



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA DE NEGOCIOS INTERNACIONALES**

**TEMA:**

**Predicción de la demanda de tabaco en los mercados internacionales mediante  
machine learning.**

**AUTOR:**

**Ortega García, Juan Sebastián**

**TRABAJO DE INTEGRACIÓN CURRICULAR PREVIO A LA OBTENCIÓN  
DEL TÍTULO DE LICENCIADO EN NEGOCIOS INTERNACIONALES**

**TUTOR:**

**Ing. Carrera Buri, Félix Miguel Mgs.**

**Guayaquil, Ecuador**

**23 de agosto del 2024**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA DE NEGOCIOS INTERNACIONALES**

**CERTIFICACIÓN**

Certificamos que el presente trabajo de titulación fue realizado en su totalidad por **Ortega García, Juan Sebastián** como requerimiento para la obtención del título de **LICENCIADO EN NEGOCIOS INTERNACIONALES**.

**Guayaquil, a los 23 días del mes de agosto del año 2024**

f. \_\_\_\_\_

**Carrera Buri, Félix Miguel**

TUTOR

f. \_\_\_\_\_

**Ing. Hurtado Cevallos, Gabriela Elizabeth**

DECANO O DIRECTOR DE CARRERA



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA NEGOCIOS INTERNACIONALES**

**DECLARACIÓN DE RESPONSABILIDAD**

**Yo, Ortega García Juan Sebastián**

**DECLARO QUE:**

El Trabajo de Integración Curricular, **Predicción de la demanda de tabaco en los mercados internacionales mediante Machine Learning**, previo a la obtención del título de Licenciadas en Negocios Internacionales, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de mi total autoría.

En virtud de esta declaración, me responsabilizo del contenido, veracidad y alcance del Trabajo de Titulación referido.

**Guayaquil, a los 23 días del mes de agosto del año 2024**

**EL AUTOR:**

f. \_\_\_\_\_

**Ortega García Juan Sebastián**



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA DE NEGOCIOS INTERNACIONALES**

**AUTORIZACIÓN**

**Yo, Ortega García Juan Sebastián**

Autorizo a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución del Trabajo de Titulación **Predicción de la demanda de tabaco en los mercados internacionales mediante Machine Learning**, cuyo contenido, ideas y criterios son de mi exclusiva responsabilidad y total autoría.

**Guayaquil, a 23 días los del mes de agosto del año 2024**

**EL AUTOR:**

f. \_\_\_\_\_

**Ortega García Juan Sebastián**

# REPORTE COMPILATIO



## Tesis Final - Ortega Garcia Juan Sebastian.

1%  
Textos sospechosos

- 5% Similitudes (ignoradas)
  - < 1% similitudes entre similitudes
  - entre las fuentes mencionadas
- 1% Idiomas no reconocidos
- 30% Textos potencialmente generados por la IA (ignorada)

Nombre del documento: Tesis Final - Ortega Garcia Juan Sebastian.pdf  
 ID del documento: e6a27f52025e5294e2a6a6660b989e5e4b607e  
 Tamaño del documento original: 1,34 MB  
 Autores: []

Depositante: Felix Miguel Carrera Buri  
 Fecha de depósito: 29/8/2024  
 Tipo de carga: Interface  
 fecha de fin de análisis: 29/8/2024

Número de palabras: 20.260  
 Número de caracteres: 141.892

Ubicación de las similitudes en el documento:



### Fuentes principales detectadas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	<a href="http://www.gub.ec">www.gub.ec</a> <a href="http://www.gub.ec/informacion/leyes/leyes-regulaciones/2020-11/ACTUALIZADO-COMIS-LOGANVED-IR-...">http://www.gub.ec/informacion/leyes/leyes-regulaciones/2020-11/ACTUALIZADO-COMIS-LOGANVED-IR-...</a> 8 fuentes similares	1%		Palabras similares: 1% (30 palabras)
2	<a href="http://www.gub.ec">www.gub.ec</a> <a href="http://www.gub.ec/informacion/leyes/leyes-regulaciones/2019-01/Reglamento-Def-Regimen-Oficial-De-...">http://www.gub.ec/informacion/leyes/leyes-regulaciones/2019-01/Reglamento-Def-Regimen-Oficial-De-...</a> 6 fuentes similares	1%		Palabras similares: 1% (30 palabras)
3	Documento de otro usuario <no se sabe> El documento proviene de otro grupo 4 fuentes similares	1%		Palabras similares: 1% (32 palabras)
4	Documento de otro usuario <no se sabe> El documento proviene de otro grupo 1 fuente similar	< 1%		Palabras similares: < 1% (16 palabras)
5	<a href="http://repositorio.ucsg.edu.ec">repositorio.ucsg.edu.ec</a> <a href="http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-CIAC-C16-24.pdf">http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-CIAC-C16-24.pdf</a> 8 fuentes similares	< 1%		Palabras similares: < 1% (122 palabras)

### Fuentes con similitudes fortuitas

Nº	Descripciones	Similitudes	Ubicaciones	Datos adicionales
1	<a href="http://repositorio.ucsg.edu.ec">repositorio.ucsg.edu.ec</a> <a href="http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-ED-CICA-172.pdf">http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-ED-CICA-172.pdf</a>	< 1%		Palabras similares: < 1% (6 palabras)
2	<a href="http://repositorio.ucsg.edu.ec">repositorio.ucsg.edu.ec</a> <a href="http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-AD-954-82.pdf">http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-AD-954-82.pdf</a>	< 1%		Palabras similares: < 1% (32 palabras)
3	<a href="http://repositorio.ucsg.edu.ec">repositorio.ucsg.edu.ec</a> <a href="http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-MS-113.pdf">http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-MS-113.pdf</a>	< 1%		Palabras similares: < 1% (32 palabras)
4	documento.com   <a href="http://repositorio.ucsg.edu.ec">repositorio.ucsg.edu.ec</a>   <a href="http://repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-ED-CICA-172.pdf">repositorio.ucsg.edu.ec/bitstream/handle/11363/6157/UCSG-PRC-ED-CICA-172.pdf</a> <a href="https://documento.com/105-2019-2-21-an-afidit-de-ata-de-las-comisiones-representativas-del-comite...">https://documento.com/105-2019-2-21-an-afidit-de-ata-de-las-comisiones-representativas-del-comite...</a>	< 1%		Palabras similares: < 1% (32 palabras)
5	<a href="https://wikis.fel.com.es/">wikis.fel.com.es</a>   Aprendizaje automático - FdWiki ELP <a href="https://wikis.fel.com.es/5/Plataforma_de_gestio...">https://wikis.fel.com.es/5/Plataforma_de_gestio...</a>	< 1%		Palabras similares: < 1% (36 palabras)



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA DE NEGOCIOS INTERNACIONALES**

**TRIBUNAL DE SUSTENTACIÓN**

f. \_\_\_\_\_

Ing. Carrera Buri, Félix Miguel Mgs.

TUTOR

f. \_\_\_\_\_

Ing. Wendy Arias

OPONENTE

f. \_\_\_\_\_

**Ing. Hurtado Cevallos, Gabriela Elizabeth Mgs.**

DECANO O DIRECTOR DE CARRERA

f. \_\_\_\_\_

**Econ. Freire Quintero Cesar Enrique Mgs.**

COORDINADOR DEL ÁREA O DOCENTE DE LA CARRERA



UNIVERSIDAD CATÓLICA  
DE SANTIAGO DE GUAYAQUIL

**FACULTAD DE ECONOMÍA Y EMPRESA**  
**CARRERA DE NEGOCIOS INTERNACIONALES**

**CALIFICACIÓN**

---

**Ortega García, Juan Sebastián**

## TABLA DE CONTENIDO

1.	Introducción.....	2
2.	Problemática.....	9
3.	Justificación.....	13
4.	Alcance.....	16
5.	Objetivos.....	19
5.1.	Objetivo general.....	19
5.2.	Objetivos específicos.....	19
6.	Marco teórico.....	20
6.1.	Introducción al Machine Learning.....	20
6.1.1.	Historia y evolución.....	23
6.2.	Machine Learning en la industria tabacalera.....	24
6.2.1.	Beneficios y desafíos del Machine Learning en la producción de tabaco 25	
6.2.2.	Aplicaciones específicas de Machine Learning en la industria tabacalera 26	
6.2.3.	Impacto de la tecnología en la industria tabacalera.....	27
6.2.4.	Automatización de procesos.....	28
6.2.4.1.	Perspectivas a largo plazo para la industria tabacalera.....	30
6.2.4.2.	Posibles desarrollos y su impacto en el mercado global.....	31
6.3.	Modelos predictivos y de regresión lineal.....	33
6.3.1.	Modelos predictivos.....	33
6.3.2.	Modelo de regresión lineal.....	34
6.3.2.1.	Métodos de construcción del modelo de regresión simple.....	34
6.3.2.2.	Métodos de construcción del modelo de regresión lineal múltiple	36
6.4.	Aprendizaje automático.....	37
6.5.	Tipos de algoritmos de aprendizaje automático.....	38



6.5.1.	Algoritmo de clasificación.....	38
6.5.2.	Predicción de mercados .....	38
6.5.3.	Algoritmo de valores numéricos continuos .....	39
6.6.	Clustering.....	39
6.7.	Redes neuronales .....	40
6.7.1.	Redes neuronales convolucionales (CNN) .....	40
6.7.2.	Redes neuronales recurrentes (RNN) .....	41
6.8.	Deep Learning .....	42
6.8.1.1.	Aplicaciones de la regresión lineal en la predicción de demanda ..	43
6.8.1.2.	Ventajas y limitaciones de la regresión lineal.....	44
6.8.1.3.	Herramientas y software para modelos de regresión.....	45
6.9.	Cadena de suministro.....	47
6.9.1.	Optimización de inventarios y logística .....	47
6.9.2.	Estrategias de gestión de la cadena de suministro .....	48
6.10.	Optimización de la cadena de suministro mediante la regresión lineal.....	49
6.11.	Análisis competitivo .....	50
7.	Marco legal.....	52
8.	Metodología.....	54
9.	Análisis de resultados .....	59
9.1.	Proceso de R.....	65
10.	Conclusiones.....	72
11.	Bibliografía.....	74
12.	Anexos.....	79

## Índice de figuras

<b>Figura 1.....</b>	<b>59</b>
<b>Figura 2.....</b>	<b>60</b>
<b>Figura 3.....</b>	<b>62</b>
<b>Figura 4.....</b>	<b>63</b>
<b>Figura 5.....</b>	<b>64</b>
<b>Figura 6.....</b>	<b>67</b>
<b>Figura 7.....</b>	<b>68</b>
<b>Figura 8.....</b>	<b>71</b>
<b>Figura 9.....</b>	<b>79</b>
<b>Figura 10.....</b>	<b>79</b>
<b>Figura 11.....</b>	<b>80</b>

## Resumen

El presente trabajo utiliza el concepto de machine learning para mejorar la competitividad de las empresas tabacaleras, basado en esto se desarrolló un modelo predictivo basados en algoritmos de aprendizaje automático, que anticipan fluctuaciones en la demanda y optimizan la cadena de suministro, lo cual es crucial para la toma de decisiones estratégicas en marketing y ventas. La metodología incluyó la recopilación de datos históricos del mercado tabacalero, el procesamiento de estos datos mediante algoritmos de machine learning, y la validación de los modelos predictivos a través de técnicas de regresión lineal y análisis competitivo. Los resultados muestran que el uso de machine learning mejora significativamente la precisión de las predicciones de demanda, permitiendo a las empresas ajustar sus estrategias comerciales y optimizar la gestión de inventarios. Además, la implementación de estas tecnologías facilita la identificación de oportunidades de crecimiento en mercados internacionales. En conclusión, la aplicación de machine learning en la industria tabacalera ofrece un enfoque innovador para predecir la demanda, contribuyendo a la eficiencia operativa y a la competitividad de las empresas en un entorno globalizado.

**Palabras clave:** Industria tabacalera, machine learning, mercados internacionales, predicción de demanda.

## **1. Introducción**

Este proyecto tiene como objetivo implementar tecnología predictiva para impulsar la expansión del tabaco ecuatoriano en mercados internacionales. Estas tecnologías se basan en estrategias para optimizar la cadena de suministro, modelos predictivos de demanda y análisis competitivo. El Machine Learning ha emergido como una herramienta poderosa que desarrolla algoritmos y modelos capaces de aprender de datos y tomar decisiones informadas. En lugar de seguir instrucciones predefinidas, estos algoritmos identifican patrones y tendencias a partir de grandes volúmenes de datos históricos y proyecciones de mercado, permitiendo predecir futuras órdenes de compra y mejorar su precisión a medida que se incorporan más datos.

En este contexto, el Machine Learning juega un papel vital en la transformación digital de las industrias modernas, incluidas las empresas tabaqueras. Los modelos predictivos de demanda generados a través de esta técnica son esenciales para prever fluctuaciones en la demanda de tabaco, lo cual es crucial para la planificación estratégica y la toma de decisiones relacionadas con la producción y comercialización (Siegel, 2017). Esto permite a las empresas ajustar con mayor precisión sus estrategias de marketing y ventas, maximizando así sus oportunidades de crecimiento en mercados internacionales.

Adicionalmente, Shalev y Ben (2014) destacan que el Machine Learning puede optimizar la cadena de suministro mediante una mejor gestión de inventarios y planificación logística. Esto no solo reduce los costos operativos, sino que también incrementa la eficiencia al asegurar la disponibilidad de productos en el lugar y momento correctos. La capacidad de analizar grandes volúmenes de datos en tiempo real permite a las empresas detectar rápidamente cualquier interrupción en la cadena de suministro y tomar decisiones informadas para mitigar sus efectos.

El análisis competitivo se ve significativamente beneficiado por el Machine Learning, ya que mediante el análisis de datos de mercado y de la competencia, las empresas tabacaleras pueden identificar oportunidades y amenazas, lo que les permite desarrollar estrategias más efectivas para posicionarse en el mercado global (Barrueta y Castillo, 2018). Este análisis incluye la evaluación de tendencias de consumo, preferencias de los consumidores y movimientos de los competidores, proporcionando una visión integral que ayuda a las empresas a mantenerse competitivas (Lozada, 2017). Con el uso del Machine Learning, las empresas obtienen información valiosa y detallada que facilita la toma de decisiones estratégicas, fortaleciendo su capacidad para adaptarse y prosperar en un entorno comercial en constante evolución.

A nivel global, la producción de tabaco se concentra principalmente en varias regiones, siendo Asia la líder con un 63% de la producción debido a sus suelos y condiciones climáticas favorables (Velásquez y Meza, 2020). África se sitúa en segundo lugar a nivel mundial, contribuyendo con el 17% de la producción, seguida por América con un 16% (Segovia, 2020). Europa representa solo el 4% de la producción global, aunque España destaca como el tercer mayor productor a nivel mundial.

El tabaco es uno de los cultivos comerciales más importantes a nivel mundial y, aunque no es comestible, tiene un impacto económico significativo en varios países como un producto clave de exportación. La industria tabacalera depende de características específicas de las hojas para satisfacer sus demandas de suministro y contribuir al proceso de industrialización. Es crucial realizar todas las actividades necesarias para mejorar las condiciones que favorecen el crecimiento y desarrollo de los cultivos de tabaco, ya que una gestión agronómica deficiente puede obstaculizar la obtención de hojas de alta calidad destinadas a la industria (Calero et. al., 2019).

En general, las empresas dedicadas a la producción agrícola de tabaco carecen de una cultura empresarial que promueva la gestión estructurada y organizada de sus procesos productivos (Valderrama, 2017). Esta falta de enfoque conlleva rendimientos subóptimos, una competitividad reducida y una disminución en la rentabilidad proyectada. Por tanto, es imperativo que estas empresas se comprometan en una planificación eficaz y en la correcta gestión de los recursos de producción para lograr sus objetivos organizacionales (Bueno, 2021).

La producción de tabaco ha sido una actividad agrícola arraigada en las Américas desde tiempos remotos, comenzando con el cultivo de la *Nicotiana tabacum* L (Calero et al., 2019). Estas hojas se consumen de diversas maneras, generalmente mediante combustión que genera humo. Aunque su consumo está limitado en muchos países debido a sus posibles efectos sobre la salud, el tabaco sigue estando disponible para la venta legal en todo el mundo (Arauz, 2021).

Es interesante notar que el tabaco tiene su origen en las tierras de la cultura maya, donde su consumo data desde el año 2000 a.C. hasta el 987 d.C., en regiones como México, Guatemala y Honduras (Sánchez, 2019). En este contexto, los mayas, expertos navegantes, comerciaban a lo largo del Golfo de México, incluyendo islas del Caribe como Cuba, República Dominicana y Jamaica, llevando productos como cacao, henequén y tabaco, el cual fumaban (Vera, 2020). Para algunos pueblos precolombinos, el tabaco tenía un uso ritual y medicinal, creyendo que su humo podía expulsar los malos espíritus causantes de enfermedades.

Investigaciones de Velásquez y Meza (2020) indican que, a finales del siglo XII, los aztecas conquistaron el territorio maya y adoptaron la costumbre de fumar tabaco, dándole un enfoque más social que religioso. Esta práctica se mantuvo hasta la llegada de los españoles a principios del siglo XVI. Para los habitantes del Viejo Mundo, el tabaco

fue descubierto por marineros españoles que exploraban la isla de Cuba por orden de Colón, encontrándolo en las playas de San Salvador, donde fueron recibidos por nativos con frutas, jabalinas y hojas secas de tabaco.

Inicialmente, el tabaco era consumido principalmente por marineros y se fue extendiendo gradualmente a otros sectores marginales y grupos sociales desfavorecidos. No obstante, a medida que su popularidad crecía, también lo hacía el interés de los grupos de mayor poder adquisitivo, lo que provocó un aumento exponencial en su consumo (Sánchez, 2019). Este auge no pasó desapercibido para los gobiernos, que vieron en el tabaco una oportunidad para generar ingresos importantes mediante la imposición de impuestos y otras regulaciones. La capacidad del tabaco para generar ingresos fiscales significativos llevó a la implementación de gravámenes específicos sobre su producción y venta.

Según Holguín (2021), para maximizar estos ingresos, los gobiernos no solo impusieron impuestos, sino que también establecieron una serie de normativas para controlar y supervisar la producción de tabaco. Estas regulaciones incluyeron medidas estrictas sobre la calidad del producto, las condiciones de cultivo y las técnicas de procesamiento. La industrialización de la fabricación de tabaco facilitó el control gubernamental, permitiendo una supervisión más eficiente de la producción y distribución.

En América Latina, y especialmente en Cuba, la producción de tabaco tiene un impacto económico significativo (Sánchez, 2019). El tabaco no solo es un producto agrícola, sino también una fuente crucial de empleo y divisas, destacándose en la economía de países como Cuba, la República Dominicana y Nicaragua. La producción del tabaco habano, en particular, depende de proveedores de hojas de alta calidad, especialmente en regiones cubanas como Vuelta Abajo, conocida por sus condiciones excepcionales para el cultivo.

Mantener relaciones con estos proveedores es esencial para preservar la calidad y autenticidad del producto. Además, la disponibilidad de recursos como semillas, fertilizantes y técnicas de cultivo sostenibles es vital para asegurar la consistencia del tabaco habano.

En el mercado de tabacos premium, el habano enfrenta la competencia de otros productores de alta calidad, principalmente de la República Dominicana, Nicaragua y Honduras (León et.al., 2020). Aunque los habanos gozan de una ventaja en términos de reputación y tradición, sus competidores han avanzado significativamente en técnicas de cultivo y manufactura, ofreciendo productos que rivalizan en calidad.

Bueno (2021) sostiene que las empresas cubanas deben innovar continuamente en sus estrategias de marketing y distribución para mantener su liderazgo. Distribuidores y minoristas especializados juegan un papel clave en llevar los habanos al consumidor final. La eficiencia en la cadena de distribución y la capacidad para garantizar que los productos lleguen frescos y en óptimas condiciones a los mercados internacionales son aspectos críticos. La participación en ferias y eventos internacionales también es fundamental para la promoción de los habanos.

Las regulaciones internacionales sobre el tabaco, incluidas las restricciones en publicidad, empaquetado y etiquetado, así como las políticas de control, varían de un país a otro y afectan significativamente la industria del habano (Arauz, 2021). Normativas que impactan las exportaciones cubanas, como embargos y restricciones comerciales impuestas por algunos países, también influyen en la estrategia de entrada a mercados internacionales.

El mercado del tabaco habano está influenciado por tendencias globales como el consumo consciente y la preferencia por productos de alta calidad y artesanales. Los consumidores están cada vez más interesados en la trazabilidad del producto, deseando



conocer el origen y el proceso de producción de sus puros. Además, el turismo tabacalero, donde los aficionados visitan plantaciones y fábricas en Cuba, está en crecimiento, promoviendo tanto la cultura del habano como su consumo (León et.al., 2020).

El sector del tabaco habano contribuye significativamente a la economía cubana a través de la generación de empleo y divisas por exportaciones. La industria debe equilibrar el incremento de la producción con la sostenibilidad, asegurando que las prácticas agrícolas y de manufactura no comprometan la calidad del producto ni el medio ambiente (Calero et. al., 2019). Esto implica la adopción de técnicas de cultivo y procesos de manufactura que sean ambientalmente responsables, manteniendo la reputación y calidad del habano en el mercado global.

En Ecuador, el cultivo de tabaco (*Nicotiana tabacum* L.) se extiende por aproximadamente 6,433 hectáreas, favorecido por condiciones climáticas óptimas que permiten el desarrollo de hojas de alta calidad (INEC, 2019). Las provincias costeras de Guayas y Los Ríos son los principales centros de producción. Las empresas tabacaleras ecuatorianas han implementado estrategias específicas en producción, marketing y ventas para lograr un posicionamiento competitivo en el mercado.

La industrialización del tabaco en Ecuador comenzó en 1997, con la exportación de cigarrillos producidos localmente. Un año después, en 1998, marcas ecuatorianas como La Casona, Quevedo y Solemne, de la Compañía Aray & Sons, entraron en el mercado internacional. No obstante, estas empresas enfrentaron grandes desafíos en mercados dominados por alianzas comerciales entre distribuidores y productores centroamericanos, lo que las obligó a buscar socios externos para acceder a los canales de distribución necesarios (Holguín, 2021).

El comportamiento del consumidor ecuatoriano refleja una clara preferencia por ciertas marcas y tendencias de consumo. La popularidad del tabaco ecuatoriano se debe

en gran medida a las características distintivas del producto, que se manifiestan en la calidad del tabaco cultivado en las regiones costeras. Para competir en mercados internacionales como el norteamericano y el europeo, los productores ecuatorianos han adaptado sus estrategias de marketing y ventas, alineándose con las preferencias y demandas específicas de los consumidores extranjeros (Segovia, 2020).

El cultivo y procesamiento del tabaco en Ecuador tienen impactos significativos en los ámbitos ambiental, social y económico. A nivel local, la industria tabacalera es una fuente importante de empleo y contribuye considerablemente a la economía regional (Santillán, 2017). Sin embargo, también enfrenta desafíos ambientales, como el uso intensivo de tierras agrícolas y el impacto de las prácticas de cultivo en el ecosistema. En el plano social, la industria proporciona empleo en comunidades rurales, pero también plantea problemas de salud y bienestar debido al manejo y procesamiento del tabaco.

Las leyes y regulaciones locales que rigen la venta y el consumo de tabaco juegan un papel crucial en la configuración del mercado tabacalero ecuatoriano. Ecuador ha implementado una serie de leyes, convenios y reglamentaciones que reflejan su compromiso con programas integrales para combatir el tabaquismo (INEC, 2019). Estas medidas buscan reducir la demanda de tabaco mediante estrategias como impuestos, educación, comunicación y concientización pública, y controlar la oferta regulando el comercio ilícito, la venta a menores de edad y apoyando alternativas económicas.

La industria tabacalera ecuatoriana ha tenido que adaptarse a estas normativas para competir eficazmente en el mercado global. La introducción de marcas ecuatorianas en mercados internacionales, como el norteamericano y el europeo, demuestra la capacidad de la industria para evolucionar y crecer. A pesar de las dificultades iniciales en un mercado dominado por grandes alianzas comerciales, empresas como Cubanfire y Aray & Sons han logrado éxito mediante asociaciones con distribuidores en Miami (Bueno,

2021). Otros productores han optado por mercados europeos, como España y Alemania, donde la estabilidad de precios y las preferencias de los consumidores ofrecen oportunidades a largo plazo.

## **2. Problemática**

El tabaco, a pesar de no ser un cultivo comestible, se posiciona como uno de los más relevantes a nivel global debido a su impacto económico, particularmente como producto de exportación. La industria tabacalera demanda hojas de una calidad muy específica, lo cual es esencial para añadir valor durante su procesamiento industrial y mantener la competitividad en un mercado global exigente.

A pesar de los conocidos riesgos para la salud asociados con el consumo de tabaco, este sigue siendo una práctica ampliamente extendida, impulsada por la intensa comercialización de cigarrillos durante el siglo pasado (Velásquez y Meza, 2020). Esta alta demanda impone a la industria la necesidad de mantener una producción eficiente y de alta calidad. Sin embargo, se observa una significativa falta de organización y gestión estructurada en los procesos productivos actuales, lo que resulta en bajos rendimientos, poca competitividad y una rentabilidad disminuida. Esto subraya la urgencia de una planificación estratégica y una gestión de recursos más efectiva (Valderrama, 2017; Bueno, 2021).

En el contexto del mercado global, las empresas tabacaleras ecuatorianas enfrentan desafíos considerables en su intento de expansión internacional. A pesar del potencial del tabaco producido en Ecuador, estas empresas se ven frecuentemente limitadas por ineficiencias en la optimización de sus cadenas de suministro, la previsión de la demanda y la ejecución de estrategias de marketing efectivas. La falta de precisión en estas áreas puede traducirse en costos operativos elevados y pérdida de oportunidades de mercado,

lo que destaca la necesidad de mejorar la eficiencia y efectividad en la gestión empresarial.

La industria tabacalera mundial opera bajo un entorno regulatorio cada vez más estricto, con gobiernos que han impuesto impuestos y regulaciones para controlar la producción, asegurar la calidad del tabaco, y regular las condiciones de cultivo y técnicas de procesamiento (Holguín, 2021). En América Latina, países como Cuba, República Dominicana y Nicaragua han desarrollado economías sólidas alrededor de la producción de tabaco premium, destacándose por la alta calidad de sus hojas y las avanzadas técnicas de cultivo (Sánchez, 2019; León et al., 2020). En particular, la industria cubana ha mantenido su liderazgo global gracias a la innovación en marketing y distribución, garantizando la frescura y calidad de sus productos en los mercados internacionales (Bueno, 2021).

Sin embargo, los agricultores familiares, que constituyen una parte importante del sector agrícola, carecen del acceso a las herramientas tecnológicas avanzadas que utilizan los grandes productores. Esto limita su capacidad para competir en igualdad de condiciones, especialmente en la selección de los cultivos más rentables. En este punto, los algoritmos de aprendizaje automático (Machine Learning) ofrecen un potencial considerable para mejorar la predicción de precios y la gestión agrícola, aunque los estudios sobre su aplicación en cultivos no comestibles, como el tabaco, son aún limitados. El sector agrícola en general ha experimentado una falta de avances tecnológicos, y es aquí donde las ciencias de la computación pueden ofrecer herramientas como el big data y el aprendizaje automático para mejorar la eficiencia y sostenibilidad en la producción agrícola.

La tecnología predictiva, y en particular el Machine Learning, se presenta como una solución potencial al permitir el análisis de grandes volúmenes de datos, la identificación

de patrones y la toma de decisiones informadas. No obstante, la adopción de estas tecnologías en la industria tabacalera ecuatoriana ha sido limitada, impidiendo que las empresas locales aprovechen plenamente sus beneficios. Una implementación efectiva de Machine Learning podría transformar la planificación estratégica y operativa, aunque su integración conlleva desafíos técnicos y económicos significativos que deben abordarse.

Además, las empresas tabacaleras ecuatorianas enfrentan la presión de competir en mercados internacionales dominados por productores con una larga tradición y alta calidad. A pesar de contar con condiciones de cultivo favorables, la industria local se ve restringida por la falta de integración de tecnologías avanzadas como el Machine Learning, que podrían optimizar la producción, mejorar la calidad del producto y fortalecer las estrategias de comercialización. La ausencia de inteligencia artificial en la cadena de valor del tabaco ecuatoriano limita la capacidad de estas empresas para predecir la demanda, gestionar eficientemente la cadena de suministro y analizar el comportamiento del mercado, factores cruciales para competir en un entorno global altamente competitivo.

La gran demanda internacional y las oportunidades de crecimiento, la producción de tabaco en Ecuador no ha integrado plenamente tecnologías avanzadas como el Machine Learning. Esta tecnología tiene el potencial de optimizar la cadena de suministro, predecir la demanda y realizar análisis competitivos, lo que podría mejorar significativamente la eficiencia y la competitividad de la industria tabacalera ecuatoriana en el mercado global. Sin embargo, la falta de uso de modelos predictivos y otras herramientas de inteligencia artificial limita la capacidad de las empresas para anticipar tendencias de mercado, ajustar la producción y mejorar las estrategias de marketing.

La industria tabacalera en Ecuador ha tenido que adaptarse a normativas estrictas para competir eficazmente en el mercado global. La introducción de marcas ecuatorianas en

mercados internacionales, como el norteamericano y el europeo, demuestra la capacidad de la industria para adaptarse y crecer. A pesar de las dificultades iniciales en un mercado dominado por grandes alianzas comerciales, empresas como Cubanfire y Aray & Sons han logrado éxito al asociarse con distribuidores de Miami (Bueno, 2021). Esta expansión hacia mercados internacionales resalta la necesidad de adoptar herramientas avanzadas como el Machine Learning en la industria tabacalera ecuatoriana. Aquí se vislumbra la falta de integración de estas tecnologías como una barrera significativa para la optimización de la cadena de valor, la anticipación de tendencias de mercado y la competencia global. La implementación de un modelo predictivo no solo mejoraría la eficiencia operativa y la calidad del producto, sino que también fortalecería la posición de Ecuador en el competitivo mercado global del tabaco.

El propósito de esta investigación es explorar y aprovechar las oportunidades que ofrecen las tecnologías avanzadas, específicamente los modelos predictivos basados en Machine Learning, para la industria tabacalera ecuatoriana. La implementación de estos modelos puede transformar radicalmente la manera en que las empresas tabacaleras gestionan su cadena de suministro, predicen la demanda del mercado y optimizan sus estrategias de marketing. Al integrar estas tecnologías, las empresas pueden mejorar su eficiencia operativa, reducir costos y maximizar la calidad del producto, lo que les permitiría competir más eficazmente en el mercado global del tabaco.

Esto lo fundamenta Siegel (2017) donde en sus investigaciones expone que el uso de modelos predictivos puede proporcionar a las empresas una ventaja competitiva significativa al permitirles anticipar y adaptarse rápidamente a las fluctuaciones del mercado. Esto no solo ayudaría a satisfacer de manera más efectiva la demanda de los consumidores internacionales, sino que también fortalecería la posición de la empresa, en este caso, las empresas tabaqueras de Ecuador, como un productor de tabaco de alta

calidad en el ámbito global. La investigación se centra en identificar las mejores prácticas para la integración de Machine Learning y desarrollar un marco que las empresas puedan seguir para implementar estas tecnologías de manera eficaz y eficiente.

### **3. Justificación**

La industria tabacalera en Ecuador ha sido un pilar significativo de la economía nacional durante varias décadas, con una producción anual que supera las 12,000 toneladas métricas de tabaco habano (ESPAC, 2023). Esta cifra posiciona a Ecuador como uno de los principales productores a nivel mundial. No obstante, para mantener esta posición en el competitivo mercado global, las empresas tabacaleras del país deben continuar innovando y perfeccionando sus procesos. En un entorno de intensa competencia, solo aquellas empresas que se adapten a los avances tecnológicos y respondan a las tendencias del mercado podrán conservar su liderazgo.

En este contexto, la adopción de tecnologías avanzadas como el Machine Learning es esencial para la supervivencia y expansión de la industria tabacalera ecuatoriana. Estas herramientas permiten a las empresas procesar grandes volúmenes de datos, identificar patrones y tendencias, y tomar decisiones informadas que mejoren su competitividad. La integración del Machine Learning en las operaciones de las empresas tabacaleras les permitiría anticipar las fluctuaciones del mercado, ajustar su producción en consecuencia y desarrollar estrategias de marketing más efectivas. Este enfoque no solo les ayudaría a mantener su posición en el mercado, sino también a avanzar hacia una gestión más sostenible y competitiva.

Así, se hace evidente la necesidad de implementar tecnologías predictivas avanzadas, como el Machine Learning, para enfrentar los desafíos actuales y mejorar significativamente la competitividad de la industria tabacalera ecuatoriana en el escenario global. Ecuador, como uno de los principales productores de tabaco habano, cuenta con

una superficie cultivada de 12,811.19 hectáreas (véase figura 1) y una producción anual de 20,862.45 toneladas métricas (véase figura 2), de las cuales el 91% se destina al mercado internacional (ESPAC, 2023). Esta producción considerable posiciona al país como un actor clave en el mercado mundial del tabaco. Sin embargo, la industria enfrenta retos importantes en cuanto a la gestión eficiente y la optimización de los procesos productivos.

Para maximizar el potencial de esta producción y aumentar su competitividad en los mercados internacionales, es fundamental adoptar tecnologías avanzadas como el Machine Learning. Estas herramientas ofrecen soluciones innovadoras para la gestión de la cadena de suministro, la predicción de la demanda y el análisis competitivo. Al integrar el Machine Learning, las empresas tabacaleras de Ecuador pueden anticipar las fluctuaciones del mercado, ajustar la producción según sea necesario y desarrollar estrategias de marketing más efectivas. Esto no solo incrementará la rentabilidad, sino que también reforzará la posición de Ecuador como un proveedor de tabaco habano de alta calidad en el mercado global.

La implementación de tecnologías predictivas avanzadas, como el Machine Learning, es esencial para abordar los desafíos que enfrenta la industria tabacalera ecuatoriana y mejorar significativamente su competitividad en el mercado mundial. Uno de los principales problemas que enfrenta el sector es la falta de estandarización y control de calidad en los procesos de producción, lo que dificulta el mantenimiento de estándares uniformes y la satisfacción de las exigentes demandas de los mercados internacionales. Además, la industria enfrenta la dificultad de adaptarse rápidamente a los cambios en las preferencias de los consumidores, ya que la falta de herramientas avanzadas de análisis de datos les impide anticipar estas tendencias y ajustar su producción y estrategias de manera oportuna. La integración de tecnologías predictivas como el Machine Learning



permitiría a las empresas tabacaleras ecuatorianas analizar grandes volúmenes de datos, identificar patrones y tendencias, y tomar decisiones informadas para mejorar la calidad y consistencia de sus productos, así como para adaptarse con agilidad a las fluctuaciones del mercado. De esta manera, la implementación de estas tecnologías se convierte en una solución estratégica para que la industria tabacalera ecuatoriana mantenga su posición competitiva en el dinámico mercado global del tabaco.

Por otra parte, la capacidad de analizar grandes volúmenes de datos mediante Machine Learning puede contribuir a la mejora en la calidad del tabaco producido, al identificar patrones y tendencias en los datos de cultivo y procesamiento que permitan optimizar las prácticas agrícolas y producir hojas de tabaco de mayor calidad. La adopción de estas tecnologías permitirá a la industria tabacalera ecuatoriana no solo mantener su volumen de producción y ventas, sino también avanzar hacia una gestión más sostenible y competitiva. La capacidad de predecir con precisión las tendencias de demanda en distintos mercados internacionales permitirá a las empresas ajustar sus estrategias de producción y distribución de manera proactiva, asegurando que las necesidades de los consumidores globales sean satisfechas con productos de alta calidad en el momento oportuno.

Además, el uso de Machine Learning puede optimizar la gestión de la cadena de suministro, reduciendo costos operativos y mejorando la eficiencia logística. Con el 91% de la producción dirigida a mercados internacionales, la capacidad de anticipar y mitigar interrupciones en la cadena de suministro es crucial. Las tecnologías predictivas pueden identificar patrones y prever problemas potenciales antes de que ocurran, permitiendo a las empresas tomar decisiones informadas y rápidas para mantener la continuidad del negocio.

El sector agrícola, uno de los más relevantes para la economía, pero a su vez de los más tradicionales de nuestro país debido a diferentes factores, no se ha quedado atrás en esta ola digital. De acuerdo con Tong et. al. (2023) expone que, a pesar del rechazo presentado por parte de los agentes intervinientes en este sector, los beneficios ofrecidos por las herramientas tecnológicas y los incentivos gubernamentales por optar hacia soluciones de carácter digital en la ejecución de las operaciones agrícolas y ganaderas han fomentado la adopción de la tecnología por un gran número de empresas agroalimentarias. Las herramientas digitales, el Big data, la Inteligencia Artificial o el Machine Learning son ya una realidad en este sector. Sin embargo, a pesar de las más que probadas mejoras que estas herramientas ofrecen al sector agroalimentario desde un punto de vista de eficiencia y productividad, existe un importante debate acerca del impacto económico de estas soluciones

La implementación de técnicas predictivas representa una oportunidad estratégica para que la industria tabacalera ecuatoriana aborde los desafíos actuales y mejore significativamente su competitividad en el mercado global (Changshou et. al. 2011). Al integrar estas tecnologías, las empresas podrán analizar grandes volúmenes de datos, identificar patrones y tendencias, y tomar decisiones informadas para optimizar sus procesos y estrategias. Esto les permitirá no solo mantener su posición en el mercado, sino también avanzar hacia una gestión más sostenible y competitiva, fortaleciendo la reputación del tabaco habano ecuatoriano a nivel internacional.

#### **4. Alcance**

En la actualidad, la industria tabacalera ecuatoriana enfrenta desafíos considerables en términos de eficiencia operativa y competitividad global. La falta de integración de tecnologías avanzadas ha limitado la capacidad de estas empresas para adaptarse rápidamente a las demandas del mercado y para optimizar sus procesos de producción y

distribución. Por lo tanto, la introducción de herramientas de Machine Learning representa una oportunidad crucial para mejorar la toma de decisiones estratégicas y operativas, así como para optimizar la gestión de recursos y la planificación de la producción. Es por esto que, este proyecto está dirigido específicamente a la industria tabacalera ecuatoriana, incluyendo a las empresas productoras y exportadoras de tabaco en el país.

La implementación de Machine Learning no solo permitirá una toma de decisiones más informada y ágil, sino que también facilitará la identificación de patrones y tendencias en los datos históricos y en tiempo real. Esto proporcionará a las empresas tabacaleras ecuatorianas una ventaja competitiva al adaptar sus estrategias de producción y comercialización a las fluctuaciones del mercado global.

Los beneficiarios directos de este proyecto incluyen a los ejecutivos y directivos de las empresas tabacaleras ecuatorianas, así como a los investigadores y académicos interesados en el desarrollo tecnológico y la competitividad empresarial. El uso de tecnologías predictivas no solo mejorará la eficiencia operativa de estas empresas, sino que también posicionará al tabaco ecuatoriano como un líder en calidad y eficiencia en el mercado internacional. Además, este proyecto fomentará la colaboración entre la industria y las instituciones académicas, promoviendo la innovación y el desarrollo de nuevas soluciones tecnológicas adaptadas a las necesidades específicas del sector.

La implementación de tecnología predictiva, como Machine Learning, permitirá a las empresas tabacaleras ecuatorianas realizar análisis predictivos de la demanda, optimizar la planificación de la producción y mejorar la gestión de inventarios. Esto no solo reducirá los costos operativos, sino que también mejorará la satisfacción del cliente al asegurar una oferta constante de productos de alta calidad. Además, la capacidad de prever la demanda con mayor precisión permitirá a las empresas ecuatorianas responder de manera

más ágil y efectiva a las fluctuaciones del mercado global, asegurando así su posición competitiva a largo plazo.

La tecnología predictiva, como Machine Learning, también proporcionará a las empresas tabacaleras ecuatorianas capacidades avanzadas de análisis de datos, lo que les permitirá identificar patrones y tendencias ocultas en grandes volúmenes de información. Esta capacidad de análisis avanzado no solo mejorará la eficiencia operativa y la toma de decisiones estratégicas, sino que también impulsará la innovación dentro del sector tabacalero ecuatoriano. La capacidad de analizar grandes cantidades de datos en tiempo real facilitará la identificación de oportunidades de mejora continua y la optimización de procesos clave.

Por tal motivo, este proyecto tiene como objetivo principal mejorar la competitividad y eficiencia operativa de la industria tabacalera ecuatoriana a través de la implementación de tecnología predictiva avanzada, específicamente Machine Learning. Al dirigirse a las empresas productoras y exportadoras de tabaco en Ecuador, este proyecto no solo mejorará la capacidad de adaptación de las empresas a las fluctuaciones del mercado global, sino que también fortalecerá su posición competitiva y promoverá la innovación dentro del sector.

## **5. Objetivos**

### **5.1. Objetivo general**

Predecir la demanda del tabaco en los mercados internacionales mediante machine learning.

### **5.2. Objetivos específicos**

- Examinar literatura existente sobre Machine Learning aplicado a la industria tabacalera y otras industrias relevantes.
- Identificar principales mercados internacionales para la exportación de tabaco habano ecuatoriano.
- Implementar el modelo predictivo desarrollado para optimizar la cadena de suministro de las empresas tabacaleras ecuatorianas.

## 6. Marco teórico

### 6.1. Introducción al Machine Learning

Según Abuduweili et al. (2021), el *Machine Learning* (ML) es una disciplina de la inteligencia artificial que se basa en la idea de que los sistemas pueden aprender de los datos, tomar decisiones con mínima intervención humana e identificar patrones. ML se considera un método de análisis de datos que automatiza la construcción de modelos analíticos. A medida que estos modelos se exponen a nuevos datos, pueden adaptarse de manera autónoma. Los sistemas aprenden de la programación previa para generar decisiones, resultados consistentes y confiables. Los algoritmos de ML se pueden clasificar en diferentes tipos, cada uno con aplicaciones específicas y características particulares.:

**1) Aprendizaje supervisado:** Este algoritmo realiza predicciones en base a un conjunto de ejemplos. Con esta técnica se tiene una variable de entrada que consta de un conjunto de training data etiquetados y una variable de salida deseada (Budd y Robinson, 2019). Para analizar el training data utilizan un algoritmo para aprender la función que asigna la entrada a la salida. Esta función inferida mapea ejemplos nuevos y desconocidos generalizando a partir del training data para anticipar resultados en situaciones no vistas. Estos a su vez tienen la siguiente clasificación según lo explica (Bayona y Rino, 2021):

- **Algoritmos de clasificación:** Cuando los datos se utilizan para predecir una variable categórica, en este caso se asigna una etiqueta o indicador, cuando existe una sola clasificación se denomina binaria y cuando hay más de dos los problemas se llaman de clasificación múltiple (Zhang et. al. 2020).
- **Algoritmos de Regresión:** Cuando se predice valores continuos (Rakhra et. al. 2021).

- **Algoritmos de pronóstico:** En este proceso se basan en situación futuras basadas en datos pasados y presentes, usualmente se utiliza para analizar tendencias (Bueno, 2021).
- 2) **Aprendizaje semi-supervisado:** Permite la utilización de ejemplos sin etiqueta con una pequeña cantidad de datos etiquetados para mejorar la precisión de aprendizaje
  - 3) **Aprendizaje no supervisado:** En este caso la maquina se presenta con datos sin etiquetar. Se le solicita que descubra los patrones intrínsecos que subyacen a los datos, como una estructura de agrupación. Dentro de esta categoría los datos se pueden clasificar en los siguientes:
    - **Agrupamiento:** Según lo explica Subasi (2020) consiste en la agrupación de un conjunto de ejemplos de datos para que los ejemplos en un grupo sean más similares según unos determinados criterios que los de otros grupos. El análisis se puede realizar en cada grupo para ayudar a los usuarios a encontrar patrones intrínsecos.
    - **Reducción de dimensión:** Consiste en la reducción del número de variables consideradas. En muchas aplicaciones, los datos brutos tienen características dimensionales muy altas y algunas funciones son redundantes o irrelevantes para la tarea. Reducir la dimensionalidad ayuda a encontrar la verdadera relación latente (Liakos y Busato, 2018).
  - 4) **Aprendizaje Reforzado:** Esta técnica analiza y optimiza el comportamiento de un agente en función de los comentarios del entorno. Las maquinas prueban diferentes escenarios para descubrir que acciones producen mayor recompensa, en lugar de que se diga que acciones tomar.

Campos (2018) investigó el uso de técnicas de aprendizaje automático, específicamente mediante un modelo de regresión basado en bosques aleatorios, para prever las ventas de medicamentos falsificados. Los hallazgos de su estudio se utilizaron para enfrentar este problema al identificar qué productos probablemente se vendan de manera fraudulenta. El autor ofrece referencias que son fundamentales para proyectos de aprendizaje automático enfocados en la predicción de ventas. Además, Huertas (2017) destaca la importancia de evaluar varios modelos de aprendizaje automático para seleccionar el más adecuado según la problemática a resolver. Con la ayuda de estas técnicas, es posible prever la venta de medicamentos falsificados y contrarrestar este tipo de actividades ilegales, especialmente en países en desarrollo, donde representan una gran amenaza para la industria farmacéutica. ML simplifica la toma de decisiones y ofrece resultados confiables basados en el conocimiento adquirido mediante el análisis de datos.

Por otro lado, Huertas (2017) explica que, es importante evaluar diferentes modelos de aprendizaje automático, y de los resultados se debe elegir el que mejor se ajuste para la solución del problema, en este caso el algoritmo de regresión de bosque aleatorio. Se concluyó que, con la ayuda de técnicas de aprendizaje automático, es posible predecir las ventas de medicamentos falsificados y, por lo tanto, contrarrestar estas actividades ilegales e inseguras en los países en desarrollo, porque se han convertido en una gran amenaza para las industrias farmacéuticas.

El aprendizaje automático es el futuro y la toma de decisiones es más simple, ya que el conocimiento de la máquina es robusto y se puede confiar en los resultados dados por ella con base en el conocimiento previamente adquirido mediante el análisis de los datos (Ackermann, 2019)



### 6.1.1. Historia y evolución

Todas las ciencias están evolucionando con el auge de las nuevas tecnologías informáticas, y el marketing no es la excepción. En la era de la información, gran parte del esfuerzo se centra en gestionar la inmensa cantidad de datos generados, lo que ha dado lugar a subcampos como la inteligencia de negocios, que incluye tecnologías como Big Data y *Machine Learning* (Rajeswari y Suthendran, 2019). Para entender la relevancia de ML en el desarrollo global, es esencial explorar sus orígenes y entender su base, ya que esta herramienta se deriva de la inteligencia artificial.

El concepto de inteligencia artificial fue formalizado en 1943 por Walter Pitts y Warren McCulloch, quienes propusieron analizar el cerebro humano como un sistema computacional. Años más tarde, en 1950, Alan Turing desarrolló el famoso "Test de Turing" para medir la inteligencia de las máquinas. El progreso en ML continuó, con Arthur Samuel desarrollando en 1952 un software que podía aprender a jugar damas, mejorando sus habilidades con cada partida. Este avance marcó el comienzo de una revolución en la capacidad de aprendizaje de las máquinas.

Aunque los primeros avances en inteligencia artificial y ML fueron prometedores, la década de 1970 vio un declive debido a la falta de financiamiento. No obstante, algunos avances clave, como el desarrollo del "Stanford Car" en 1979, mostraron el potencial de ML, utilizando el algoritmo *Nearest Neighbor* para la identificación de patrones, que sentó las bases para el futuro desarrollo de ML. En la década de 1980, la creación de modelos y sistemas expertos revitalizó el campo, con software como el de Gerald DeJong en 1981 que introdujo el concepto de "Explanatory Based Learning" (EBL). Aunque la década de 1990 fue menos prolífica en descubrimientos, hubo avances significativos, como la victoria de la computadora Deep Blue de IBM sobre el campeón mundial de ajedrez Garry Kasparov en 1997. A principios de la década de 2000, ML comenzó a ganar

tracción nuevamente, con empresas como IBM y Microsoft impulsando su desarrollo, culminando en hitos como el lanzamiento de Azure *Machine Learning* por Microsoft en 2008 y el éxito del ordenador Watson de IBM en el concurso Jeopardy en 2011.

Abdel et al. (2016) definen el aprendizaje automático como un proceso que permite identificar patrones complejos en grandes volúmenes de datos obtenidos a través de ejemplos, experiencias o instrucciones predefinidas. La esencia del aprendizaje se basa en un algoritmo que revisa estos datos, posibilitando la predicción de comportamientos futuros a medida que se incorpora nueva información, ajustando así los resultados. Por otro lado, la inteligencia de negocios abarca un conjunto de conceptos, técnicas y herramientas destinadas a analizar y transformar los datos en información útil, lo cual permite tomar decisiones estratégicas, tácticas y operativas de manera más efectiva (Khan y Rahmani, 2018).

## **6.2. Machine Learning en la industria tabacalera**

De acuerdo con Bayona y Rino (2021) señalan que el aprendizaje automático es un proceso continuo que utiliza datos de entrenamiento o experiencias para mejorar progresivamente su rendimiento en una tarea específica o para realizar predicciones precisas. Los datos empleados en el aprendizaje automático consisten en conjuntos de ejemplos, donde cada ejemplo se describe mediante un conjunto de atributos, también conocidos como características o variables.

El aprendizaje automático se clasifica generalmente en aprendizaje supervisado, no supervisado y de refuerzo (Shalev y Ben, 2014). Por otra parte, Durán (2017) explica que, en el aprendizaje supervisado, el modelo utiliza una serie de muestras etiquetadas como datos de entrenamiento para predecir todas las instancias no vistas. En el aprendizaje no supervisado, el modelo recibe únicamente datos no etiquetados y se utiliza principalmente para la reducción de dimensionalidad y análisis exploratorio de datos. El aprendizaje

automático ha ganado popularidad debido a su eficacia en la clasificación y predicción, destacando modelos como los de Regresión, Clustering, Bayesiano, Árbol de Decisión y Redes Neuronales Artificiales.

### **6.2.1. Beneficios y desafíos del Machine Learning en la producción de tabaco**

Al aplicar modelos predictivos y técnicas avanzadas de análisis de datos, las empresas tabacaleras pueden optimizar la cadena de suministro, aumentar la eficiencia en la producción y desarrollar estrategias de marketing más efectivas. Sin embargo, existen desafíos importantes, como la necesidad de contar con infraestructura adecuada, gestionar grandes volúmenes de datos y superar la resistencia al cambio por parte del personal. La implementación de Machine Learning en la industria tabacalera ofrece múltiples beneficios, incluyendo la optimización de la cadena de suministro mediante la mejora en la gestión de inventarios y la eficiencia logística (Effendy et al., 2017).

Los modelos predictivos permiten anticipar la demanda de tabaco, ajustando la producción y el almacenamiento de manera más precisa, lo que reduce los costos operativos. Además, Machine Learning puede aumentar la productividad al automatizar procesos e identificar patrones en el crecimiento de las plantas, permitiendo a las empresas maximizar los rendimientos de cultivo y mejorar la calidad del producto. Los sistemas de Machine Learning pueden monitorear continuamente la calidad del tabaco, identificando y corrigiendo problemas en tiempo real, lo que contribuye a una mejora constante en el producto final (Arauz, 2021).

A pesar de estos beneficios, la implementación de Machine Learning enfrenta desafíos, como la necesidad de inversión en tecnología e infraestructura adecuada, la gestión de grandes volúmenes de datos y la resistencia al cambio dentro de la organización. Además, la seguridad y privacidad de los datos son aspectos críticos que deben ser manejados

cuidadosamente, cumpliendo con normativas y regulaciones sobre el uso y protección de datos (Siegel, 2017).

### **6.2.2. Aplicaciones específicas de Machine Learning en la industria tabacalera**

Según Rajeswari y Suthendran (2019), el Machine Learning ha encontrado aplicaciones específicas en la industria tabacalera, transformando significativamente los procesos de producción y comercialización. Una de las aplicaciones más destacadas es el análisis predictivo para la gestión de cultivos, donde los modelos de Machine Learning utilizan datos históricos y en tiempo real sobre condiciones meteorológicas, calidad del suelo, producción y exportación de materia prima, entre otras variables, para prever los rendimientos de las cosechas y optimizar las prácticas agrícolas.

Esto no solo mejora la eficiencia y productividad, sino que también ayuda a reducir costos operativos y minimizar riesgos asociados con factores impredecibles. Además, los algoritmos pueden analizar patrones de oferta y demanda, predecir fluctuaciones en el mercado y sugerir estrategias para gestionar inventarios de manera más eficiente, lo que es especialmente útil para las empresas tabacaleras que operan en mercados internacionales con condiciones cambiantes. La capacidad de anticiparse y responder a estos cambios proporciona una ventaja competitiva significativa, permitiendo a las empresas satisfacer la demanda sin acumular inventario excesivo (Huy y Asif, 2019).

Los algoritmos pueden detectar anomalías en tiempo real y recomendar ajustes para corregir cualquier desviación, reduciendo el desperdicio y mejorando la eficiencia del proceso de producción. Además, Machine Learning facilita la personalización del producto final, ajustándolo a las preferencias específicas de los consumidores, lo que aumenta la satisfacción del cliente y puede fortalecer la lealtad y repetición de compra, beneficiando la competitividad de la empresa en el mercado (Shalev y Ben, 2014).

### **6.2.3. Impacto de la tecnología en la industria tabacalera**

Según lo explica Barrueta y Castillo (2018), la tecnología ha tenido un impacto profundo en la industria tabacalera, transformando todos los aspectos del negocio desde la producción hasta la comercialización. La adopción de tecnologías avanzadas ha permitido a las empresas tabacaleras mejorar significativamente su eficiencia operativa. Por ejemplo, la automatización de procesos ha reducido la necesidad de mano de obra manual, disminuyendo costos y aumentando la velocidad de producción. Equipos avanzados y robots ahora manejan tareas que antes eran realizadas manualmente, lo que no solo acelera la producción sino también mejora la consistencia y calidad del producto.

La digitalización también ha mejorado la trazabilidad y transparencia dentro de la cadena de suministro de tabaco. Con tecnologías como blockchain y sistemas de rastreo en tiempo real (Lin, 2018), las empresas pueden monitorear cada etapa del proceso de producción y distribución. Esto no solo asegura que los productos cumplan con las normativas de seguridad y calidad, sino que también permite a las empresas responder rápidamente a cualquier problema que pueda surgir, como contaminaciones o problemas logísticos. Esta capacidad de trazabilidad también es crucial para cumplir con las normativas internacionales y ganar la confianza de los consumidores.

En términos de comercialización, la tecnología ha permitido una mayor personalización y segmentación del mercado (Tong et. al., 2023). A través del análisis de datos de consumo y comportamientos de los clientes, las empresas pueden desarrollar estrategias de marketing más precisas y efectivas. Plataformas de marketing digital y análisis de big data permiten a las empresas tabacaleras identificar y dirigirse a segmentos específicos del mercado con ofertas personalizadas. Esto no solo aumenta la efectividad de las campañas de marketing, sino que también mejora la experiencia del cliente y fortalece la lealtad de la marca.

La implementación de tecnologías avanzadas ha permitido a las empresas en general a innovar en sus productos. Nuevos desarrollos, como los dispositivos de tabaco sin combustión y los productos de nicotina de liberación controlada, han surgido gracias a los avances en tecnología. Estas innovaciones no solo responden a las demandas de los consumidores por productos menos dañinos, sino que también abren nuevas oportunidades de mercado para las empresas. La capacidad de desarrollar y lanzar productos innovadores es esencial para mantenerse competitivos en un mercado global en constante cambio.

#### **6.2.4. Automatización de procesos**

La automatización de procesos ha tenido un impacto significativo en la industria tabacalera, mejorando la eficiencia operativa y reduciendo los costos. Al implementar tecnologías avanzadas como robots y sistemas automatizados, las empresas tabacaleras pueden realizar tareas repetitivas y laboriosas de manera más rápida y precisa (Ackermann, 2019). Esto no solo aumenta la velocidad de producción, sino que también mejora la calidad y consistencia del producto final. Además, la automatización reduce la dependencia de la mano de obra manual, lo que puede ayudar a mitigar los desafíos asociados con la escasez de trabajadores y los costos laborales crecientes.

De acuerdo con Kaur et. al., (2014) explica que, en la fase de cultivo, la automatización puede incluir el uso de drones y maquinaria autónoma para realizar tareas como la siembra, el riego y la cosecha. Estos sistemas pueden ser programados para operar de manera eficiente y precisa, asegurando que las plantas reciban el cuidado adecuado en el momento correcto. Además, los sensores y sistemas de monitoreo automatizados pueden recopilar datos en tiempo real sobre las condiciones del suelo y las plantas, permitiendo a los agricultores tomar decisiones informadas y ajustar sus prácticas agrícolas para maximizar el rendimiento y la calidad del tabaco.

Durante el proceso de manufactura, la automatización permite un control más preciso y consistente de los parámetros de producción. Los sistemas automatizados pueden monitorear y ajustar continuamente variables como la temperatura, la humedad y la presión, asegurando que el tabaco se procese de acuerdo con los estándares de calidad establecidos. Esto no solo mejora la calidad del producto final, sino que también reduce el desperdicio y los costos asociados con la producción defectuosa. Además, la automatización permite una mayor flexibilidad en la producción, permitiendo a las empresas cambiar rápidamente entre diferentes productos y lotes según la demanda del mercado.

En la logística y distribución, la automatización puede mejorar la eficiencia y la precisión del manejo y transporte de productos. Los sistemas automatizados de almacenamiento y recuperación pueden organizar y mover productos de manera rápida y precisa, reduciendo los tiempos de manipulación y los errores. Además, los sistemas de seguimiento y monitoreo automatizados pueden proporcionar visibilidad en tiempo real del estado y la ubicación de los productos, mejorando la gestión de inventarios y asegurando que los productos lleguen a los clientes en condiciones óptimas y a tiempo.

La automatización también puede mejorar la sostenibilidad y la responsabilidad social de las empresas tabacaleras. Al reducir el desperdicio y optimizar el uso de recursos, la automatización puede ayudar a las empresas a reducir su impacto ambiental. Además, al mejorar la eficiencia y reducir los costos, la automatización puede permitir a las empresas reinvertir en prácticas más sostenibles y responsables, como el uso de materiales y procesos más ecológicos. Esto no solo mejora la reputación de la empresa y su relación con los consumidores, sino que también puede proporcionar beneficios económicos a largo plazo.

#### **6.2.4.1. Perspectivas a largo plazo para la industria tabacalera**

Según lo explica Kaur et. al., (2014), las perspectivas a largo plazo para la industria tabacalera están marcadas por la continua adopción de tecnologías avanzadas y la evolución de las regulaciones y preferencias de los consumidores. A medida que las empresas tabacaleras siguen implementando tecnologías como el Machine Learning, Big Data, IoT y la automatización, se espera que la eficiencia operativa y la calidad del producto sigan mejorando. Estas tecnologías permiten a las empresas optimizar sus procesos, reducir costos y responder más rápidamente a las demandas del mercado, lo que es crucial para mantenerse competitivas en un entorno global en constante cambio.

Uno de los mayores desafíos a largo plazo para la industria tabacalera será adaptarse a las crecientes regulaciones y restricciones sobre el consumo de tabaco. Las políticas de salud pública y las campañas contra el tabaco están impulsando una disminución en el consumo de productos tradicionales de tabaco en muchos países. Para enfrentar este desafío, las empresas tabacaleras están diversificando sus productos, desarrollando alternativas como los productos de tabaco sin combustión y los dispositivos de nicotina de liberación controlada. Estas innovaciones no solo responden a las preocupaciones de salud, sino que también abren nuevas oportunidades de mercado (Calero et. al., 2019).

La sostenibilidad y la responsabilidad social corporativa también serán factores clave para el éxito a largo plazo de la industria tabacalera. A medida que los consumidores y los reguladores demandan prácticas más sostenibles y responsables, las empresas deberán invertir en tecnologías y procesos que reduzcan su impacto ambiental y mejoren sus prácticas laborales (Abdullah et. al., 2018). La adopción de tecnologías limpias y la implementación de programas de sostenibilidad no solo mejoran la reputación de las empresas, sino que también pueden proporcionar beneficios económicos al reducir costos y mejorar la eficiencia.



Según Sánchez (2019) explica que, la globalización y la digitalización seguirán siendo tendencias importantes que moldearán el futuro de la industria tabacalera. Las empresas deberán adaptarse a los cambios en las dinámicas del mercado global, incluyendo la fluctuación de los precios de las materias primas, las variaciones en la demanda y la competencia internacional. La capacidad de utilizar datos y análisis avanzados para predecir y responder a estos cambios será crucial para el éxito. Además, la digitalización permitirá a las empresas mejorar sus operaciones y ofrecer productos más personalizados y de mayor calidad a sus consumidores.

El papel de la innovación continuará siendo fundamental para la evolución de la industria tabacalera. La inversión en investigación y desarrollo permitirá a las empresas desarrollar nuevos productos y tecnologías que respondan a las necesidades y preferencias cambiantes de los consumidores. La colaboración con instituciones académicas y otros socios industriales también será crucial para fomentar la innovación y mantener la competitividad en el mercado global. En resumen, las perspectivas a largo plazo para la industria tabacalera estarán determinadas por su capacidad para adaptarse, innovar y responder a un entorno dinámico y desafiante.

#### **6.2.4.2. Posibles desarrollos y su impacto en el mercado global**

La industria tabacalera está experimentando una transformación significativa impulsada por la innovación tecnológica y los cambios en las preferencias de los consumidores. Uno de los desarrollos más notables es la introducción y expansión de productos de tabaco de nueva generación, como los dispositivos de tabaco calentado y los productos de nicotina de liberación controlada (Valderrama, 2017). Estos productos están diseñados para ofrecer alternativas menos perjudiciales que los cigarrillos tradicionales, lo que responde a las crecientes preocupaciones de salud pública. A medida que estos productos ganan aceptación, pueden transformar el mercado global al reducir

la prevalencia del consumo de tabaco combustible y abrir nuevas oportunidades para las empresas tabacaleras.

La adopción de tecnologías avanzadas como el Machine Learning y el Big Data continuará desempeñando un papel crucial en la evolución de la industria. De acuerdo con las investigaciones de Effendy et. al., (2017), Subasi (2020) explican que, estas tecnologías permiten a las empresas tabacaleras analizar grandes volúmenes de datos para obtener información valiosa sobre las tendencias del mercado, las preferencias de los consumidores y la eficiencia operativa. Con esta información, las empresas pueden personalizar sus productos y estrategias de marketing para satisfacer mejor las necesidades de sus clientes, optimizar sus procesos de producción y mejorar la gestión de la cadena de suministro. Esta capacidad de adaptarse rápidamente a los cambios del mercado será esencial para mantener la competitividad en un entorno global dinámico.

Asimismo, la automatización y la robótica están cambiando la forma en que se producen y distribuyen los productos de tabaco. La implementación de sistemas automatizados en la manufactura y la logística puede mejorar la eficiencia, reducir los costos y aumentar la consistencia y calidad de los productos. A medida que la tecnología avanza, es probable que veamos una mayor integración de la robótica y la inteligencia artificial en todas las etapas de la cadena de valor del tabaco, desde el cultivo hasta la distribución (Vera, 2020). Esto no solo optimiza las operaciones, sino que también permite a las empresas responder más rápidamente a las fluctuaciones del mercado y las demandas de los consumidores.

La globalización sigue siendo una fuerza importante en el mercado tabacalero. Las empresas tabacaleras están expandiendo sus operaciones a nuevos mercados, especialmente en regiones donde la demanda de tabaco está creciendo. Sin embargo, esta expansión viene con desafíos, como la necesidad de cumplir con regulaciones variadas y

a menudo estrictas en diferentes países. La capacidad de utilizar datos y análisis avanzados para predecir y responder a estos desafíos regulatorios será crucial para el éxito de las empresas en el mercado global. Además, la globalización facilita el intercambio de tecnologías y mejores prácticas, lo que puede impulsar aún más la innovación en la industria.

De acuerdo con Durán (2017) expone que, el papel de la sostenibilidad y la responsabilidad social corporativa se está volviendo cada vez más importante en la industria tabacalera. Las empresas están invirtiendo en tecnologías y prácticas que reduzcan su impacto ambiental y mejoren las condiciones laborales. Esto incluye el uso de materiales más sostenibles, la reducción de emisiones y el desarrollo de programas de reciclaje y gestión de residuos. La adopción de estas prácticas no solo mejora la reputación de las empresas, sino que también puede proporcionar beneficios económicos al reducir costos y aumentar la eficiencia (Huy y Asif, 2019). A medida que los consumidores y los reguladores exigen cada vez más responsabilidad ambiental y social, las empresas que lideren en estas áreas estarán mejor posicionadas para tener éxito en el mercado global.

### **6.3. Modelos predictivos y de regresión lineal**

#### **6.3.1. Modelos predictivos**

Autores como Guo et al. (2018) llevaron a cabo un estudio en el que analizaron seis algoritmos de aprendizaje automático para determinar cuál es más efectivo en la predicción de la clasificación de estudiantes basada en su rendimiento académico. Identificaron que se pueden emplear varios algoritmos de clasificación para esta tarea, tales como Naive Bayes, Regresión Logística, Máquina de Vectores de Soporte, Perceptrón Multicapa, Árbol de Decisión y Bosque Aleatorio. Después de procesar los

datos recolectados, los autores concluyeron que el Perceptrón Multicapa y el Bosque Aleatorio son los más adecuados para predecir el desempeño estudiantil.

### **6.3.2. Modelo de regresión lineal**

De acuerdo con Durán (2017) explica que el objetivo principal al construir un modelo de regresión es evaluar cómo las variaciones en ciertas características (variables independientes) afectan a una característica específica (variable dependiente). Este enfoque es conocido como modelo con fines explicativos. Alternativamente, el objetivo podría ser estimar o predecir el valor de una variable dependiente en función de un conjunto de variables independientes, lo cual se denomina modelo con fines predictivos. Barrueta y Castillo (2018) resaltan que existen diversas técnicas para estimar un modelo de regresión, siendo las más comunes por su simplicidad y facilidad de interpretación la regresión lineal y la regresión logística. Según Abdullah et al. (2018), cuando la variable dependiente es continua, se suele utilizar la regresión lineal, mientras que para variables dicotómicas (como sí/no o hombre/mujer), es más adecuado aplicar la regresión logística.

#### **6.3.2.1. Métodos de construcción del modelo de regresión simple**

La regresión lineal simple es una técnica estadística que se utiliza para modelar la relación entre dos variables: una variable independiente  $x$  y una variable dependiente  $y$ . Por lo que siempre va en función a una hipótesis como se modela a continuación:

$$h_{\theta}(x) = \theta_0 + \theta_{1x}$$

Donde

$h_{\theta}(x)$  = es la predicción de  $y$

$\theta_0$  = es la intersección (término independiente)

$\theta_{1x}$  = es la pendiente de la línea

El método de los mínimos cuadrados consiste en calcular la suma de las distancias al cuadrado entre los puntos reales y los puntos definidos por la recta estimada a partir de

las variables introducidas en el modelo, de forma que la mejor estimación será la que minimice estas distancias. Para poder decidir qué modelo es el que mejor se adecua a los datos de los que disponemos en el modelo de regresión lineal se comparan la F parcial obtenida en cada uno de los modelos de regresión construidos, tal como se muestra en la ecuación 1 (Calero et. al., 2019).

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \epsilon$$

Donde:

$y$  = Variable dependiente

$x_1, x_2, \dots, x_n$  = Variables independientes

$\beta_0$  = Ordenada al origen o intercepto

$\beta_1, \beta_2, \dots, \beta_n$  = Coeficientes de regresión que representan la magnitud y dirección de la relación entre cada variable independiente y dependiente.

A su vez, se pueden estimar sus coeficientes  $\beta_0, \beta_1, \beta_2, \dots$ , utilizando el método de mínimos cuadrados de la siguiente manera:

$$y = X\beta + \epsilon$$

Donde:

$y$  = vector de valores de la variable dependiente

$X$  = matriz de diseño (incluyendo un vector de unos para el término independiente).

$\beta$  = vector de coeficientes a estimar

$\epsilon$  = vector de errores

La solución se obtiene minimizando la suma de los errores cuadrados, lo cual (Kaur et.al., 2014) lo explica a continuación:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

Al aplicar cualquier técnica de selección de variables, el coeficiente se recalcula cada vez que se elimina o introduce una variable, lo que en esencia implica la estimación de

nuevos modelos de regresión, como se menciona en Bayona y Rino (2021). Si se utiliza un paquete estadístico, este proceso suele realizarse automáticamente, a menos que se utilice la técnica de incluir todas las variables, en cuyo caso, la estimación de los modelos debe hacerse manualmente. Kaur et al. (2014) advierten que la inclusión de valores extremos en el modelo de regresión, aunque sean reales, puede distorsionar los resultados debido a que la regresión lineal se basa en el método de mínimos cuadrados, el cual depende del cálculo de distancias entre puntos. Por lo tanto, es crucial considerar estos valores al ajustar el modelo, estimando uno que los incluya y otro que los excluya, para decidir cuál se ajusta mejor a los propósitos del estudio.

### 6.3.2.2. Métodos de construcción del modelo de regresión lineal múltiple

La regresión lineal múltiple es una extensión de la regresión lineal simple que permite la predicción de una variable dependiente  $y$ , utilizando múltiples variables independientes  $x_1, x_2, \dots, x_n$ . A su vez, tal como la regresión simple, va en función de una hipótesis que se expresa de la siguiente manera:

$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \dots + \theta_n x_n$$

Donde

$h_{\theta}(x)$  = es la predicción de  $y$

$\theta_0$  = es la intersección (término independiente)

$\theta_1, \theta_2, \dots, \theta_n$  = son los coeficientes (pendientes) que multiplican a cada variable independiente  $x_1, x_2, \dots, x_n$ .

Sin embargo, la función de costo para la regresión lineal múltiple es el error cuadrático medio (MSE), que mide la diferencia promedio entre las predicciones  $h_{\theta}(x)$  y los valores reales  $y$ , tal como lo expone Kaur et. al. (2014) se expresa de la siguiente forma:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Donde:

$m$  = es el número de muestras

$x^{(i)}$  = es el vector de características de la  $i$  -ésima muestra

$y^{(i)}$  = es el valor real de la  $i$  -ésima muestra

Sin embargo, para minimizar la función de costo  $J(\theta)$ , se utiliza el algoritmo de gradiente descendente. Los parámetros  $\theta$  se actualizan iterativamente:

$$\theta_j = \theta_j - \alpha \frac{\partial J(\theta)}{\partial \theta_j}$$

Donde:

$\alpha$  = tasa de aprendizaje

$\frac{\partial J(\theta)}{\partial \theta_j}$  = gradiente de la función de costo con respecto a  $\theta_j$

De tal manera que, los parámetros son:

$$\theta_0 = \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$\theta_1 = \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_1^{(i)}$$

$$\theta_2 = \theta_2 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})x_2^{(i)}$$

#### 6.4. Aprendizaje automático

Según Abuduweili et al. (2021) y Cui et al. (2018), el aprendizaje automático, un subcampo de la inteligencia artificial, se centra en el desarrollo de algoritmos que permiten hacer predicciones precisas a partir de datos. Este proceso de aprendizaje se basa en la mejora continua a través de datos de entrenamiento o experiencia, utilizando ejemplos que se describen mediante un conjunto de atributos, también conocidos como

características o variables. El aprendizaje automático se clasifica en aprendizaje supervisado, no supervisado y de refuerzo. En el aprendizaje supervisado, el sistema utiliza datos etiquetados para predecir instancias desconocidas, mientras que, en el aprendizaje no supervisado, trabaja con datos sin etiquetar para realizar predicciones. Este último se emplea principalmente en la reducción de dimensionalidad y análisis exploratorio. Loola et al. (2020) señalan que los modelos más conocidos en este ámbito son la regresión, el clustering, los modelos bayesianos, los árboles de decisión y las redes neuronales artificiales.

## **6.5. Tipos de algoritmos de aprendizaje automático**

El Machine Learning comprende una variedad de algoritmos diseñados para diferentes tareas, desde clasificación y regresión hasta agrupamiento y redes neuronales.

### **6.5.1. Algoritmo de clasificación**

Los algoritmos de clasificación se utilizan para predecir una variable categórica o etiqueta. Este tipo de problemas implica un conjunto de datos de entrenamiento con variables de entrada y una variable de salida que es discreta, como lo describe Lin (2018). Algunos ejemplos de algoritmos de clasificación incluyen Naive Bayes, Máquinas de Vectores de Soporte, y Redes Neuronales, entre otros.

### **6.5.2. Predicción de mercados**

La predicción de mercados se apoya en algoritmos de aprendizaje automático para anticipar tendencias futuras en los mercados financieros, siendo cruciales para la toma de decisiones en comercio e inversión. Chi et al. (2014) explican que entre los algoritmos más utilizados se encuentran los modelos de regresión, empleados para prever el valor numérico de una variable dependiente en función de una o más variables independientes, y las redes neuronales artificiales, que son modelos inspirados en el cerebro humano y



aplicados para predecir tendencias de mercado basadas en análisis de sentimiento y patrones históricos. Siendo el modelado para una red neuronal de regresión el siguiente:

$$\hat{y} = \sum_{j=1}^h v_j \sigma \left( \sum_{i=1}^n w_{ij} x_i + b_j \right) + c$$

Donde:

$x_i$  = Características de entrada

$w_{ij}$  = Pesos de las conexiones entre la capa de entrada y la de salida

$b_j$  = Sesgos de las neuronas en la capa oculta

$\sigma$  = Función de activación

$c$  = Sesgo de la neurona en la capa de salida

### **6.5.3. Algoritmo de valores numéricos continuos**

Lin (2018) destaca la utilidad de estos algoritmos en la predicción de demanda y la optimización de la cadena de suministro, permitiendo modelar relaciones entre variables que influyen en la demanda de productos y la eficiencia operativa. En la predicción de demanda, los modelos de regresión ayudan a estimar la cantidad de productos o servicios requeridos en el futuro, basándose en datos históricos y otras variables relevantes. Esto permite a las empresas planificar mejor la producción y gestionar inventarios.

Además, en la optimización de la cadena de suministro, estos modelos permiten a los fabricantes prever la demanda de materias primas, mejorando la eficiencia y reduciendo costos, como lo menciona Sánchez (2019).

## **6.6. Clustering**

El clustering o agrupamiento es una técnica de Machine Learning utilizada para agrupar conjuntos de datos similares en clústeres. Estos algoritmos ayudan a descubrir estructuras ocultas en datos no etiquetados. Algunos ejemplos de algoritmos de clustering los describe (Ackermann, 2019):

- **K-Means:** Algoritmo de agrupamiento que particiona  $n$  observaciones en  $k$  clústeres donde cada observación pertenece al clúster con el valor medio más cercano. Usado para la segmentación de clientes en grupos de comportamiento similares.
- **DBSCAN:** Algoritmo de agrupamiento que agrupa puntos que están cerca unos de otros en regiones de alta densidad. Comúnmente utilizado para detección de anomalías en datos.
- **Clustering jerárquico:** Agrupa los datos en un árbol jerárquico de clústeres. Generalmente empleado para el análisis de secuencias genéticas para identificar similitudes.

## 6.7. Redes neuronales

Las redes neuronales son modelos computacionales inspirados en el sistema nervioso central de los animales. Estos modelos se utilizan para reconocimiento de patrones, clasificación y aproximación de funciones (Cui et. al., 2018).

### 6.7.1. Redes neuronales convolucionales (CNN)

Las redes neuronales convolucionales (CNNs) son un tipo especializado de redes neuronales artificiales ampliamente utilizadas en tareas como el análisis y reconocimiento de imágenes. Según Khan y Rahmani (2018), aunque las CNNs funcionan de manera similar a las redes neuronales convencionales, su diferenciación principal radica en su capacidad para aprender patrones específicos en imágenes. Las capas convolucionales dentro de estas redes son esenciales, pues funcionan como un conjunto de filtros que se aplican a diferentes partes de una imagen, generando un mapa de características en la salida. Arauz (2021) menciona que este proceso de aplicar filtros a secciones específicas de la imagen facilita su procesamiento y análisis., se puede expresar matemáticamente de la siguiente manera:

$$y = W^T x + b$$

Donde:

$x$  = Vector de característica aplanado

$W$  = Pesos de la capa completamente conectada

$b$  = Sesgo

$y$  = Salida (puede ser una clase o un valor continuo)

Es así que, las primeras capas de estas NNs se encargan de identificar los patrones más básicos como puede ser una línea recta vertical, o cierta curva, y a medida que cada capa pasa su output a la siguiente, se van identificando patrones cada vez más complejos a partir de los anteriores y así hasta llegar a identificar cosas más abstractas, como puede ser un gato en una imagen (Guo et. al., 2018).

### **6.7.2. Redes neuronales recurrentes (RNN)**

Las redes neuronales recurrentes (RNNs), otra variante de las redes neuronales profundas (DNNs), destacan por sus aplicaciones en áreas como el reconocimiento de escritura a mano, la composición musical y el reconocimiento de voz. Según Tong et al. (2023), las RNNs poseen una arquitectura que les permite realizar predicciones futuras, lo que las hace aptas para tareas creativas como la composición musical. Una de sus características distintivas es que, a diferencia de otras redes, las neuronas recurrentes envían su salida no solo a la siguiente neurona, sino también a sí mismas. Esto crea un mecanismo de retroalimentación que permite a las neuronas recibir tanto la entrada de la neurona anterior como su propio estado previo., tal como se explica en la siguiente ecuación:

La salida en cada paso temporal  $t$  puede calcularse como:

$$y_t = W_{hy}h_t + b_y$$

Donde:

$y_t$  = Salida en el tiempo  $t$

$W_{hy}$  = Pesos de la capa oculta a la capa de salida

$b_y$  = Sesgo

Este mecanismo de retroalimentación en las redes neuronales recurrentes es lo que les otorga una memoria, permitiéndoles recordar y aprender dinámicas temporales complejas en secuencias, como explican Budd y Robinson (2019). Estas redes, denominadas redes neuronales de retropropagación (backward neural networks), pueden así procesar información secuencial y mantener la información durante varias épocas, lo que es crucial para tareas que dependen de la secuencia de datos.

## 6.8. Deep Learning

El aprendizaje profundo (Deep Learning) es una rama avanzada del aprendizaje automático que ha transformado la capacidad de las máquinas para aprender y ejecutar tareas complejas. A diferencia de los modelos tradicionales de aprendizaje automático, que requieren la intervención humana para seleccionar y extraer manualmente las características relevantes, el Deep Learning permite a las máquinas aprender automáticamente estas características a través de la construcción de representaciones jerárquicas. Subasi (2020) destaca que las redes neuronales profundas pueden aprender patrones de datos no estructurados o de alta dimensionalidad, como imágenes, audio y texto. Un ejemplo notable es el reconocimiento de imágenes y videos, donde las CNNs, un tipo específico de red profunda diseñada para procesar datos en malla, son capaces de identificar objetos con alta precisión. Esta capacidad ha sido aplicada en sistemas de reconocimiento facial, diagnósticos médicos basados en imágenes, y conducción autónoma., por lo que se expone la siguiente modelo:

$$y_k = \sum_{j=1}^{h_L} w_{jk}^{(L+1)} a_j^{(L)} + b_k^{(L+1)}$$

Donde:

$L$  = Número de capas ocultas

$w_{jk}^{(L+1)}$  = Pesos de la conexión entre la neurona  $j$  en la última capa oculta y la neurona  $k$  en la capa de salida

$b_k^{(L+1)}$  = Sesgo de la neurona  $k$  en la capa de salida

Otra área en la que el Deep Learning ha mostrado un avance significativo es en el procesamiento del lenguaje natural (NLP). A través de redes neuronales recurrentes (RNNs) y modelos de atención, el Deep Learning ha permitido grandes avances en la traducción automática, análisis de sentimientos, generación de texto y respuestas a preguntas. Estos avances han dado lugar a la creación de asistentes virtuales como Siri, Alexa y Google Assistant, que son capaces de comprender y responder comandos de voz de manera eficaz (Zhang et al., 2020).

#### **6.8.1.1. Aplicaciones de la regresión lineal en la predicción de demanda**

La regresión lineal es una técnica estadística fundamental que se utiliza para modelar la relación entre una variable dependiente y una o más variables independientes. En el contexto de la producción de tabaco, la regresión lineal se aplica eficazmente para la predicción de la demanda, permitiendo a las empresas anticipar las necesidades del mercado y ajustar sus operaciones en consecuencia (Budd y Robinson, 2019)

La predicción de la demanda es crucial para la planificación y gestión de la producción en cualquier industria, y el sector tabacalero no es una excepción. La regresión lineal ayuda a identificar patrones históricos en las ventas de tabaco y otros factores relevantes que influyen en la demanda, como las tendencias de consumo, la estacionalidad, las campañas de marketing, los precios y la competencia.

- **Análisis de datos históricos:** El análisis de datos históricos implica la recopilación y examen de datos pasados para identificar patrones y tendencias que

puedan ser utilizados para hacer predicciones futuras. En el contexto de la producción de tabaco, esto incluye la evaluación de datos de ventas anteriores, precios, promociones y otros factores relevantes que afectan la demanda (Ackermann, 2019).

- **Modelado predictivo:** El modelado predictivo utiliza técnicas estadísticas para crear modelos que pueden prever resultados futuros basados en datos históricos. En el caso de la demanda de tabaco, los modelos predictivos ayudan a anticipar cuánto tabaco se necesitará en un período futuro, basándose en factores como el comportamiento del mercado y las tendencias de consumo (Arauz, 2021).
- **Validación y ajuste del modelo:** La validación y el ajuste del modelo son pasos cruciales para garantizar la precisión y confiabilidad del modelo predictivo. La validación implica probar el modelo con datos que no se utilizaron durante su desarrollo para ver cómo se desempeña, mientras que el ajuste del modelo implica modificarlo para mejorar su precisión. Los modelos predictivos se validaron utilizando técnicas como la validación cruzada y el análisis de residuos. Se ajustaron los modelos para mejorar su precisión y capacidad predictiva, asegurando que las predicciones sean lo más exactas posible (Campos, 2018).

#### **6.8.1.2. Ventajas y limitaciones de la regresión lineal**

La regresión lineal es una técnica estadística popular por su sencillez y efectividad. Bayona y Rino (2021) destacan que una de sus principales ventajas es su capacidad para modelar de forma clara la relación entre variables dependientes e independientes. Esto facilita la interpretación de los resultados y la comprensión de cómo las variables dependientes responden a los cambios en las independientes. Además, su eficiencia computacional permite su aplicación en una amplia gama de contextos, desde investigaciones académicas hasta aplicaciones industriales. Ackermann (2019) señala que

la regresión lineal permite identificar y medir la magnitud y dirección de las relaciones entre variables, lo cual es especialmente beneficioso en la producción y comercialización de tabaco, donde entender los factores que afectan la demanda puede optimizar la producción y estrategias de marketing. Por ejemplo, aplicando la regresión lineal a datos históricos de producción y venta de tabaco, las empresas pueden prever variaciones en la demanda y ajustar sus operaciones. Calero et al. (2019) también destacan la utilidad de estas técnicas en la agricultura, donde la predicción precisa puede mejorar la eficiencia y rentabilidad de los cultivos. Sin embargo, una limitación importante de la regresión lineal es la suposición de que la relación entre variables es lineal, lo que no siempre es el caso en situaciones reales (Ackermann, 2019). Además, la regresión lineal es sensible a los valores atípicos, que pueden distorsionar los resultados. Por ello, es fundamental realizar un análisis de datos exhaustivo y considerar métodos más robustos o no lineales si es necesario (Zhang et al., 2020).

Por otro lado, la regresión lineal enfrenta dificultades en el manejo de la multicolinealidad, que ocurre cuando las variables independientes están altamente correlacionadas entre sí. Esto puede complicar la estimación precisa de los coeficientes y disminuir la fiabilidad de las conclusiones del modelo (Rajeswari y Suthendran, 2019). Aunque es útil para identificar relaciones entre variables, la regresión lineal no establece causalidad. Por lo tanto, sus resultados deben interpretarse con cautela y complementarse con análisis adicionales y pruebas empíricas. Campos (2018) enfatiza que, en estudios predictivos más complejos, como la exportación de productos agrícolas, es crucial considerar múltiples métodos y enfoques para obtener una visión más completa y precisa.

### **6.8.1.3. Herramientas y software para modelos de regresión**

Actualmente, hay una variedad de herramientas y software que facilitan la implementación y el análisis de modelos de regresión, permitiendo realizar análisis

complejos con mayor eficiencia y precisión. Google Sheets, por ejemplo, es una aplicación web que ofrece robustas capacidades para el análisis de datos y la regresión lineal. Esta plataforma permite realizar cálculos estadísticos avanzados mediante funciones integradas y complementos específicos, como "XLMiner Analysis ToolPak", que simplifica el análisis de regresión y otras técnicas estadísticas. Además de Google Sheets, Microsoft Excel Online también incluye herramientas avanzadas para el análisis de datos, incluyendo regresión lineal mediante el complemento Analysis ToolPak. Estas plataformas no solo permiten realizar análisis detallados, sino que también facilitan la colaboración y el acceso compartido a datos y resultados, lo cual es crucial en entornos de trabajo colaborativos.

Otra herramienta destacada es Tableau, conocida principalmente por sus capacidades de visualización de datos, que también ofrece funciones de análisis estadístico, incluyendo la regresión lineal. Tableau permite integrar datos de diversas fuentes y aplicar análisis de regresión lineal, proporcionando una visualización clara de los resultados. Además, herramientas especializadas como IBM SPSS Statistics y SAS son ampliamente utilizadas en la industria y la academia por sus potentes capacidades de análisis y manejo de grandes volúmenes de datos. Estas plataformas ofrecen interfaces intuitivas y funcionalidades extensivas para el análisis de regresión lineal y otros métodos estadísticos. En el ámbito de la programación y análisis avanzado, Python y R son esenciales. Python, con bibliotecas como scikit-learn y statsmodels, y R, con el paquete lm para modelos lineales, proporcionan herramientas sofisticadas para el desarrollo y análisis de modelos de regresión. La versatilidad de estos lenguajes los hace adecuados para una amplia gama de aplicaciones, desde la investigación académica hasta el desarrollo industrial.



## **6.9. Cadena de suministro**

Según Arauz (2021), la cadena de suministro en la industria tabacalera abarca todos los procesos que transforman la materia prima en productos terminados y los entregan a los consumidores finales. Este sistema complejo incluye desde la producción y procesamiento del tabaco hasta su distribución y venta. Una gestión eficiente de la cadena de suministro es crucial para asegurar que los productos lleguen a los consumidores de manera oportuna y a un costo razonable. Una cadena de suministro bien gestionada permite reducir costos operativos, minimizar desperdicios y optimizar el uso de recursos, lo que se traduce en menores costos de producción y distribución, mejorando los márgenes de beneficio. Además, garantiza que los productos estén disponibles cuando y donde los consumidores los necesitan, aumentando la satisfacción del cliente. La capacidad de adaptarse rápidamente a la demanda del mercado y a las preferencias cambiantes es fundamental para mantener la competitividad (He y Chen, 2018). También, una cadena de suministro robusta puede adaptarse a cambios en la demanda y a interrupciones en el suministro, lo que es esencial para la continuidad del negocio en un entorno global dinámico. Finalmente, la optimización de la cadena de suministro puede ofrecer precios más competitivos y mejorar la posición en el mercado (Changshou et al., 2011)

### **6.9.1. Optimización de inventarios y logística**

La optimización de inventarios y logística es clave para una gestión efectiva de la cadena de suministro. La gestión de inventarios implica controlar y supervisar los niveles de stock para mantenerlos en niveles óptimos, evitando tanto el exceso como la escasez de inventario (Abdullah et al., 2018). Una organización y gestión eficientes de los almacenes permiten una rápida respuesta a las necesidades del mercado. La implementación de sistemas automatizados para el seguimiento de inventarios ayuda a

reducir los costos de almacenamiento y mejorar la eficiencia operativa (Abuduweili et al., 2021). La planificación logística, que abarca la coordinación de todas las actividades relacionadas con el transporte y distribución de productos, incluye la planificación de rutas y programación de entregas para asegurar una distribución eficiente. El uso de modelos de regresión lineal y técnicas de Machine Learning puede optimizar la planificación de rutas y entregas, reduciendo tiempos de entrega y costos de transporte (He y Chen, 2018). La monitorización en tiempo real, utilizando tecnologías avanzadas, proporciona datos actualizados sobre el estado de la cadena de suministro, lo que permite ajustar rápidamente las operaciones en respuesta a cambios en la demanda o en el mercado. Huertas (2017) explica que mejorar la coordinación en la cadena de suministro implica una integración efectiva entre todos los actores, desde proveedores hasta distribuidores, para asegurar una operación fluida. La utilización de tecnologías de comunicación avanzadas y la integración de datos permiten optimizar la logística y mejorar la eficiencia general.

### **6.9.2. Estrategias de gestión de la cadena de suministro**

Holguín (2021) señala que una gestión eficiente de la cadena de suministro es crucial para la competitividad y sostenibilidad en la industria tabacalera. Las estrategias de gestión buscan optimizar todos los procesos, desde la adquisición de materias primas hasta la entrega del producto final al consumidor. La integración vertical es una estrategia clave que implica coordinar y controlar diferentes etapas de producción y distribución dentro de la misma empresa, proporcionando mayor visibilidad y control sobre calidad, costos y tiempos de entrega, lo que mejora la eficiencia operativa. Otra estrategia importante es la gestión de inventarios, que incluye técnicas como el Just-in-Time (JIT) y la planificación de requerimientos de materiales (MRP) (Abuduweili et al., 2021).

El JIT reduce inventarios al mínimo necesario, produciendo solo lo requerido en el momento justo, disminuyendo costos de almacenamiento y aumentando flexibilidad para responder a cambios en la demanda. El MRP, por su parte, se centra en la planificación y control de inventarios, asegurando disponibilidad para la producción y entrega en el momento adecuado. La colaboración con proveedores y socios estratégicos es esencial para una comunicación y coordinación eficaces, reduciendo riesgos y mejorando la eficiencia. El uso de contratos a largo plazo y alianzas estratégicas asegura un suministro constante de materias primas de alta calidad a precios competitivos y fomenta la innovación (Arauz, 2021).

La adopción de tecnologías de la información y comunicación (TIC), como sistemas de planificación de recursos empresariales (ERP) y sistemas de gestión de la cadena de suministro (SCM), facilita la integración y automatización de procesos, proporcionando datos en tiempo real para una toma de decisiones más informada (Loola et al., 2020). Estas tecnologías ayudan en la gestión de inventarios, planificación de producción y coordinación logística, mejorando la eficiencia operativa y reduciendo costos.

Por otro lado, Sánchez (2019) destaca que la sostenibilidad se ha convertido en una estrategia esencial en la gestión de la cadena de suministro, con prácticas sostenibles que minimizan el impacto ambiental y cumplen con regulaciones estrictas. Esto incluye procesos de producción eficientes, reducción de emisiones de carbono y prácticas agrícolas sostenibles, así como transparencia y trazabilidad en la cadena de suministro para demostrar el compromiso con la sostenibilidad ante consumidores y reguladores.

#### **6.10. Optimización de la cadena de suministro mediante la regresión lineal**

De acuerdo con la investigación de Guo et. al., (2018) la regresión lineal, como una técnica de Machine Learning, puede jugar un papel crucial en la optimización de la cadena de suministro. Esto debido a que, al analizar datos históricos y patrones de demanda, la

regresión lineal permite predecir con precisión la demanda futura de productos de tabaco. Esto permite a las empresas tabacaleras ajustar sus niveles de inventario y planificación logística de manera más efectiva. Mediante la modelación de la relación entre variables independientes (como las ventas históricas, las campañas de marketing, y las tendencias estacionales) y la demanda futura, las empresas pueden anticipar mejor las necesidades del mercado. Esta capacidad predictiva facilita una mejor planificación de la producción y la distribución, reduciendo los costos de almacenamiento y evitando tanto los excesos de inventario como los desabastecimientos.

Además, la regresión lineal puede ayudar a optimizar las rutas de distribución al prever las necesidades de reabastecimiento en diferentes puntos de venta. Esto permite una planificación más eficiente de las entregas y la utilización óptima de los recursos logísticos (Abdel, Deng, y Yu, 2016). El análisis en tiempo real de los datos logísticos, combinado con las predicciones obtenidas mediante regresión lineal, permite ajustar rápidamente las operaciones para responder a cambios en la demanda o en las condiciones del mercado. En resumen, la implementación de modelos de regresión lineal en la gestión de la cadena de suministro proporciona una ventaja competitiva significativa, permitiendo a las empresas tabacaleras mejorar su eficiencia operativa, reducir costos y satisfacer mejor las necesidades de sus clientes.

### **6.11. Análisis competitivo**

La competitividad es un aspecto que adquiere cada vez mayor relevancia en el campo de las empresas, lo cual se deriva de las exigencias del entorno económico actual enmarcado en el proceso de globalización (Abdullah et. al., 2018) (Budd y Robinson, 2019) han señalado que las que compiten son las empresas no las naciones; a un país lo hacen competitivo las empresas competitivas que hay en este; por lo tanto, son estas la base de la competitividad. La competitividad de las empresas ha sido analizada

teóricamente por Porter en el llamado “Diamante de la Competitividad” (Arauz, 2021), compuesto por los siguientes atributos: Las condiciones de los factores. Las condiciones de la demanda. Las industrias relacionadas y de apoyo. El contexto para la estrategia y rivalidad empresarial.

Estos factores determinantes crean el ambiente adecuado para que las empresas se ajusten y aprendan a competir. Existen cuatro niveles de competitividad donde las empresas desarrollan sus actividades comerciales. El nivel metaeconómico de la competitividad está conformado por la estructura política y economía orientada al desarrollo, estructura competitiva de la economía, visiones estratégicas, planes nacionales de desarrollo (Calero et. al., 2019). Las condiciones que ofrece el nivel macroeconómico dan las oportunidades, pero son las empresas las que “crean bienes y servicios con valor agregado usando métodos eficientes. Solamente de esta forma una nación puede mantener altos salarios y atractivos retornos de capital para apoyar una inversión sostenida” (Khan y Rahmani, 2018)

El resultado de la cohesión social se ve reflejado en las acciones que emprenden varias empresas para cooperar y competir en un espacio geográfico; esta cohesión al interior de la región puede generar clústeres, en la medida en que las empresas se agrupan en una industria específica. Para el modelo sistémico, los pilares del ambiente mesoeconómico son el vínculo entre los niveles macro y micro en la estructura industrial, las exportaciones/ importaciones regionales, la infraestructura, el medio ambiente, la tecnología, la educación y el trabajo (Effendy et. al., 2017). El nivel microeconómico de la competitividad contiene los procesos internos de la empresa que permiten crear ventajas competitivas, la capacidad de gestión de las empresas, sus estrategias empresariales, gestión e innovación (Cui et. al., 2018). Es el que se deriva de la ventaja

competitiva que tiene una empresa a través de sus métodos de producción y de organización.

## 7. Marco legal

El marco legal de la presente investigación se fundamenta en diversas normativas nacionales e internacionales que regulan aspectos relacionados con la producción y comercialización de tabaco, así como normativas específicas para la exportación. Estas normativas incluyen disposiciones legales que rigen la investigación científica y el uso de datos, garantizando el cumplimiento ético y legal en la recopilación, tratamiento y análisis de datos históricos de producción de tabaco y cualquier otro dato relevante para la predicción de la demanda en el mercado de tabaco habano.

En Ecuador, la investigación científica está regulada por la Constitución de la República del Ecuador y la Ley Orgánica de Educación Superior. Estas normativas establecen los principios éticos y legales que deben guiar la investigación, así como la protección de los derechos de los participantes y el uso adecuado de los datos obtenidos. De acuerdo con el Código Orgánico de la Producción, Comercio e Inversiones (COPCI) se expone:

**Art. 4.- Fines.** - La presente legislación tiene, como principales, los siguientes fines:

- A. Transformar la Matriz Productiva, para que esta sea de mayor valor agregado, potenciadora de servicios, basada en el conocimiento y la innovación; así como ambientalmente sostenible y eficiente.
- B. Democratizar el acceso a los factores de producción, con especial énfasis en las micro, pequeñas y medianas empresas, así como de los actores de la economía popular y solidaria.

- C. Fomentar la producción nacional, comercio y consumo sustentable de bienes y servicios, con responsabilidad social y ambiental, así como su comercialización y uso de tecnologías ambientalmente limpias y de energías alternativas.
- D. Generar trabajo y empleo de calidad y dignos, que contribuyan a valorar todas las formas de trabajo y cumplan con los derechos laborales.
- E. Generar un sistema integral para la innovación y el emprendimiento, para que la ciencia y tecnología potencien el cambio de la matriz productiva; y para contribuir a la construcción de una sociedad de propietarios, productores y emprendedores.
- F. Garantizar el ejercicio de los derechos de la población a acceder, usar y disfrutar de bienes y servicios en condiciones de equidad, óptima calidad y en armonía con la naturaleza.
- G. Incentivar y regular todas las formas de inversión privada en actividades productivas y de servicios, socialmente deseables y ambientalmente aceptables.
- H. Regular la inversión productiva en sectores estratégicos de la economía, de acuerdo con el Plan Nacional de Desarrollo.
- I. Promocionar la capacitación técnica y profesional basada en competencias laborales y ciudadanas, que permita que los resultados de la transformación sean apropiados por todos. (Código Orgánico de Producción Comercio e Inversiones, 2011)

Por otro lado, la Ley Orgánica para la Regulación y Control del Tabaco explica:

**Art. 6.- Responsabilidad tributaria y aduanera.** - Será responsabilidad del Servicio de Rentas Internas promover y adoptar medidas para el control tributario de los productos de tabaco; y, del Servicio Nacional de Aduanas combatir, de conformidad con la Ley, todas las formas de comercio ilícito y contrabando de tales productos.

**Art. 144.- Control Aduanero.** - El control aduanero se aplicará al ingreso, permanencia, traslado, circulación, almacenamiento y salida de mercancías, unidades de carga y medios de transporte hacia y desde el territorio nacional, inclusive la mercadería que entre y salga de las Zonas Especiales de Desarrollo Económico, por cualquier motivo. (Código Orgánico de Producción Comercio e Inversiones, 2011).

## **8. Metodología**

Esta investigación es de tipo aplicada y descriptiva, con componentes tanto bibliográficos como experimentales. Su objetivo fue desarrollar un modelo predictivo de regresión lineal basado en Machine Learning, destinado a mejorar la competitividad de las empresas tabacaleras ecuatorianas y apoyar su expansión hacia mercados internacionales. El estudio se organiza en tres etapas: una revisión exhaustiva de la literatura, el desarrollo del modelo predictivo y su implementación en la cadena de suministro.

La fase bibliográfica incluye la revisión de estudios previos que exploran el uso de modelos predictivos en la industria tabacalera y sectores relacionados. A partir de esta base teórica, se desarrolla un modelo de regresión lineal utilizando R Studio, aprovechando datos históricos y económicos para predecir la demanda de tabaco en mercados internacionales. Este modelo es evaluado mediante técnicas de validación cruzada y métricas como  $R^2$  y el MSE.

El análisis de datos en R comenzó con la preparación del entorno de trabajo y la organización de la información disponible. Este proceso inicial fue fundamental para estructurar adecuadamente los datos históricos de ventas y demanda de tabaco, permitiendo así que el análisis se desarrolle de manera fluida y precisa. Al cargar los datos y asignar identificadores claros a cada observación, se establecieron las bases para un



análisis riguroso, donde cada paso fue diseñado para optimizar la interpretación de los resultados y garantizar la fiabilidad de las predicciones generadas por el modelo.

El modelo de regresión lineal múltiple se expone Calero et. al., (2019) y se expresa matemáticamente de la siguiente manera:

$$y = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n + \epsilon$$

Donde:

$y$  = Variable dependiente

$x_1, x_2, \dots, x_n$  = Variables independientes

$\beta_0$  = Ordenada al origen o intercepto

$\beta_1, \beta_2, \dots, \beta_n$  = Coeficientes de regresión que representan la magnitud y dirección de la relación entre cada variable independiente y dependiente.

Para estimar coeficientes  $\beta_0, \beta_1, \beta_2, \dots, \beta_n$  se utiliza el método de los mínimos cuadrados ordinarios, que minimiza la suma de los cuadrados de las diferencias entre los valores observados y los valores predichos por el modelo

$$\min_{\beta} = \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

$$\min_{\beta} = \sum_{i=1}^m (y_i - (\beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n))^2$$

Donde:

$y_i$  = Valor observado de la variable dependiente

$\hat{y}_i$  = Valor predicho por el modelo, es decir,  $\hat{y}_i = \beta_0 + \beta_1x_1 + \beta_2x_2 + \dots + \beta_nx_n$

Este problema puede ser reformulado utilizando algebra matricial. Definimos  $X$  como la matriz de diseño que incluye un vector de unos para el término constante, y  $y$  como el vector de las observaciones de la variable dependiente:

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & x_{1n} \\ 1 & x_{21} & x_{22} & x_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{m1} & x_{m2} & x_{mn} \end{pmatrix}, y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

El problema de la minimización se puede reescribir en términos matriciales como:

$$m_\beta = \|y - X\beta\|^2$$

Donde:

$y$  = Vector columna de valores observados de la variable dependiente (dimensión  $m \times 1$ )

$X$  = Matriz de diseño que incluye un vector de unos para el término constante y las variables independientes  $x_1, x_2, \dots, x_n$  (dimensión  $m \times (n + 1)$ ).

$\beta$  = Vector de coeficientes a estimar (dimensión  $(n + 1) \times 1$ )

$\varepsilon$  = Vector de errores

La solución óptima para  $\beta$  se obtiene derivando la función de costo respecto a  $\beta$  y resolviendo el sistema de ecuación normal:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

El término de error  $\varepsilon$  se asume que sigue una distribución normal con media cero y varianza constante  $\sigma^2$  se puede obtener como:

$$\hat{\sigma}^2 = \frac{1}{m - n - 1} \sum_{i=1}^m (y_i - \hat{y}_i)^2$$

Cada coeficiente estimando  $\hat{\beta}_j \pm t_{\alpha/2(m-n-1)} \sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}$

Donde  $t_{\alpha/2(m-n-1)}$  es el valor crítico de la distribución t de student con  $(m - n - 1)$  grados de libertad.

Para cada coeficiente  $\beta_j$ , podemos realizar una prueba de hipótesis para verificar si su valor es significativamente diferente de cero (es decir, si  $x_j$  tiene un efecto significativo sobre  $y$ ). La hipótesis nula y la alternativa son:

$$H_0: \beta_j = 0, H_A: \beta_j \neq 0$$

El estadístico de prueba se calcula como:

$$t = \frac{\hat{\beta}_j}{\sqrt{\hat{\sigma}^2 (X^T X)^{-1}_{jj}}}$$

El coeficiente de determinación  $R^2$  mide la proporción de la variabilidad en la variable dependiente que es explicada por las variables independientes:

$$R^2 = 1 - \frac{\sum_{i=1}^m (y_i - \hat{y}_i)^2}{\sum_{i=1}^m (y_i - \bar{y})^2}$$

Donde  $\bar{y}$  es la media de los valores observados de la variable dependiente.

El ajuste del modelo no sólo se evalúa mediante  $R^2$ , sino también mediante la raíz del error cuadrático medio (RMSE) y la prueba F para la significancia global del modelo:

$$RMSE = \sqrt{\frac{1}{m} \sum_{i=1}^m (y_i - \hat{y}_i)^2}$$

$$F = \frac{\frac{SSR}{n}}{\frac{SSE}{m - n - 1}}$$

Donde SSR es la suma de los cuadrados de regresión y SSE es la suma de cuadrados de los errores. Una vez estimados los coeficientes  $\hat{\beta}$ , el modelo lineal ajustado sería:

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_n x_n$$

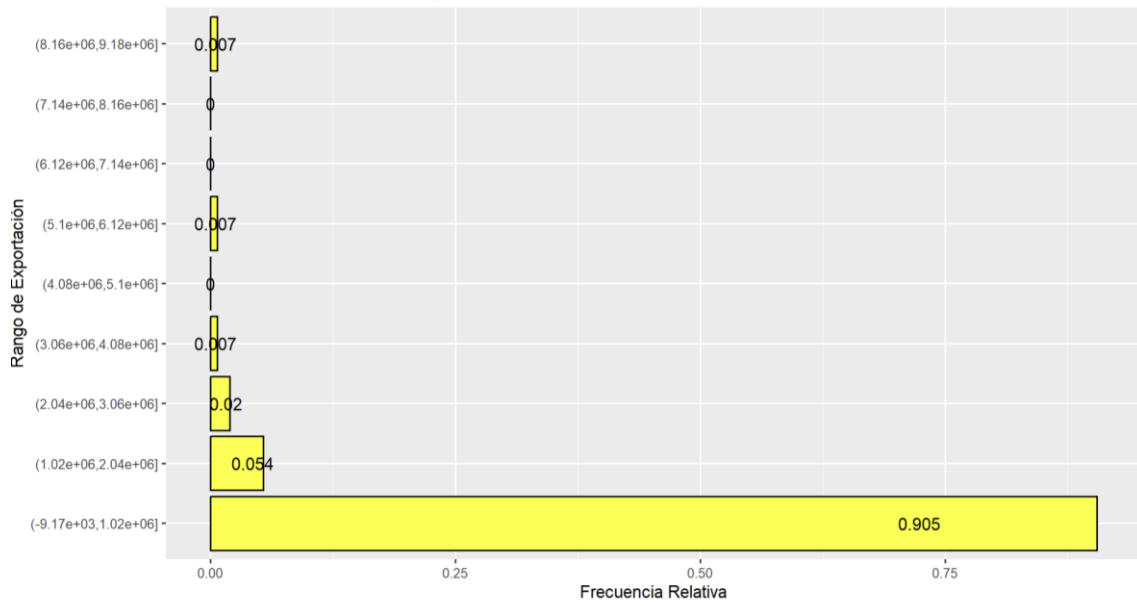
Donde  $\hat{y}$  representa la predicción de la variable dependiente basada en las variables independientes y sus coeficientes estimados. Por lo que es importante verificar la calidad del modelo utilizando métodos de diagnóstico como:

- **Gráficos de residuos:** Para identificar cualquier patrón no explicado.
- **Prueba de significancia de coeficientes:** Utilizando el estadístico t para determinar si los coeficientes son significativamente diferentes de cero.
- **Validación cruzada:** Para asegurar que el modelo generaliza bien a datos no observados.

## 9. Análisis de resultados

**Figura 1.**

*Distribución de la variable de exportación*



El gráfico muestra la distribución de la variable "Exportación" a través de un histograma de frecuencias relativas, que utiliza barras horizontales para ilustrar cómo se distribuyen los valores dentro de ciertos rangos. Este análisis visual permite observar que la mayoría de los datos de exportación están altamente concentrados en un rango específico, lo que sugiere una fuerte tendencia hacia valores bajos en la variable estudiada.

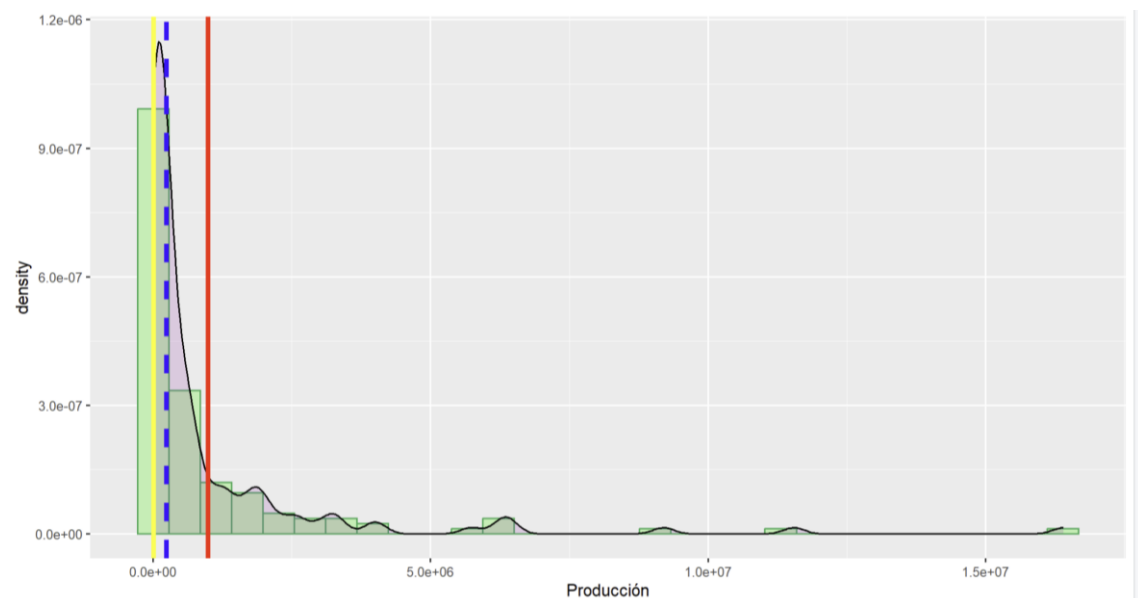
En el eje Y se presentan los diferentes rangos de exportación, donde cada rango abarca un intervalo específico de valores. Esto proporciona una clara división de la variable en segmentos manejables para la interpretación. El eje X, por su parte, muestra la frecuencia relativa, que representa la proporción de datos que caen dentro de cada rango en relación con el total de datos. Lo más destacable es que la barra más grande en el histograma, con una frecuencia relativa de 0.905, indica que el 90.5% de los valores de exportación se encuentran en el rango más bajo del eje Y, lo que refleja una concentración significativa en este intervalo.

La visualización de estas barras amarillas sugiere que la mayoría de los datos están agrupados cerca del origen, lo que es indicativo de una distribución desigual. Las otras barras del gráfico, con frecuencias relativas mucho menores (como 0.007 y 0.02), señalan que fuera del rango más bajo, los valores de exportación están considerablemente más dispersos y son menos comunes. Este patrón revela un sesgo en la distribución hacia la izquierda, lo que significa que la mayoría de las observaciones corresponden a valores de exportación bajos, mientras que solo una minoría se encuentra en rangos más altos.

La distribución asimétrica observada indica que la producción en la mayoría de los casos es baja, lo que podría sugerir una predominancia de pequeños productores o países con capacidades productivas limitadas.

## Figura 2.

### *Histograma de producción*



El gráfico presentado es un histograma de densidad que representa la distribución de la variable "Producción" en el conjunto de datos. En este gráfico, las barras verdes muestran cómo se distribuyen los datos de producción a lo largo de diferentes intervalos, destacando la concentración de los valores. A diferencia de un histograma de frecuencias, las barras se han normalizado para mostrar la densidad, de manera que el área total bajo

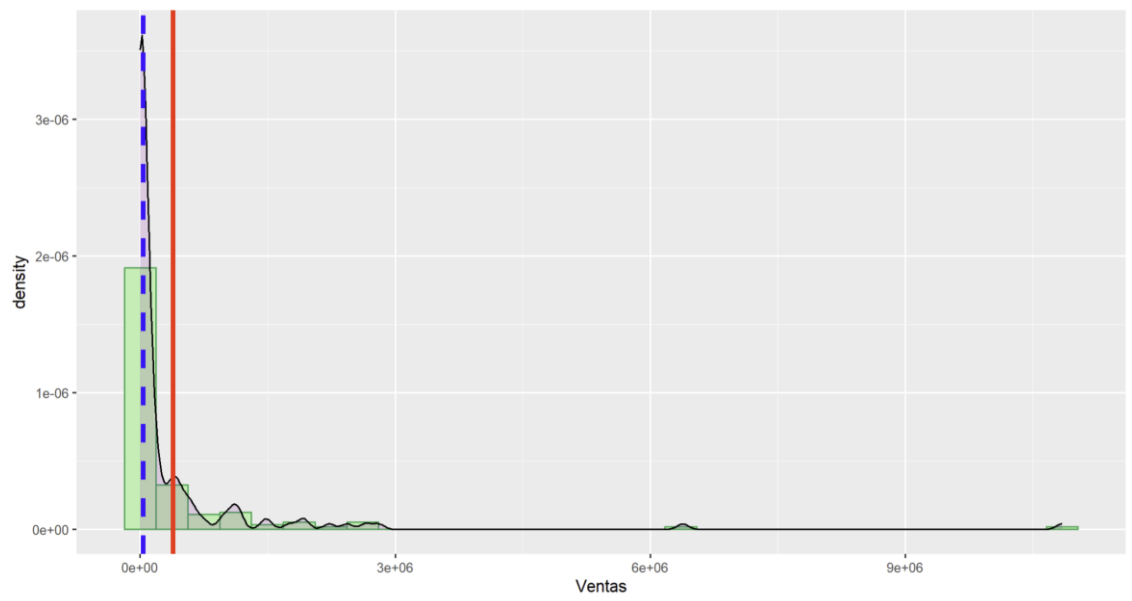
la curva es igual a 1, lo que permite apreciar con mayor claridad cómo se distribuyen los valores de producción en términos relativos.

La curva de densidad, representada por la línea negra, proporciona una visión suavizada de la tendencia general de la distribución. Se observa que la mayor densidad se concentra en los valores cercanos al cero, indicando que la mayoría de las entidades o países tienen producciones relativamente bajas. Esto se complementa con las líneas verticales que marcan las medidas de tendencia central: la media (línea roja), la mediana (línea azul punteada) y la moda (línea amarilla). La media está desplazada hacia la derecha en comparación con la mediana, lo que evidencia un sesgo positivo en la distribución; es decir, existen algunos valores de producción significativamente altos que están influyendo en el promedio general.

La moda, situada muy cerca del cero, refuerza la conclusión de que la mayoría de los valores de producción son bajos. Este sesgo hacia la derecha sugiere que, aunque la mayoría de los productores tienen capacidades limitadas, hay un pequeño número que alcanza niveles de producción mucho más altos, lo cual es un aspecto relevante para considerar en el análisis de la demanda en los mercados internacionales de tabaco.

**Figura 3.**

*Histograma de ventas*



El gráfico presentado es un histograma de densidad de la variable "Ventas". Este gráfico permite observar cómo se distribuyen los valores de ventas en el conjunto de datos, proporcionando una visión clara de la tendencia central y la dispersión de los datos.

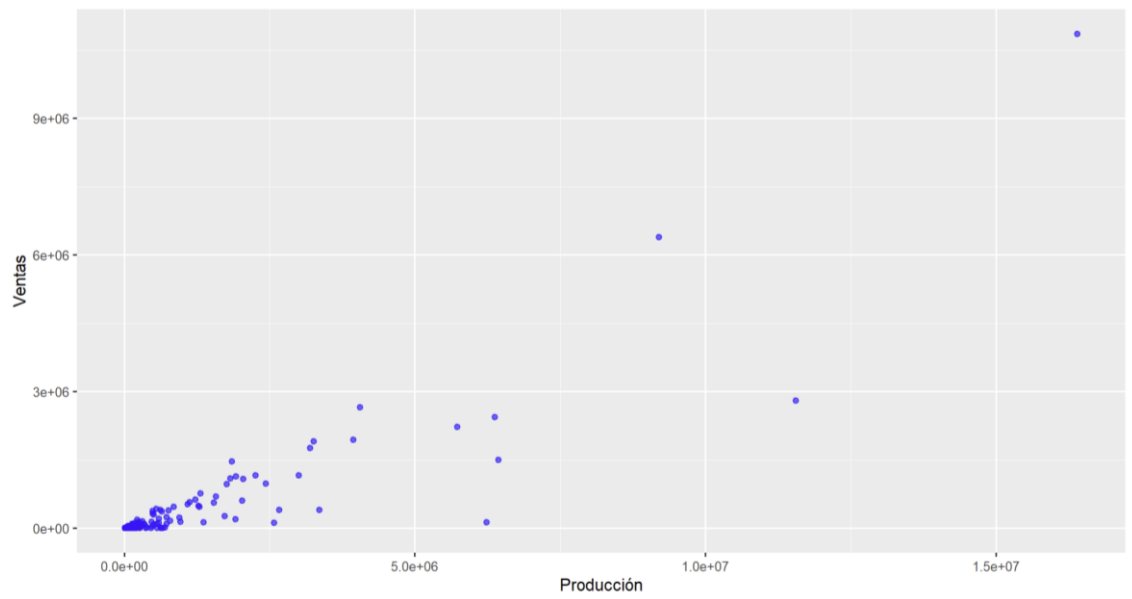
El histograma, representado por las barras verdes, muestra que la mayor densidad de ventas se concentra en los valores bajos, lo que significa que la mayoría de los valores están en rangos pequeños. La curva de densidad, representada por la línea negra, suaviza la distribución y destaca que la mayoría de las ventas se concentran cerca del cero, confirmando la predominancia de valores bajos.

Las líneas verticales en el gráfico señalan la media, mediana y moda de los datos. La media (línea roja) está desplazada hacia la derecha en relación con la mediana (línea azul punteada), lo que indica un sesgo positivo en la distribución; esto sugiere que hay algunos valores de ventas muy altos que influyen en el promedio. La moda (línea amarilla), ubicada cerca del cero, confirma que los valores de ventas más frecuentes son extremadamente bajos.



**Figura 4.**

*Relación entre ventas y producción*



El gráfico presentado es un diagrama de dispersión que ilustra la relación entre las variables "Producción" y "Ventas". En este gráfico, se observa cómo estas dos variables se comportan en conjunto, proporcionando una visión clara sobre la posible relación entre ellas.

En términos generales, la mayoría de los puntos se concentran cerca del origen, lo que indica que la mayoría de los valores tanto de producción como de ventas son bajos. A medida que los valores de producción aumentan, también lo hacen las ventas, sugiriendo una relación positiva entre ambas variables. Sin embargo, esta tendencia ascendente presenta cierta dispersión, lo que implica que la relación no es perfectamente lineal.

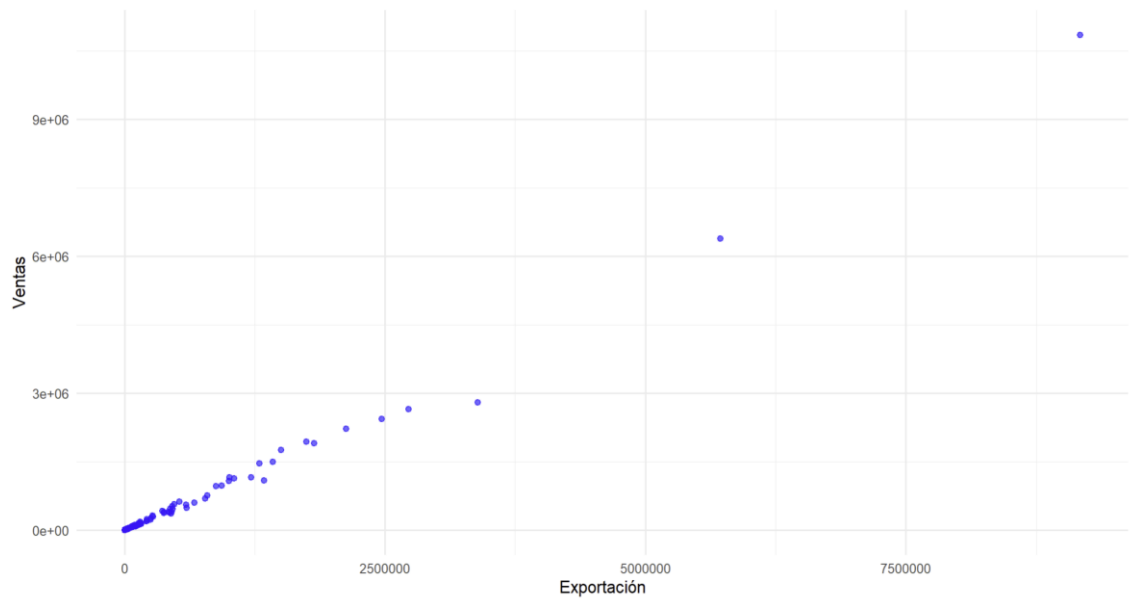
Además, se destacan algunos valores atípicos que se alejan considerablemente del resto de los puntos. Estos outliers, que muestran valores de producción y ventas significativamente más altos, podrían influir de manera desproporcionada en cualquier análisis posterior.

Es así que el diagrama de dispersión sugiere una relación positiva entre la producción y las ventas: a mayor producción, generalmente se observan mayores ventas. No obstante,

la dispersión observada, especialmente en niveles altos de producción, indica que no siempre un aumento en la producción conlleva un incremento proporcional en las ventas, sugiriendo una relación más compleja entre ambas variables.

### **Figura 5.**

*Relación entre ventas y exportación*



El diagrama de dispersión presentado explora la relación entre las variables "Exportación" y "Ventas". Este gráfico revela una tendencia clara: a medida que aumentan las exportaciones, también lo hacen las ventas, lo que sugiere una fuerte correlación positiva entre ambas variables.

La mayoría de los puntos en el gráfico siguen una línea ascendente, lo que indica una relación lineal clara entre exportación y ventas. Sin embargo, destaca la presencia de un valor atípico que se sitúa significativamente más arriba en la escala de exportaciones y ventas, separándose del grupo principal de datos. Este outlier podría corresponder a un país o entidad con exportaciones y ventas excepcionalmente altas.

La concentración de puntos cerca del origen sugiere que la mayoría de los países o entidades registran tanto exportaciones como ventas bajas. Sin embargo, a medida que

las exportaciones aumentan, se observa una ligera dispersión en los valores de ventas, aunque la tendencia general sigue siendo positiva.

El gráfico sugiere una fuerte relación lineal positiva entre exportación y ventas, lo que implica que un aumento en las exportaciones generalmente se traduce en un aumento en las ventas. Esta relación es especialmente relevante para cualquier análisis o modelo predictivo que considere ambas variables. Sin embargo, es esencial tener en cuenta los valores atípicos y la posible variabilidad a mayores niveles de exportación.

### 9.1. Proceso de R

Se estableció el directorio de trabajo donde se ubican los archivos de datos relevantes, asegurando que todos los archivos sean accesibles sin necesidad de especificar una ruta completa. Posteriormente, se cargaron los datos de un archivo CSV delimitado por punto y coma (;) en un *dataframe* denominado “**Tabaco**”, que contiene los datos históricos de demanda y ventas para el entrenamiento y prueba del modelo de predicción.

```
Tabaco <- read.csv2("../DATA/Datos limpios nuevo.csv")
```

Inicialmente, se utilizó el comando *rownames* para asignar la columna **Pais** como identificadores de las filas en el *dataframe* **Tabaco**, lo que facilita la referencia a cada observación. Luego, se eliminó la columna **Pais** del *dataframe*, dado que la información ya está almacenada en los nombres de las filas y no es necesaria como una columna separada.

```
rownames(Tabaco) <- Tabaco$Pais:
```

Además, se procedió a eliminar la primera columna del *dataframe*, correspondiente a la importación, ya que no aporta información relevante al análisis.

```
Tabaco$Pais <- NULL
```

Para garantizar la reproducibilidad del proceso de particionamiento, se fijó una semilla utilizando **set.seed(2018)**. A continuación, se cargó la librería **caret**, que facilita el proceso de particionamiento de datos y otras tareas de modelado. Los datos se dividieron en un conjunto de entrenamiento (80%) y uno de prueba (20%), utilizando el índice numérico generado por **createDataPartition**.

```
Entrenamiento <- createDataPartition(Tabaco$Ventas, p=0.80, list= FALSE)
```

Se ajustó un modelo de regresión lineal donde **Ventas** es la variable dependiente y el resto de las columnas del *dataframe* **Tabaco** son las variables predictoras. Este modelo, denominado **TabacoM**, se entrenó utilizando solo el conjunto de entrenamiento. La evaluación del modelo incluyó la revisión de coeficientes, errores estándar, valores p, R-cuadrado, y R-cuadrado ajustado, proporcionando una visión clara de la calidad del ajuste del modelo y la importancia de los predictores.

```
TabacoM <- lm(Ventas~., data= Tabaco[Entrenamiento,])
```

Para diagnosticar el modelo, se generaron gráficos clave, como residuos versus valores ajustados, gráficos Q-Q, y gráficos de *leverage*, evaluando la linealidad, normalidad y homocedasticidad de los residuos. Estos gráficos son esenciales para verificar las suposiciones del modelo y detectar posibles problemas. El comando **Summary** muestra un resumen del modelo, incluyendo los coeficientes, errores estándar y métricas de ajuste.

```
Summary(TabacoM)
```

- **Coefficientes:** Muestran la magnitud y dirección del impacto de cada predictor en las ventas.

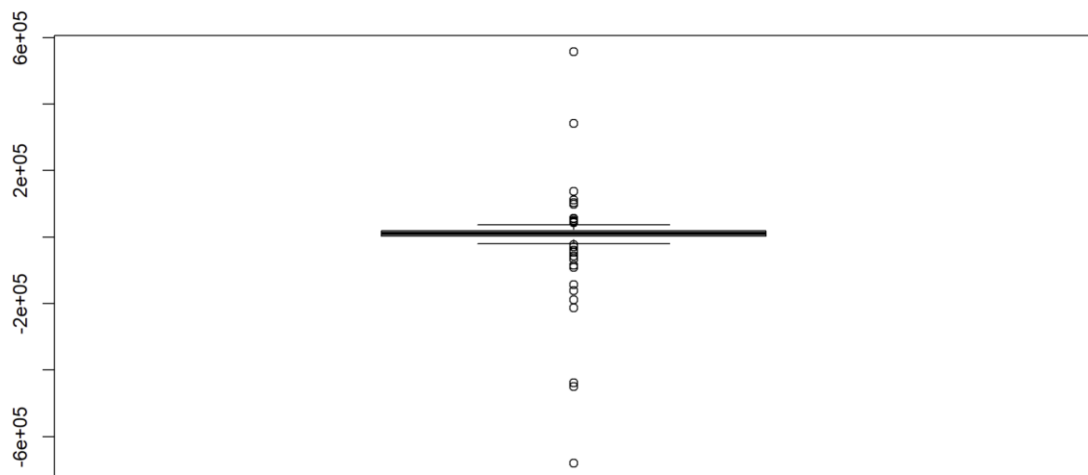
- **Errores estándar:** Indican la precisión de las estimaciones de los coeficientes.
- **Valores p:** Ayudan a determinar si los coeficientes son significativamente diferentes de cero (lo que indicaría que la variable predictora tiene un efecto significativo en las ventas).
- **R-cuadrado:** Mide la proporción de la variabilidad en las ventas que es explicada por el modelo.
- **R-cuadrado ajustado:** Similar al R-cuadrado, pero ajustado por el número de predictores en el modelo, es útil para comparar modelos con diferentes números de variables.

Este código muestra un diagrama de caja de los residuos del modelo (diferencias entre los valores observados y los valores predichos por el modelo). Esto es útil para detectar posibles valores atípicos y evaluar si los residuos están distribuidos simétricamente alrededor de cero. En un buen modelo de regresión, los residuos deben ser aleatorios y no mostrar patrones obvios.

**boxplot(TabacoM\$residuals)**

**Figura 6.**

*Boxplot de los residuos del modelo lineal ajustado.*



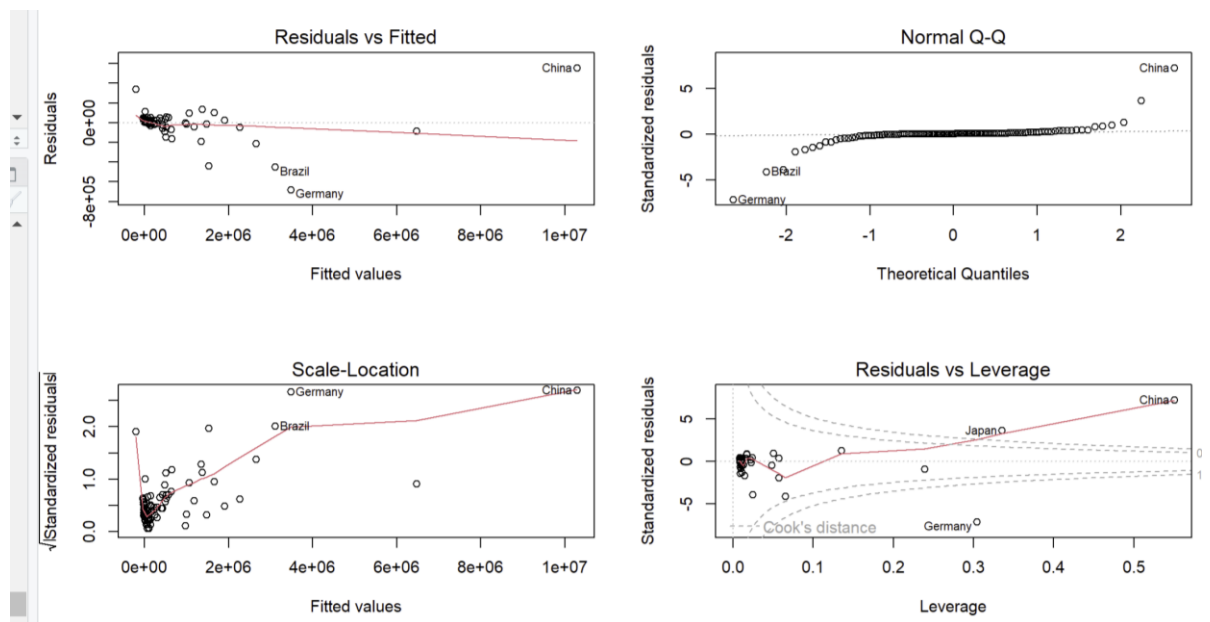
El boxplot muestra que la mayoría de los residuos están concentrados cerca de cero, lo cual es esperado en un buen modelo lineal. Sin embargo, hay una cantidad significativa de puntos fuera del rango intercuartílico (IQR), que son considerados outliers.

Este código facilita la configuración de la ventana gráfica para la presentación de cuatro gráficos dispuestos en una cuadrícula de 2x2. Esta disposición resulta beneficiosa para la visualización simultánea de múltiples gráficos de diagnóstico del modelo, permitiendo así una evaluación más exhaustiva y detallada.

**par(mfrow= c(2,2)):**

**Figura 7.**

*Diagnóstico del modelo: Análisis de residuos y evaluación de supuestos*



La figura 2 permite visualizar de una manera más detallada "Residuals vs Fitted" muestra si los residuos se distribuyen aleatoriamente, lo que indicaría que el modelo es adecuado; en este caso, se observa cierta variación en las ventas altas. El gráfico "Normal Q-Q" evalúa la normalidad de los residuos; aquí, los puntos se alinean en su mayoría con la línea, sugiriendo que los residuos siguen una distribución normal. El gráfico "Scale-Location" ayuda a verificar la homocedasticidad; una dispersión constante indicaría que el modelo mantiene un error constante, pero hay cierta variabilidad. Finalmente, el gráfico

"Residuals vs Leverage" identifica puntos influyentes; algunos países, como China y Japón, muestran una alta influencia, lo que sugiere que tienen un impacto significativo en el modelo.

El siguiente código genera 4 gráficos claves para el análisis del modelo:

- **Residuos vs. valores ajustados:** Permite verificar la linealidad y la homocedasticidad (uniformidad de la varianza) de los residuos.
- **Gráfico Q-Q:** Evalúa si los residuos siguen una distribución normal.
- **Escala de localización:** Evalúa la homocedasticidad.
- **Gráfico de leverage:** Detecta puntos de influencia que podrían estar afectando el modelo de manera desproporcionada.

**plot(TabacoM)**

El modelo entrenado se utilizó para predecir las ventas en el conjunto de entrenamiento, almacenando las predicciones en la variable **Prediccion**. También se calcularon las predicciones para todo el conjunto de datos, incluyendo tanto entrenamiento como prueba, y se evaluó el rendimiento mediante la raíz del error cuadrático medio (RMSE), que mide la precisión del modelo.

**Prediccion <- predict(TabacoM, Tabaco[Entrenamiento,])**

Se realizaron predicciones para la totalidad del conjunto de datos, abarcando tanto el conjunto de entrenamiento como el de prueba. Las predicciones se registraron en una nueva columna denominada **pred** dentro del dataframe **Tabaco**, utilizando el siguiente código:

**Tabaco\$pred <- predict(TabacoM, Tabaco)**

Para evaluar la precisión del modelo, se calculó la raíz del error cuadrático medio (RMSE), una métrica que mide la magnitud promedio de los errores cometidos por el modelo en sus predicciones. Un valor inferior de RMSE sugiere un ajuste más preciso del modelo, indicando que las predicciones están más cercanas a los valores observados.

```
Tabaco$rmse <- sqrt(mean((Tabaco$Ventas - Tabaco$pred)^2))  
par(mfrow= c(1,1))
```

Posteriormente, se generó un gráfico de dispersión que contrasta las ventas reales (**Tabaco\$Ventas**) con las ventas predichas (**Tabaco\$pred**). En un modelo óptimamente ajustado, se esperaría que los puntos del gráfico se alineen a lo largo de una línea con pendiente 1, lo que reflejaría una alta precisión en las predicciones realizadas por el modelo.

```
plot(Tabaco$Ventas, Tabaco$pred)
```

Se añadió una línea de referencia con pendiente 1 y origen en el punto cero para representar la situación ideal en la cual las predicciones coinciden exactamente con los valores reales. Esta visualización permite una rápida evaluación del rendimiento del modelo, siendo más favorable cuanto más próximos estén los puntos a dicha línea.

```
abline(0,1)
```

Se calcularon los intervalos de predicción para cada observación, considerando el RMSE del modelo. Estos intervalos representan el rango dentro del cual se espera que se sitúen las ventas reales, conforme a las predicciones realizadas.

```
Tabaco$L.inferior <- Tabaco$pred - Tabaco$rmse
```

```
Tabaco$L.superior <- Tabaco$pred + Tabaco$rmse
```



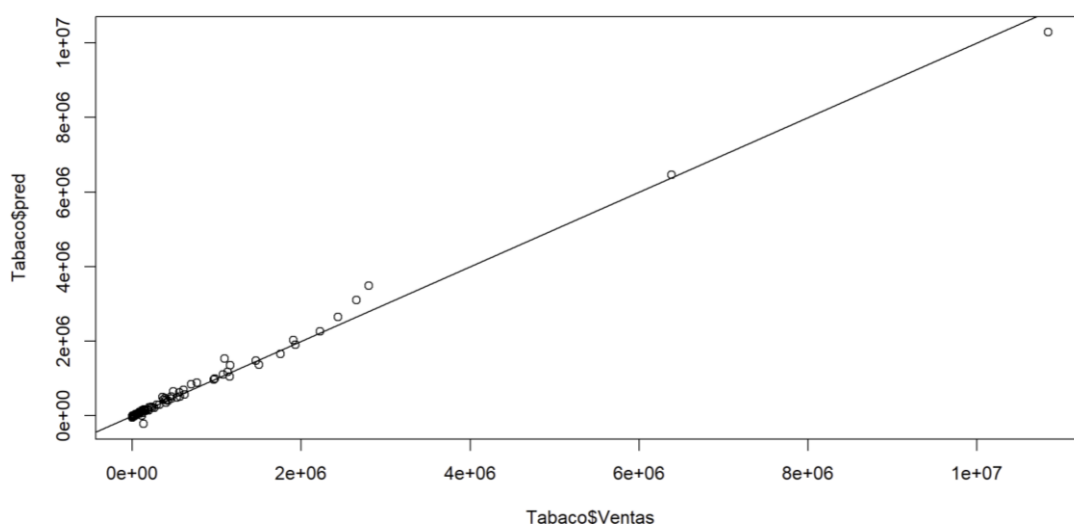
Se compararon las medias de las ventas reales y las predichas, con el propósito de detectar cualquier sesgo sistemático en el modelo, como una posible tendencia a subestimar o sobreestimar las ventas.

**mean(Tabaco\$Ventas)**

**mean(Tabaco\$pred)**

**Figura 8.**

*Relación entre ventas reales y las predichas por el modelo*



El gráfico que presentas muestra la relación entre las ventas reales y las predichas por el modelo en un contexto de predicción de ventas de tabaco. La recta de ajuste, representada por la línea diagonal en el gráfico, indica cómo se alinean las ventas predichas con las ventas reales. Los puntos dispersos alrededor de esta línea representan las observaciones individuales.

En general, cuanto más cercanos estén los puntos a la línea diagonal, más preciso es el modelo en sus predicciones. En este caso, la mayoría de los puntos se encuentran muy cerca de la línea, lo que sugiere que el modelo tiene un buen desempeño al predecir las ventas de tabaco. Sin embargo, algunos puntos se desvían más, especialmente en las ventas más altas, lo que puede indicar ligeras discrepancias en la precisión del modelo.

## 10. Conclusiones

El presente trabajo tuvo como objetivo principal predecir la demanda del tabaco en los mercados internacionales mediante técnicas de Machine Learning, con el fin de optimizar la cadena de suministro de las empresas tabacaleras ecuatorianas. Para ello, se desarrolló un modelo predictivo que permitió alcanzar este objetivo y arrojar conclusiones significativas.

- A través del examen de la literatura existente, se concluyó que del concepto Machine Learning en la industria tabacalera no solo es viable sino altamente beneficioso. Se identificaron múltiples estudios que resaltan cómo el aprendizaje automático facilita la optimización de procesos, mejora la eficiencia operativa y ofrece predicciones precisas en contextos de demanda incierta.
- El análisis de los datos de exportación y de las características de los mercados permitió identificar que el mercado europeo y asiático son los principales destinos para el tabaco habano ecuatoriano. Estos mercados se caracterizan por una alta demanda y precios competitivos, lo cual justifica la necesidad de implementar tecnologías predictivas que permitan a las empresas ecuatorianas posicionarse de manera estratégica y competitiva.
- La implementación del modelo predictivo fue un éxito. El modelo logró predecir las ventas con una Raíz del Error Cuadrático Medio (RMSE), lo cual refleja una alta precisión en las predicciones. Este valor indica que las diferencias entre las ventas predichas y las ventas reales son mínimas, validando así la eficacia del modelo desarrollado. Además, al comparar las medias de las ventas reales y las predichas, no se detectó un sesgo significativo,

lo que sugiere que el modelo es robusto y generaliza bien en distintos conjuntos de datos.

- Gráficamente, se observó que las predicciones del modelo están alineadas con las ventas reales, lo cual se visualizó en un gráfico de dispersión donde la mayoría de los puntos se alinearon a lo largo de una línea de referencia, confirmando la alta precisión del modelo. Los intervalos de predicción calculados a partir del RMSE proporcionaron una estimación confiable del rango dentro del cual se espera que se sitúen las ventas reales.
- El análisis de mercados internacionales a través de herramientas predictivas facilita la identificación de oportunidades de exportación, lo que incrementa la competitividad del tabaco ecuatoriano en el ámbito global. Este enfoque permite a las empresas adaptar sus estrategias de comercialización a las fluctuaciones del mercado internacional.
- La adopción de tecnologías avanzadas como el machine learning no solo impulsa la eficiencia operativa, sino que también fomenta la innovación en el sector tabacalero. Las empresas pueden adaptarse de manera más ágil a las dinámicas del mercado, fortaleciendo su posición competitiva a largo plazo.

## 11. Bibliografía

- Abdel, O., Deng, L., y Yu, D. (2016). Exploring convolutional neural network structures and optimization techniques for speech recognition. *In Interspeech*, 113-157.
- Abdullah, M., Alshannaq, A., Balamash, A., y Almabdy, S. (2018). Enhanced Intrusion Detection System using Feature Selection Method and Ensemble Learning Algorithms. *International Journal of Computer Science and Information Security* .
- Abuduweili, A., Li, X., Shi, H., y Dou, D. (2021). Adaptive Consistency Regularization for Semi-Supervised Transfer Learning, *Computer Vision and Pattern Recognition*. *Elsevier*.
- Ackermann, N. (2019). Introduction to 1D Convolutional Neural Networks in Keras for Time Sequences. *Elsevier*.
- Arauz, M. (2021). Costos de producción de tabaco de pequeños productos en las comunidades Tastali y Teotecacinte. *Doctoral dissertation*.
- Barrueta, R., y Castillo, V. (2018). Modelos de análisis predictivo para determinar clientes con tendencia a la deserción en bancos peruanos. *Sci Big Data*.
- Bayona, S., y Rino, C. (2021). Machine Learning for Price Prediction for Agricultural Products. *WSEAS TRANSACTIONS on BUSINESS and ECONOMICS*.
- Budd, S., y Robinson, B. (2019). A Survey on Active Learning and Human-in-the-Loop Deep Learning for Medical Image Analysis. *Scopus*.
- Bueno, F. (2021). Modelo para la administración de la producción en la Empresa Agrícola Tabaquera. *Master thesis of Agroindustrial Engineering*.
- Calero, A., Quintero, E., Olivera, D., y Peña, K. (2019). Influencia de dos bioestimulantes en el comportamiento agrícola del cultivo del tabaco (*Nicotiana tabacum* L.). *Revista de la Facultad de Ciencias*.

- Campos, A. (2018). Modelamiento matemático predictivo del comportamiento de la exportación de aguaymanto (*Physalis peruviana* L.) fresco del Perú a un país destino frente a factores internos y externos. *UPN*.
- Changshou, L., Qingfeng, W., Zhou, L., Zhang, J., y Sun, S. (2011). Prediction of Vegetable Price Based on Neural Network and Genetic Algorithm . *International Federation for Information Processing*.
- Chi, C., Wee, T., y G, H. (2014). Extreme learning machines for intrusion detection. *The 2012 International Joint Conference on Neural Networks*.
- Cui, J., Long, J., y Min, E. (2018). Comparative Study of CNN and RNN for Deep Learning Based Intrusion Detection System. *International Conference on Cloud Computing and Security*.
- Durán, G. (2017). Redes neuronales convolucionales en R: Reconocimiento de caracteres escritos a mano.
- Effendy, D., Kusriani, K., y Sudarmawan, S. (2017). Classification of intrusion detection system (IDS) based on computer network. *2nd International conferences on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*.
- ESPAC. (2023). Encuesta de superficie y producción agropecuaria continua. *Instituto Nacional de Estadística y Censos*.
- Guo, Y., Liu, Y., y Bakker, E. (2018). CNN-RNN: a large-scale hierarchical image classification framework. *Multimedia Tools and Applications*.
- He, D., y Chen, Z. (2018). An Improved Kernel Clustering Algorithm Used in Computer Network Intrusion Detection. . *International Symposium on Circuits and Systems*

- Holguín, L. (2021). Propuesta de políticas públicas para el fortaleciendo de la cadena de valor del tabaco en rama del Ecuador 2021-2025. *Scielo Brasil*.
- Huertas, A. (2017). Modelos predictivos para el mercado FOREX. *Universidad de Murcia, España*.
- Huy, V., y Asif, N. (2019). A smart system for short-term price prediction using time series models . *Computers and Electrical Engineering*.
- INEC. (2019). Producción agropecuaria 2018-2019. *INEC*.
- Kaur, M., Heena, G., y Harish, K. (2014). Data mining in Agriculture on crop price prediction: Techniques and Applications. *International Journal of Computer Applications*.
- Khan, S., y Rahmani, H. (2018). A guide to convolution neural networks for computer vision. *Synthesis Lectures on Computer Vision*.
- León, C., Coronado, R., y Forero, C. (2020). Modelo productivo de tabaco(Nicotiana tabacum) variedades Burley y Negro en Santander. *Revista Científica Agrosiv*.
- Liakos, K., y Busato, P. (2018). Machine Learning in Agriculture: A review. *Sensors*.
- Lin, X. (2018). Association Rules algorithm based on MapReduce. *International Conference on Software Engineering and Service Science*.
- Loola, P., Ouazzani, K., y Souissi, N. (2020). Predictive analysis using machine learning: Review of trends and methods. *Research Gate*.
- Lozada, C. (2017). La Logística Integral y su incidencia en la Mejora Continua de la empresa. *Tesis Doctoral de la Universidad técnica de Ambato*.
- Nasira, G., y Hemegeetha, N. (2016). “Vegetable price prediction using data mining classification technique. *International Conference on Pattern Recognition, Informatics and Medical Engineering*.

- Rajeswari, S., y Suthendran, K. (2019). Developong an Agricultural Product Price Prediction Model usaing HADT Algorithm. *Internatioanl Journal of Engineering and Advanced Technology*.
- Rakhra, M., Soniya, P., y D, T. (2021). Crop price prediction Using random forest and decision tree regression: A review. . *Material Today: Proceedings, In Press*.
- Sánchez, O. (2019). Manejo agronpomico del cultivo de tabaco (Nicotiana Tabacum L.) en una empresa tabaquera. *Tesis Doctoral. Universidad Autónoma de México*.
- Santillan, H. (2017). Agroindustrialización dle tabaco como materia prima premium para exportación. *Revista Agroindustrial de la Facultad de Ciencias Agrícolas*.
- Segovia, S. (2020). El proceso de adopción tecnológica en productores tabacaleros. *Sci Food Che*.
- Shalev, S., y Ben, D. (2014). Understanding Machine Learning: From Theory to Algorithms. *Cambrigde University Press*.
- Siegel, E. (2017). Predictive Analytics. The power to predict who will click, buy, lie or die. *Indiana University. Journal of Experimental Psychology*.
- Subasi, A. (2020). Machine learning techniques. Practical machine learning for data analysis using Python. *Academic Press*.
- Tong, S., Chaoqun, X., Yang, C., y Luyao, Z. (2023). Cost-effective identification of the field maturity of tobacco leaves based on deep semi-supervised active learning and smartphone photograph. *Scopus*.
- Valderrama, A. (2017). Eficiencia técnica en la industria manufacturera en México. *Investigación económica*.
- Velásquez, R., y Meza, Q. (2020). Desarrollo de alternativas de MIP para la producción sostenible del cutlivo de tabaco en Lima. *Universidad Politecnica de Lima*.

Vera, C. (2020). Efecto de dos insecticidas químicos en el control de la mosca blanca (*Bemisia tabaci*) en tabaco (*Nicotiana tabacum*). *Rev Sci Food*.

Zhang, D., Chen, S., Liwen, L., y Xia, Q. (2020). Forecasting agricultural commodity prices using model selection framework with time series features and forecast horizons. *IEEE Access*.



## 12. Anexos

**Figura 9.**

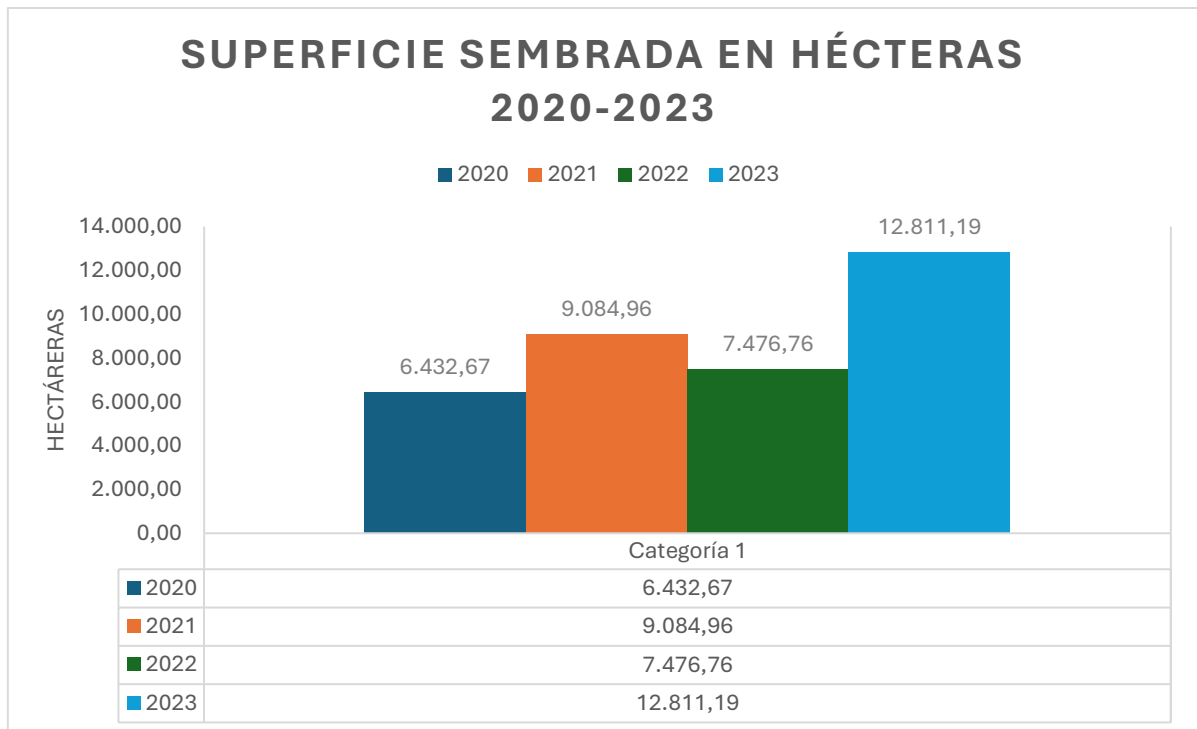
*Exportaciones de tabaco habano 2000-2023*



**Fuente:** (Banco Central del Ecuador, 2024)

**Figura 10.**

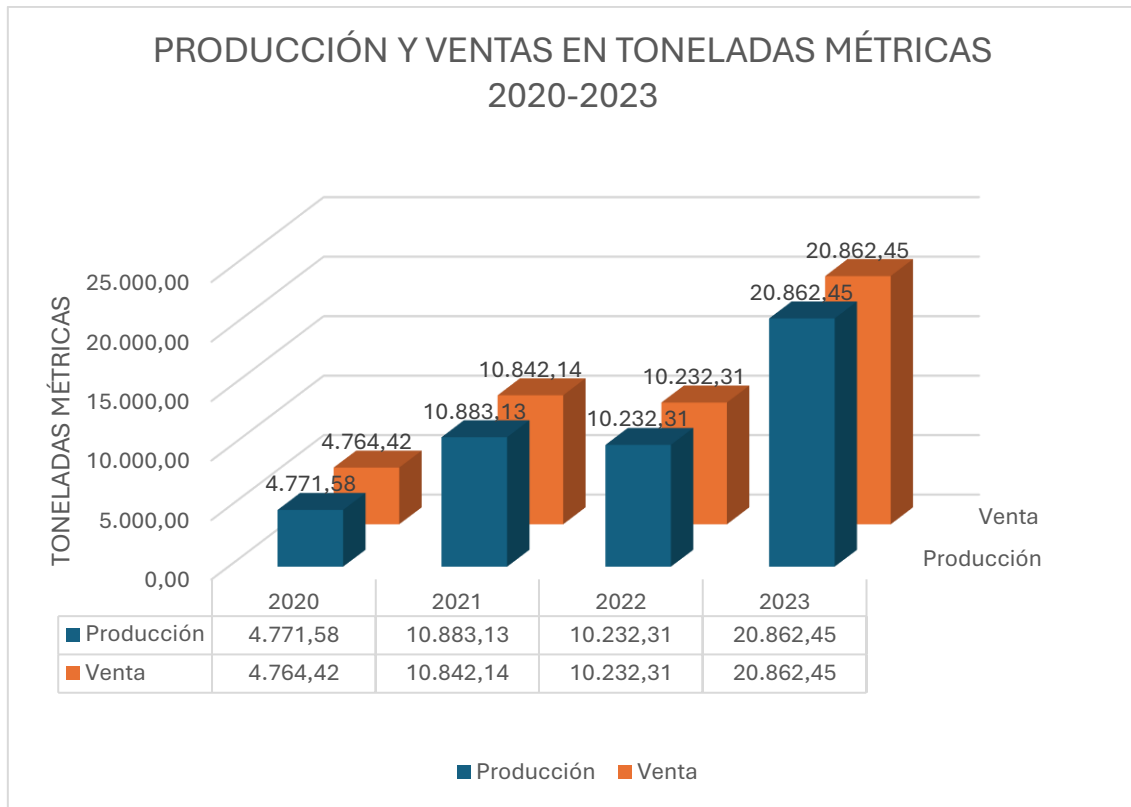
*Superficie sembrada en hectáreas 2020-2023.*



**Fuente:** (ESPAC, 2023)

**Figura 11.**

*Producción y ventas en toneladas métricas 2020-2023*



**Fuente:** (ESPAC, 2023)

## Base de datos

	Exportación	Producción	Ventas
<b>Germany</b>	3392844	11553155	2801866
<b>Japan</b>	128974	6231567	131045
<b>United States of America</b>	1422932	6435076	1500186
<b>Italy</b>	2468190	6377607	2433835
<b>China</b>	9174946	16402778	10844900
<b>Spain</b>	454271	3361905	400727
<b>United Kingdom</b>	133849	2582364	117939
<b>Belgium</b>	2127757	5732084	2223156
<b>France</b>	422938	2667893	403489
<b>United Arab Emirates</b>	210022	1915428	191627
<b>Poland</b>	5717572	9200701	6382305
<b>Netherlands</b>	1008597	3003305	1152078
<b>Indonesia</b>	1748221	3939430	1933389
<b>Czech Republic</b>	1505430	3197853	1754425
<b>Türkiye</b>	932836	2432968	980521
<b>Russian Federation</b>	256411	1727693	260339
<b>Hong Kong, China</b>	671165	2025718	604609
<b>Viet Nam</b>	132185	1363001	124254
<b>Dominican Republic</b>	1214147	2260621	1159908
<b>Singapore</b>	592760	1546491	561070
<b>Canada</b>	162595	970024	140493
<b>Greece</b>	1004706	2045737	1072638
<b>Hungary</b>	438978	1290916	463134
<b>Korea, Republic of</b>	1052594	1923846	1130050
<b>Romania</b>	1819126	3258920	1905227
<b>Ukraine</b>	86279	728916	88106
<b>Norway</b>	2271	627729	2563
<b>Australia</b>	16854	700398	14174
<b>Bulgaria</b>	247768	952317	227374
<b>Saudi Arabia</b>	279	664643	263
<b>Austria</b>	20169	634623	16566
<b>Croatia</b>	458023	1090971	526313
<b>Iraq</b>	6	570296	537
<b>Switzerland</b>	475382	1124103	565551

Showing 1 to 35 of 148 entries, 7 total columns

	Exportación	Producción	Ventas
Portugal	879816	1768523	968468
Luxembourg	152856	787810	160523
Philippines	525018	1224279	619984
Algeria	8	461066	9
Egypt	110125	557549	94380
Kazakhstan	132037	592569	113247
Slovakia	37488	492199	33401
Sweden	774792	1576458	697442
Serbia	594602	1278959	485044
Taipei, Chinese	214531	726730	237345
Lithuania	796333	1316831	764372
Cambodia	11324	395767	9124
Israel	272	371757	246
Morocco	106329	505688	86021
Thailand	215571	597180	200663
South Africa	145257	475684	139297
Ireland	21452	278951	17416
Slovenia	5275	271768	5173
Denmark	434720	765844	388521
Malaysia	79180	342499	86761
Tunisia	71664	358364	71654
Jordan	29600	257225	26773
Latvia	41972	250389	49109
Kuwait	194	194256	162
Libya, State of	3	214876	3
Finland	4809	200270	4313
New Zealand	358	206559	352
Cyprus	44386	243367	52426
Colombia	6495	195779	6042
Nicaragua	462982	848876	463718
Nigeria	81871	302588	86634
Azerbaijan	21391	171251	20395
Paraguay	30993	208925	26434
Iran, Islamic Republic of	3071	163092	2946
Costa Rica	2040	157540	2004

Showing 35 to 69 of 148 entries, 7 total columns

	Exportación	Producción	Ventas
Georgia	2948	157510	3094
Côte d'Ivoire	65411	250160	62035
Armenia	380657	625830	398915
Bosnia and Herzegovina	1803	125639	1474
Albania	6615	142179	7828
Lebanon	11122	145410	10632
Mongolia	3174	125658	3346
Macao, China	21415	138305	22376
Mauritius	4187	107033	4864
Panama	65968	216725	58752
Congo, Democratic Republic of the	5127	119603	4662
Yemen	14	109901	13
Brazil	2729478	4059878	2651482
India	1342326	1828172	1091412
Argentina	134984	308414	150844
Kyrgyzstan	2991	100868	3210
Sri Lanka	95758	243738	93736
Honduras	275963	511424	291190
Mexico	271977	486883	321760
Aruba	58320	179100	63557
Estonia	16195	109373	17675
Burkina Faso	10966	90644	10250
Zimbabwe	1297280	1849449	1462417
Belize	14202	86150	14681
Moldova, Republic of	8668	89254	9632
Malawi	449206	652425	359955
Macedonia, North	137082	269166	113459
Qatar	39	60280	41
Nepal	580	58411	691
Sudan	18	45265	15
Guinea	68	52822	55
Andorra	273	49595	226
Montenegro	1217	43541	1139
Guatemala	73980	151075	86507

Showing 69 to 103 of 148 entries, 7 total columns

row names	Exportación	Producción	Ventas
Namibia	1370	45005	1531
Uzbekistan	50499	102891	51493
Myanmar	29784	90790	31838
Iceland	2931	48472	2753
Chile	26490	68713	30393
Malta	26	41097	24
Kenya	100473	188161	111566
Bahrain	172	42219	204
Congo	179	38762	212
Tanzania, United Republic of	378437	492427	370007
Senegal	9983	42338	10298
Mali	17	37288	14
Curaçao	1968	36439	2170
Belarus	40	31494	44
Uruguay	17610	52471	14512
Venezuela, Bolivarian Republic of	254	28850	303
Bolivia, Plurinational State of	27	24562	28
El Salvador	131	26900	125
Cameroon	7673	39144	7496
Costa Rica	7074	38736	7498
Mozambique	154209	250968	137398
Afghanistan	680	25706	777
Oman	92	19621	103
Lao People's Democratic Republic	15801	41959	13966
Haiti	18	20074	19
Angola	192	19235	160
Pakistan	90675	135186	87567
Peru	3334	22498	3371
Cuba	361900	549835	425282
Ghana	2	16676	2
Trinidad and Tobago	25367	47943	27792
Botswana	2300	17893	2077
Bangladesh	151527	227470	180099
Niger	1	13073	1

Showing 103 to 137 of 148 entries, 7 total columns

<b>Tajikistan</b>	328	12054	327
<b>Ethiopia</b>	181	11352	173
<b>Ecuador</b>	101063	145757	91710
<b>Mauritania</b>	19	11484	18
<b>Suriname</b>	44165	69645	43136
<b>Togo</b>	119	10139	142
<b>South Sudan</b>	5	9808	5
<b>Uganda</b>	48337	78122	49071
<b>Sint Maarten (Dutch part)</b>	241	9696	236
<b>Guyana</b>	5	8801	5
<b>Madagascar</b>	34	8318	28
<b>New Caledonia</b>	20	7286	18

Showing 114 to 148 of 148 entries, 7 total columns

## Proceso R (Análisis Descriptivo)

```
1 setwd("C:/Users/jseba/OneDrive/Desktop/B.I A.2023 N.I/DATA")
2 Tabaco <- read.csv2("../DATA/Datos limpios nuevo.csv")
3
4 Tabaco <- Tabaco[,-c(2)]
5
6
7 # VARIABLE NUMÉRICA "Exportación"
8 x1 <- Tabaco$Exportación
9 # Aplicar la regla de Sturges
10 k1 <- nclass.Sturges(x1)
11 # Calcular la amplitud
12 Int1 <- cut(x1, breaks = k1)
13 Int1
14
15 # Construcción de Tabla de Frecuencias para "Exportación"
16 Fre_Exportacion <- as.data.frame(table(Int1))
17 Fre_Exportacion
18 Tab_Fre_Exportacion <- transform(Fre_Exportacion,
19                                 Frel = round(prop.table(Fre_Exportacion$Freq), 3),
20                                 FAcu = cumsum(Fre_Exportacion$Freq),
21                                 FAcuR = round(cumsum(prop.table(Fre_Exportacion$Freq)), 3))
22 Tab_Fre_Exportacion
23
24
25 # Gráfico de barras para "Exportación"
26 G2 <- ggplot(Tab_Fre_Exportacion, aes(x=Int1, y=Frel)) +
27   geom_bar(stat = "identity", fill="yellow", colour="black", size=0.5) +
28   geom_text(aes(label=paste0(Frel)), position = position_stack(vjust = 0.8)) +
29   coord_flip() +
30   labs(title = "Distribución de la Variable Exportación",
31        x = "Rango de Exportación",
32        y = "Frecuencia Relativa")
33
34 G2
35
36 ##### Medidas de tendencia central #####
37 # Para la variable "Producción"
38 mean(Tabaco$Producción, na.rm = TRUE) # Media
39 median(Tabaco$Producción, na.rm = TRUE) # Mediana
40 library(modeest)
41 mfv(Tabaco$Producción, na.rm = TRUE) # Moda
42 library(moments)
43 skewness(Tabaco$Producción, na.rm = TRUE) # Coeficiente de asimetría
44 var(Tabaco$Producción, na.rm = TRUE) # Varianza
45 sd(Tabaco$Producción, na.rm = TRUE) # Desviación estándar
46 sd(Tabaco$Producción, na.rm = TRUE) / mean(Tabaco$Producción, na.rm = TRUE) # Coeficiente de variación
47 kurtosis(Tabaco$Producción, na.rm = TRUE) # Curtosis
48
49 moda <- mfv(Tabaco$Producción, na.rm = TRUE)
50 moda <- moda[1]
51
52 # Gráfico de normalidad para "Producción"
53 GR3 <- ggplot(Tabaco, aes(x=Producción)) +
112:1 [?] Medidas de tendencia central
```



```

52 # Gráfico de normalidad para "Producción"
53 GR3 <- ggplot(Tabaco, aes(x=Producción)) +
54   geom_histogram(aes(y=..density..), fill="#B6F0B0", colour="#23AB57") +
55   geom_density(alpha=.2, fill="#AF41B2") +
56   geom_vline(aes(xintercept=mean(Producción, na.rm = TRUE)), color="red", size=1.5) +
57   geom_vline(aes(xintercept=median(Producción, na.rm = TRUE)), color="blue", size=1.5, linetype="dashed") +
58   geom_vline(aes(xintercept=moda), color="yellow", size=1.5) + # Usando solo una moda
59   labs(title = "Histograma de Producción")
60
61 GR3
62
63 # Medidas de tendencia central para "Ventas"
64 mean(Tabaco$Ventas, na.rm = TRUE) # Media
65 median(Tabaco$Ventas, na.rm = TRUE) # Mediana
66 mfv(Tabaco$Ventas, na.rm = TRUE) # Moda
67 moda <- moda[1]
68
69 skewness(Tabaco$Ventas, na.rm = TRUE) # Coeficiente de asimetría
70 var(Tabaco$Ventas, na.rm = TRUE) # Varianza
71 sd(Tabaco$Ventas, na.rm = TRUE) # Desviación estándar
72 sd(Tabaco$Ventas, na.rm = TRUE) / mean(Tabaco$Ventas, na.rm = TRUE) # Coeficiente de variación
73 kurtosis(Tabaco$Ventas, na.rm = TRUE) # Curtosis
74
75 # Gráfico de normalidad para "Ventas"
76 GR4 <- ggplot(Tabaco, aes(x=Ventas)) +
77   geom_histogram(aes(y=..density..), fill="#B6F0B0", colour="#23AB57") +
78   geom_density(alpha=.2, fill="#AF41B2") +
79   geom_vline(aes(xintercept=mean(Ventas, na.rm = TRUE)), color="red", size=1.5) +
80   geom_vline(aes(xintercept=median(Ventas, na.rm = TRUE)), color="blue", size=1.5, linetype="dashed") +
81   labs(title = "Histograma de Ventas")
82
83 GR4
84
85 ## Gráficas Especiales ##
86 # Relación entre "Producción" y "Ventas"
87 GRES1 <- ggplot(Tabaco, aes(x=Producción, y=Ventas)) +
88   geom_point(color="blue", alpha=0.6) +
89   labs(title = "Relación entre Producción y Ventas",
90         x = "Producción",
91         y = "Ventas")
92
93
94 GRES1
95

```

```

96
97 # Relación entre "Exportación" y "Ventas"
98 GRES2 <- ggplot(Tabaco, aes(x = Exportación, y = Ventas)) +
99   geom_point(color = "blue", alpha = 0.6) +
100   labs(title = "Relación entre Exportación y Ventas",
101         x = "Exportación",
102         y = "Ventas") +
103   theme_minimal()
104
105 GRES2
106
107
108 # Tablero de control
109 install.packages("gridExtra")
110 library(gridExtra)
111 Tablero_De_Control <- grid.arrange(G2, GR3,GR4,GRES1,GRES2)
112
113

```

112:1 # Medidas de tendencia central ↕

```

1 setwd("C:/Users/jseba/OneDrive/Desktop/B.I A.2023 N.I/DATA")
2 Tabaco <- read.csv2("../DATA/Datos limpios nuevo.csv")
3
4 rownames(Tabaco) <- Tabaco$Pais
5 Tabaco$Pais <- NULL
6
7 Tabaco <- Tabaco[,-c(1)]
8 set.seed(2018)
9 library(caret)
10
11 Entrenamiento <- createDataPartition(Tabaco$Ventas, p=0.80, list= FALSE)
12 Entrenamiento
13
14 TabacoM <- lm(Ventas~., data= Tabaco[Entrenamiento,])
15 summary(TabacoM)
16
17 boxplot(TabacoM$residuals)
18
19 par(mfrow= c(2,2))
20
21 plot(TabacoM)
22
23 #Prediccion de tendencia
24
25 Prediccion <- predict(TabacoM, Tabaco[Entrenamiento,])
26 Prediccion
27
28
29 #####Evaluacion de la raiz del error cuadrático medio#####
30
31 Tabaco$pred <- predict(TabacoM, Tabaco)
32
33 #ejecuto la raiz del error cuadratico medio
34
35 Tabaco$rmse <- sqrt(mean(Tabaco$Ventas - Tabaco$pred)^2)
36
37 par(mfrow= c(1,1))
38
39 plot(Tabaco$Ventas, Tabaco$pred)
40 abline(0,1)
41
42 #Crear los intervalos
43
44 Tabaco$L.inferior <- Tabaco$pred - Tabaco$rmse
45 Tabaco$L.superior <- Tabaco$pred + Tabaco$rmse
46
47 mean(Tabaco$Ventas)
48 mean(Tabaco$pred)
49

```



**Presidencia  
de la República  
del Ecuador**



**Plan Nacional  
de Ciencia, Tecnología,  
Innovación y Saberes**



**SENESCYT**  
Secretaría Nacional de Educación Superior,  
Ciencia, Tecnología e Innovación

## DECLARACIÓN Y AUTORIZACIÓN

Yo, **Ortega García, Juan Sebastián** con C.C: **#0931458087** autor del trabajo de integración curricular: **Predicción de la demanda de tabaco en los mercados internacionales mediante machine learning**, previo a la obtención del título de **Licenciado en Negocios Internacionales** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, **23 de agosto** de 2024

f. \_\_\_\_\_

**Ortega García, Juan Sebastián**

**C.C: #0931458087**



## REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

### FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

<b>TEMA Y SUBTEMA:</b>	Predicción de la demanda de tabaco en los mercados internacionales mediante machine learning.		
<b>AUTOR(ES)</b>	Ortega García, Juan Sebastián		
<b>REVISOR(ES)/TUTOR(ES)</b>	Carrera Buri, Félix Miguel Mgs.		
<b>INSTITUCIÓN:</b>	Universidad Católica de Santiago de Guayaquil		
<b>FACULTAD:</b>	<b>Economía y Empresa</b>		
<b>CARRERA:</b>	<b>Negocios Internacionales</b>		
<b>TÍTULO OBTENIDO:</b>	<b>Licenciado en Negocios Internacionales</b>		
<b>FECHA DE PUBLICACIÓN:</b>	23 de agosto de 2024	<b>No. DE PÁGINAS:</b>	88
<b>ÁREAS TEMÁTICAS:</b>	Comercio exterior, estrategia de marketing		
<b>PALABRAS CLAVES/ KEYWORDS:</b>	Industria tabacalera, machine learning, mercados internacionales, predicción de demanda.		
<b>RESUMEN/ABSTRACT (150-250 palabras):</b>	<p>El presente trabajo utiliza el concepto de machine learning para mejorar la competitividad de las empresas tabacaleras, basado en esto se desarrolló un modelo predictivo basados en algoritmos de aprendizaje automático, que anticipan fluctuaciones en la demanda y optimizan la cadena de suministro, lo cual es crucial para la toma de decisiones estratégicas en marketing y ventas. La metodología incluyó la recopilación de datos históricos del mercado tabacalero, el procesamiento de estos datos mediante algoritmos de machine learning, y la validación de los modelos predictivos a través de técnicas de regresión lineal y análisis competitivo. Los resultados muestran que el uso de machine learning mejora significativamente la precisión de las predicciones de demanda, permitiendo a las empresas ajustar sus estrategias comerciales y optimizar la gestión de inventarios. Además, la implementación de estas tecnologías facilita la identificación de oportunidades de crecimiento en mercados internacionales. En conclusión, la aplicación de machine learning en la industria tabacalera ofrece un enfoque innovador para predecir la demanda, contribuyendo a la eficiencia operativa y a la competitividad de las empresas en un entorno globalizado.</p>		
<b>ADJUNTO PDF:</b>	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
<b>CONTACTO CON AUTOR/ES:</b>	<b>Teléfono:</b>	<b>E-mail:</b> juan.ortega07@cu.ucsg.edu.ec	
<b>CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UIC):</b>	<b>Nombre:</b> César Enrique, Freire Quintero		
	<b>Teléfono:</b> +593 990090702		
	<b>E-mail:</b> cesar.freire@cu.ucsg.edu.ec		
<b>SECCIÓN PARA USO DE BIBLIOTECA</b>			
<b>Nº. DE REGISTRO (en base a datos):</b>			
<b>Nº. DE CLASIFICACIÓN:</b>			
<b>DIRECCIÓN URL (tesis en la web):</b>			