

**MODELO DE PREDICCIÓN DE LOS PRECIOS DEL GANADO EN
SUBASTAS GANADERAS DE CASANARE**

Elaborado por:

Eymi Nallive González Tobón

Julio Cesar Rodríguez Cristancho

Paola Andrea Numpaque Merchán

Universidad EAN

Escuela de Formación en Investigación

Seminario de Investigación de Posgrado

Bogotá

01/06/2025

Contenido

Resumen	4
Planteamiento del Problema	4
Antecedentes del problema.....	5
Descripción del problema.....	7
Pregunta de investigación	7
Objetivo general	7
Marco Teórico	9
Diseño Metodológico	23
Referencias.....	42

Lista de tablas

Tabla 1 Categorías de ganado en la subasta del Casanare. Elaboración propia	25
Tabla 2 Modelo-MAE-RMSE-R2-MAPE. Elaboración propia.....	41

Lista de figuras

Figura 1 Colombia-Casanare. Fuente: Wikipedia	23
Figura 2 Ubicación Geográfica. Fuente: Google Maps.....	23
Figura 3 Top frecuencias de Sexo (Original Limpio, excluyendo nulos). Elaboración propia.....	32

Figura 4 Descomposición de precio promedio semanal (Estacionalidad Mensual).
Elaboración propia..... 34

Figura 5 Conteo Categorías de sexo (Post-Procesamiento). Elaboración propia ... 35

Figura 6 Distribución de Precio-Final por sexo (Datos Finales) Elaboración propia
..... 36

Figura 7 Matriz de correlación, (Variables numéricas finales y target). Elaboración
propia 37

Figura 8 XGBost (Opt) Reales vs Predicciones (Test). Elaboración propia..... 39

Figura 9 Random Forest (Opt): Reales vs predicciones (Test). Elaboración propia39

Figura 10 Importancia de Features (XGBoost (Opt)-Top 20). Elaboración propia 40

Resumen

En el departamento del Casanare, la ganadería no solo es una tradición, sino que también es una actividad fundamental para la economía de la región y lo que más preocupa, son los precios en las subastas ganaderas ya que cambian sin previo aviso y no es clara la forma para este comportamiento, afectando a la población. Los posibles factores pueden estar relacionados al tema de las estaciones, peso, raza y edad del ganado, siendo esto, un reto constante al momento de las subastas. Realizando la investigación se ve la oportunidad de aplicar un aprendizaje de machine learning, con el fin, de diseñar un modelo predictivo que ayude a obtener unos precios competitivos y más precisos del mercado ganadero de la región.

Planteamiento del Problema

En el sector ganadero colombiano, no es fácil encontrar información sobre su proceso comercial, estos datos no se encuentran organizados lo que no permite un análisis eficaz para un beneficio estratégico. Esta situación conlleva a los productores, comercializadores y gobierno a tomar decisiones no efectivas sobre la compra y venta de ganado, generando incertidumbre en los precios, lo que afectando la estabilidad del mercado (Ministerio de Aricultura y Desarrollo Rural, 2023)

La carencia de una base de datos consolidada para reunir, desarrollar y analizar datos del sector ganaderas impide predecir el comportamiento en los precios y dificulta el desarrollo de estrategias de comercialización basadas en información precisa.

Frente a este panorama, resulta prioritario desarrollar herramientas para análisis de bases de datos y modelos machine learning, que permitan transformar la información recolectada en una base de conocimiento, con el fin de mejorar el proceso de toma de decisiones en el sector ganadero y también apoyar la sostenibilidad económica, fortaleciendo su capacidad de respuesta ante el proceso de comercialización (compra y venta).

Antecedentes del problema

Actualmente en Casanare los ganaderos toman las decisiones basadas en su experiencia o intuición, debido a que la falta de información comercial no es clara, teniendo en cuenta que los productores no cuentan con una base de datos estructurada y de fácil acceso sobre los precios del ganado en las subastas. Esta situación limita a los ganaderos a desarrollar y planificar estrategia en sus ventas y compras, disminuyendo su rentabilidad.

En la Actualidad, no hay datos consolidados sobre los precios del ganado, esta información no ha sido organizada ni analizada de manera efectiva, lo que impide la predicción de precios (Ministerio de Agricultura y Desarrollo Rural, 2023).

En el estudio de Fausto Camilo Moreno Vásquez, docente en la UNAL y coordinador de un laboratorio en análisis de datos, resalta lo importante que es estructurar los datos ganaderos en una base de datos para tomar decisiones estratégicas en este sector. Aunque hay una gran cantidad de datos generados, por ejemplo, durante las campañas de vacunación, estos no sirven de nada si no se convierten en conocimiento útil.

Para enfrentar estos desafíos, el laboratorio utiliza herramientas de inteligencia artificial que ayudan a crear modelos descriptivos, predictivos, prospectivos y prescriptivos. Esto facilita la toma de decisiones basadas en pruebas concretas. También hay expertos que ayudan a encontrar las mejores soluciones a problemas específicos.

Con respecto a la sostenibilidad, Moreno Vásquez destaca que esta debe ajustarse a las características de cada lugar, ya que no hay una única manera de lograr un desarrollo ganadero sostenible. La analítica de datos se convierte entonces en una clave para personalizar estrategias que aborden los aspectos productivos, económicos y ambientales.

Por último, el laboratorio ha formado alianzas con organizaciones como FEDEGAN y el Ministerio MINITIC para mejorar la asistencia técnica y facilitar el acceso a servicios de extensión tecnológica, principalmente para los medianos y pequeños productores en Colombia. (Contexto Ganadero, 2024; Contexto Ganadero, 2024) . La falta de

estructuración y análisis de datos en el sector ganadero colombiano limita las decisiones y la capacidad de los productores para anticipar fluctuaciones del mercado. Según Fausto Camilo Moreno Vásquez, coordinador del laboratorio de analítica de datos de la Universidad Nacional de Colombia, es crucial transformar la información dispersa en conocimiento aplicable mediante herramientas de inteligencia artificial (Contexto Ganadero, 2024) En este contexto, el desarrollo de un modelo de Machine Learning para la predicción de precios en subastas ganaderas se presenta como una solución innovadora. Utilizando datos históricos y variables clave que influyen en la formación del precio, este modelo permitirá a los ganaderos acceder a información precisa y oportuna para optimizar sus decisiones comerciales. La integración de modelos descriptivos, predictivos y prescriptivos garantizará que los productores, comités ganaderos y entidades gubernamentales puedan anticiparse a las dinámicas del mercado y mejorar la competitividad del sector.

La falta de estructuración y análisis de datos en el sector ganadero colombiano limita la toma de decisiones estratégicas y la capacidad de los productores para anticipar fluctuaciones del mercado. Según Fausto Camilo Moreno Vásquez, coordinador del laboratorio de analítica de datos de la Universidad Nacional de Colombia, es crucial transformar la información dispersa en conocimiento aplicable mediante herramientas de inteligencia artificial (Contexto Ganadero, 2024) En este contexto, el desarrollo de un modelo de Machine Learning para la predicción de precios en subastas ganaderas se presenta como una solución innovadora. Al construir una base con información histórica, se podrá determinar variables que permitan la formación del precio, con esto, los ganaderos podrán acceder a información clara y oportuna para lograr tomar mejores decisiones comerciales. La integración de modelos de machine learning garantizará que todos los involucrados en la cadena comercial, puedan anticiparse a la fluctuación del mercado, mejorando la competitividad del sector.

Descripción del problema

En el departamento del Casanare, no existen herramientas para los ganaderos que permitan tener una visibilidad en el comportamiento de los precios del ganado con una tendencia histórica, por lo cual, no les permite una precisión al momento de realizar un proceso comercial de compra y venta del ganado. Se ha identificado que los modelos tradicionales de estimación de precios no contemplan las diferencias categorías de ganado como machos, hembras, si son de levante o de ceba, etc., lo que minimiza la utilidad práctica en el mercado actual. (FAO, 2022)

Pregunta de investigación

¿De qué manera se puede desarrollar e implementar un modelo de Machine Learning utilizando técnicas de series de tiempo y aprendizaje supervisado, que permita predecir los precios del ganado de forma clara en subastas ganaderas del departamento del Casanare?

Objetivo general

Analizar la viabilidad de implementar un modelo de Machine learning para predecir los precios del ganado en la subasta ganadera del departamento del Casanare, teniendo en cuenta una base de información histórica y las diferentes variables del mercado.

Objetivos específicos

1. Investigar el comportamiento del sector, identificando los factores climáticos, estacionales, económicos y la categorización del ganado, que influyen en el precio.
2. Analizar la composición de los precios del ganado estableciendo los criterios de la base de datos con el fin de depurar e integrar la información que no se encuentra estructurada.

3. Evaluar el desempeño de distintos modelos Machine learning que se usarán para el pronóstico de los precios en las subastas ganaderas en el Casanare.

Justificación

Esta investigación busca conocer los diferentes aspectos que conllevan a la falta de información en la composición de los precios para las subastas ganaderas y analizar una solución para tomar decisiones acertadas con un método de machine learning.

1. Conveniencia

La investigación sirve como base para futuras aplicaciones digitales en el sector del campo, permitiendo que con la aplicación de los modelos de machine learning se optimice la toma de decisiones en la comercialización del ganado, por medio de herramientas predictivas.

2. Impacto Social

Al brindar una herramienta de predicción de precios, se beneficiará todos los involucrados en la cadena de venta, debido a que se podrá obtener información más clara de la composición de precios, obteniendo precios justos, adicional, puede ser clave, en la implementación de políticas públicas ajustadas a la realidad del sector.

3. Implicaciones Prácticas

Al usar herramientas de Machine learning, los ganaderos podrán tener acceso para la toma de decisiones, disminuyendo la afectación en su negocio por la fluctuación del sector, planificando estrategias más efectivas.

4. Valor Teórico

Este proyecto aporta conocimiento de los diferentes métodos de uso analítico en bases de datos para predecir los precios en el sector ganadero. Adicional, deja bases para

próximas investigaciones de uso de inteligencia artificial, modelos predictivos, economía sostenible.

5. Utilidad Metodológica

La Metodología de esta investigación, permite poner a prueba los modelos de análisis de datos en un sector nuevo en el uso de estas tecnologías, logrando la modernización y optimización en sus procesos de producción hasta la comercialización.

Campo, Grupo y Línea de Investigación

El proyecto va en línea con la universidad EAN, ya que maneja temas de economía y productividad, buscando que el sector ganadero sea más rentable y competitivo con el uso de las nuevas tecnologías. También, maneja Innovación y desarrollo debido a que se busca que el sector ganadero use herramientas como machine learning, haciendo sus procesos más eficientes.

Marco Teórico

La ganadería es una de las principales actividades económicas en Colombia, con departamentos que se destacan por su aporte significativo a la producción bovina. Entre ellos, Córdoba se consolida como una de las regiones ganaderas más importantes del país, no solo por su inventario bovino, sino también por su papel en la comercialización y exportación de ganado a otras zonas del territorio nacional. Sin embargo, diversos factores, como la inseguridad en las carreteras, han impactado las dinámicas de comercialización, promoviendo nuevas formas de negociación en el sector.

Según Hoz (Hoz, 2003) el departamento ganadero en Córdoba refleja una transformación impulsada por factores externos como la inseguridad en las vías. Esto ha derivado en una reducción de la participación de la Feria Ganadera de Medellín y un aumento en las subastas locales como una alternativa más segura y eficiente para la comercialización del ganado. Este fenómeno ha cambiado la estructura del mercado,

beneficiando a los productores locales y promoviendo una mayor estabilidad en la cadena de suministro. Además, las subastas han permitido optimizar la transparencia y eficiencia en las transacciones, consolidándose como el modelo de negociación predominante en Córdoba y Sucre.

Una investigación de (Martínez, Elías-Caro, & Escobar, 2022) hace un recorrido sobre los diferentes problemas desde el siglo XIX. La ganadería ha sido importante en la economía colombiana, debido a que aporta un gran porcentaje en el PIB del país. La sostenibilidad fue un elemento importante en su estudio, la investigación de Martínez tiene como finalidad conocer la sostenibilidad económica, social y ambiental ganadera a partir del siglo XIX.

Según (Martínez, Elías-Caro, & Escobar, 2022), a pesar del papel económico que ha jugado la ganadería en Colombia, su desarrollo ha estado marcado por profundas asimetrías. La sostenibilidad ha sido entendida principalmente desde una lógica empresarial, sin considerar con igual peso los impactos negativos sobre los ecosistemas naturales ni sobre las comunidades rurales. Este análisis histórico permite comprender que la sostenibilidad en el sector ganadero no puede evaluarse únicamente en términos de rentabilidad, sino que requiere una mirada multidimensional que contemple las relaciones entre lo económico, lo ambiental y lo social.

El trabajo articulado de (ICA, 2022), los gremios y las autoridades locales evidencia el compromiso del sector ganadero con la legalidad y competitividad en Casanare. A través de la revisión de normativas como la Resolución ICA 075495 de 2020, se busca garantizar la sanidad animal y la salud pública, asegurando que los productores cuenten con herramientas adecuadas para enfrentar los retos sanitarios y mejorar la calidad del hato ganadero en la región.

Subasta Ganadera

Según Muñoz, se ha incorporado nuevas tecnologías para la comercialización del ganado, una de estas, es la venta por subastas en línea, con esto, se ha visto un avance grande hacia la digitalización. (Muñoz, Díaz, & Mendoza, 2018)

Las subastas ganaderas han ayudado a reducir costos de comercialización, reduciendo tiempos de traslados, tanto para el ganadero como el comprador, adicionalmente, ayuda realizar seguimiento a las transacciones financieras porque usan medios digitales para el pago.

Según (Palacio, 2025), las subastas ganaderas no solo representan un método eficiente de comercialización, sino que también han logrado consolidarse como un pilar esencial en la economía rural de Colombia. Su evolución y expansión reflejan la creciente importancia de este sistema en la dinamización del sector pecuario, promoviendo la transparencia y la competitividad en el mercado. Con el respaldo de Aso subastas, estas subastas continúan fortaleciendo el gremio ganadero, garantizando espacios de negociación seguros y eficientes que benefician a productores, compradores e inversionistas, asegurando así el crecimiento sostenible de esta actividad en el país.

Tecnologías del sector agropecuario

En Colombia según (Electrónico, CMS Rodriguez-Azuero, 2017), resalta la importancia de la integración tecnológica en el sector agrícola, ya que mejora la eficiencia y competitividad y contribuye a la adaptación frente a los desafíos ambientales y económicos. La digitalización del agro puede ser clave para garantizar la seguridad alimentaria y la sostenibilidad en el país. La innovación tecnológica ha revolucionado el sector agrícola en varios países, permitiendo mayor eficiencia en la producción de alimentos. La implementación de tecnologías emergentes en el agro no solo mejora la productividad, sino que también ayuda a superar los desafíos ambientales y de escasez de recursos.

Este caso demuestra cómo la combinación de innovación, apoyo gubernamental como el Ministerio de agricultura y colaboración entre distintos actores del sector puede

generar un modelo agrícola altamente eficiente. Aplicar estrategias similares en Colombia podría potenciar el desarrollo del agro, mejorar la seguridad alimentaria y hacer más competitivo al país en el mercado global.

Con este sistema, se centralizan y estandarizan procesos y bases de datos que antes estaban dispersos. Esto permite a las autoridades sanitarias contar con información única, confiable y puntual para la vigilancia epidemiológica, la seguridad alimentaria y la optimización de la gestión pecuaria. Según el ministro de Agricultura, Rodolfo Zea Navarro, asegurar la trazabilidad del ganado mantener el buen estatus sanitario del país y facilitar el acceso a mercados internacionales que son cada vez más exigentes.

Deyanira Barrero León, la gerente general del ICA, señaló que la implementación de SINIGAN es un esfuerzo compartido entre el gobierno, los gremios y los productores, y busca lograr una identificación completa del hato bovino colombiano. Este sistema permitirá ofrecer servicios en las oficinas del ICA, subastas, alcaldías municipales y organizaciones gremiales, asegurando el acceso a trámites esenciales como el registro de predios, transportadores, usuarios y establecimientos, además de la expedición de guías de movilización, bonos de venta y el reporte de eventos clave en la vida de los animales.

Con SINIGAN, Colombia está dando pasos hacia una ganadería más eficiente, competitiva y alineada con los estándares internacionales, lo que fortalece la trazabilidad y garantiza la seguridad en el comercio de productos pecuarios (ICA Instituto Colombiano Agropecuario, 2022)

Por otra parte, Bayona-Oré (Universidad Autónoma del Perú, 2021), realizó una revisión al tema de predicción de precios en la agricultura con machine learning, este análisis evidencia que las investigaciones usan un enfoque cuantitativo basado en series de tiempos en un método positivista. (Universidad Autónoma del Perú, 2021) este estudio destaca que con técnicas de aprendizaje automático mejora la toma de decisiones debido a que la predicción es más exacta en los precios agrícolas, Se examinan diversos modelos, y la regresión lineal, junto con las redes neuronales, emerge como la técnica más popular en

la predicción de precios agrícolas. Además, se consideran múltiples factores que afectan los precios, como las condiciones climáticas, la demanda del mercado, las políticas gubernamentales y los datos históricos de precios. Al final, los hallazgos sugieren que los modelos de aprendizaje profundo suelen ser los más precisos, aunque requieren un mayor coste computacional, mientras que algoritmos más sencillos pueden ser apropiados para ciertas aplicaciones, según el contexto. Aparte de los modelos estadísticos tradicionales, están los support vector machines (SVM), modelos bayesianos y árboles de decisión (Universidad Autónoma del Perú, 2021). Este enfoque refleja una tendencia a comparar métodos de inteligencia artificial con enfoques econométricos convencionales para determinar cuál ofrece mayor precisión en diferentes contextos. Sin embargo, Bayona-Oré et al. también subrayan que la adopción de machine learning en este campo todavía está en sus inicios en comparación con otros sectores. Por esta razón, sugieren profundizar en investigaciones que integren múltiples fuentes de datos y algoritmos avanzados para mejorar la capacidad predictiva. (Universidad Autónoma del Perú, 2021). Sin embargo, Bayona-Oré et al. también subrayan que la adopción de machine learning en este campo todavía está en sus inicios en comparación con otros sectores. Por esta razón, sugieren profundizar en investigaciones que integren múltiples fuentes de datos y algoritmos avanzados para mejorar la capacidad predictiva.

En el sector pecuario, se han realizado investigaciones pioneras aplicando machine learning para predecir precios de ganado. Un estudio de (Rahmani, 2024) exploró diversas técnicas de ML para modelar el precio de ganado bovino en Canadá, mediante enfoques univariados ARIMA (Modelo autorregresivo integrado de medias móviles), SARIMA (Modelo autorregresivo integrado de medias móviles estacional), SARIMAX (Modelo autorregresivo integrado de medias móviles estacional con variables exógenas), y multivariados (máquinas de soporte vectorial, bosques aleatorios y AdaBoost), “Using probabilistic machine learning methods to improve beef cattle price modeling and promote beef production efficiency and sustainability in Canada” (Pre prints Org, 2023), este estudio aborda la incertidumbre e imprevisibilidad en los mercados agrícolas, que pueden verse perjudicados por diversos factores, como los precios de insumos agrícolas, condiciones de

producción, desastres naturales, el cambio climático, la economía global y cuestiones políticas. Debido a esta variabilidad, la modelización precisa y oportuna de los precios de productos agrícolas asegurar la sostenibilidad y viabilidad económica del sector, proporcionando información más precisa sobre el comportamiento de los precios.

Viendo otros mercados como el pecuario, en este se ha empezado a usar modelos híbridos que combinan las series temporales con redes neuronales, en el ejemplo descrito por (Ma, 2025) puede separar los componentes estacionarios de una serie temporal usando la descomposición modal empírica con ensamble complementario (CEEMD), esta técnica permite diferenciar patrones de tendencia, estacionalidad y ruido. posteriormente se analiza la información por intervalos que permite la cuantificación de la incertidumbre en los datos puntuales (Ma, 2025)

El uso de redes neuronales LSTM para predicción ha sido nombrado constantemente en los estudios, debido a que han demostrado captar comportamientos no lineales en datos económicos (Zhao, 2023) ,estas herramientas han sido útiles para contextos de alta fluctuación donde con otras técnicas pierden precisión.

La implementación de lógica difusa en estos tipos de modelos de predicción ayuda a mejorar la interpretación de los resultados, según García y Torres (García, 2024) esta lógica permite una integración de conocimiento y medición de la incertidumbre en series temporales, logrando que se ajusten a predicciones en entornos dinámicos, lo cual es vital en mercados pecuarios.

Otros autores indican que el enfoque híbrido fortalece los resultados al compararlos con los datos obtenidos en modelos tradicionales, en un estudio de Wang, (Wang, 2022) se evidenció que alguna combinación de técnicas disminuye los errores en la predicción financiera, generalmente en comportamientos no estacionarios. Por otro lado, (Cheng, 2023) habla que los modelos híbridos también pueden integrar potencialmente niveles de información, logrando una mejor interpretación en los diferentes escenarios de los precios en productos agrícolas y pecuarios.

En términos operativos, (Ma, 2025) resalta la capacidad de integrar diferentes técnicas de inteligencia artificial y estadísticas avanzadas, con el fin de obtener una mejor predicción de precios cuando hay complejidad en los datos y duda. ((García, 2024); (Cheng, 2023)).

Se ha probado para predecir precios agrícolas, pero los resultados han sido variados. Su éxito depende mucho de elegir el 'kernel' adecuado (que define la forma de la función) y de ajustar bien sus parámetros. En escenarios donde la relación entre las variables no es una línea recta, un SVR con un kernel específico (como el RBF) podría detectar patrones que una regresión lineal simple pasaría por alto. Sin embargo, hay que tener cuidado: con conjuntos de datos pequeños, a veces puede 'sobre ajustarse', es decir, aprenderse demasiado bien los datos de entrenamiento, incluido el ruido, y luego no funcionar tan bien con datos nuevos.

Vale la pena mencionar que las redes neuronales parecen ser una opción bastante popular para este tipo de desafíos, como señalaron (Bayona-Oré, Cerna, & Tirado Hinojoza, 2021) en su revisión sobre el uso de machine learning para predicción de precios agrícolas. Además de las redes recurrentes (RNN) enfocadas en series de tiempo, también se han usado redes más simples de tipo 'feed-forward', como el perceptrón multicapa. Estas a menudo se entrenan con datos en formato de tabla, donde se incluyen como características los precios de momentos anteriores (rezagos). Estas redes son potentes y pueden aproximar relaciones muy complejas, pero también vienen con muchas decisiones de configuración (número de capas, neuronas por capa, función de activación, etc.). Por lo general, se recomienda compararlas siempre con modelos más sencillos para asegurarse de que la complejidad añadida realmente aporta una mejora significativa en el rendimiento.

Mirando el panorama general, contamos con un rico conjunto de métodos supervisados, que van desde los modelos lineales básicos hasta técnicas sofisticadas de ensamble y redes neuronales profundas. Cada uno tiene sus puntos fuertes y sus propias asunciones. Por eso, nuestro plan de probar varios enfoques distintos no es arbitrario;

sigue la recomendación científica de que no existe un modelo único que sea el mejor para todos los problemas (el famoso "Teorema del No Almuerzo Gratis"). La literatura científica respalda esto, mostrando cómo diferentes algoritmos han tenido éxito según las características específicas de los datos y las dinámicas del mercado estudiado.

La oferta está muy ligada al ciclo ganadero (por ejemplo, los destetes estacionales que aumentan la oferta de terneros en ciertas épocas) y también a eventos climáticos (una sequía prolongada puede obligar a vender animales por falta de pasto, aumentando la oferta momentáneamente y presionando los precios a la baja). La demanda, por su parte, puede variar según el consumo interno de carne, la necesidad de ganado para engorde en otras regiones, e incluso por las exportaciones (en Colombia, la apertura o cierre de mercados internacionales tiene un impacto notable en el precio interno del ganado gordo). Los costos de producción también influyen indirectamente; si criar un animal se vuelve mucho más caro (por el precio de los concentrados o insumos veterinarios), algunos productores podrían optar por liquidar parte de su ganado, afectando los precios. Otro elemento clave es el propio animal: su peso y categoría. Generalmente, el precio por kilo varía mucho según la categoría (un ternero de levante no vale lo mismo por kilo que un novillo gordo listo para sacrificar). Por eso, es crucial segmentar los datos por categoría o incluirla como una variable en el modelo (como se menciona en la Estudios del sector (FAO, 2022) también señalan que la calidad del ganado (raza, condición corporal) y factores sanitarios pueden influir en el precio de subasta, aunque estas variables suelen ser más difíciles de cuantificar de forma estándar en un modelo.

El tiempo, por supuesto, es una dimensión crítica. En los Llanos Orientales, existen ciclos anuales bien conocidos asociados a la estacionalidad climática (verano/invierno). Tradicionalmente, en la época seca ('verano'), la escasez de forraje suele llevar a una mayor oferta de ganado flaco, lo que deprime los precios. En cambio, después de buenas lluvias, cuando los pastos se recuperan y los animales ganan peso, la necesidad de venta urgente disminuye y los precios pueden subir por una menor oferta. Estas hipótesis deben poder reflejarse en el modelo, ya sea incluyendo indicadores por

trimestre o mes, o dejando que modelos específicos para series de tiempo como Prophet capturen automáticamente esta estacionalidad anual. Incorporar datos externos ('exógenos') al modelo –más allá de los precios pasados y las características básicas del animal– tiene el potencial de mejorar significativamente las predicciones si esos datos aportan información relevante. Pensemos en esto como la 'X' en los modelos ARIMAX, o simplemente como variables adicionales en los enfoques de machine learning. Por ejemplo, incluir el índice de vegetación NDVI promedio de la región (un indicador del verdor de los pastos obtenido por satélite) podría ayudar a anticipar aumentos en la oferta de ganado flaco. Sin embargo, conseguir, limpiar y sincronizar este tipo de datos diversos puede ser un desafío considerable, quizás fuera del alcance de este primer estudio. En la literatura algunos trabajos multivariados (como el de (Rahmani, 2024)) han integrado factores macroeconómicos (precios internacionales, inventario ganadero nacional, etc.) y han obtenido mejores resultados. Para este proyecto, se dará prioridad a variables locales de más fácil acceso, como pueden ser las precipitaciones mensuales (como indicador indirecto de la disponibilidad de pasto), los índices de precio del ganado a nivel nacional (publicados por FEDEGAN, que sirven de referencia) o el IPC de alimentos (que puede reflejar el poder adquisitivo del consumidor), siempre y cuando contemos con series históricas fiables para estas variables.

Una parte fundamental al construir un modelo predictivo es validar rigurosamente su funcionamiento, se necesita una base teórica-metodológica sólida para esta evaluación y se podría usar el MAPE (Error Porcentual Absoluto Medio) para tener una idea del error en términos porcentuales, aunque hay que tener cuidado si los precios pueden acercarse a cero, ya que se vuelve inestable.

Otro concepto importante es el de 'sobreajuste' (overfitting). Esto ocurre cuando un modelo se ajusta 'demasiado bien' a los datos históricos con los que fue entrenado, incluyendo el ruido aleatorio, y como resultado pierde su capacidad de generalizar y hacer buenas predicciones sobre datos futuros. Los modelos muy complejos (como redes neuronales profundas o bosques de decisión con muchos árboles sin regularización) corren este riesgo. Por eso, aplicaremos técnicas de regularización (como limitar la

profundidad de los árboles en XGBoost o usar 'parada temprana' al entrenar redes neuronales) y siempre reservaremos datos exclusivamente para la prueba final. Un buen modelo logrará un equilibrio entre ajustarse a los datos y poder generalizar, lo que se reflejará en que su error durante el entrenamiento y su error en la prueba sean ambos bajos y relativamente cercanos.

Finalmente, dado que el objetivo último es facilitar la toma de decisiones, también nos interesa la 'interpretabilidad' del modelo, es decir, qué tan fácil es entender por qué hace las predicciones que hace. Modelos como la regresión lineal o ARIMA son bastante transparentes: sus coeficientes nos dicen directamente cómo se espera que cambie el precio ante cambios en las variables. En cambio, modelos más complejos como XGBoost o las redes neuronales a menudo funcionan como 'cajas negras', dificultando ver la lógica interna. Afortunadamente se han desarrollado técnicas de interpretabilidad (como valores SHAP o las medidas de importancia de variables en los bosques aleatorios) que nos permiten extraer cierta comprensión incluso de estos modelos complejos. Por ejemplo, podríamos usar la importancia de variables calculada por un Random Forest o XGBoost para ver qué factores (quizás el mes del año, o la precipitación reciente) tuvieron más peso en sus predicciones. Esto conecta con el valor teórico del estudio: no solo buscamos predecir, sino también entender mejor la dinámica de los precios.

Haciendo un resumen, se ha visto que la predicción de precios ganaderos puede abordarse desde una perspectiva de series de tiempo univariadas (como ARIMA o Prophet) y también desde el aprendizaje automático supervisado multivariante (con modelos como XGBoost o redes neuronales), o incluso combinando ambos enfoques. La literatura sugiere que probar y comparar múltiples métodos es beneficioso, ya que distintas técnicas pueden capturar diferentes aspectos de la realidad de los datos (como se vio en el estudio de Rahmani et al., 2024).

Actualmente, el Machine Learning se ha utilizado en sectores productivos y en la sostenibilidad ambiental, este estudio identifica su uso en áreas de predicción

meteorológica, en el monitoreo de plagas y diferentes enfermedades. Su metodología recopiló y realizó un análisis sistemático de fuentes científicas logrando manejar la tecnología de datos en la agricultura. En América, esta investigación destaca la importancia de trabajar en la implementación de estas herramientas disminuyendo los problemas tecnológicos y económicas que tienen la agricultura (Alfonso Gamboa, 2024) Si bien los avances en Machine Learning han demostrado un gran potencial para transformar la agricultura en América Latina, su adopción a gran escala sigue enfrentando desafíos significativos. La falta de acceso a tecnologías digitales, la escasez de personal capacitado y la necesidad de políticas gubernamentales que promuevan la innovación son factores que deben abordarse para garantizar un uso efectivo y equitativo de estas herramientas. Además, es fundamental fomentar la cooperación entre gobiernos, instituciones académicas y el sector privado para desarrollar estrategias que faciliten la implementación de soluciones basadas en inteligencia artificial. La creación de plataformas de datos abiertos, la inversión en infraestructura digital y la capacitación de los agricultores en el uso de estas tecnologías son pasos clave para aprovechar al máximo los beneficios que el Machine Learning puede ofrecer al sector agropecuario. Diversos estudios han explorado cómo los algoritmos de aprendizaje automático pueden mejorar la salud animal mediante el análisis de grandes volúmenes de datos clínicos, ambientales y de comportamiento. La capacidad del ML para reconocer patrones complejos permite detectar signos tempranos de enfermedades, facilitando intervenciones oportunas y reduciendo el impacto en la producción ganadera. Esta integración de la inteligencia artificial en la ganadería representa una transformación significativa en la manera en que se diagnostican y manejan las enfermedades, optimizando el uso de recursos y mejorando el bienestar de los animales. Por otra parte (Vieta-Vega, 2024) menciona en su artículo la implementación de Machine Learning en la producción ganadera, incluyendo herramientas, modelos y técnicas para el manejo sanitario del ganado, proponiendo oportunidades de mejora para lograr el proceso ganadero más eficiente.

En la ganadería, la implementación de tecnología como el Machine learning ha ayudado en la detección temprana de enfermedades, estos modelos analizan los datos clínicos y ambientales que permiten dar alertas sobre posibles brotes de enfermedades, de manera que los ganaderos puedan ser ágiles en su intervención. (Elicit, 2024), logrando disminuir pérdidas económicas y de ganado.

En otro sector de campo como el avícola, el uso de Machine learning se ha implementado para predecir las condiciones ambientales adecuadas dentro de los galpones, un ejemplo es que estos modelos permiten ajustar automáticamente los sistemas de ventilación y calefacción, logrando una reducción de costos en el proceso de producción (Elicit, 2023). En el cultivo de banano, también se han usado modelos de predicción como Random Forest y XGBoost para predecir el peso de los racimos, teniendo en cuenta las condiciones ambientales, ayudando a tener un costo más preciso de la producción y la ganancia del racimo por su peso. (Elicit, 2023)

Por otra parte, la literatura destaca para que el modelo predictivo funcione, debe contar con una base de datos depurada y confiable, con el fin de que los resultados al usar estos modelos sean efectivos. (Elicit, 2024).

Las técnicas como Random Forest, Support Vector Machines (SVM), redes neuronales profundas (Deep Learning) y XGBoost han demostrado eficacia en sus primeros usos logrando identificar de forma temprana enfermedades en los animales como bovinos, porcinos y aves de corral. (Elicit, 2024), esto ha ayudado a disminuir el consumo de antibióticos, dado que ha detectado de forma temprana las enfermedades. (Elicit, 2024)

La revista de la universidad de Montevideo, indica que lo más importante para estos modelos es, contar con una base de información clara y verídica, adicional, el uso debe ser de forma ética, porque cuenta con información sensible de los animales y sus procesos de producción, su correcta implementación ayudaría de forma radical a los

problemas que hoy en día se enfrentan como es el caso de las enfermedades. (Revistas Um, 2024).

Analizando otras investigaciones, los algoritmos de aprendizaje como regresión lineal, árboles de decisión, Random Forest, máquinas de vectores de soporte (Support Vector Machines) y redes neuronales artificiales, con los modelos más complejos, especialmente Random Forest y redes neuronales, se evidenció un nivel superior para la predicción de precios con un menor margen de error en comparación con modelos más sencillos (wseas, 2021) .

También se vio que incorporar datos de diversas fuentes (meteorológicos, comerciales y de mercado) mejoró el rendimiento de los modelos, destacando la importancia de integrar diferentes dimensiones del contexto agrícola en el proceso de predicción.

En conclusión, el estudio resalta que el uso de machine learning es una herramienta poderosa para la agricultura moderna. Implementar estos modelos puede contribuir a mejorar la estabilidad de los mercados, reducir pérdidas económicas y fortalecer la seguridad alimentaria. Sin embargo, su efectividad depende de la disponibilidad de datos confiables y actualizados. Se sugiere continuar investigando en esta área, añadiendo nuevas variables y ajustando los modelos a distintas regiones agrícolas para mejorar su generalización

Revisando diferentes estudios independiente de que sector del agro pertenezca, es algo importante revisarlo, ya que da otro punto de vista del uso de ML aplicado, El informe “Índice de Precios en Origen y Destino de los Alimentos (IPOD)” elaborado por COAG (Coordinadora de Organizaciones de Agricultores y Ganaderos), tiene como objetivo dejar en evidencia las grandes diferencias de precios que existen entre lo que se paga a los productores agroalimentarios en el campo y lo que finalmente abona el consumidor en los puntos de venta. Este estudio denuncia cómo, a lo largo de la cadena de distribución, se generan márgenes excesivos que perjudican tanto a los agricultores

como a los compradores, mostrando cómo la especulación intermedia afecta gravemente el valor real de los alimentos (SCIELO, 2023).

El IPOD se calcula dividiendo el precio de venta al público entre el precio que recibe el productor, lo que permite obtener un valor que refleja cuántas veces se multiplica el precio original. Este índice se aplica a productos agrícolas como frutas, verduras, carnes, cereales y legumbres, y ha demostrado en múltiples ocasiones que el precio en el mercado puede ser cinco, seis o incluso más de siete veces superior al que recibe el agricultor. Ejemplos concretos que se mencionan incluyen el caso del ajo o del limón, con índices que superan el 600%, lo cual evidencia un grave desequilibrio (SCIELO, 2023).

Los resultados del estudio muestran que los productos frescos son los más afectados por estas diferencias, especialmente las frutas y hortalizas; mientras que los productos procesados o carnes tienden a tener márgenes más contenidos, aunque también significativos. En general, los agricultores no solo perciben precios muy bajos en origen, sino que muchas veces esos ingresos ni siquiera cubren los costos de producción, poniendo en riesgo la viabilidad económica de sus explotaciones y la sostenibilidad del modelo agrario tradicional.

El informe también apunta que gran parte del problema radica en el poder que ejercen las grandes cadenas de distribución, que imponen condiciones de compra a los productores y aplican aumentos desproporcionados en los precios de venta (SCIELO, 2023).

Segmentación geográfica del departamento del Casanare.

El mapa geográfico de Colombia muestra la ubicación del departamento del Casanare, se encuentra en la región de Orinoquía, al oriente del país.

Ubicación subasta ganadera SUBACASANARE



Figura 1 Colombia-Casanare. Fuente: Wikipedia

La subasta Ganadera del Casanare, está ubicada en la zona rural del municipio de Yopal, con temperaturas medias de 30°C a una latitud 5.3366386 con longitud 72.3502606.

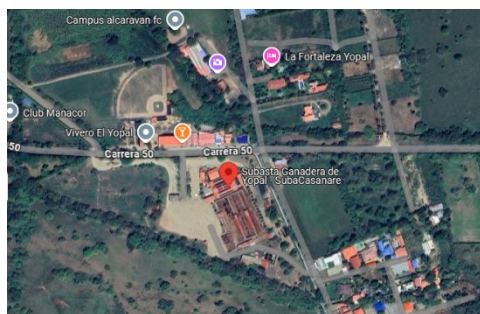


Figura 2 Ubicación Geográfica. Fuente: Google Maps

Diseño Metodológico

1. Enfoque de la investigación

Para el problema planteado que busca predecir los precios del ganado en las subastas en el departamento del Casanare, se seleccionó como método de investigación cuantitativo teniendo en cuenta que se analizarán las variables que influyen en el precio por medio de bases de datos de las subastas ganaderas del año 2024, con el objetivo de determinar los resultados que permitan implementar un modelo de Machine learning.

1. Tipo y alcance de la investigación

El tipo de estudio es correlacional y predictivo, porque pretende identificar con las variables investigadas, la relación que existe entre estas, para predecir el comportamiento del precio del ganado dentro de una subasta.

1.1. Alcance

El alcance de esta investigación es explicativo, debido a que busca entender qué variables influyen en los precios del Ganado dentro de la subasta, explicando por qué ocurre y si tiene alguna relación entre las diferentes variables para determinar el precio.

1.2 Variables

1.2.1 Variables dependientes

La variable dependiente que se estudiará será el precio del ganado en las subastas en el departamento del Casanare, buscando predecir el precio de venta por kilogramo, teniendo en cuenta que este, es influenciado por diferentes factores.

1.2.2 Variables Independientes

Categorías de ganado en la subasta del Casanare: El ganado está clasificado en categorías según su sexo, edad, el propósito y condición corporal.

Sigla	Descripción
MC	Macho Ceba
VG	Vaca Gorda
VC	Vaca de Cria
XX	Categoría no especificada / mixta
HL	Hembra Levante
NC	Novillo de Ceba
ML	Macho Levante
NG	Novilla Gorda
VP	Vaca Preñada
MG	Macho Gordo

Tabla 1 Categorías de ganado en la subasta del Casanare. Elaboración propia

Cantidad

Esta Variable corresponde al número de reses que ingresan a la subasta en un solo lote.

Peso promedio

Corresponde al peso promedio de las reses de un mismo lote, permitiendo comparar esta información con los diferentes lotes.

Hora de entrada del lote

Es la hora en la que el lote ingresa a la subasta.

Fecha de subasta (Estacionalidad)

Corresponde a la fecha en la cual se lleva a cabo la subasta y donde se podrá medir el comportamiento de los efectos estacionales; dependiendo de la fecha de la subasta, se

presentan periodos de alta y baja demanda, para ello se crearon variables adyacentes a la fecha para poder medir las estacionalidades.

- Mes de la feria: Mes del año de la subasta (1 a 12).
- Día de la semana de la feria: Día de la semana de la subasta
- Día del Mes de la Feria: Día del mes en que se realizó la subasta
- Semana del año de la feria: Número de la semana del año en que se realizó la subasta.
- Tendencia del precio de la semana: Tendencia general de los precios, que se calcula a partir de promedios semanales.
- Precios estacionales del mes: Precio estacional, estimado a partir de variaciones que se repiten aproximadamente cada cuatro semanas.

1.3 Población, muestra y técnica de muestreo

Población

La población está constituida por los registros históricos de precios de las subastas ganaderas en el departamento del Casanare del año 2024.

Muestra y técnica de muestreo

En esta investigación se trabajará con base de datos histórica de precios, depurando los registros que no generan valor e inconsistentes, con el fin de garantizar su validez para el análisis cuantitativo.

Una de las herramientas que se utilizó para los objetivos específicos fue el análisis de fuentes documentales, donde se encontraron diferentes resultados. La técnica empleada fue la recolección de datos mediante web scraping (Rogelio Mijangos-Espinosa, 2022), el cual, logra extraer información de las diferentes fuentes digitales públicas y privadas, donde han publicado los datos de cada subasta realizada, facilitando la construcción de una base sólida para desarrollar un modelo predictivo de precios.

2. Diseño de segundo nivel

2.1 Modelos y referentes

Se utilizarán modelos Random Forest y XGBoost con regresión de tiempo con base a IMF (Intrinsic Mode Functions) para realiza la predicción del ganado por lote según las variables definidas, estos son modelos de aprendizaje para crear una predicción final. La principal diferencia entre ellos es como se construyen y combinan estos modelos individuales. El objetivo de un modelo de machine learning es ayudar a entender los factores que afectan toda la formación del precio final de venta del ganado. Y con la aproximación a la regresión de tiempo de IMF, el objetivo es predecir los precios futuros en ciclo o tiempos establecidos como, por ejemplo, el precio promedio de un mes en específico o de una categoría en un mes.

2.2 Elementos funcionales del modelo aplicado

Para realizar el estudio de la predicción de precios del ganado, se ha estructurado el trabajo desde la preparación inicial de algunas bases de datos hasta obtener un modelo funcional. A continuación, se explica el flujo de trabajo a realizar.

Entradas

La información usada son bases de datos de subastas ganaderas, extraída mediante un proceso de web scraping, publicados en SUBACASANARE en PDF y posteriormente convertidos a Excel. Estas bases son el conjunto de variables para el desarrollo de algunos modelos. Luego, sigue la fase de limpieza y validación, donde se selecciona las variables como sexo, cantidad, peso promedio, hora entrada, etc., y excluyendo otras como por ejemplo peso-total, precio-base y procedencia, debido a inconsistencias en datos que puede afectar la calidad del modelo predictivo.

Proceso

Esta es una de las etapas donde serán convertidos los datos crudos a un modelo predictivo. Teniendo en cuenta que los datos están en archivo PDF, puede que traiga ciertas inconsistencias, por lo cual, se debe realizar pasos adicionales antes de pasarlo a un modelo predictivo:

Adecuación y Preparación de los datos (Preprocesamiento):

El primer paso consiste en una revisión detallada de los datos para identificar y corregir los datos erróneos, luego se identifican las variables sin datos definidos para reemplazarlos según el proceso que se establece como, ej: para Peso-Promedio, se utiliza la mediana para completar el campo vacío. Además, como el modelo predictivo es un modelo matemático, se requiere una conversión de texto a número. Por ello, la variable categórica Sexo se transformó utilizando la técnica de One-Hot Encoding. Debemos también realizar la equiparación de escalas (Normalización) Las variables numéricas pueden presentar rangos de valores muy dispares, por lo cual, para evitar que afecte el dato de predicción, estos datos son modificados a una escala común con media de cero y con desviación estándar de uno.

Para la construcción y ajuste de modelos utilizaremos el entrenamiento con Validación Cruzada donde en primer lugar realizaremos la segmentación de datos, este proceso se realiza en dos etapas, la primera es de entrenamiento de datos, seleccionando el 80% de la base y la segunda etapa es la de prueba el 20% restante, donde se usa la base restante para probar los escenarios planteados en el modelo predictivo, con el objetivo de ser lo más realista posible.

Ahora se entrenan los modelos predictivos seleccionados que en este caso son XGBoost y Random Forest Regressor, pero para realizar la aproximación a modelos IMF tendrá Componentes de Series Temporales (Ej: los promedios de las últimas 4 semanas). La evaluación se realiza con el fin de garantizar la validez de los datos y asegurar que el rendimiento no sea casual. Se observan los datos en orden confirmando que los eventos pasados influyan a los futuros, pero no al contrario. Se emplea la estrategia TimeSeriesSplit

de Scikit-learn (Sebastian Raschka, 2022), que consiste en dividir los datos en pliegues de manera secuencial, donde se revisa la primera iteración con la siguiente y así sucesivamente de manera secuencial.

Para medir los resultados se realiza una prueba final, donde se evalúa el rendimiento del modelo, luego se generan métricas e indicadores claves para compararlos como el Error Absoluto Medio (MAE), Error Cuadrático Medio (MSE), Coeficiente de Determinación (R^2) Error Absoluto Medio Porcentual (MAPE), priorizaremos MAPE a R^2 similares y por último se presentan los resultados de los dos modelos. Al final del proyecto se contará con los modelos predictivos validados, donde se confirmará la viabilidad de uso, acompañado de un análisis que confirme la capacidad predictiva y de los factores que inciden en la formación de precios.

2.3 Instrumentos de recolección y validación:

Se obtendrá las bases de datos de las subastas ganaderas del año 2024 de SUBACASANARE (SUBACASANARE, s.f.), donde se encuentran los registros de las ferias ganaderas con las variables mencionadas anteriormente, estas bases las obtendremos en archivo PDF. Adicional, se consultarán diferentes fuentes públicas y privadas como IDEAM, DANE, MinAgricultura y FEDEGAN para identificar el comportamiento de variables.

Para poder analizar los datos, se realiza una depuración de datos, donde se elimine los valores atípicos (outliers), datos inconsistentes, y se realizará una homologación de los valores de las variables con codificación.

2.4 Técnicas de análisis de datos

En esta investigación de modelo de predicción de los precios del ganado en subastas ganaderas en el Casanare, se utilizan técnicas de análisis de datos con el objetivo de construir modelos predictivos de precios, se apoyó en un conjunto de técnicas de análisis de datos con lenguaje de programación Python (Rojas, 2020) y sus librerías especializadas que

permite el desarrollo de modelos capaces de predecir los precios del ganado de las subastas del Casanare. Adicionalmente se usa una técnica de estadística descriptiva en el análisis de la base de datos de las subastas y de fuentes documentales, artículos de revista y bibliográficas donde se encuentran diferentes resultados de la aceptación de estos productos en el mercado.

3. Resultados

3.1. Comportamiento de variables que influyen en los precios

Para el desarrollo de este objetivo específico se analizó el comportamiento de diferentes factores que influyen en el precio del ganado como: los factores climáticos, la estacionalidad del mercado, los elementos económicos del entorno y la categorización del ganado.

(FEDEGÁN, 2024) ha mostrado que el sector ganadero especialmente el de Casanare es vulnerable a factores climáticos, estacionales, económicos y segmentación del ganado, los cuales influyen en la variación de precios. En cuanto al clima, el fenómeno del Niño afecta el precio debido a que provoca la muerte de miles de bovinos por sequías intensas y por la escasez de agua, mientras que en el fenómeno de la Niña genera inundaciones, afectaciones en el transporte y aumento de enfermedades. Adicional, en la temporada seca disminuye la oferta de alimento y agua, aumentando los precios de producción del ganado y en temporadas de lluvia, aunque hay mejoras en la producción, se presentan problemas de transporte para su comercialización.

En el plano económico, (Federación Colombiana de Ganaderos – FEDEGÁN, 2024-2025) el precio del ganado cambia por nuevas políticas sanitarias, por exportaciones donde se ha llegado a vender más de 238.000 animales en el año y por el aumento de precio de los insumos; un ejemplo de ello fue en marzo de 2025, el kilo de novillo gordo alcanzó los \$8.429/kg. Independiente a estos factores económicos, el sector ganadero continúa siendo estratégico para el país.

La categorización del ganado también influye en el precio, por su edad, sexo, condición corporal y objetivo de venta, en las subastas se valoran más los animales que se observen con mejor salud, las diferentes características morfológicas, que sean para exportación o con mayor rendimiento en canal.

Revisando el artículo (ganadero, 2024) el comportamiento del precio según la estacionalidad se observa que la mayor oferta ocurre entre los meses de abril y julio, mientras que desde agosto al mes de diciembre se observa que disminuye, lo que provoca una fluctuación constante en los precios durante todo el año. Adicionalmente, las categorías del ganado también influyen en su precio, donde se clasifican como, por ejemplo, macho levante (ML), macho ceba (MC), hembra levante (HL) y hembra vientre (HV), siendo estas categorías importantes para fijar su precio, pero van de la mano con el objetivo de la venta, es decir, su uso, si es para reproducción, exportación o sacrificio. Estos factores son esenciales para la toma de decisiones adecuada con respecto a la venta y compra del ganado.

3.2. Composición de precios del ganado

La información primaria se extrajo de informes en formato PDF emitidos por SUBACASANARE, utilizando técnicas de web scraping. Inicialmente, este conjunto de datos presentó variaciones significativas, por lo cual, se realizó una revisión detallada de las variables disponibles (tales como sexo del animal, cantidad, peso y promedio, fecha de feria y precio final), donde se evaluó su calidad e identificando inconsistencias o valores atípicos.

A continuación, se detalla el paso a paso empleado para el análisis de los precios del ganado dentro de la subasta. La fase inicial consistió en la depuración y estructuración de 19000 filas de datos correspondientes a los lotes que llegaron a la feria en SUBACASANARE durante el primer año desde mayo de 2024 a mayo de 2025, donde originalmente el formato no era homogéneo, por lo cual, se logró consolidar una base de

datos limpia y validada, donde se pudo identificar los factores en la variación de los precios y así poder realizar modelos para la predicción de precios futuros.

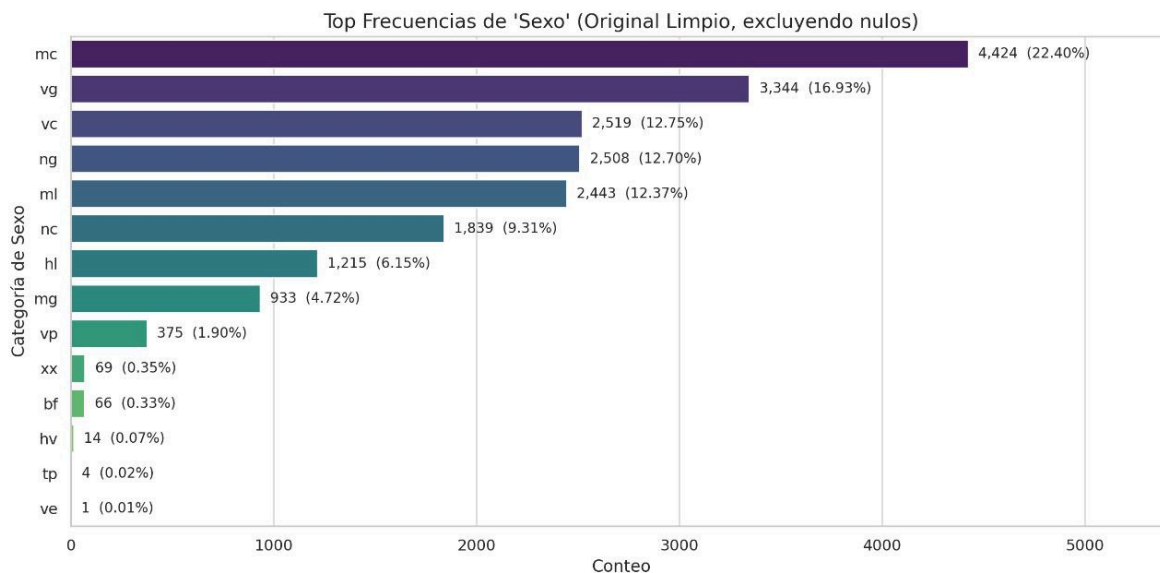


Figura 3 Top frecuencias de Sexo (Original Limpio, excluyendo nulos). Elaboración propia

Después de un proceso de limpieza, estandarización e imputación en la figura #3 se muestra la distribución de frecuencias de la variable “Sexo”, principalmente, esta variable presentaba datos dispersos con errores y algunos con valores faltantes. Por lo cual, se realizó una estandarización convirtiendo los valores a minúsculas y eliminando los espacios. La unificación de valores nulos se unificó al valor estándar NaN, también se agruparon las categorías con una frecuencia muy baja en una etiqueta general llamada “sexo-otro”, con el fin de poder mejorar los análisis y los rendimientos de los modelos futuros. Por ultimo los NaN que quedaron se clasificaron como “sexo-indeterminado”, logrando conservar los registros.

Para este resultado, la variable “Sexo” se convirtió en un grupo de categorías más claras y de fácil manejo. La figura #3 muestra el total de animales por categorías depuradas principalmente macho y hembra, explicando la composición del ganado post-procesamiento.

Este proceso de depuración y estructuración se realizó en las diferentes variables; en el tratamiento de Fechas. Las fechas de feria (FECHA-FERIA) se estandarizaron a un formato datetime uniforme y se ordenaron cronológicamente. Posteriormente se llevó a cabo el procesamiento de variables numéricas, donde se realizó la corrección e imputación de valores para variables cuantitativas como precio-final, peso-promedio, cantidad y hora – entrada, asimismo, se generaron nuevas variables a partir de los datos existentes para mejorar el análisis. También se descompuso la fecha en componentes como mes y día, para poder generar indicadores de tendencia y estacionalidad de los precios. Se realizó la verificación para eliminar registros que siguen teniendo información incompleta en las variables seleccionadas. Este proceso concluyó con la formación de un conjunto de datos estructurados, para realizar un análisis posterior e implementación de un modelo predictivo.

Para estudiar cómo ha cambiado los precios del ganado (Precio-Final), se identificaron dos datos clave: la tendencia (price-trend-weekly) y la estacionalidad (price-seasonal-monthly). Estos datos reflejan cambios a largo plazo y fluctuaciones periódicas. La elaboración de estas variables generó los promedios generales, determinando el precio promedio semanal del ganado y separando la serie de precios semanales en sus componentes de tendencia (dirección general a largo plazo) y estacionalidad mensual (patrones repetitivos mensuales), esto permite acercarse a los modelos IMF, de una manera sencilla y poder ver la efectividad del modelo con los precios más recientes y la

estacionalidad en la que se encuentra.

Descomposición de Precio Promedio Semanal (Estacionalidad Mensual)

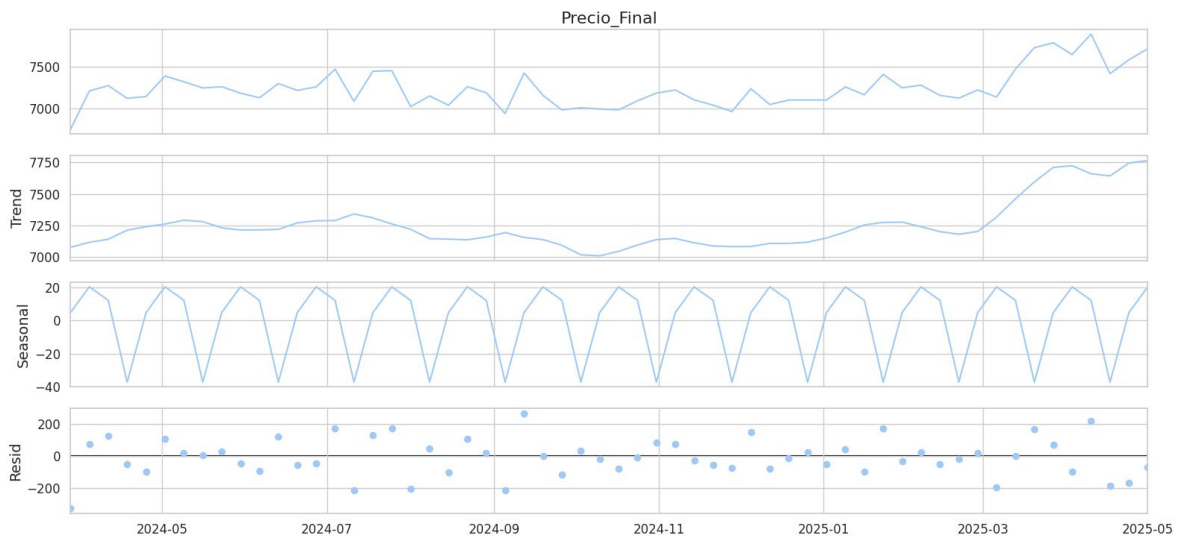


Figura 4 Descomposición de precio promedio semanal (Estacionalidad Mensual). Elaboración propia.

Se determinaron promedios semanales a partir de los precios diarios con el fin de poder suavizar la serie y calcular el tren como se ve en la figura 4. Se completaron valores que faltaban en la serie semanal. La serie de precios promedio semanales (Y_t) se descompuso usando seasonal-decompose bajo un modelo aditivo, que asume $Y_t =$ Tendencia t + Estacionalidad t + Residuo t . Estimación de Tendencia (T_t): Se estimó aplicando una media móvil centrado a la serie Y_t , esto generó la variable price-trend-weekly, lo cual en pocas palabras es el promedio de las últimas dos subastas de la semana. Por otro lado, la estimación de Estacionalidad (S_t): Sobre la serie sin tendencia ($Y_t - T_t$), se determinaron los promedios de los valores para cada punto del ciclo estacional (un período

de 4 semanas), realizando un ajuste para que el resultado de cero en un ciclo y así, generar la variable price-seasonal-monthly.

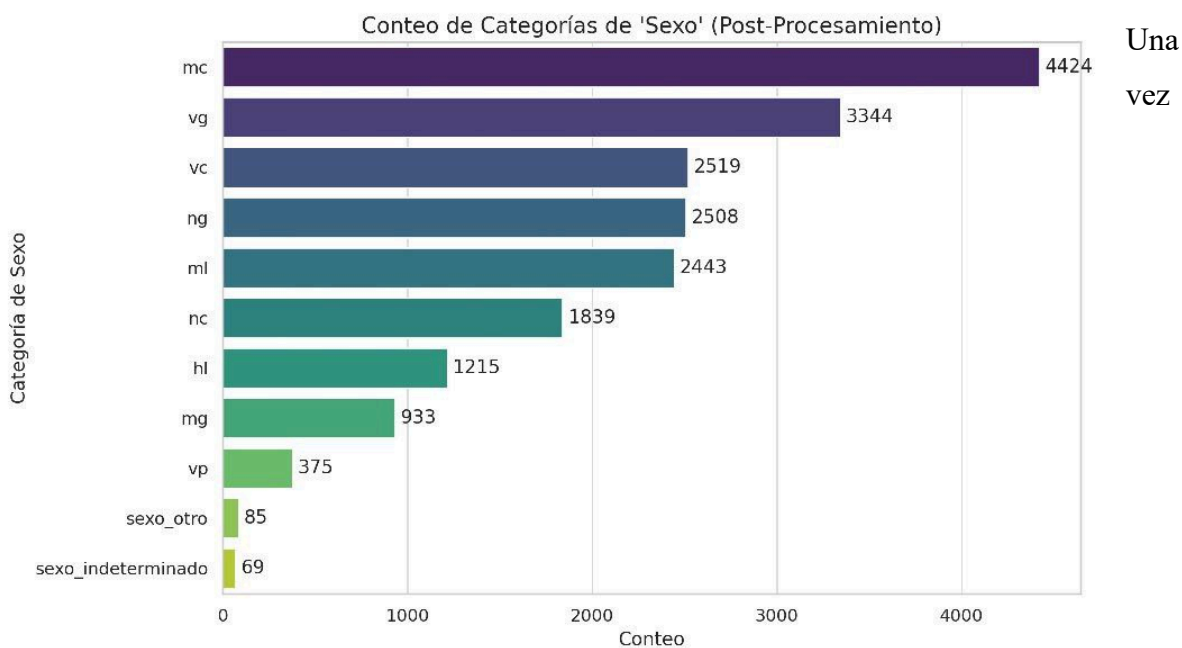


Figura 5 Conteo Categorías de sexo (Post-Procesamiento). Elaboración propia

depurada y consolidada la base de datos, se procedió a analizar las relaciones entre las distintas variables y el Precio-Final. Se identificó que las variables como el Peso-Promedio y el Sexo del animal presentaban una relación importante, así como los componentes temporales y las tendencias generales de los precios, como lo muestra la figura 5.

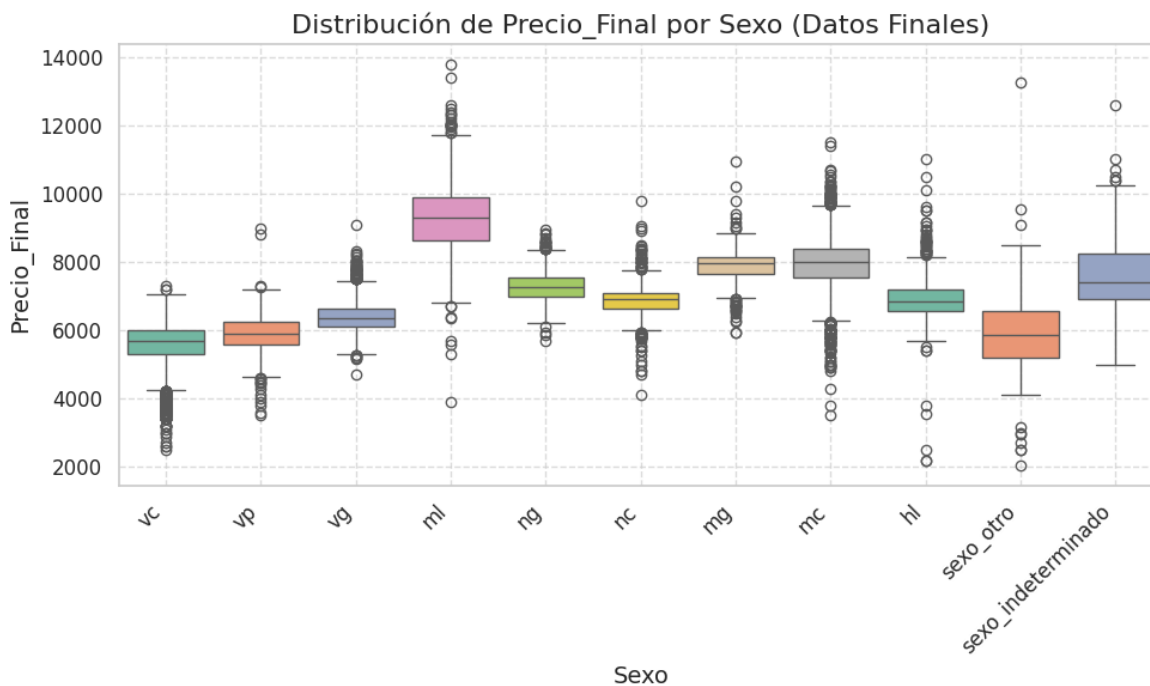


Figura 6 Distribución de Precio-Final por sexo (Datos Finales) Elaboración propia

En la figura 6, se observa diferencias importantes en los niveles de precio mediano como en las diferentes categorías de Sexo. Por ejemplo, la categoría ml (macho de levante) muestra la mediana de Precio-Final más alta (alrededor de 9000) y una gran variación entre las categorías como mc (macho de cebo), mg (macho godó) y sexo-indeterminado, también muestran medianas altas (entre 7500-8000). Al contrario de las categorías vc (vaca de cría), vp (vaca preñada), hl (hembra de levante) y sexo-otro, tienden a tener una mediana de precio inferior (entre 5000-6500) y, en ciertos casos, menor fluctuación interna. Se puede observar varios valores atípicos en la mayoría de las categorías, especialmente en los precios más elevados.

Estas variaciones visuales, complementadas por la figura 7 Matriz de correlación, confirman que el Sexo es un factor diferenciador importante en la composición del Precio-Final.

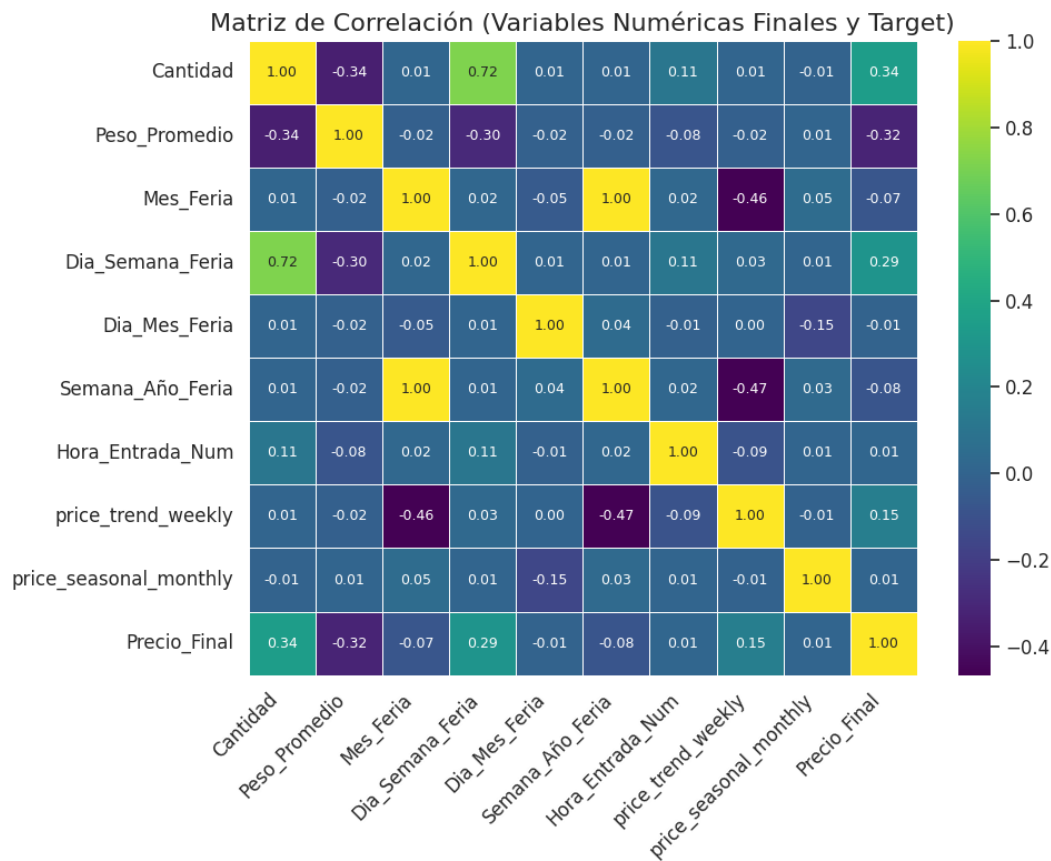


Figura 7 Matriz de correlación, (Variables numéricas finales y target). Elaboración propia

En la Figura 7 Matriz correlación final, se visualiza las correlaciones directas entre variables numéricas claves, incluyendo aquellas directamente generadas a partir de la fecha y los componentes de tendencia y estacionalidad.

La Cantidad de animales por lote presenta una correlación positiva moderada (aproximadamente 0.34) con el precio Final, sugiriendo que lotes con más animales tienden a alcanzar precios finales más altos en conjunto, lo cual tiene sentido ya que generalmente

se busca llenar el cupo del camión que transporta el ganado a la finca del comprador por eso en general lotes más grandes tienen más demanda.

El Peso-Promedio muestra una relación negativa moderada (alrededor de -0.32) con el Precio-Final. Este estudio es inesperado, debido a que generalmente se pensaría que los animales más pesados tienen un precio más alto, sin embargo, como se habla de peso por kilo, los animales como macho de levante (ML) son los más costosos por kilo debido a su gran potencial y calidad genética, esto quiere decir, que no es tan relevante su peso

La Hora-Entrada-Num tiene una correlación mínima, cerca a cero (aproximadamente 0.01) con el Precio-Final, lo que indica que la variable hora de entrada, no parece tener un efecto lineal por sí sola, ni tampoco afecta en el precio final del lote.

La Cantidad y el Peso-Promedio son posibles variables dominantes y predictoras, aunque es necesario investigar a profundidad la relación del Peso-Promedio con el Precio-Final y revisando la variable Hora-Entrada-Num parece tener un impacto menor en la relación con el precio.

3.3. Desempeños modelo machine learning

Para predecir los precios en las subastas ganaderas en Casanare, se analizó los rendimientos en los dos modelos de Machine Learning que son conocidos por su eficiencia en tareas de regresión: Random Forest y XGBoost. El análisis se realizó mediante los siguientes procesos metodológicos.

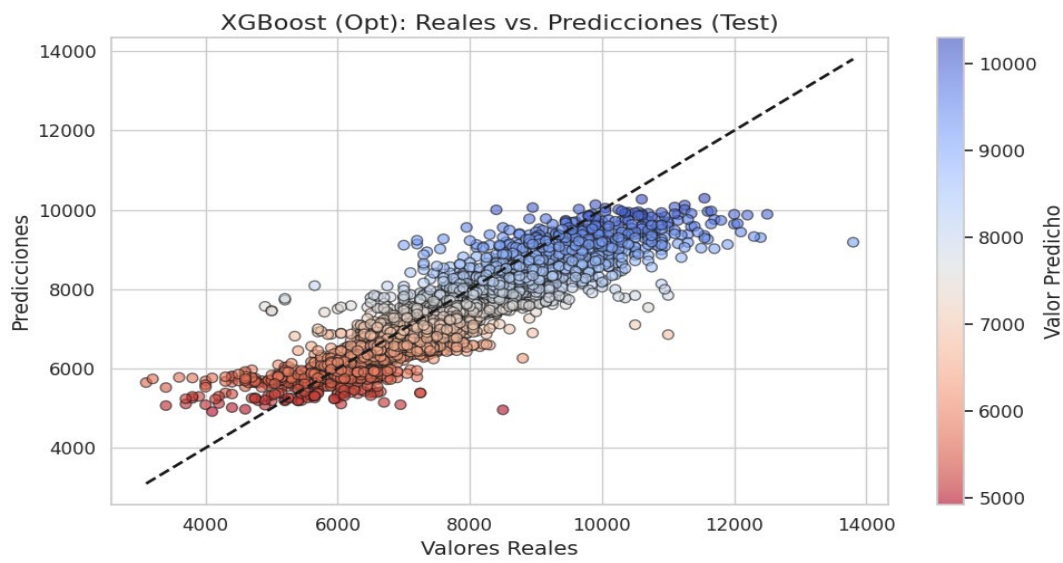


Figura 8 XGBost (Opt) Reales vs Predicciones (Test). Elaboración propia

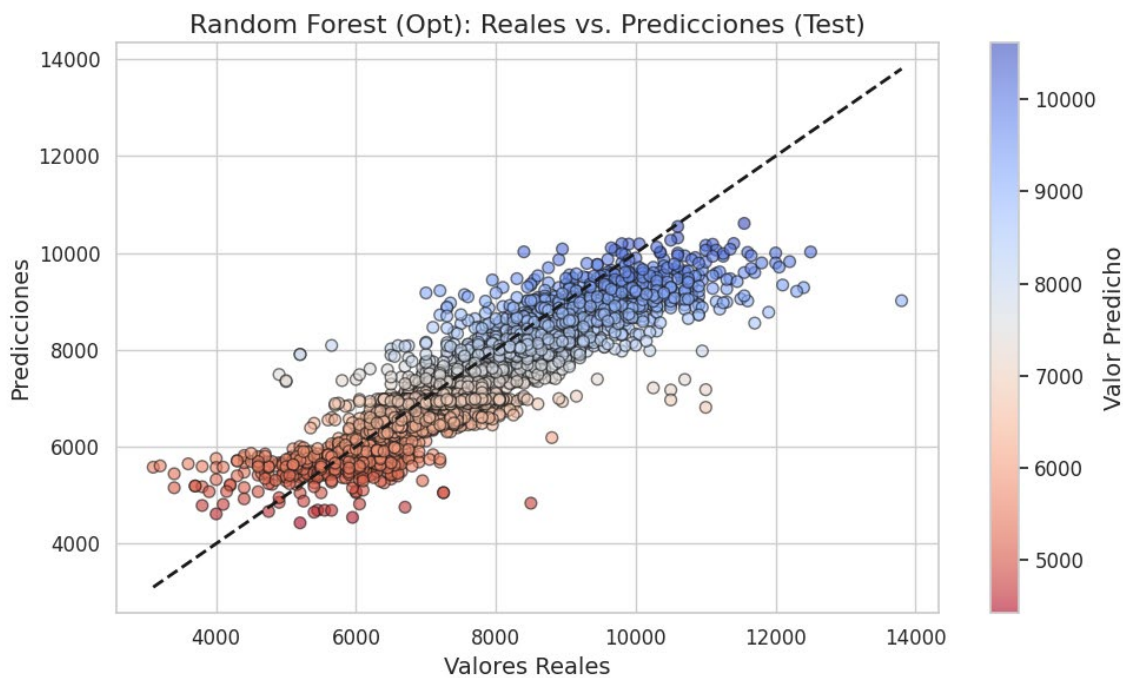


Figura 9 Random Forest (Opt): Reales vs predicciones (Test). Elaboración propia

Primero se realizó la preparación de datos para los modelos, donde se llevó a cabo una división cronológica de la base de datos depurada en un conjunto de entrenamiento (para optimizar los modelos) y un conjunto de prueba (para analizar su capacidad de generalización sobre datos no vistos). Se utilizó un pipeline para preparar el procesamiento que contemplo el ajuste de variables numéricas (StandardScaler) y la transformación de variables categóricas (OneHotEncoder) para garantizar la compatibilidad con los algoritmos.

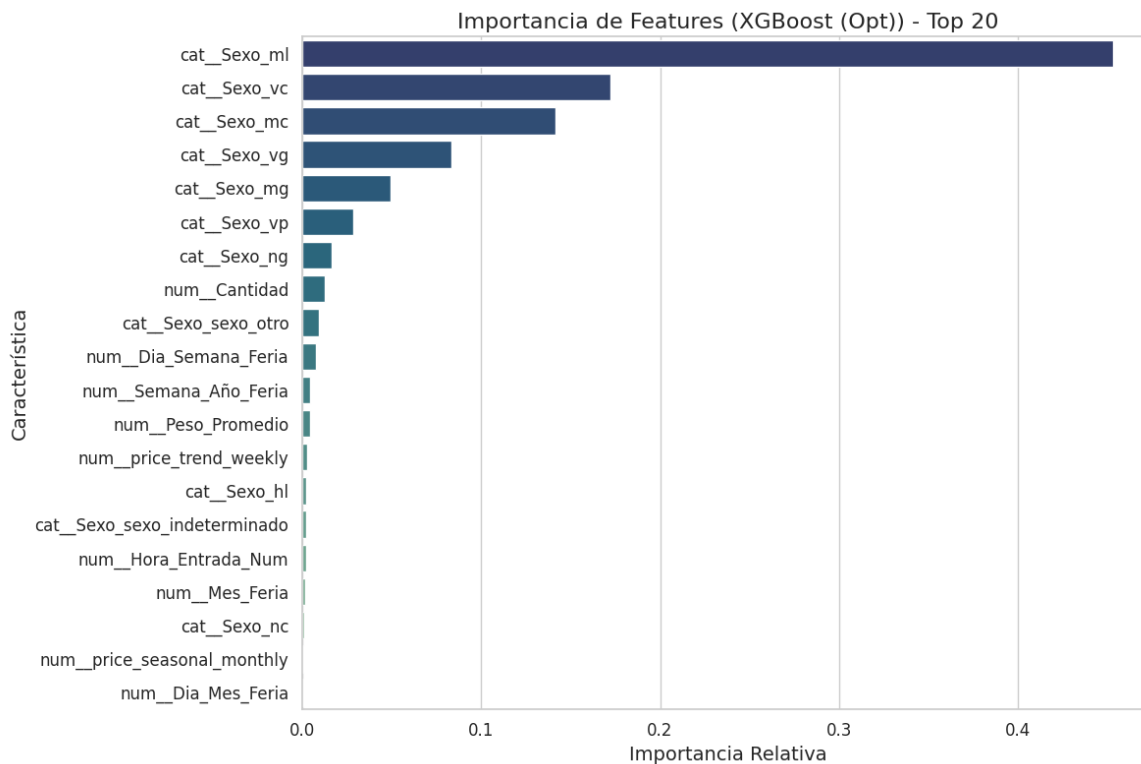


Figura 10 Importancia de Features (XGBoost (Opt)-Top 20). Elaboración propia

Luego siguió el entrenamiento y optimización de los modelos, Random Forest y XGBoost, estos fueron entrenados con el conjunto de datos de entrenamiento. Donde se realizó una búsqueda de los parámetros más adecuados mediante GridSearchCV, junto con una validación cruzada específica para series temporales (TimeSeriesSplit). Esto facilitó la configuración óptima para cada modelo, minimizando el error absoluto medio (MAE) durante el proceso de validación. En cuanto a las métricas, se evalúa los rendimientos de

los modelos optimizados sobre el conjunto de prueba utilizando un conjunto común de métricas de regresión:

Se elaboró una tabla donde compara las métricas de rendimiento de cada modelo en el conjunto de datos de prueba. El modelo que presenta menor Error Absoluto Medio (MAE) es el conjunto de prueba elegido, debido a que contiene el mejor desempeño para la predicción de precios. Adicionalmente, se generaron visualizaciones para un análisis más detallado, que incluye gráficos de valores reales en comparación con los valores predichos.

Este procedimiento sistemático facilitó la determinación del modelo más adecuado y poder cuantificar su capacidad predictiva para los precios de las subastas ganaderas en el departamento del Casanare.

Modelo	MAE	RMSE	R2	MAPE (%)
XGBoost (Opt)	455.74	638.86	0.77	5.97
Random Forest (Opt)	463.05	644.54	0.77	6.05

Tabla 2 Modelo-MAE-RMSE-R2-MAPE. Elaboración propia

Conclusiones

El Desarrollo de este proyecto permitió realizar un análisis basado en la realidad de la comercialización del ganado en las subastas del departamento del Casanare, donde se evidencia la falta de información consolidada para que los ganaderos puedan conocer el comportamiento de los precios dentro de la subasta.

En la investigación, se determinó que los factores como el clima, la estacionalidad, la categoría del ganado y las condiciones del mercado influyen de manera significativa en la formación de precios, adicional, no se cuenta aún con una herramienta eficaz que les ayude a los ganaderos a anticiparse a esas variaciones.

Se logró construir una base de datos organizada y confiable a partir de las diferentes fuentes no estructuradas, realizando procesos de limpieza, validación y estandarización de los datos. Esta base fue fundamental para poder aplicar modelos de aprendizaje de machine learning con el fin de evaluar su capacidad para predecir los precios del ganado con mayor precisión. Comparando los modelos como Random Forest y XGBoost, se pudo evidenciar que es factible contar con un sistema de predicción que no solo sea preciso, si no que pueda servir como herramienta útil para la toma de decisiones comerciales por parte de los ganaderos.

Este trabajo deja una reflexión significativa comprendiendo que la tecnología no transforma el sector agropecuario por si sola, lo que hace la diferencia es la forma como la usamos para responder a las necesidades reales. En este caso, el modelo predictivo que se plantea aparte de buscar optimizar precios busca aportar a la estabilidad económica de una región donde la ganadería es parte de su identidad.

Referencias

Machine Learning for Price Prediction for Agricultural Products . (n.d.). Retrieved from
Machine Learning for Price Prediction for Agricultural Products :
<https://wseas.com/journals/bae/2021/b865107-1274.pdf>

Alfonso Gamboa, J. A. (2024, 12 20). *Una exploración por los avances del Machine Learning que optimizan la agricultura latinoamericana*. Retrieved from Repositorio Institucional UNAD: <https://repository.unad.edu.co/handle/10596/67132>

Bayona-Oré, S., Cerna, R., & Tirado Hinojoza, E. (2021). Machine Learning for Price Prediction for Agricultural Products. *Wseas Transactions on Business and Economics*.

Cheng, L. X. (2023). Forecasting livestock commodity prices using a hybrid deep learning and fuzzy logic framework. *Journal of Agricultural Economics*, 45–68.

Contexto Ganadero. (2024). Retrieved from Contexto Ganadero Sostenibilidad: <https://www.contextoganadero.com/ganaderia-sostenible/por-que-es-importante-la-analitica-de-datos-para-la-ganaderia-colombiana>

Electrónico, C. C. (2017). *CMS Rodriguez-Azuero*. Retrieved from El rol de las tecnologías digitales en el agro: <https://ccce.org.co/wp-content/uploads/2017/06/Una-oportunidad-para-Colombia-Policy-Paper-CMS.pdf>

Electrónico, C. C. (2017). *CMS Rodriguez-Azuero*. Retrieved from El rol de las tecnologías digitales en el agro: <https://ccce.org.co/wp-content/uploads/2017/06/Una-oportunidad-para-Colombia-Policy-Paper-CMS.pdf>

Elicit. (2023). Retrieved from Machine learning aplicado al análisis de un set de datos de parámetros ambientales en galpones de pollos de engorde: <https://elicit.com/notebook/4166bc55-a23f-42c8-80b2-7edd410ddc13#18333673c8ae24f428ef5852308bcdf2>

Elicit. (2024). Retrieved from Predicción de Precios de Ganado en Casanare mediante Machine Learning: <https://elicit.com/notebook/4166bc55-a23f-42c8-80b2-7edd410ddc13#18333673c8b48bde39aea260cbd235c4>

Elicit. (2024). Retrieved from Modelo predictivo basado en algoritmos de machine learning para la estimación del peso de racimos de banano en una hacienda:

<https://elicit.com/notebook/4166bc55-a23f-42c8-80b2-7edd410ddc13#18333673c8cbf29d776abb167f065bca>

Elicit. (2024). Retrieved from Modelo predictivo basado en algoritmos de machine learning para la estimación del peso de racimos de banano en una hacienda:

<https://elicit.com/notebook/4166bc55-a23f-42c8-80b2-7edd410ddc13#18333673c8b48bde39aea260cbd235c4>

Elicit. (2024). Retrieved from Modelo predictivo basado en algoritmos de machine learning para la estimación del peso de racimos de banano en una hacienda:

<https://elicit.com/notebook/4166bc55-a23f-42c8-80b2-7edd410ddc13#18333673c8cbf29d776abb167f065bca>

FAO. (2022). Retrieved from Oficina Regional de la FAO para América Latina y el Caribe:

<https://www.fao.org/americas/regional-initiatives/top-pages/sustainable-livestock-farming-in-latin-america-and-the-caribbean/es>

García, R. &. (2024). Fuzzy logic applications in economic forecasting: An integrative approach. *Expert Systems with Applications*, 114–128.

Hoz, J. V. (2003). *La economía ganadera en el departamento de Córdoba*.

doi:10.32468/dtseru.43

ICA Instituto Colombiano Agropecuario. (2022, 08 04). Retrieved from ICA Instituto Colombiano Agropecuario: <https://www.ica.gov.co/noticias/ica-sinigan-nueva-herramienta-para-ganaderos>

ICA, I. C. (2022, febrero 14). Ganadería en Casanare, de calidad y competitiva en el mundo.

Ma, X. Z. (2025). Interval price prediction of livestock product based on fuzzy mathematics and improved LSTM. *PLOS One*.

Martínez, A. P., Elías-Caro, J. E., & Escobar, Á. E. (2022). Sostenibilidad del sector ganadero colombiano: una aproximación historiográfica. *Revista de Ciencias Sociales*, pp. 28(4), 245–263. Retrieved from *Revista de Ciencias Sociales*, 28(4), 245–263: <https://dialnet.unirioja.es/servlet/articulo?codigo=8703842>

Ministerio de Aricultura y Desarrollo Rural. (2023). Retrieved from Ministerio de Agricultura y Desarrollo Rural: Ministerio de Agricultura y Desarrollo Rural. (2023). Informe de mercado ganadero en Colombia. Bogotá, Colombia

Muñoz, J. S., Díaz, L. M., & Mendoza, B. J. (2018). Las TIC como herramienta de acceso a la comercialización de ganado bovino a través de la subasta ganadera. *Revista de Tecnología*, 17-45-43. doi:<https://doi.org/10.18270/rt.v17i2.3332>

Palacio, C. J. (2025, 02 26). Las subastas ganaderas: más allá del martillo, conozca quién es quién en este negocio. *Agronegocios*, pp. <https://www.agronegocios.co/finca/las-subastas-ganaderas-mas-alla-del-martillo-conozca-quien-es-quien-en-este-negocio-4068979>.

Pre prints Org. (2023, 12). Retrieved from Pre prints Org:
<https://www.preprints.org/manuscript/202312.1090/v1>

Rahmani. (2024). Using Probabilistic Machine Learning Methods to Improve Beef Cattle Price Modeling and Promote Beef Production Efficiency and Sustainability in Canada. *ResearchGate*.

Revistas Um. (2024). Retrieved from Machine Learning en la detección y predicción de enfermedades del ganado:
<https://revistas.um.edu.uy/index.php/ingenieria/article/view/1412>

Rogelio Mijangos-Espinosa, A. M.-R. (2022). *rsc*. Retrieved from Research in Computing Science:

https://rsc.cic.ipn.mx/2022_151_5/Uso%20de%20tecnicas%20de%20Web%20Scraping%20para%20obtencion%20automatica%20de%20bases%20de%20datos%20en%20la%20Web.pdf

Rojas, E. M. (2020). *Proquest*. Retrieved from Machine Learning: análisis de lenguajes de programación y herramientas para desarrollo:

<https://www.proquest.com/openview/c7e24c997199215aa26a39107dd2fe98/1?cbl=1006393&pq-origsite=gscholar>

RTYC. (n.d.). Retrieved from RTYC :

<https://rtyc.utn.edu.ar/index.php/ajea/article/view/1040/936>

SCIELO. (2023). Retrieved from SCIELO:

<https://www.scielo.br/j/ar/a/9ZnXbFS6zFbWWryTGSKbYJb/?lang=en>

Sebastian Raschka, Y. (. (2022). *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning*. Birminham: Packt.

SUBACASANARE. (n.d.). Retrieved from SUBACASANARE:

<https://subacasanare.com/Precios>

Universidad Autonoma del Perú. (2021). Retrieved from Universidad Autonoma del Perú:

<https://repositorio.autonoma.edu.pe/handle/20.500.13067/1687#:~:text=work%20aims%20to%20review%20the,the%20most%20commonly%20used%20algorithms>

Vieto-Vega, M. (2024, 12). Machine Learning en la detección y predicción de enfermedades del ganado: una visión general. *Memoria Investigaciones en Ingeniería*, 27 46 59. Retrieved from Memoria Investigaciones en Ingeniería:

<https://revistas.um.edu.uy/index.php/ingenieria/article/view/1412>

Wang, S. L. (2022). Hybrid modeling of non-stationary time series using CEEMD and LSTM: An application to financial forecasting. *Neural Computing and Applications*, 1564–1575.

wseas. (2021). Retrieved from wseas: <https://wseas.com/journals/bae/2021/b865107-1274.pdf>

Zhao, Q. &. (2023). Attention-enhanced LSTM networks for volatile market prediction. *IEEE Transactions on Neural Networks and Learning Systems*. *IEEE Transactions on Neural Networks and Learning Systems*.