

**Aplicación de Tecnologías de la
Información y Comunicaciones
Avanzadas y Accesibilidad
ATICA2021**

OBRAS COLECTIVAS
TECNOLOGÍA 34

Luis Bengochea
António M. Teixeira
José Ramón Hilera
(Editores)

UAH

Ciencia de datos aplicada a la agricultura. Un estudio del comportamiento del cultivo de la mostaza blanca (*Sinapsis alba L.*) en la Patagonia Argentina

Melisa Isaja¹, Paola Pizzingrilli^{1,2,6}, Pamela Britos¹, Maximiliano Donadio², Giuliana Fois², Gustavo Agüero², Pablo Enrique Argañaraz^{2,5}, Martín René Vilugrón^{2,5}, Lina María Montoya Suárez⁴, Paola Britos², Gastón Di Bonis¹, Rodrigo Arce²

¹ Universidad Nacional de Río Negro. Río Negro. Argentina. melisa.isaja@gmail.com, gdibonis@unrn.edu.ar

² Universidad Nacional de Río Negro. Laboratorio de Informática Aplicada. Río Negro. Argentina. ppizzingrilli@unrn.edu.ar, maxdonadio21@gmail.com, gfois@unrn.edu.ar, gustavo54ar@gmail.com, pbritos@unrn.edu.ar, parganaras@unrn.edu.ar, mvilugron@unrn.edu.ar

³ Sede Regional Sur: Metán – Rosario de la Frontera, Universidad Nacional de Salta. Argentina. pamebritos@gmail.com

⁴ Universidad Católica Luis Amigó, Grupo de Investigación SISCO Medellín, Colombia lina.montoyasu@amigo.edu.co

⁵ Comisión Nacional de Energía Atómica, Centro Atómico Bariloche, banda@cab.cnea.gov.ar, martinvilu@cab.cnea.gov.ar

⁶ Universidad Nacional de Río Negro. Instituto de Investigaciones de Recursos Naturales, Agroecología y Desarrollo Rural (IRNAD). Río Negro. Argentina. ppizzingrilli@unrn.edu.ar

Resumen. En este artículo se pudo analizar las respuestas climáticas y fenológicas presentadas de los cultivares de mostaza Delfina y local, mediante el análisis del comportamiento, a través de algoritmos de inducción de rendimiento, en el paraje Las Golondrinas, Lago Puelo, formando parte en la Comarca Andina del Paralelo 42° S, territorio que abarca ambas provincias de Río Negro, y Chubut. Se pudo constatar que ambas variedades mostraron una respuesta al fotoperíodo, ya que han denotado un alto rendimiento en la siembra de otoño en comparación con la realizada durante la temporada de primavera.

Palabras clave: Cultivo, Agro, Mostaza, Ciencia de datos, Explotación de datos, Rendimiento, Ciencia de datos.

1. Introducción

1.1 Cultivo de la mostaza

La mostaza es una de las principales hierbas aromáticas del comercio mundial en volumen de las tres variedades actualmente conocidas. Este cultivo es originario del suroeste de Asia y Europa, siendo la mostaza blanca (*Sinapsis alba L.*) la variedad con mayor relevancia para fines productivos, posicionándose en el mercado canadiense de

granos con el 50% de la producción y/o exportación mundial, seguida por otras variedades [1][2]. En Argentina, si bien se cultiva en distintos puntos del país, la mayor producción se centraliza en el sudeste de la provincia de Buenos Aires, seguida por la provincia de Santa Fe y Entre Ríos, siendo considerada la zona más óptima para su desarrollo. En general, la mostaza es un cultivo alternativo de invierno que se extiende en las zonas templadas del planeta. Argentina es uno de los pocos países que tiene como ventaja comparativa, la importación de grano y la exportación de mostaza realizada para el abastecimiento del mercado externo ya que, tiene un alto índice de demanda del grano, que hasta el momento no puede ser cubierto por los cultivos de la zona. Para satisfacer esta demanda, el Instituto Nacional de Tecnología Agropecuaria (INTA) creó cultivares de mostaza nacionales y de calidad para ser más competitivos de acuerdo a los países líderes, con un rendimiento de más de 1000 Kg. ha⁻¹ mediante el cultivo de mostaza blanca y entre otras variedades [3] [4]. El ciclo de la mostaza se divide en tres etapas como son la vegetativa, la reproductiva y el llenado de granos. Con respecto a la primera etapa, se puede identificar el desarrollo y diferenciación de las hojas en el estado de roseta, la parte reproductiva comienza con el alargamiento del botón floral hasta que se produce la floración y producción de ramas florales, formando los siliquios (frutos) en los que se produce la última etapa de llenado de los granos donde se desarrollan las semillas. [1] [5] [6]. Por otro lado, cabe destacar que la mostaza es una planta de día largo, de clima templado-húmedo, resistente a las bajas o altas temperaturas según el estado fenológico del ciclo ontogénico en el que se encuentre. El manejo del calendario de siembra es importante ya que nos permite ajustar el ciclo de producción y evaluar cuáles son los puntos críticos en el rendimiento. Para ello, es necesario conocer el efecto de los factores ambientales sobre los cambios morfológicos y fisiológicos en cada una de sus etapas analizando la duración del día (fotoperiodo) y la temperatura óptima de crecimiento de la planta. [1] [6]. Para esta evaluación se analizaron los siguientes parámetros: altura en plena floración, número de ramas florales por planta, siliquios por planta y el peso de 1000 semillas. El ciclo ontogénico de la mostaza blanca, su época de siembra y las interacciones ecológicas que se producen en el cultivo permiten determinar si dichos parámetros condicionan su rendimiento final.

1.2. Ciencia de los datos

La ciencia de los datos es una disciplina dentro del campo de la Informática que permite el estudio de masas de información, está compuesta por diversas fuentes, según las citadas por EMC [7], en los últimos siete años generamos más información que en toda la historia de la humanidad, y la duplicamos cada dos años. Para el estudio de los datos, Britos [8] propone, entre otros, el descubrimiento de reglas de comportamiento como parte del análisis en este campo. Cada proceso tiene asociadas técnicas de minería de datos para descubrir patrones a través de su aplicación. Por otro lado, Fois, Agüero y Britos [9] realizan una comparativa de metodologías ágiles para aplicar en este tipo de proyectos, de las cuales se han tomado las actividades más inherentes para planificar cómo abordar el problema y organizar las actividades para abordar este caso de estudio. Como trabajos previos, podemos citar [10], que analiza los tipos de rendimiento en Argentina y [11] en el que se propone una metodología de modelización estadística

para el cultivo de la mostaza, en Pakistán, utilizando técnicas de modelización de regresión.

2. Problemas

En el paraje Las Golondrinas, perteneciente a Lago Puelo presenta el escenario de condiciones climáticas ideales para el desarrollo del cultivo de mostaza, es allí donde se evaluó el rendimiento de 2 tipos de cultivos de mostaza blanca (*Sinapis Alba L.*), el tiempo térmico y la descripción fenológica [10], siendo los cultivares: INTA-Delfina (Res. 67/12 INASE) y Local, que fue cedido por un productor de la zona. El objetivo de este trabajo es obtener datos que permitan determinar rangos óptimos de rendimiento para nuestra zona de mostaza variedad Delfina y Local, a través del proceso que determina el comportamiento propuesto en [8].

3. Resultados obtenidos y discusión

Los datos fueron tomados de [10] y en su primera versión se separaron en 3 hojas de cálculo (Fases fenológicas, Temperaturas medias y Rendimiento), después de seleccionar con los expertos los atributos más significativos, la tabla consta de 479 registros, con los siguientes campos (Tabla 1: Campos utilizados), en los que se pueden visualizar 241 registros del cultivar de mostaza tipo Delfina, y 238 registros del cultivar de mostaza local.

Tabla 1. Campos utilizados

<ul style="list-style-type: none"> • Siembra • Hojas de roseta sin botón • Cultivar • Hojas de roseta Max-Temp sin botón • F. Estación de siembra • Hojas de roseta Min-Temp sin botón • Estación de germinación • Roseta de hojas de temperatura media sin botón • Estación True Leaves • Altura Start-Repro-St Cm • Estación de Raleo • Repro-St de inicio de Max-Temp • Estación de salida de la roseta • Inicio de minitemperatura Repro-St • Temporada de botones florales • Repro-St de arranque a media temperatura • Estación de alargamiento • Altura total Floración Cm • Alargamiento en la floración • Temp- Max Full Bloom • Estación de juego de frutas 	<ul style="list-style-type: none"> • Temp-Min Full Bloom • Estación de la cosecha • Temperatura media de floración • Días entre la siembra de F. y la germinación • Rf In Full Bloom • Días entre la germinación y las hojas verdaderas • Temp- Max Rf Full Bloom • Días entre las hojas verdaderas y el adelgazamiento • Temp-Medium Rf Full Bloom • Días entre el aclareo y la hoja de roseta • Temp-Min Rf Full Bloom • Días entre la roseta y el botón floral • Altura de fructificación • Días entre el botón floral y el alargamiento • Fructificación Temp-Max • Días entre la elongación y la elongación en flor • Fructificación Temp-Min • Días entre el alargamiento de la floración y el cuajado de los frutos • Fructificación Temp-Medio • Días entre el cuajado y la cosecha • Rendimiento Kg/Ha
--	--

Para realizar un mejor análisis, se decidió discretizar el Atributo Rendimiento Kg. Ha-1 en 5 intervalos de igual frecuencia ($< 637,5$, $637,5 - 1600$, entre $1600 - 2412,5$;

entre 2412,5 - 3087,5 y mayor de 3087,5), y transformar el atributo siembra de nominal a categórico ya que el mismo indica la identificación de la siembra y no una cantidad en sí. A continuación, en la Figura 1: Relación entre Cultivo - Siembra y Rendimiento se ven los distintos tipos de segmentos de rendimiento y su relación con el momento de la siembra y el tipo de cultivo.

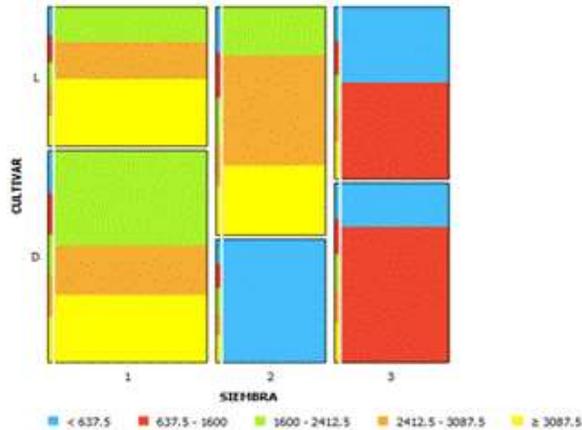


Fig. 1. Relación entre cultivo - siembra y rendimiento

En este gráfico se pueden observar, los tipos de cultivo para cada siembra por rango de rendimiento evaluado. En ella se observa una disminución significativa del rendimiento a medida que avanza la siembra, denotando el menor valor en la siembra 2 para el cultivar Delfina y los mayores valores de rendimiento en la siembra 1 para ambos cultivares. Los diferentes rangos de rendimiento que se dieron en cada siembra y cultivar se debe a la influencia de la época de siembra en las fases fenológicas y componentes de rendimiento de ambos cultivares. En este gráfico se puede observar, los tipos de cultivo para cada siembra por rango de rendimiento evaluado. En ella se observa una disminución significativa del rendimiento a medida que avanza la siembra, denotando el menor valor en la siembra 2 para el cultivar Delfina y los mayores valores de rendimiento en la siembra 1 para ambos cultivares. Los diferentes rangos de rendimiento que se presentaron en cada siembra y cultivar se debe a la influencia de la época de siembra en las fases fenológicas y componentes del rendimiento de ambos cultivares. Posteriormente se realizaron diferentes tipos de análisis con diversos parámetros. En las siguientes tablas se muestran las comparaciones de estos parámetros, a través de la matriz de confusión. El modelo de comparación se realizó con validación cruzada estratificada de 10 veces, siendo el objetivo de cada tipo de mostaza determinar el más eficiente. (Tabla 2 y 3 para la Mostaza Delfina, Tabla 4 y 5 para la Mostaza Local).

Tabla 2. Mostaza Delfina

Modelo	Precisión	Recall
Naive Bayes	0.7488633857269136	0.7053941908713693
CN2 inductor de reglas	0.7028517862685415	0.7053941908713693

Tabla 3. Matriz de confusión para las diversas técnicas (mostrando el número de instancias) para Mostaza Delfina

		Predicción					
		637.5 - 1600	1600 - 2412.5	2412.5 - 3087.5	< 637.5	≥ 3087.5	Σ
Naive Bayes	637.5 - 1600	55	0	0	0	0	55
	1600 - 2412.5	0	27	18	0	9	54
	2412.5 - 3087.5	0	6	14	0	8	28
	< 637.5	18	0	0	48	0	66
	≥ 3087.5	0	3	9	0	26	38
	Σ	73	36	41	48	43	241
CN2	637.5 - 1600	43	0	0	12	0	55
	1600 - 2412.5	0	35	11	0	8	54
	2412.5 - 3087.5	0	16	9	0	3	28
	< 637.5	8	0	0	58	0	66
	≥ 3087.5	0	10	3	0	25	38
	Σ	51	61	23	70	36	241

Tabla 4. Mostaza local

Modelo	Precisión	Recall
Naive Bayes	0.42777556290455570	0.5042016806722689
CN2 inductor de reglas	0.44788901862114716	0.4537815126050420

Tabla 6. Matriz de confusión para las diversas técnicas (mostrando el número de instancias) para la Mostaza Local

		Predicción					
		637.5 - 1600	1600 - 2412.5	2412.5 - 3087.5	< 637.5	≥ 3087.5	Σ
C2N	637.5 - 1600	18	0	0	21	0	39
	1600 - 2412.5	0	9	12	0	18	39
	2412.5 - 3087.5	0	7	41	0	16	64
	< 637.5	17	0	0	14	0	31
	≥ 3087.5	0	19	20	0	26	65
	Σ	35	35	73	35	60	238
Naive Bayes	637.5 - 1600	26	0	0	13	0	39
	1600 - 2412.5	0	0	19	0	20	39
	2412.5 - 3087.5	0	1	43	0	20	64
	< 637.5	17	0	0	14	0	31
	≥ 3087.5	0	1	27	0	37	65
	Σ	43	2	89	27	77	238

Según el análisis presentado anteriormente, se observa la falta de precisión en estos modelos. Los expertos mencionan que el comportamiento del ciclo fisiológico en la mostaza y las fases que configuran el rendimiento, son claves para determinar qué cultivar se comporta mejor en el ambiente estudiado. Por lo anterior, se investigaron las reglas de decisión para cada tipo de mostaza: A continuación, se presentan las reglas obtenidas para el tipo de mostaza Delfín, así como el análisis de las mismas:

Rendimiento < 637,5

Temporada de botones florales = Primavera
| Siembra = 2: <637,5

Rendimiento entre 1600 - 2412,5

Temporada de botones florales = Invierno
| Altura Star-Repro-St Cm \leq 31
|| Altura de plena floración Cm \leq 94,5 :1600-2412

Rendimiento entre 2412,5 - 3087,5

No hay resultados.

Rendimiento entre 637,5 – 1600

Temporada de botones florales = Primavera
| Siembra = 3: 637,5 - 1600

Rendimiento \geq 3087,5

Temporada de botones florales = Invierno
| Altura Star-Repro Cm $>$ 31: \geq 3087,5
| Altura Star-Repro Cm \leq 31
|| Altura de floración Cm $>$ 94.500: \geq 3087,5

Análisis: La mostaza Delfina mostró diferentes comportamientos en cada una de las siembras y los rangos evaluados. Los menores rendimientos correspondientes al rango < 637,5 kg. ha-1 que se observaron en la Figura 1 y en las normas se presentan en la siembra 2, ya que el cultivo se sembró a finales del otoño 1, presentando un retraso de 63 días en la germinación, como se refiere en [10] con el consiguiente acortamiento de la etapa vegetativa, lo que condiciona su rendimiento a diferencia de lo observado en la mostaza Local. Este retraso puede haberse producido por una susceptibilidad a las bajas temperaturas para la germinación y también por una menor adaptación de Delfina a las condiciones climáticas de la zona donde se realizó el ensayo. Teniendo en cuenta que la mostaza es una planta fisiológicamente de día largo, es de esperar que el rendimiento no sea óptimo a medida que avanza hacia la mitad de la primavera como se observa en la Figura 1 para la siembra 3 Este retraso en la siembra hace que el crecimiento y desarrollo se acelere y provoque un acortamiento de su ciclo productivo, con la consiguiente disminución de su potencial productivo. Sin embargo, se observa que los rangos de mayor rendimiento se obtuvieron en la siembra 1, realizada al inicio de la temporada de otoño [10], permitiendo una mayor amplitud de la etapa vegetativa y reproductiva acompañada del incremento de la biomasa que se traduce en un aumento del rendimiento del cultivar. Este comportamiento se refleja en las reglas obtenidas, donde la altura al inicio de la etapa reproductiva y la altura en plena floración, son indicadores de su productividad. Esta situación también se puede observar en el gráfico 1, donde los mayores rendimientos se presentan en la siembra 1 para dicho cultivar.

A continuación, se presentan las reglas obtenidas para la Mostaza Local y la discusión de las mismas:

Rendimiento < 637,5

F. Estación de siembra = Primavera
| Altura de fructificación $>$ 92,5
|| Altura Inicio-Repro-St Cm $>$ 6
|| Altura completa Floración Cm \leq 71
||| Altura Start-Repro-St Cm $>$ 8,5
||| Rf en plena floración $>$ 5,5
||| Rf In Full Bloom $>$ 9,5: <637,5
||| Rf en plena floración \leq 8,5: <637,5
|| Altura Inicio-Repro-St \leq 6: <637,5

Rendimiento entre 1600 - 2412,5

F. Estación de siembra = Otoño
| Altura completa Floración Cm $>$ 93: 1600 - 2412,5
| Altura completa Floración Cm \leq 93
|| Altura de floración Cm \leq 89,5
|| Altura completa Floración Cm $>$ 88,5: 1600-2412,5

Rendimiento entre 637,5 - 1600

F. Estación de siembra = Primavera
| Altura de fructificación $>$ 92,5
|| Altura Inicio-Repro-St Cm $>$ 6
|| Altura de floración Cm $>$ 71: 637,5 - 1600
|| Altura completa Floración Cm \leq 71
||| Altura Start-Repro-St Cm $>$ 8,5
||| Rf en plena floración $>$ 5,5
||| Rf en plena floración \leq 9,5
||| Rf In Full Bloom $>$ 8,5: 637,5-1600
||| Rf en plena floración \leq 5,5: 637,5-1600
||| Altura Inicio-Repro-St Cm \leq 8,5: 637,5-1600

Rendimiento \geq 3087,5

F. Estación de siembra = Otoño
| Altura completa Floración Cm \leq 93
|| Altura de floración Cm $>$ 89,5: \geq 3087,5
|| Altura de floración Cm \leq 89,5
|| Altura de floración Cm \leq 88,5

Altura de floración $Cm \leq 88,5$	Rf en plena floración $> 1,5$
Rf en plena floración $> 1,5$	Altura Inicio-Repro-St $Cm \leq 37,5 \geq 3087,5$
Altura Start-Repro-St $> 37,5: 1600-2412,5$	
Rf En plena floración $\leq 1,5$	
Hojas de roseta sin botón $> 3,5: 1600-2412,5$	

Rendimiento entre 2412,5 - 3087,5

No hay resultados.

Análisis: Se puede observar, tanto en las normas como en la Figura 1, que el cultivar Local tuvo un comportamiento similar al de Delfina, observándose que los mejores rendimientos se encuentran también en las siembras de otoño y los menores rendimientos en la 3ª siembra correspondiente a la época de primavera [10]. En la siembra de primavera, el ciclo vegetativo se acorta condicionando la fase de plena floración correspondiente a la etapa reproductiva con una menor producción de semillas. Por el contrario, los mayores rendimientos se obtuvieron en la siembra de otoño, porque permitió a la planta cumplir con su etapa vegetativa y reproductiva en condiciones ambientales favorables. Esto es observable en las reglas, con mayores alturas en las etapas reproductivas de la especie, y mayor número de ramas florales (Rf). Como se observa en los datos, cuando el rendimiento es superior a 3087,5 kg corresponde a una altura de plena floración inferior a 93 cm. También se observa que cuando su altura de plena floración está entre 88,5 y 93 cm el rendimiento disminuye, posiblemente porque la altura está relacionada con una menor tasa fotosintética por parte de la planta. La situación del rendimiento se puede reafirmar en el gráfico 1.

Finalmente, el estudio de las redes bayesianas se realizó para cada especie de mostaza, pero previamente se discretizaron los atributos en 5 intervalos, cada uno con la misma frecuencia. A continuación, en el link <https://docs.google.com/document/d/1mhxoGENiTVOLJWHFErDupGB6PYMVF-T3kzoYBvGddU/edit?usp=sharing> se pueden observar los resultados.

Como se puede observar, las frecuencias probabilísticas siguen afirmando las mismas conclusiones que indican los expertos y las reglas.

4. Conclusión

La caracterización de los rangos de rendimiento de cada cultivar de mostaza fue posible con la aplicación de las herramientas y técnicas descritas. A partir de las condiciones del estudio de origen, podemos identificar estas mismas características, dejando establecida la metodología de aplicación.

El estudio infiere las relaciones entre la altura de inicio del ciclo reproductivo, la altura en plena floración y el mayor rendimiento en las siembras de otoño para ambos cultivares, observándose los mejores rendimientos en la primera fecha de otoño. Este rendimiento se destaca por encima de la siembra de primavera y de la segunda siembra de otoño, similar a las cosechas encontradas en la provincia de Buenos Aires [4]. Además, se puede observar que el cultivar Local presenta un mejor comportamiento ante las condiciones ambientales dadas durante las diferentes siembras, no así con Delfina que demostró susceptibilidad en momentos de siembras cercanas a la época invernal. Este análisis se basa en la observación de la correlación de atributos, las reglas de comportamiento identificadas y el gráfico de relación semilla-cultivar. Las líneas de trabajo futuro que quedan planteadas en él, pasan por considerar rangos más limitados

de temperatura y análisis de suelos, y predecir el rendimiento aplicando técnicas de regresión.

5. Referencias

- [1] Curioni, A, Alfonso, W y Arizio, O. (2010). Mostaza blanca (*Sinapis alba* L. syn. *Brassica hirta*). Agrotecnología, calidad y mercados. Universidad Nacional de Luján. pp. 2-22.
- [2] Arizio O. P. & Curioni A. O. (2016). Mostaza, un grano oleaginoso, aromático y medicinal. Análisis del mercado mundial y principales importadores regionales. Departamento de Tecnología. Universidad Nacional de Luján, Bs. As. *Horticultura Argentina* 35 (87), pp. 5-18.
- [3] Piola, M. (2012). Delfina INTA: La primera mostaza Argentina. Instituto Nacional de Tecnología Agropecuaria, INTA. <http://inta.gov.ar/noticias/delfinainta-la-primera-mostaza-argentina>. Último acceso octubre 2020.
- [4] Paunero I. (2015). Evaluación de material genético de mostaza (*Sinapis alba*). Curso de Postgrado: Complejo cultivos aromáticos y medicinales. Economía, mercados, agrotecnología y calidad. Universidad Nacional de Luján, Bs.As.
- [5] Cirera, L. y Jara, S. (2011). Determinación de Índices Bioclimáticos de mostaza blanca (*Sinapis alba* L.) VI Congreso Cubano de Meteorología. La Habana, Cuba.
- [6] García, M., Cañón, H., Alfonso, C., Cavallero, M. y Curioni, A. (2017). Efecto de la fecha de siembra sobre la fenología y el rendimiento en un cultivo de mostaza blanca (*Sinapis alba* L.) en Luján, provincia de Buenos Aires. *Horticultura Argentina* 36 (89): ISSN de la edición online 1851-9342, pp.17-27.
- [7] EMC - "El universo digital de las oportunidades: Rich Data y el creciente valor del Internet de las cosas" - (2014). Publicado en: <https://www.emc.com/leadership/digital-universe/2014iview/executive-summary.htm>. Último acceso Octubre 2020.
- [8] Britos, P. (2008). Procesos de explotación de información basados en sistemas inteligentes. Tesis doctoral. Universidad Nacional de La Plata. Argentina.
- [9] Fois, G; Aguero, 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.
- [10] Isaja M., Sobrero C. (2020). Evaluación preliminar del rendimiento de dos variedades de mostaza blanca (*Sinapis alba* L.): Comparación de fechas de siembra y fenología del cultivo para la Comarca Andina del Paralelo 42°. Universidad Nacional de Río Negro.
- [11] Saleem, A., Abbas, K., Asad, K., & Anjum, M. S. (2013). Best Statistical Model Estimation for Mustard Yield in Azad Kashmir, Pakistan. *Pakistan Journal of Science*, 65(1), 77-82.
- [12] Rapidminer (2020). Rapidminer. <https://rapidminer.com/>. Último acceso Octubre 2020.