



Modelos predictivos para la estimación de oferta de fresas en Cundinamarca: un enfoque comparativo con aprendizaje automático y regresión.

Oscar Mauricio Gelves Alarcón¹

Enlace ORCID: <https://orcid.org/0000-0003-0557-775X>

Nataly Lorena Guarín Cortés²

Enlace ORCID: <https://orcid.org/0000-0002-8703-6211>

María Paula Peña Martínez³

Enlace ORCID: <https://orcid.org/0009-0003-9795-3325>

María Fernanda Rebolledo Marulanda⁴

Enlace ORCID: <https://orcid.org/0009-0003-0859-8211>

Fecha de Recepción: 12 de agosto, 2025

Fecha de Aprobación: 19 de octubre, 2025

Fecha de Publicación: 19 de noviembre, 2025

Resumen:

El cultivo de fresas en Cundinamarca representa una actividad agrícola clave con alto impacto económico regional. Este estudio tiene como objetivo evaluar y comparar diferentes modelos de pronóstico para estimar la oferta futura de fresas, considerando variables como área sembrada, rendimiento y producción. Para ello, se aplicaron métodos de series de tiempo (Suavización exponencial Holt y Brown), el algoritmo AdaBoost y un modelo de regresión lineal múltiple, empleando datos históricos del Ministerio de Agricultura y Agronet. Los resultados obtenidos indican que el modelo de regresión lineal múltiple es el modelo con mejor desempeño, obteniendo un coeficiente de determinación de 0,9947 y superando en precisión a los métodos de series de tiempo, los cuales presentaron errores promedio superiores al 17%. Las pruebas de normalidad y homocedasticidad validaron estadísticamente la robustez del modelo. Se concluye que el modelo propuesto proporciona una herramienta efectiva para la planificación agrícola en Cundinamarca, al permitir proyectar incrementos en la producción asociados al aumento de área cultivada y del rendimiento por hectárea. Se recomienda para futuras investigaciones, integrar variables climáticas y políticas agrícolas para mejorar la precisión de los pronósticos.

Palabras clave: Fresas, Regresión lineal múltiple, Series de tiempo, Modelos, Predictivos.

¹ Magister en Ingeniería de Dirección Industrial, Universidad de Buenos Aires, Argentina. Docente Ocasional, Facultad de Ingeniería, Programa de Ingeniería Industrial, Universidad Militar Nueva Granada, Cundinamarca, Colombia. oscar.gelves@unimilitar.edu.co

² Magister en Gerencia Integral de Proyectos, Universidad Militar Nueva Granada, Cundinamarca, Colombia. Docente de carrera, Facultad de Ingeniería, Programa de Ingeniería Industrial, Universidad Militar Nueva Granada, Cundinamarca, Colombia. nataly.guarin@unimilitar.edu.co

³ Estudiante, Facultad de Ingeniería, Programa de Ingeniería Industrial, Universidad Militar Nueva Granada, Cundinamarca, Colombia. est.mariap.pena@unimilitar.edu.co

⁴ Estudiante, Facultad de Ingeniería, Programa de Ingeniería Industrial, Universidad Militar Nueva Granada, Cundinamarca, Colombia. est.maria.rebolledo@unimilitar.edu.co

Predictive Models for Strawberry supply estimation in Cundinamarca. A comparative Approach Using Machine Learning and Regression

Abstract:

Strawberry cultivation in Cundinamarca represents a key agricultural activity with a significant regional economic impact. This study aims to evaluate and compare different forecasting models to estimate the future supply of strawberries, considering variables such as cultivated area, yield, and production. Time series methods (Holt and Brown exponential smoothing), the AdaBoost algorithm, and a multiple linear regression model were applied, using historical data from the Ministry of Agriculture and Agronet. The results indicate that the multiple linear regression model exhibited the best performance, achieving a coefficient of determination of 0.9947 and outperforming the time series methods, which presented average errors exceeding 17%. Normality and homoscedasticity tests statistically validated the robustness of models. It is concluded that the proposed model provides an effective tool for agricultural planning in Cundinamarca allowing for projections of production increases associated with the expansion of cultivated areas and improvements in yield per hectare. Future research is recommended to integrate climatic variables and agricultural policy factors to enhance forecasting accuracy.

Keywords: Strawberries, Multiple linear regression, Time series, Models, Predictive.

Modelos Preditivos para Estimativa da Oferta de Morangos em Cundinamarca: Uma Abordagem Comparativa com Aprendizado de Máquina e Regressão

Resumo:

O cultivo de morangos em Cundinamarca representa uma atividade agrícola fundamental, com alto impacto econômico regional. Este estudo tem como objetivo avaliar e comparar diferentes modelos de previsão para estimar a oferta futura de morangos, considerando variáveis como área plantada, rendimento e produção. Foram aplicados métodos de series temporais (Suavização exponencial de Holt e Brown), o algoritmo AdaBoost e um modelo de regressão linear múltipla, utilizando dados históricos do Ministério da Agricultura e do Agronet. Os resultados obtidos indicam que o modelo de regressão linear múltipla apresentou o melhor desempenho, alcançando um coeficiente de determinação de 0,9947 e superando em precisão os métodos de series temporais, que apresentaram erros médios superiores a 17%. Os testes de normalidade e homoscedasticidade validaram estaticamente a robustez do modelo. Conclui-se que o modelo proposto oferece uma ferramenta eficaz o planejamento agrícola em Cundinamarca permitindo projetar aumentos na produção associados a expansão da área cultivada e melhoria do rendimento por hectare. Recomenda-se que pesquisas futuras integrem variáveis climáticas e políticas agrícolas para aprimorar a previsão das previsões

Palavras-chave: Morangos, Regressão linear múltipla, Séries temporais, Modelos, Preditivos.

1. INTRODUCCION

El cultivo de fresas en la región de Cundinamarca es un producto principalmente cultivado por pequeños y medianos agricultores el cual enfrenta diversos desafíos provenientes de la volatilidad del mercado, las condiciones climáticas y la gestión de recursos. En este contexto la capacidad de determinar la oferta de fresas con precisión es esencial para mejorar la planeación de cultivos, la reducción de pérdidas y mejorar la competitividad del mercado en un entorno de incertidumbre. La implementación de modelos de pronóstico se presenta con una herramienta determinante para facilitar la toma de decisiones informadas en el ámbito de la agricultura.

La literatura científica ha explorado ampliamente el uso de diferentes metodologías de pronóstico en la agricultura, destacando enfoques relacionados con los modelos de series de tiempo y de regresión. Por ejemplo, los métodos de suavización exponencial como Holt, Brown, o ARIMA han demostrado ser útiles para capturar tendencias en datos históricos, mientras que los modelos de regresión se han enfocado a la identificación de relaciones entre las variables que componen un sistema. Sin embargo, existe una necesidad de evaluar el desempeño de estos métodos en contextos específicos, como el cultivo de fresas en Cundinamarca donde las características del departamento pueden influir en la precisión.

Este estudio tiene como objetivo ofrecer una herramienta de pronóstico de la oferta de fresas en el departamento de Cundinamarca, para ello se utilizaron datos históricos proporcionados por el Ministerio de Agricultura y Agronet. Con ello, se desarrollaron y evaluaron diferentes modelos de series de tiempo y regresión lineal múltiple para estimar la oferta futura. Adicionalmente, se realizaron pruebas estadísticas como Shapiro-Wilk y Breusch-Pagan, para garantizar la validez del modelo de regresión lineal múltiple. Este enfoque permitió no solo encontrar las metodologías más adecuadas, sino también, proporcionar información práctica para los tomadores de decisiones en el sector agrícola.

La investigación destaca la necesidad de considerar variables adicionales como las condiciones climáticas y políticas agrícolas que pueden afectar significativamente los resultados de las proyecciones.

Esto subraya la necesidad de enfoques más integrales y adaptativos que combinen técnicas tradicionales con herramientas avanzadas como el aprendizaje automático, para mejorar la precisión y utilidad de los modelos. En este sentido, los resultados de este estudio no solo contribuyen a la literatura académica, sino que ofrecen un recurso valioso para agricultores, instituciones gubernamentales y otros actores interesados en la sostenibilidad del sector agrícola.

A pesar de los avances en modelo predictivos agrícolas, se evidencia la necesidad de reconocer la metodología más precisa para predecir la oferta de fresas en Cundinamarca considerando variables como la variabilidad y las limitaciones existentes en datos históricos. Por tanto, la pregunta de investigación es:

¿Cuál es el modelo de pronóstico más preciso para estimar la oferta de fresas en el departamento de Cundinamarca considerando variables históricas de área sembrada, rendimiento y producción?

2. MARCO TEORICO

En el ámbito agrícola, investigaciones como las de García-Mora et al (2023) desarrolla un modelo de regresión lineal múltiple para predecir el rendimiento del cultivo de papa en Ecuador durante el periodo 2002-2019. El documento propone como variables: el uso de semillas, prácticas de cultivo, volumen de venta y áreas sembradas. El estudio encontró que el rendimiento está significativamente influenciado por cinco factores principales: el uso de semilla mejorada, el riego, la aplicación de fertilizantes y otros factores menores. El modelo mostró un buen ajuste con un coeficiente de determinación del 0.86 lo cual valida la capacidad predictiva del modelo propuesto. También realizó un análisis espacial y temporal del rendimiento, encontrando tendencias positivas en seis provincias y estabilidad en otras cuatro; factores como la disponibilidad de agua y el uso de semillas mejoradas fueron identificados como determinantes en el incremento del rendimiento. Sin embargo, se destaca la limitada tecnificación en riego y acceso restringido a fertilizantes efectivos los cuales reducen de manera considerable las oportunidades de maximizar la productividad en ciertas regiones.



El modelo de regresión lineal múltiple empleó una metodología de regresión por pasos para eliminar variables no significativas manteniendo aquellas con mayor impacto en el rendimiento. Las variables seleccionadas incluyeron la superficie sembrada con semilla mejorada, el área bajo riego, la aplicación de fertilizantes y la incidencia de las heladas. Los resultados confirmaron que las semillas mejoradas tienen un efecto positivo en el rendimiento, mientras que las condiciones de sequía o heladas impactan de forma negativa.

Balaji (2018) analiza la efectividad de modelos predictivos en la gestión agrícola destacando la implementación paralela como una solución para el manejo de grandes volúmenes de datos, el estudio se enfoca en la falta de sincronización entre la oferta y la demanda de los productos agrícolas en la India, que, por lo general, es ocasionada por las fluctuaciones en precios, lo cual genera pérdidas entre los consumidores y los productores. Para ello se implementaron modelos de regresión lineal y series de tiempo utilizando un marco de MapReduce optimizando así el procesamiento de datos recolectados entre 2005 y 2016. Se destaca en el documento que las diferencias entre oferta y demanda se deben principalmente a la falta de planificación basada en los datos históricos. Los modelos predictivos desarrollados permitieron proyectar valores de demanda, oferta y precios para diferentes commodities agrícolas, evaluando la precisión de estos al compararlos con datos reales del mercado. Los resultados experimentales revelaron brechas significativas entre la oferta y la demanda a causa de la variabilidad de los precios.

El modelo de regresión lineal propuesto se basó en variables como la elasticidad del ingreso, el crecimiento poblacional y la demanda per-cápita en años base, mientras que el modelo Holt-Winter incorpora factores como la tendencia y la estacionalidad. Ambos métodos fueron evaluados por su capacidad para anticipar patrones en los datos; y por medio del MapReduce, se manejó de forma eficiente el volumen de datos reduciendo de manera significativa los tiempos de cálculo y aumentando la precisión de las predicciones.

Por su parte, Kumar (2023) identifica los diferentes retos existentes en la cadena de suministro, como la

cantidad de desperdicio postcosecha identificando retos en la cadena de suministro de alimentos frescos como los desperdicios postcosecha y las ineficiencias logísticas en este proceso. Propone oportunidades como la adopción de tecnologías avanzadas (blockchain y modelos predictivos), el rediseño de las cadenas de suministro y la mejora en la colaboración entre los diferentes actores que conforman la cadena logística. Este marco conceptual permite orientar a los tomadores de decisiones de la cadena logística hacia una gestión más productiva y eficiente.

De la misma manera, Mohamed-Amine (2024) profundiza en la utilización de la inteligencia artificial (IA) para mejorar la predicción de ventas en el sector agrícola con un enfoque específico en soluciones fitosanitarias en la región Sous Massa, Marruecos. El estudio recopila datos de diversas fuentes, incluyendo el sistema ERP AXAPTA, que contiene información detallada sobre fechas de venta, tipos de cultivos, condiciones climáticas y ubicaciones específicas de venta. Estos datos se emplean para la construcción de modelos de predicción de ventas utilizando técnicas de aprendizaje automático, destacando el algoritmo Gradient Boosting Regressor como el más eficaz alcanzando un error absoluto medio del 0.0035. Los resultados subrayan el impacto de factores como el clima, las prácticas agrícolas y las particularidades regionales en las predicciones de ventas. El modelo desarrollado no solo mejora la precisión en la predicción, sino que también permite gestionar de manera más eficiente el inventario, mejorando las operaciones de tipo logístico y alineándolas a la demanda de los mercados. Esto es especialmente relevante en diferentes contextos agrícolas donde las condiciones climáticas tienen una influencia directa sobre la demanda de productos fitosanitarios como insecticidas y fertilizantes.

Se destaca, además, cómo los avances en la integración de datos mediante sistemas ERP proporcionan una base sólida para los pronósticos, facilitando la toma de decisiones informadas. Este enfoque no solo beneficia a los agricultores y distribuidores al evitar excesos o déficits en la oferta, sino que también ayuda a minimizar costos de tipo operativo al anticipar la demanda con mayor precisión. La aplicación de modelos de aprendizaje automático como Gradient Boosting permite capturar relaciones complejas entre variables climáticas,

patrones de ventas históricos y características del mercado.

En conclusión, las investigaciones revisadas destacan la creciente integración de herramientas avanzadas como algoritmos de aprendizaje automático, modelos híbridos y tecnologías emergentes en el ámbito agrícola y afines. Estas investigaciones evidencian como metodologías como la regresión lineal múltiple, series de tiempo, ARIMA-Anfis y sistemas basados en inteligencia artificial han permitido mejorar la precisión en predicciones de rendimiento, demanda, oferta y otros parámetros críticos. Asimismo, establecen la importancia de factores contextuales como las condiciones climáticas, la calidad de los insumos y el contexto de las regiones en la toma de decisiones informadas.

Sin embargo, persisten retos y dificultades como la falta de estandarización de los datos, la falta de tecnificación en algunas zonas y las ineficiencias logísticas, los cuales requieren una priorización para maximizar el impacto de estas metodologías.

3. METODOLOGIA

Se presentan a continuación las estadísticas del Ministerio de Agricultura y del DANE relacionadas con la producción de fresas en Cundinamarca y sus precios desde el año 2013 al 2022. Estas estadísticas proporcionan una base sólida para el desarrollo de modelos predictivos que puedan capturar la tendencia ascendente de la oferta de fresas.

AÑO	Producción en Toneladas
2007	20432
2008	22422
2009	31117
2010	23937
2011	28544
2012	22782
2013	22562
2014	25367
2015	38752
2016	41425
2017	54124
2018	54917
2019	55996
2020	64699
2021	52480
2022	70813

Tabla 1. Producción en toneladas de fresas en Cundinamarca 2013-2022



Figura 1. Oferta de Fresas en toneladas en Cundinamarca
Fuente: (Minagricultura,2024)

De acuerdo con el comportamiento de la producción de las fresas, se observa una tendencia positiva, por tanto, se propone para la predicción modelos de series de tiempo que contengan el factor de tendencia y los métodos de regresión lineal y múltiple.

Aplicación de series de tiempo

Teniendo en cuenta el comportamiento de oferta de las fresas en el departamento de Cundinamarca, se definen dos métodos que según la literatura pueden ser apropiados para el desarrollo de la predicción de la oferta: Suavizamiento exponencial con tendencia o Holt y el método Brown. A continuación, se presenta el desarrollo correspondiente para el método Holt el cual se basa en la siguiente ecuación según Gelves (2021).

$$A_t = \text{Promedio exponencial de la serie del periodo } t$$

$$T_t = \text{promedio suavizado de la tendencia } t$$

$$A_t = (\alpha * D_t) + (1 - \alpha) * (A_{t-1} + T_{t-1})$$

$$T_t = \beta(A_t - A_{t-1}) + ((1 - \beta) * (T_{t-1}))$$

$$F_{t+1} = A_t + T_t$$

ANO	OFERTA	AT	TT	FT	ERROR	PEMA
2007	20432	32469,41607	163,7540047			
2008	22422	27474,41877	-2018,085976	32633,17008	-10211,17008	46%
2009	31117	28316,13966	-808,5605318	25456,33228	5660,667204	18%
2010	23937	25703,69871	-1571,492913	27507,57913	-3570,579133	15%
2011	28544	26361,07369	-628,816521	24132,2058	4411,7942	15%
2012	22782	24241,76755	-1259,203538	25732,25717	-2950,257167	13%
2013	22562	22770,09226	-1349,066243	22982,56401	-420,5640094	2%
2014	25367	23414,55842	-505,9225362	21421,02602	3945,973979	16%
2015	38752	30912,80923	2879,358973	22908,63589	15843,36411	41%
2016	41425,6	37648,62891	4510,408752	33792,1682	7633,431801	18%
2017	54124	48203,81653	7066,984833	42159,03767	11964,96233	22%
2018	54917,1	55092,10907	6991,408796	55270,80136	-353,7013628	1%
2019	55996,5	59008,3158	5690,7842	62083,51787	-6087,017868	11%
2020	64699,1	64699,1	5690,784201	64699,1	2,48258E-06	0%
2021	52480,22	61341,80226	1863,992453	70389,8842	-17909,6642	34%
2022	70813,2	67049,10666	3489,481097	63205,79471	7607,405287	11%

Tabla 2. Pronóstico de serie de tiempo por suavización exponencial HOLT Fuente: Elaboración propia.



Dentro de los parámetros utilizados para el desarrollo del pronóstico se utilizó un Alpha = 0,50 y una Beta = 0,2 los cuales fueron optimizados con la aplicación excel solver. De acuerdo con los resultados de la tabla 2 se determina un PEMA (error de porcentaje absoluto promedio) del pronóstico del 17%. A continuación, se presenta la ecuación de pronóstico:

$$Y = 67049,1066 + 3489,48X \quad (1)$$

Conforme a los resultados de la ecuación de pronóstico se presenta el comportamiento de la oferta real y el pronóstico correspondiente:

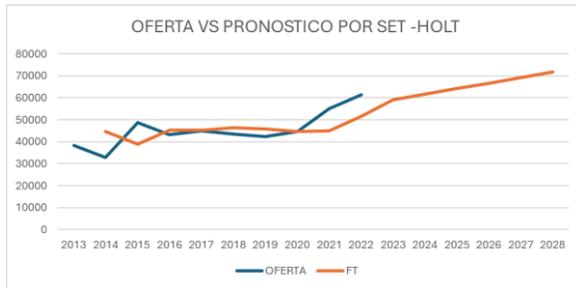


Figura 2. Relación Oferta y pronóstico de fresas por serie de tiempo suavizamiento Holt
 Fuente: Elaboración Propia.

El pronóstico realizado que utiliza el método de suavización exponencial con tendencia de Holt ha generado una ecuación de recta que modela los datos, indicando un ajuste lineal con un componente de tendencia. Según los resultados proporcionados, el promedio de error alcanza un nivel del 17%, lo cual representa un error porcentual medio absoluto. Este nivel de error sugiere que el modelo tiene una precisión baja.

Método Brown

El método Brown también se conoce como alisado exponencial doble, ya que, se realiza una doble suavización de los datos objeto de estudio y de la variable resultado de la suavización (Alarcón, 2009). La ecuación se representa de la siguiente manera:

Primera Suavización S't

$$= \alpha Xt + (1 - \alpha) + (1 - \alpha)S't - 1$$

Segunda suavización S t = \alpha S't + (1-\alpha) S t - 1

A continuación, se presentan los resultados utilizando el Método Brown.

AÑO	OFERTA	S	S'	a	b	F	error	PEMA
2007	20432	20342	20342	20342				
2008	22422	20387,63	20365,13441	20410,12559	23,13441	20433,26	1988,74	10%
2009	31117	21419,05559	20910,56277	21927,54841	522,9327741	22450,48118	8666,51882	39%
2010	23937	26335,91341	23911,9025	28759,92431	2492,846913	31252,77122	7315,771222	24%
2011	28544	25119,66431	25719,27511	24520,0535	616,6382921	23903,41521	4640,584788	19%
2012	22782	26855,8025	25999,88637	27711,71863	880,222065	28591,9407	5809,9407	20%
2013	22562	24790,38463	25808,63564	23772,13362	1047,16686	22724,96676	162,9667649	1%
2014	25367	23660,59362	24217,58059	23103,60666	572,804042	22530,80262	2836,197385	13%
2015	38752	24525,74166	24099,22368	24952,25964	438,6300523	25390,88969	13361,11031	53%
2016	41425,6	31738,45464	28182,58714	35294,32214	3656,845481	38951,16762	2474,432383	6%
2017	54124	36649,83734	34228,52567	39071,14901	2490,071028	41561,22004	12562,77996	30%
2018	54917,1	45509,23781	41141,55337	49876,92224	4491,716039	54368,63828	548,4617227	1%
2019	55996,5	50279,02394	47927,51938	52630,5285	2418,281569	55048,81007	947,6899293	2%
2020	64699,1	53177,7843	51748,69544	54606,87316	1469,671504	56076,54466	8622,555335	15%
2021	52480,22	59019,09136	56139,32698	61898,85574	2961,542679	64860,39842	12380,17842	19%
2022	70813,2	55703,88358	57338,28102	54069,48615	1680,810344	52388,6758	18424,5242	35%

Tabla 3. Pronóstico de serie de tiempo por Método Brown

Fuente: Elaboración Propia.

El parámetro Alpha de 0,507 se optimizó por medio de excel solver lo cual arrojó un valor de PEMA del 19%, lo cual representa un valor de error alto y no

Modelos predictivos para la estimación de oferta de fresas en Cundinamarca: un enfoque comparativo con aprendizaje automático y regresión

Oscar Mauricio Gelves Alarcón, Nataly Lorena Guarín Cortés, María Paula Peña Martínez, María Fernanda Rebolledo Marulanda

permite una predicción adecuada de los datos por medio de este método.

El gráfico siguiente muestra el comportamiento de la oferta vs el pronóstico:

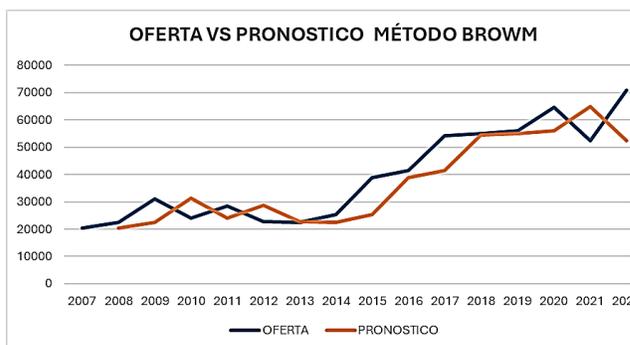


Figura 3. Relación Oferta y pronóstico de fresas por serie de tiempo Método Brown.

Fuente: Elaboración Propia.

En conclusión, dentro de los métodos de serie de tiempo propuestos no existe una alta precisión y utilizar los modelos ARIMA según la literatura no es conveniente si no existe un número mayor a 50 datos (Hancke 2010). Por lo tanto, se utiliza el Método AdaBoost el cual es aplicado por Barack (2016), el método AdaBoost según Hastie (2009) consiste en un algoritmo de aprendizaje automático diseñado para mejorar la precisión de modelos débiles como árboles de decisión de baja profundidad combinándolos con un clasificador fuerte. En cada iteración se le asignan mayores pesos a las muestras de los modelos anteriores que se clasificaron de forma incorrecta, obligando a los siguientes modelos a enfocarse en esos errores.

A continuación, se presentan los resultados obtenidos por medio del Método AdaBoost utilizando el lenguaje Python.

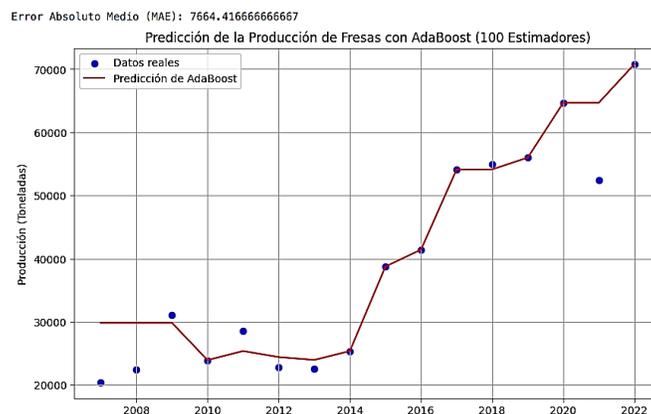


Figura 4. Método AdaBoost.

Fuente: Elaboración Propia.

El modelo de predicción mediante AdaBoost utilizó 100 estimadores los cuales logran capturar de manera general la tendencia ascendente en la producción de fresas a lo largo de los años. Se obtiene un error absoluto medio de 7644 toneladas y un error promedio relativo del 19.45%, lo cual indica que existen ciertas discrepancias en años específicos como en el 2013 y 2015, reflejando que no se está ajustando el modelo a las variaciones atípicas.

Regresión Lineal

Para el desarrollo de un método de pronóstico se propone la utilización de modelos correlacionales como la regresión lineal múltiple. Para ello, se usan las estadísticas ofrecidas por Agronet, donde se toman las variables: Oferta o producción, Área sembrada y Rendimiento en toneladas/hora. Estas estadísticas se presentan a continuación.

Año	Departamento	Producto	Área (ha)	Producción (ton)	Rendimiento (ha/ton)
2007	CUNDINAMARCA	FRESA	598,50	20.432,00	34,14
2008	CUNDINAMARCA	FRESA	622,27	22.422,77	36,03
2009	CUNDINAMARCA	FRESA	649,61	31.117,30	47,90
2010	CUNDINAMARCA	FRESA	692,05	23.937,00	34,59
2011	CUNDINAMARCA	FRESA	595,05	28.544,70	47,97
2012	CUNDINAMARCA	FRESA	487,15	22.782,75	46,77
2013	CUNDINAMARCA	FRESA	556,60	22.562,75	40,54
2014	CUNDINAMARCA	FRESA	600,90	25.367,00	42,22



2015	CUNDINAMARCA	FRESA	793,80	38.572,80	48,59
2016	CUNDINAMARCA	FRESA	804,68	41.425,60	51,48
2017	CUNDINAMARCA	FRESA	930,88	54.124,00	58,14
2018	CUNDINAMARCA	FRESA	1.001,88	54.917,10	54,81
2019	CUNDINAMARCA	FRESA	1.079,00	55.996,50	51,90
2020	CUNDINAMARCA	FRESA	1.299,50	64.699,10	49,79
2021	CUNDINAMARCA	FRESA	1.044,70	52.480,22	50,23
2022	CUNDINAMARCA	FRESA	1.349,50	70.813,20	52,47

Tabla 3. Pronóstico por Método Regresión Lineal

Fuente: Elaboración Propia.

Realizando los cálculos basados en la regresión lineal múltiple se define la siguiente ecuación de pronóstico:

$$Y = 50,0196 X1 + 742.4285 X2 - 36391.9922$$

siendo

$Y =$ producción anual en toneladas de fresa

$X1 =$ área sembrada en hectareas

$X2 =$ rendimiento dado en $\frac{\text{Toneladas}}{\text{Hectareas}}$

El valor del coeficiente de determinación indica que la variabilidad de la oferta de fresas es explicada en un 99,47% por las variables independientes área sembrada y rendimiento, lo cual sugiere un modelo con un ajuste alto.

A continuación, se realizan las pruebas de normalidad y homocedasticidad con el fin de validar el modelo propuesto. El primero con el fin de definir la normalidad de los residuos y el segundo para definir los cambios de variabilidad de los residuos. Para el desarrollo de las pruebas de normalidad se utilizan las pruebas Shapiro-Wilk y D'Agostino Pearson.

Test Shapiro – Wilk resultados:

Valor de estadístico: =0.9047, *p-valor*= 0,0957

La hipótesis nula (H_0) indica que los residuos siguen una distribución normal. Dado que $p > 0.05$ y, por tanto, no se rechaza H_0 . Se puede concluir que los residuos están normalmente distribuidos.

Test D'Agostino-Pearson resultados:

Valor de estadístico: =4.5299, *p-valor*= 0,1038

La hipótesis nula (H_0) indica que los residuos siguen una distribución normal. Dado que $p > 0.05$ y, por tanto, no se rechaza H_0 . Los resultados de la prueba refuerzan la conclusión que refleja que los residuos no violan el supuesto de normalidad.

Debido al resultado de las pruebas anteriormente nombradas se puede garantizar la validez del modelo.

Para definir el grado de Homocedasticidad que precisa los residuos de un modelo de regresión lineal, contienen una varianza constante en todas las variables independientes. Para ello, se realiza la prueba Breusch-Pagan. A continuación, se presentan los resultados.

Test Breusch-Pagan resultados:

Valor de estadístico: =2.7250, *p-valor*= 0.2560

La hipótesis nula (H_0) indica que los residuos tienen varianza constante. Dado que $p > 0.05$ y, por tanto, no se rechaza H_0 . Los resultados implican que los residuos presentan homocedasticidad y conlleva a garantizar la confiabilidad de los coeficientes estimados.

A continuación, se presentan los gráficos Q-Q y de residuos vs valores ajustados para complementar los resultados de las Pruebas propuestos en el desarrollo de la investigación.

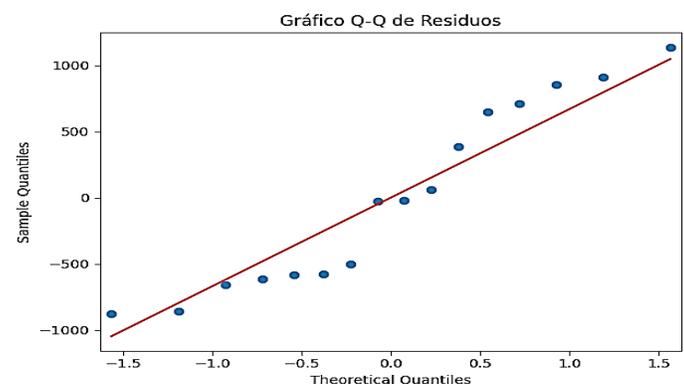


Figura 5. Gráficos Q_Q.
Fuente: Elaboración Propia.

De acuerdo con la gráfica se observa que los puntos siguen aproximadamente la línea teórica representada en rojo, lo cual sugiere que los residuos se ajustan a una distribución normal. Por otro lado, se observa que en las colas existe una desviación ligera pero necesariamente puede ser algún indicio de no normalidad.

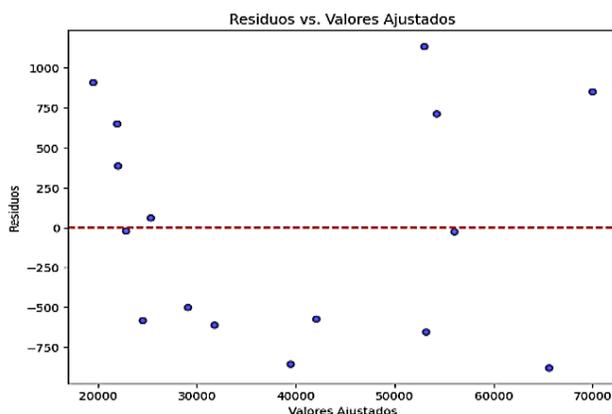


Figura 6. Residuos vs valores ajustados.
Fuente: Elaboración Propia.

Dentro de la distribución de los residuos se evidencia que los puntos se dispersan de forma aleatoria y no representa una forma cónica, lo cual indicaría una tendencia sistemática, lo cual permite respaldar la homocedasticidad resultado de la prueba Breusch-Pagan.

Para fortalecer el modelo se realizó el cálculo del indicador VIF o factor de inflación de la varianza (VIF) el cual se presenta en el siguiente cuadro:

Variable	VIF
Constante	46.89
Área (hectáreas)	1.5945
Rendimiento (toneladas/hectárea)	1.5945

Según los resultados obtenidos el VIF de la constante no es relevante para la evaluación de la multicolinealidad, mientras los resultados obtenidos del VIF de área y rendimiento se encuentran en el intervalo ($1 < VIF < 5$) lo cual indica que cada

variable contribuye de manera independiente al modelo y por tanto no están altamente correlacionados.

Realizando el pronóstico correspondiente se asume un aumento del 6 % del área cultivable y un aumento en el rendimiento del 4% teniendo por área un total de 1429,4 hectáreas y un rendimiento de tonelada/hectárea de 54,66 se obtendría una producción de fresas en el departamento de Cundinamarca en toneladas pronosticadas de 75617,38 toneladas/hectárea.

4. DISCUSIÓN

La presente investigación sobre el pronóstico de oferta de fresas en Cundinamarca puede aportar un enfoque complementario al realizado por Lim et al (2022), donde se propone un modelo predictivo basándose en principios del machine Learning para estimar la producción de fresas en invernaderos de Corea del Sur. Aunque ambos estudios tienen como objetivo la predicción de la oferta de fresas, persisten diferencias en la metodología, el entorno productivo y las variables para el estudio.

En el estudio realizado en Corea del Sur se enfocó en un entorno de agricultura de precisión e invernadero teniendo en cuenta parámetros fisiológicos de las plantas junto variables ambientales como la temperatura y la humedad. Para la modelación se empleó el método Lasso con una validación cruzada tipo K-Fold obteniendo un MAPE de 0,5111 en la predicción de flores y 0,488 en frutos.

Los resultados obtenidos en el estudio realizado de la oferta de fresas en Cundinamarca se alinean parcialmente con investigaciones, como la de Masley et al (2019). El presente documento define un modelo de regresión múltiple basado en variables macro agrícolas como área sembrada y rendimiento, el estudio realizado en California implementa modelos de machine learning específicamente redes neuronales y random forest, incorporando 27 variables climáticas y fisiológicas a nivel de campo. Ambos estudios coinciden en destacar la alta dependencia del cultivo de fresa con factores ambientales sin embargo una de las diferencias más significativas es la escala de aplicación: el modelo del departamento de Cundinamarca está orientado a la



planificación regional mientras Maskey tiene como objetivo optimizar decisiones operativas semanales fincas tecnificadas.

A pesar de las diferencias los hallazgos son complementarios. El modelo propuesto presenta un enfoque mas accesible y replicable en regiones donde la infraestructura tecnológica es de carácter limitado con un coeficiente de determinación del 0,9947, por su parte Maskey et al. Reportan un R^2 del 0,95 para la red neuronal propuesta. Estas diferencias metodológicas definen la necesidad de adaptar los modelos predictivos al entorno de aplicación.

Con respecto al estudio realizado por parte de Nassar et al. (2020) quienes desarrollaron un modelo basado a Deep Learning para predecir el rendimiento del cultivo y el precio de las fresas en California. El estudio utiliza arquitecturas complejas como CVNN_LSTM el cual tiene como objetivo capturar patrones no lineales y logra predecir el precio al agricultor, integrando el pronóstico de producción con la dinámica del mercado cual amplía el enfoque hacia la cadena de abastecimiento.

A pesar de las diferencias en complejidad y alcance, ambos estudios coinciden en la importancia de predecir el comportamiento de la oferta agrícola para optimizar la toma de decisiones. El modelo Cundinamarca se enfoca en la planificación agrícola regional a mediano plazo, logrando un R^2 de 0,99947 con datos históricos mientras Nassar et al reporta que reduce los errores significativamente tanto en rendimiento como en precio.

5. LIMITACIONES Y RECOMENDACIONES

El presente estudio se enfoca en el pronóstico de la oferta de fresas en Cundinamarca a partir de variables históricas como rendimiento y área sembrada, utilizando modelos de regresión lineal múltiple, series de tiempo y algoritmos de aprendizaje automático. Sin embargo, una de las principales limitaciones es la ausencia de variables climatológicas y económicas que podrían influir en la precisión de las proyecciones. Factores como la temperatura, la precipitación y la variabilidad en los precios no fueron considerados en este análisis, lo cual restringe la capacidad del modelo para anticipar eventos atípicos o cambios estructurales en el mercado agrícola.

Para futuras investigaciones, se recomienda la incorporación de variables exógenas relacionadas a la climatología como precipitaciones, temperatura promedio y fenómenos extremos que puedan alterar el comportamiento del cultivo de fresas. Por otro lado, cabe recalcar la integración de factores económicos como precios de mercado y costos de producción para obtener una visión mas integral que permita anticipar la oferta con las dinámicas de la cadena de suministro, una herramienta podría ser la aplicación de dinámica de sistemas. También se sugiere explorar modelos híbridos como redes neuronales o modelos basados en atención que permitan la captura de patrones no lineales y mejorar la capacidad de generalización. Por último, es importante desarrollar estudios a escala micro con datos obtenidos por sensores o sistemas inteligentes de agricultura.

6. CONCLUSIONES

El modelo de regresión lineal múltiple demostró se la técnica de pronóstico mas precisa para la predicción de la oferta alcanzando un coeficiente de determinación del 0,9947 y cumpliendo con las pruebas estadísticas de normalidad y homocedasticidad.

Los modelos de series de tiempo aplicados y el algoritmo de AdaBoost presentan mayores niveles de error lo que limita su aplicabilidad en entornos de datos históricos limitados y alta variabilidad.

Los resultados obtenidos permiten anticipar un crecimiento de la oferta de fresas en el departamento de Cundinamarca el cual es afectado notablemente por variables como incrementos en el área sembrada y el rendimiento por hectárea, lo cual puede ser una herramienta clave para la planificación agrícola de la región.

El estudio evidencia la necesidad de agregar en futuras investigaciones variables adicionales como condiciones climáticas y políticas públicas, así como modelos que combinen aprendizaje automático y técnicas estadísticas como series de tiempo para mejorar las predicciones de las proyecciones de oferta.

REFERENCIAS;

- Agronegocios. (2024, septiembre). Precios de fresa en Andina, Cundinamarca. <https://www.agronegocios.co/precios/fresa/andina/cundinamarca>
- Alarcón, M. J. (2009). Calificación del método de pronóstico de Torres (Segunda parte). *Poliantea*, 5(9).
- Barak, S., & Sadegh, S. S. (2016). Forecasting energy consumption using ensemble ARIMA–ANFIS hybrid algorithm. *International Journal of Electrical Power & Energy Systems*, 82, 92-104.
- Balaji, P., & Dakshayini, M. (2018). Performance analysis of the regression and time series predictive models using parallel implementation for agricultural data. *Procedia Computer Science*, 132, 198-207.
- García-Mora, V. M., Ilbay-Yupa, M., & Veintimilla, R. A. R. (2023). Modelo de predicción para los factores que influyen en el rendimiento del cultivo de papa (*Solanum tuberosum*) en Ecuador. *Bases de la Ciencia*, 8(2).
- Gelves Alarcón, Ó y Navarro Romero, E. (2021). Principios de la gestión de la producción. Una revisión teórica y aplicada de los conceptos. Universidad Santo Tomás.
- Hanke, J. E. (2010). Pronósticos en los negocios (9ª ed.). México: Pearson Educación.
- Hastie, T. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Ed springer
- Kumar, A., & Agrawal, S. (2023). Challenges and opportunities for agri-fresh food supply chain management in India. *Computers and Electronics in Agriculture*, 212, 108161.
- Maskey, M. L., Pathak, T. B., & Dara, S. K. (2019). Weather Based Strawberry Yield Forecasts at Field Scale Using Statistical and Machine Learning Models. *Atmosphere*, 10(7), 378. <https://doi.org/10.3390/ATMOS10070378>
- Ministerio de Agricultura de Colombia. (2024). Cifras sectoriales: Fresa. <https://sioc.minagricultura.gov.co/Fresa/Documentos/2021-0331%20Cifras%20Sectoriales.pdf>
- Mohamed-Amine, N., Abdellatif, M., & Belaid, B. (2024). Artificial intelligence for forecasting sales of agricultural products: A case study of a moroccan agricultural company. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(1), 100189.
- Nassar, L., Okwuchi, I. E., Saad, M., Karray, F., Ponnambalam, K., & Agrawal, P. (2020). Prediction of Strawberry Yield and Farm Price Utilizing Deep Learning. *International Joint Conference on Neural Network*, 1–7. <https://doi.org/10.1109/IJCNN48605.2020.9206998>
- Oh, H. R., Lim, J., Yang, S. B., Shin, C.-S., & Cho, Y. (2022). A Study on the Prediction of Strawberry Production in Machine Learning Infrastructure. *스마트미디어저널*, 11(5), 9–16. <https://doi.org/10.30693/smj.2022.11.5.9>