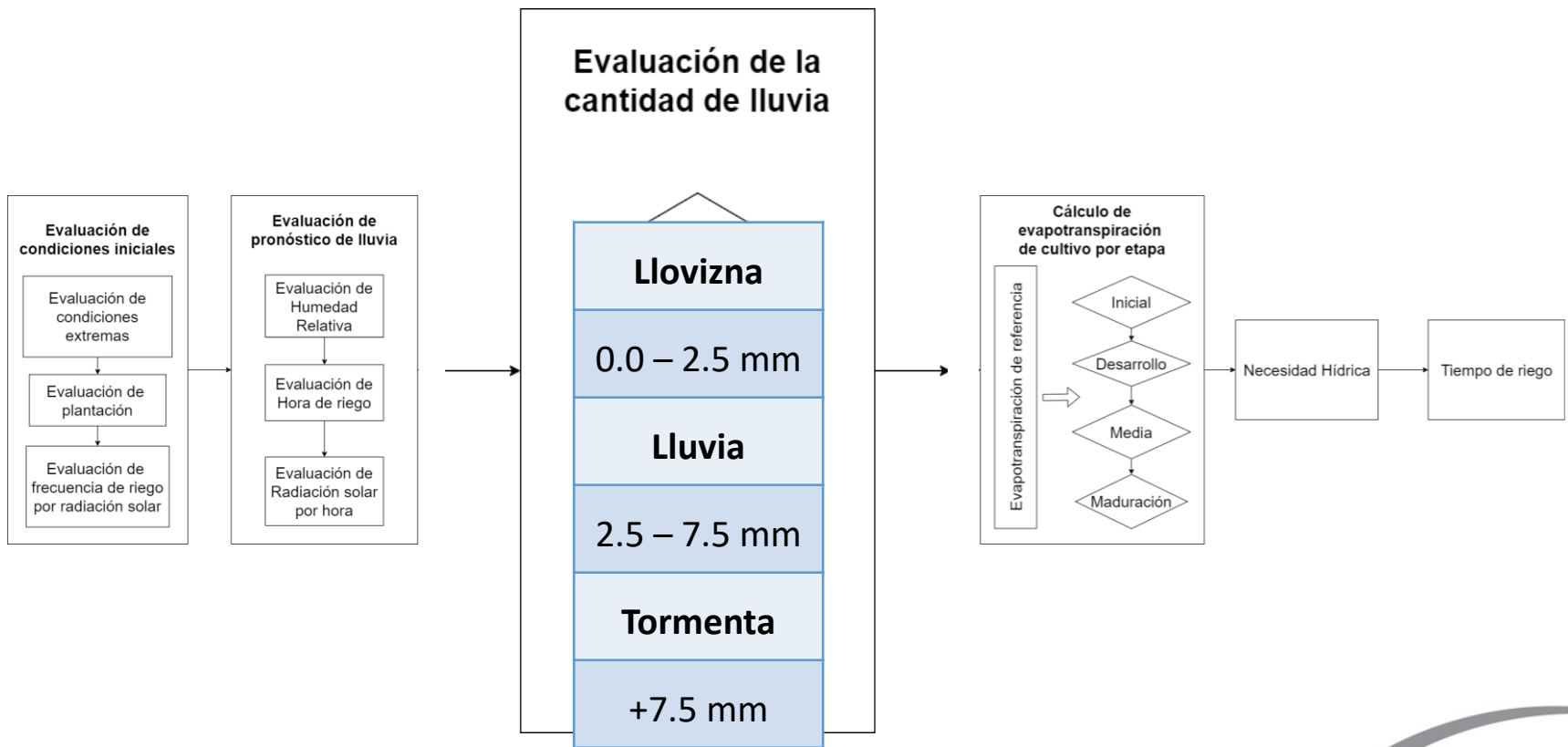
3. Resultados

3.10 Algoritmo



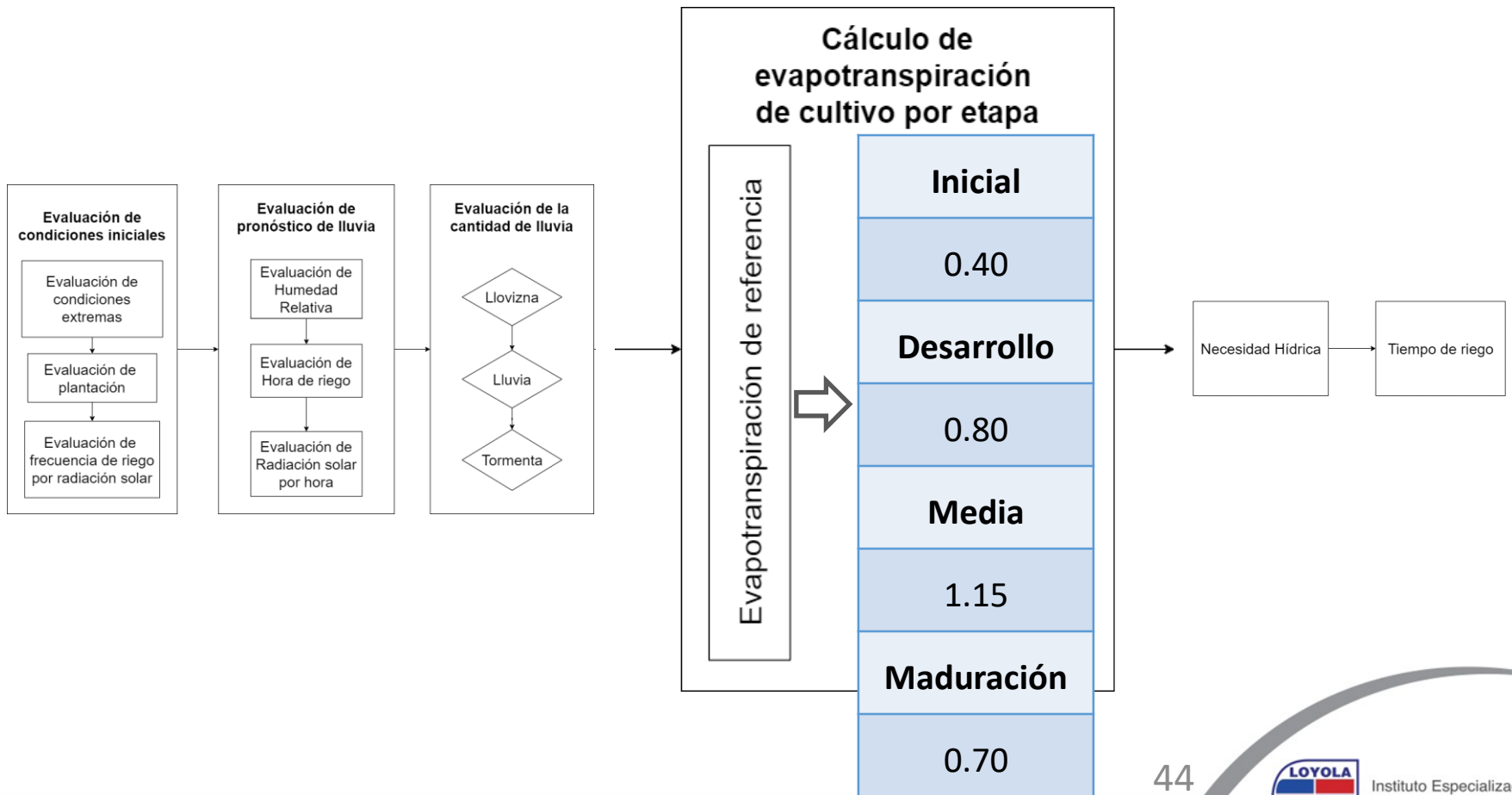
3. Resultados

3.10 Algoritmo



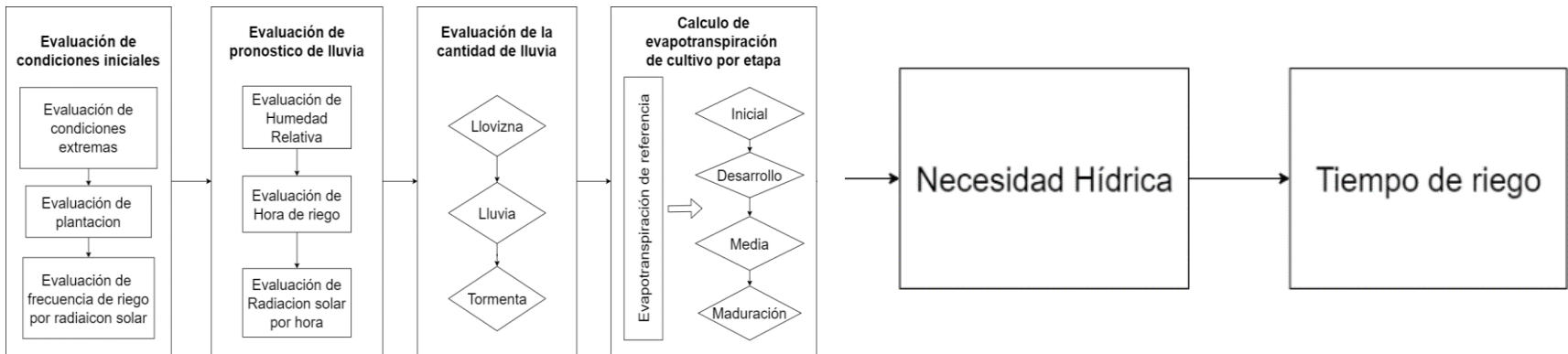
3. Resultados

3.10 Algoritmo



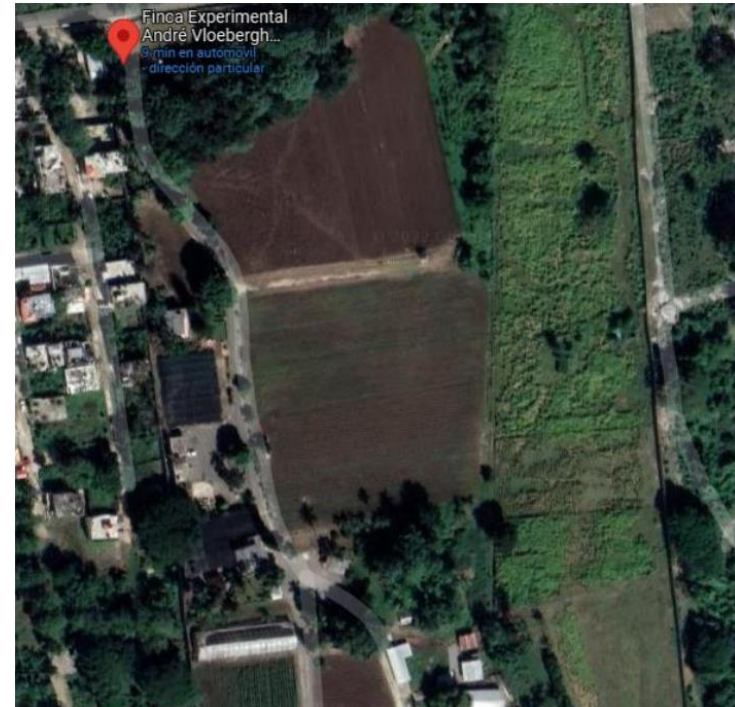
3. Resultados

3.10 Algoritmo



4. Limitaciones

La finca nunca ha realizado una recolección de datos edafoclimáticos orientada al análisis de datos, lo que obligó al desarrollo de una metodología, acompañada de la selección de herramientas adecuadas, para realizar una recolección de datos efectiva.



4. Limitaciones

Los eventos climáticos desastrosos (tormentas fuertes), registrados durante el transcurso de la investigación provocaron diferentes interrupciones en la captura de datos.



4. Limitaciones

La muestra utilizada se encontró limitada debido a que la estación meteorológica Vantage Pro2 Plus solo incluye cuatro pares de sensores de humedad y temperatura de suelo.

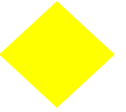


4. Limitaciones

El rango máximo disponible para la comunicación entre las estaciones y los concentradores no toma en cuenta fenómenos naturales e interferencia, por lo que la falta de recepción provocó pérdida de datos de manera ocasional.



4. Limitaciones



El horario de la finca representó un impedimento para la toma diaria de fotografías para el análisis del crecimiento del cultivo. De la misma forma, la captura de imágenes realizadas en diferentes horarios causó una diferencia en la relación de crecimiento diaria.



5. Conclusiones



La Estación Meteorológica Pro2 Plus es la herramienta ideal para la recolección de datos, permitiendo obtener datos efectivos para su utilización en procesos de análisis de datos

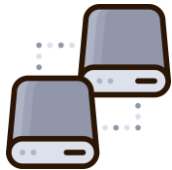


Orange Data Mining es la mejor herramienta para el análisis de los datos, permitiendo la obtención de resultados útiles a partir de este.

5. Conclusiones



Para poder obtener resultados efectivos, es necesario recolectar datos de crecimiento sobre el cultivo en intervalos de entre 15 min a 1 hora, salvo contadas excepciones.



Es necesario contar con un segundo conjunto de datos, sin importar la frecuencia (dentro de los rangos eficientes).



Es necesario asegurar respaldos de energía para los diferentes equipos de recolección de datos y confirmar la recepción entre los mismos.

5. Conclusiones



La utilización de análisis de imágenes como metodología para la recolección de datos de crecimiento sobre el cultivo resulta efectiva.



Es necesario la captura continua de imágenes de manera precisa durante su desarrollo. Esto para poder modelar curva del crecimiento de la planta.

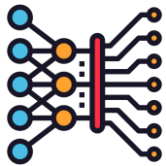


Se debe establecer una muestra más diversa tanto para la captura de imágenes como para la medición de humedad de suelo (variedad)

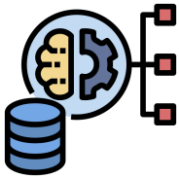
5. Conclusiones



El árbol de decisión y el bosque aleatorio funcionan de forma efectiva para el análisis de datos, ofreciendo predicciones muy precisas.



El bosque aleatorio ofrece mayor precisión de análisis que el árbol de decisión.



El **árbol de decisión** es el modelo de análisis más adecuado ofreciendo resultados con una precisión de **más del 90%** a la vez que permite evaluar la importancia de cada variable.

Gracias por su atención

55



www.ipl.edu.do



@politecnicoloyola



@loyolaipl



Instituto Especializado
de Estudios Superiores

LOYOLA