



Predicción de sacarosa utilizando Machine Learning



INGENIO PROVIDENCIA S.A.
Dulzura sostenible

Machine Learning

- El machine learning es una rama dentro del campo de la Inteligencia Artificial que proporciona a los sistemas la capacidad de aprender y mejorar de manera automática, a partir de la experiencia. Estos sistemas transforman los datos en información, y con esta información pueden tomar decisiones



Machine Learning



Supervisado

- Knn
- Árboles de decisión
- Random forest
- Regresión logística
- Support vector machine
- Árboles de regresión
- Modelos de ensamble

No supervisado

- Clustering
- Reducción de dimensiones
- Concurrencia



Antecedentes

Análisis estadísticos *realizados con Cenicaña*

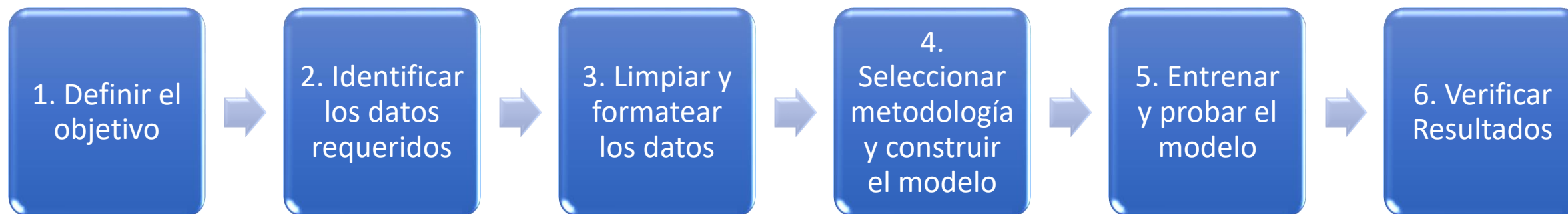
- *Definición de variables que pueden afectar el rendimiento*
- *Generación de una base de datos en **Siagri***
- *Análisis de **21 variables***
- ***24,000 datos analizados** en un periodo de 11 años (2007-2018)*



ANALISIS DE VARIANZA Y COMPARACION DE PROMEDIOS PARA RENDIMIENTO AÑOS 2007 -2018

Variable	F Value	Pr > F
Madurante	718.23	<.0001
Tipo Corte	576.53	<.0001
Variedad	268.08	<.0001
Cultivo	263.69	<.0001
Materia Extraña	252.19	<.0001
Precipitación 9 meses a cosecha	225.45	<.0001
Precipitación 6 a 9 meses	133.45	<.0001
Edad	98.13	<.0001
Distancia	89.49	<.0001
Tipo de Quema	75.9	<.0001
Precipitación de 0 a 3 meses	75	<.0001
Semmad	48.15	<.0001
Precipitación 2 Meses antes de cosecha	37.68	<.0001
Precipitación total	22.89	<.0001
Tenencia	15.44	<.0001
Agua total x Ha	8.59	0.0002
Dias dd Ult Riego	6.91	0.001
Fosfatos en Jugo	6.13	0.0022
Precipiatción de 3 a 6 meses	5.42	0.0044
Diatraea	0.63	0.5345
Vinaza aplicada	0.15	0.8639

Pasos del proyecto





Paso 1. Objetivos específicos

- Obtener un modelo para las **variedades objetivo**, por medio del uso de algoritmos que permitan aumentar la probabilidad de obtener sacarosa, mayor a la media histórica de cada una de ellas.

Variedad	Variedad Asociada	Promedio histórico de sacarosa	Ganancia de sacarosa esperada	Meta
CC01-1940	CC05-430	12,7%	0,3%	13,0%
CC85-92	CC93-4418	13,0%	0,3%	13,3%

- Predecir la sacarosa de las cosechas futuras con un grado de efectividad aceptable.



Paso 1. Objetivos específicos

- **Generar una herramienta de fácil uso**, que le permita a la **Gerencia de Cosecha** visualizar el pronóstico de sacarosa de las suertes analizadas, para optimizar la toma de decisiones en la generación del programa.
- Obtener una hoja de ruta de **buenas prácticas** que impacten positivamente el resultado final de sacarosa.





Paso 2. Identificar los datos requeridos

- **Bases de datos (SIAGRI)**

Histórico de cosecha de los últimos 6 años

- **Criterio de selección**

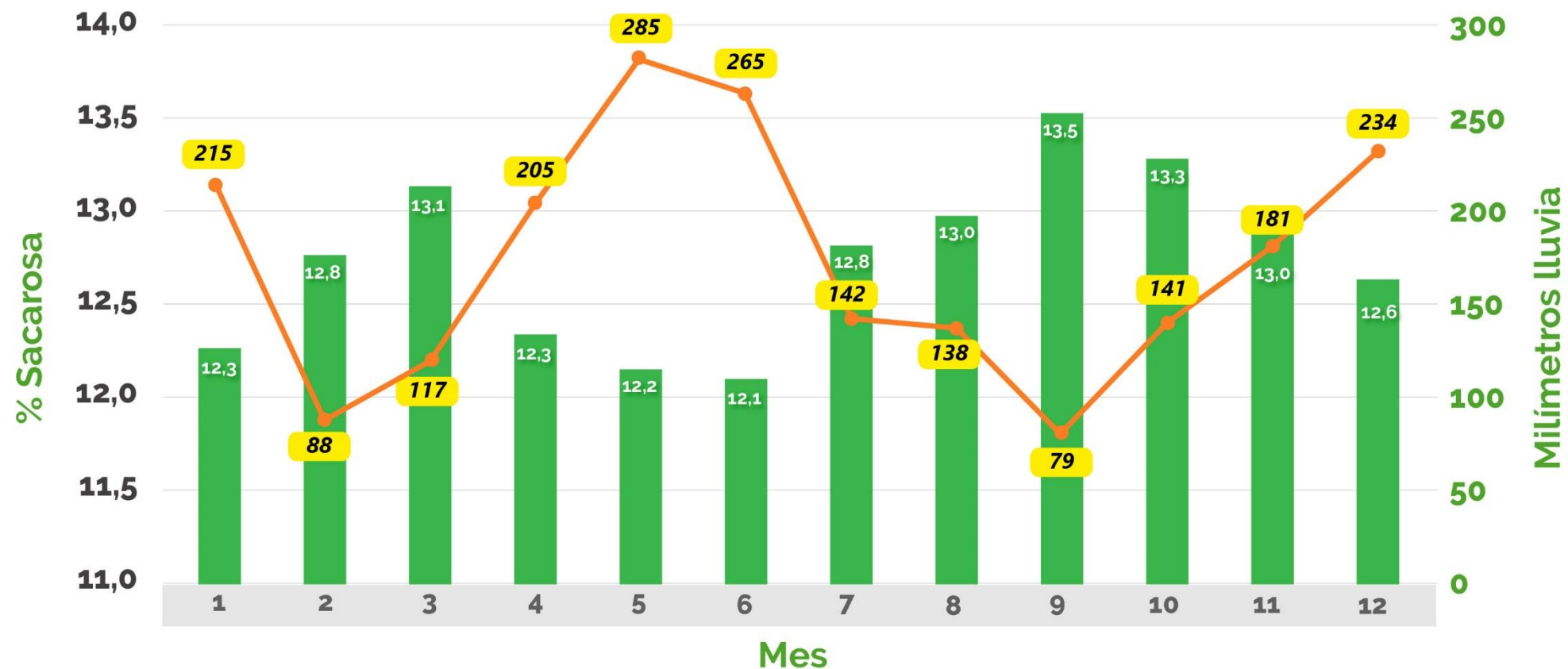
Variables que pudieran ser controladas

Variables seleccionadas

1. Variedad
2. Tipo de madurante
3. Tipo de cosecha
4. Dosis de madurante
5. Semanas de madurante
6. Edad de cosecha
7. Materia extraña total
8. Vejez
9. % Infestación de **Diatraea**
10. Precipitación (monitoreada)

Sacarosa Vs. Precipitación

Variedad CC 01-1940



Promedio de sacarosa Acumulación lluvias a partir de 10 meses



Paso 3. Limpiar y formatear los datos

- Selección de las variedades objetivo **maduradas con Bonus 250 y cosechadas mecánicamente** en verde.
- Transformación de la variable objetivo (**Sacarosa**) a categórica, de acuerdo a la media histórica.

Variedad	Categoría		
CC01-1940	> =	13,0%	Alta
CC05-430			
CC85-92	> =	13,3%	Alta
CC93-4418			

Variedad	Categoría		
CC01-1940	<	13,0%	Baja
CC05-430			
CC85-92	<	13,3%	Baja
CC93-4418			



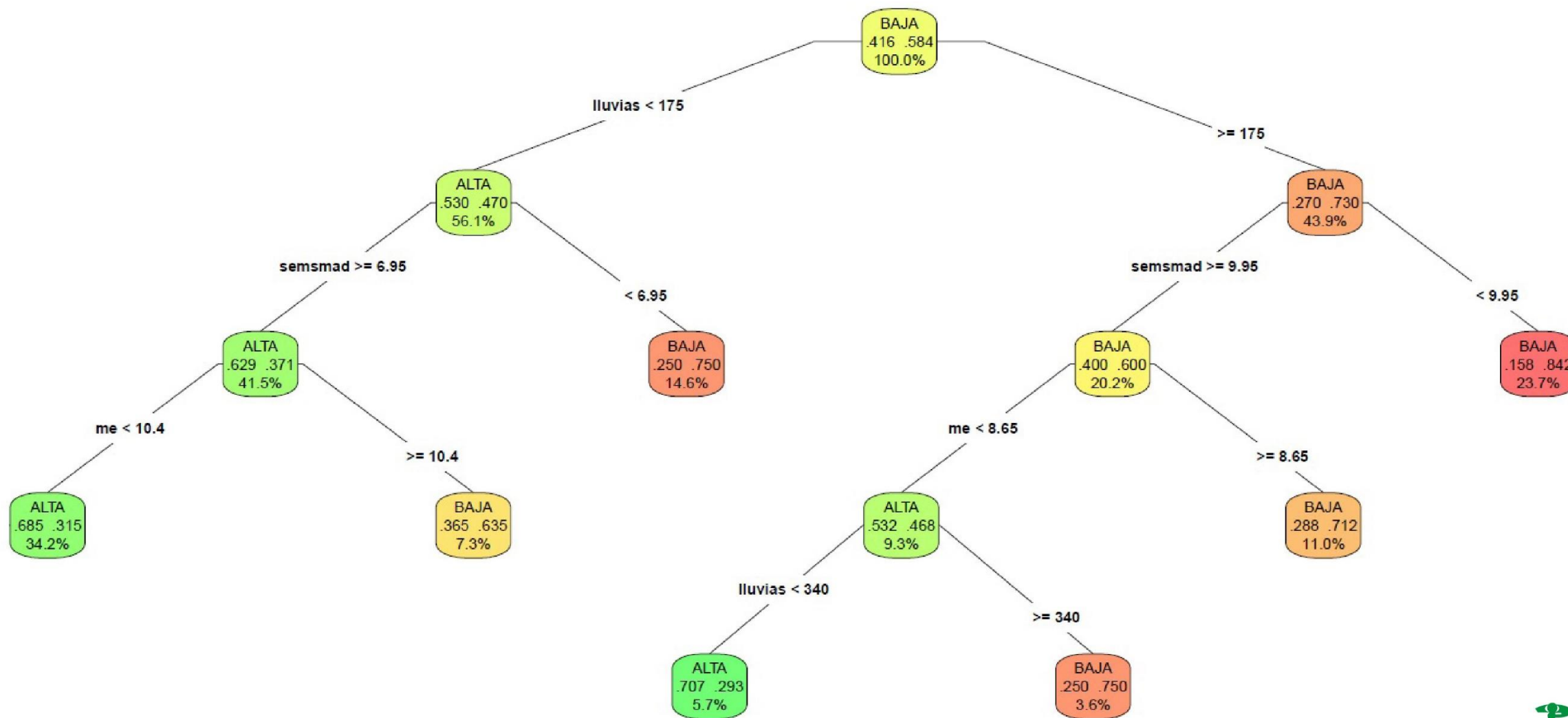
Paso 4. Seleccionar metodología y construir el modelo

- ***Algoritmos de clasificación***
Árboles de decisión



Modelo CC01-1940 bonus (13)

Efectividad media: 68,19% Efectividad mínima: 61,9% Efectividad máxima: 73,55%



Conclusiones

1. Se tienen dos modelos que proponen un **mejoramiento de la probabilidad de obtener mejor sacarosa** con respecto a la media.
2. **Las lluvias acumuladas a partir de los 10 meses** hasta la fecha de cosecha, explican en gran medida el resultado final de la sacarosa.
3. Se encontraron **variables que pueden ser controladas** para mitigar los efectos adversos de la lluvia en la sacarosa:
 - Semanas de maduración
 - Materia extraña
 - Porcentaje de *Diatraea*
 - Dosis de madurante

Fases del proyecto



MM ULT. 3 DIAS

☐ Seleccionar todo

☐ Central Ipsa

☐ Centro Norte

500,00

☐ BONUS 250 EC REG...

☐ 1 - Normal

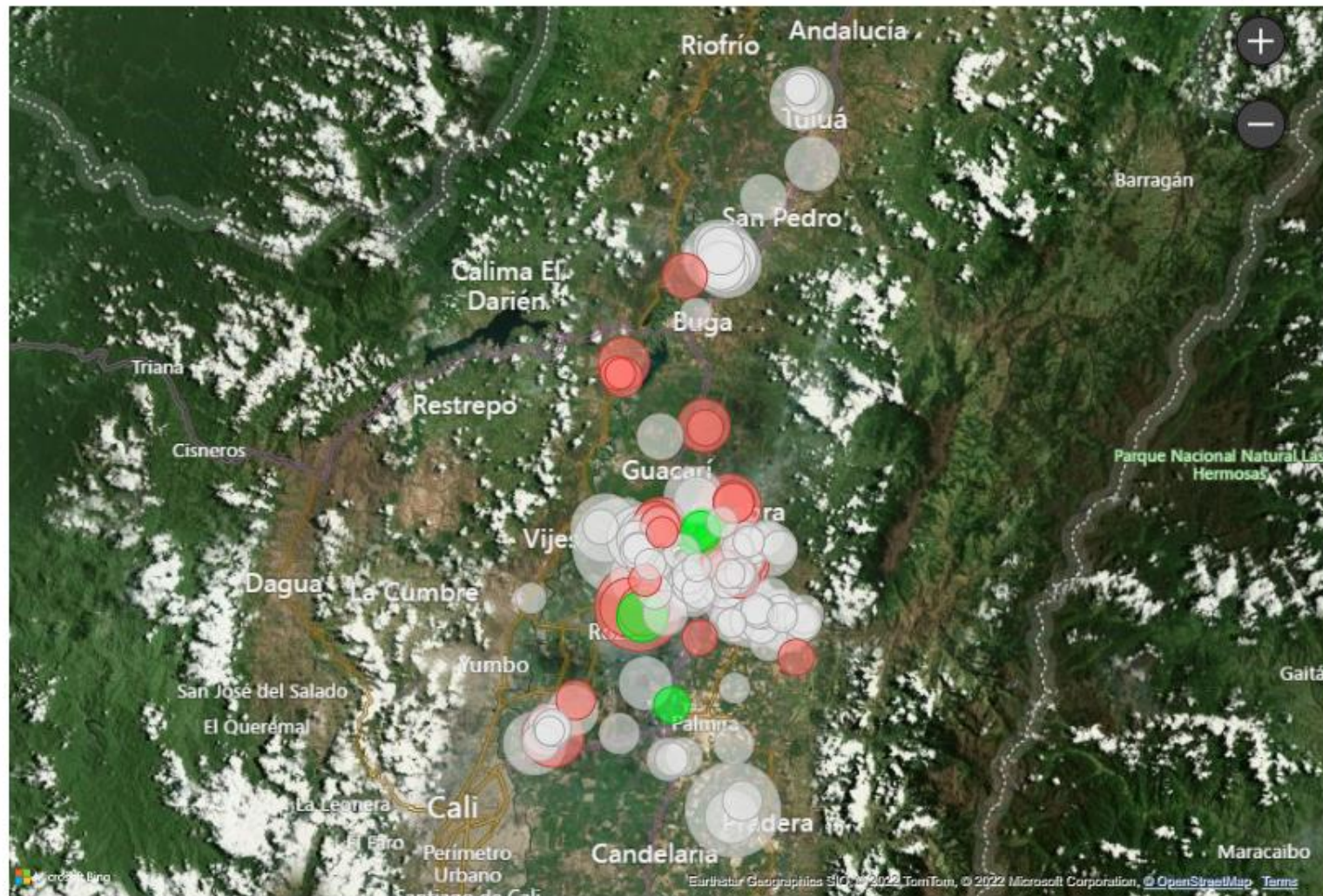
☐ 2 - Organico

☐ PLANTILLA

☐ SOCA

 Buscar

- ☐ BELEN_4
- ☐ BETANIA
- ☐ BRITANIA
- ☐ CANTACLARO
- ☐ CANTALAMOTA
- ☐ CASABLANCA
- ☐ EL_ALBION
- ☐ EL_AMPARO_GARCIA
- ☐ EL_BRILLANTE_BOR...
- ☐ EL_CANELO_M
- ☐ EL_ESPEJO
- ☐ EL_GUACHAL



162.591

☐ -
☐ ALTA
☐ BAJA

122.253,00
TONELADAS

235,03
LLUVIAS ACUMULAD...

1,35
SEMS MADURACION

BAJA

35.795,00
TONELADAS

361,72
LLUVIAS ACUMULAD...

840

Filtros

Buscar

Filtros de este objeto visual ...

PC
es (todos)

Filtros de todas las páginas ...

sociedad
es AI08

[← Volver atrás](#)

≡ PREDICCIÓN

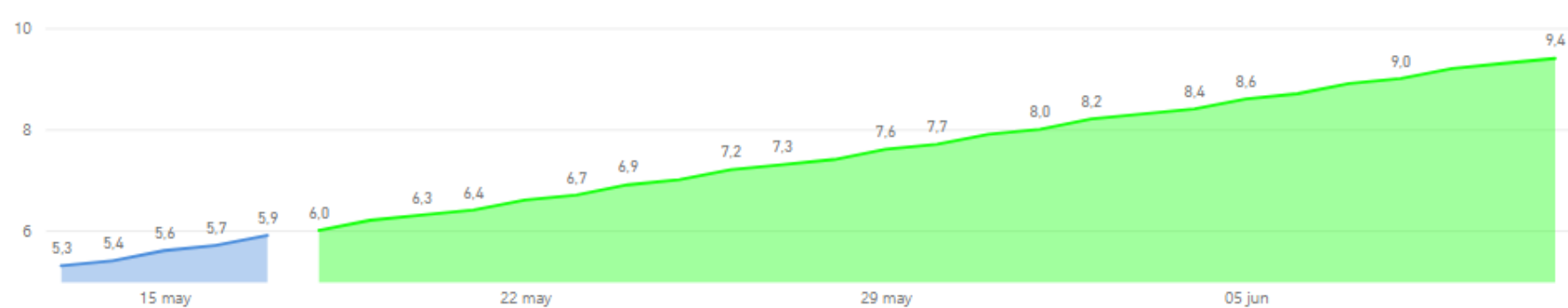
Programador de cosecha

91 % ☐

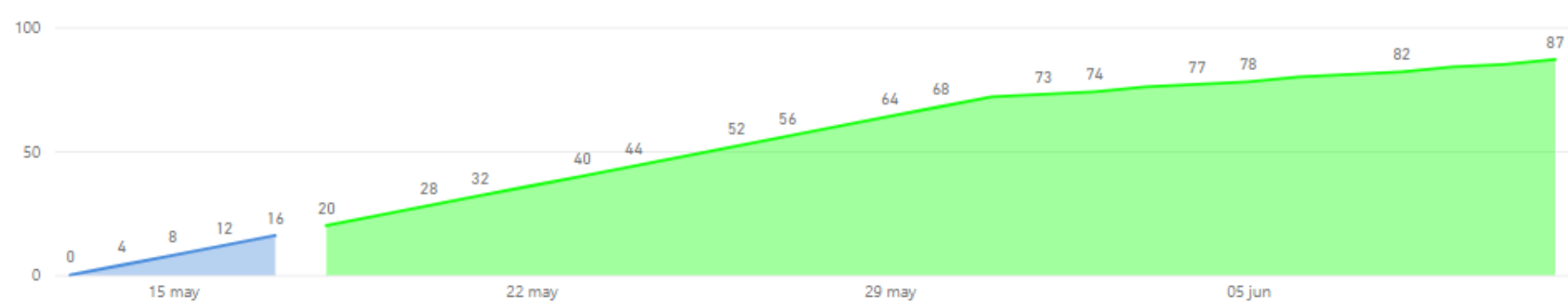
HACIENDA

- ☐ AURORA_CUCAL...
- ☐ BETANIA
- ☐ CANTACLARO
- ☐ CASABLANCA
- ☐ CHAMBURO
- ☐ EL_ANTOJO_SA...
- ☐ EL_BRILLANTE_...
- ☐ EL_ESPEJO
- ☐ EL_FARO_EMPRE...
- ☐ EL_GUABITO_EM...
- ☐ EL_HIGUERON_E...
- ☐ EL_JAGUAL
- ☐ EL_JARDIN
- ☐ EL_LIBANO_5
- ☐ EL_MANGON_D...
- ☐ EL_PALMAR_PAY...
- ☐ EL_PRADO_FERRO
- ☐ EL_RANCHO_EM...
- ☐ EL_REFUGIO_LIN...
- ☐ EL_SAMAN
- ☐ EL_TESORITO
- ☐ EL_TESORO
- ☐ EL_TREJO_2
- ☐ JAMAICA_ABADIA
- ☐ JERICO
- ☐ JOSEPILLA_II
- ☐ LA_ALDEA_CABAL
- ☐ LA_ARGELIA
- ☐ LA_CUCUITA

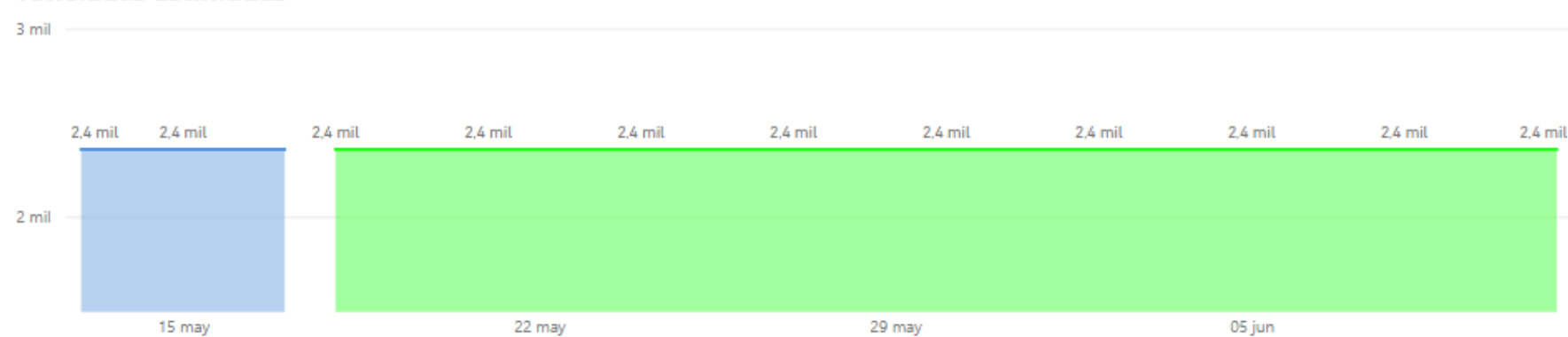
Semanas de Madurante



Lluvias



Toneladas Estimadas



Filtros

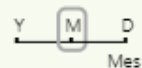
Buscar

Filtros de esta página

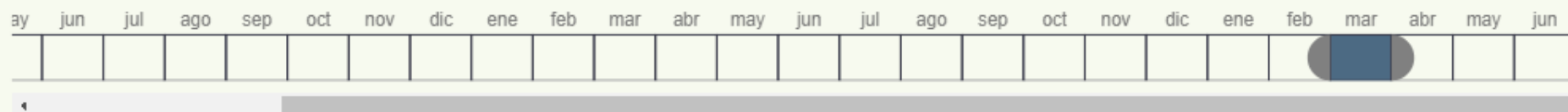
PC
es ALTA o BAJA

Filtros de todas las páginas

sociedad
es AI08



mar 2022 - mar 2022



AI01

AI08

☐ -
☐ ACIERTO
☐ ERROR

EDAD
14,09

Mex TOTAL
9,39

SDDA
9,20

SACAROSA
12,40

EFFECTIVIDAD
72,2%

HACIENDA

- ☐ AMAIME SILCA
- ☐ ARANJUEZ
- ☐ BARCELONA
- ☐ BELLA FLOR ASCENSION
- ☐ BELLA FLOR SODIEGO
- ☐ BERLIN TENORIO
- ☐ BETANIA
- ☐ CHAMBERY 2
- ☐ EL ALBION
- ☐ EL AMPARO SAA
- ☐ EL ARRANQUE
- ☐ EL BOLO MORROCO
- ☐ EL BRASIL TENORIO
- ☐ EL CARMEN EMPRESA
- ☐ EL CARMEN MEJIA
- ☐ EL CARMEN VELASCO
- ☐ EL DELIRIO PARRA
- ☐ EL GUACHAL
- ☐ EL GUAYABITO
- ☐ EL JAGUAL
- ☐ EL JAPON

MBRE	STE	VARIEDAD	EDAD	%Me	SDDA	%DTR	MM	3	7	15	TPP	TONS	%SAC	TC	ACC	REAL	PRED	R^
ESTRELLITA ABADIA	001	CC85-92	13,5	10,7		5,3	159				4,3	222	13,47			-	-	-
ANI	001	CC01-1940	14,7	8,8	7,9	4,2	356				4,3	1.548	12,56			BAJA	BAJA	✓ A
IMON	002A	CC85-92	13,7	8,4		5,8	78				4,3	296	12,31			-	-	-
IDA MARIA	002A	CC05-430	13,5	7,5	5,6	2,7	222				4,2	281	12,38			BAJA	BAJA	✓ A
REFORMA AGRICOLA	050	VARIAS	11,7	9,4	8,4		0				4,2	950	11,88			-	-	-
DECHINCHE	049	CC01-1940	13,9	11,6	8,0	5,1	770				4,2	452	12,16			BAJA	BAJA	✓ A
EMILIA 3	002	CC01-1940	13,8	7,8	7,6	6,0	148				4,2	390	11,16			BAJA	ALTA	✗ E
AURORA	120B	CC09-535	13,0	8,6	3,1	2,3	204				4,2	922	11,23			-	-	-
ESMERALDA	419	CC01-1940	12,6	11,4		3,4	159				4,2	32	9,21			-	-	-
RASIL TENORIO	001	CC01-1940	13,6	7,4	7,9	4,9	230				4,1	502	12,82			BAJA	BAJA	✓ A
RAIMA	006	CC01-1940	13,2	10,6		4,5	0				4,1	780	11,76			-	-	-
ILENCIO CRUZ	004	CC01-1940	14,0	8,9	10,9	3,3	0				4,1	244	13,81			ALTA	ALTA	✓ A
RAIMA	001B	CC01-1940	13,2	13,8		3,4	0				4,1	483	11,31			-	-	-
MARIA EL ROBLE	003	CC05-430	15,1	8,3		1,9					4,1	116	13,15			-	-	-
RSELLA	525	CC05-430	13,8	8,2	7,6	4,9	191				4,1	266	11,81			BAJA	ALTA	✗ E
BETULIA	006	CC05-430	13,4	13,4	8,3	5,8					4,1	381	13,86			-	-	-
ITA ELENA	018	CC01-1940	14,1	11,3	12,7	0,8	185				4,0	554	13,96			ALTA	ALTA	✓ A

← Volver atrás

≡ DIARIO



Diario_de_Caña

"La innovación constante

es la única forma de mantenerse competitivo, porque ninguna ventaja es sostenible en el largo plazo"

Jorge Gonzalez Moore