

**Microbiología predictiva mediante aprendizaje
automatizado para la optimización de procesos
productivos: Metanálisis.**

Autor (Author): Verónica Yepes Medina

Asesor (Advisor): Nicolás Pinel

Maestría en Biociencias (Master Programme in Biosciences)

Res. 11910 de Noviembre 14 de 2019. Válida hasta Noviembre 13 de 2026.

SNIES 108453

Escuela de Ciencias Aplicadas e Ingeniería
(School of Applied Sciences and Engineering)

Universidad EAFIT

2023

Capítulo 1

Introducción

Los alimentos de origen animal o vegetal son ecosistemas complejos ricos en nutrientes con características fisicoquímicas que permiten el crecimiento microbiano durante su procesamiento y almacenamiento. Estos ecosistemas están colonizados por comunidades microbianas que pueden incluir microorganismos patógenos, alterantes o benéficos. El consumo de alimentos contaminados con patógenos es una causa importante de morbilidad y mortalidad a nivel mundial. Cada año, aproximadamente 600 millones de personas (1 de cada 10 personas) se enferman a causa de patógenos transmitidos por alimentos, 420.000 de los cuales mueren. El daño humano causado por patógenos transmitidos por alimentos resulta en pérdidas económicas colosales que ascienden a 110 mil millones de dólares debido a la pérdida de productividad y gastos de salud. Las causas del desperdicio y la pérdida de alimentos son numerosas, pero una parte importante de la destrucción de alimentos relacionada con la contaminación microbiana se debe al incumplimiento de las reglamentaciones relacionadas con patógenos o al deterioro. A nivel de la Unión Europea, el 20% del total de alimentos disponibles se pierde o desperdicia, siendo las frutas y verduras la categoría más afectada (43,5%) por delante de los productos cárnicos y pesqueros (26,3%). En todo el mundo, se estima que un tercio de los alimentos producidos para el consumo humano, alrededor de 1.300 millones de toneladas por año, se pierden o desperdician a lo largo de la cadena de suministro de alimento, mientras que alrededor del 12% de la población mundial sufre de hambre. Por lo tanto reducir el desperdicio de alimentos es crucial no sólo por razones éticas sino también por razones económicas [1]

1. Enfoque seguridad alimentaria

El enfoque "tradicional" de la gestión de la calidad y la seguridad de los alimentos se basa en las pruebas de los productos finales. Sin embargo, las pruebas de productos finales sólo proporcionan una información muy limitada sobre el estado de seguridad de un alimento. La crisis de seguridad alimentaria de los años 90 (por ejemplo, *Listeria*, *Salmonella*, *Escherichia coli*, *Campylobacter*, dioxinas, antibióticos, acrilamida, EEB, etc.) puso de manifiesto el fracaso de este enfoque tradicional. Si se encuentra un organismo peligroso, significa algo; pero su ausencia en un número limitado de muestras no es garantía de la seguridad de todo

un lote de producción. Las pruebas del producto final suelen ser demasiado escasas y se realizan demasiado tarde. Hoy en día, los sistemas eficaces de gestión de la calidad y la seguridad de los alimentos deben basarse en programas de prerrequisitos (por ejemplo, buenas prácticas de fabricación y buenas prácticas de higiene), en el plan de Análisis de Peligros y Puntos de Control Crítico (APPCC), así como en herramientas cuantitativas, concretamente en enfoques de microbiología predictiva y evaluación de riesgos[2]–[4] [5]

Las pruebas precisas y apropiadas para detectar peligros potenciales en la cadena de suministros de alimentos son clave para garantizar la seguridad del consumidor y la calidad de los alimentos. El monitoreo y las pruebas periódicas de los ingredientes pueden revelar fluctuaciones dentro de la cadena de suministro que pueden ser un indicador de la calidad de un ingrediente o de un peligro potencial. Dicha calidad se evalúa mediante pruebas estandarizadas de composición química y microbiológica para cumplir con los requisitos y especificaciones de las agencias gubernamentales de todo el mundo. Para que las materias primas o los productos terminados cumplan con estos límites de seguridad y calidad, su composición generalmente debe tener una carga microbiológica baja y ser químicamente idénticos en macro componentes como carbohidratos, proteínas y grasas. Los métodos en este segmento deben evitar resultados falsos negativos que pueden poner en peligro a los consumidores [6] [7].

Un aspecto preocupante relativo a la gestión de la seguridad alimentaria es el recurso continuo e indiscriminado al análisis microbiológico de los productos finales, la lógica de producción asume que muestrear solo el producto final es suficiente. Es habitual encontrarse con las siguientes afirmaciones: *“si muestreo todos los lotes de alimento tengo verificado el proceso”*, *si tengo una desviación en el proceso, tomo más muestras*, *“la calidad es un gasto y no una inversión y las pruebas microbiológicas realizadas toman hasta 8 días, dado que se usan metodologías tradicionales o quién realiza los análisis no tiene mayor conocimiento de cómo funcionan los microorganismos de manera individual y dentro de un sistema biológico”* y *“la microbiología es una caja de Petri con unas colonias”*[8]. En consecuencia a los anteriores pensamientos entregan “Actas de defunción” porque el producto analizado al final de su proceso no tiene opción de ser ajustado o modificado para

que cumpla con los requisitos necesarios para su distribución, son toneladas de alimento que se van a una cava de devolución o ingresan como “desperdicios” a otros eslabones de la cadena con pérdidas millonarias. ¿Cómo hacer más eficientes los procesos? La respuesta se encuentra en la combinación de varios mundos: la **ecología microbiana**, entendida como el estudio de las interacciones de los microorganismos con su entorno, entre sí y con las especies vegetales y animales. La **minería de datos** o exploración de datos es un campo de la estadística y las ciencias de la computación referido al proceso de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático y sistemas de bases de datos. Y **estadística** como ciencia que se encarga de recopilar, analizar, presentar e interpretar los datos [9]. Suena complicado, costoso y poco viable, pero no; cuando comprendemos la forma en la que se desarrollan los microorganismos y las condiciones que favorecen o no su desarrollo podríamos predecir como pueden comportarse dentro de un sistema y hacerlo escalable usando la minería de datos para tener patrones en el tiempo con niveles de precisión significativos que complementen el trabajo realizado en los laboratorios con pruebas tradicionales, alternativas y/o moleculares. Piense en algo, para reproducirse cualquier ser vivo requiere: nutrientes, tiempo, temperatura y variables como pH, actividad de agua (aw) y oxígeno, si usted conoce dichos requisitos podrá diseñar líneas de producción donde sean potenciados o controlados para favorecer o evitar la aparición de ciertas cepas microbianas [10]. En la producción de yogurt con probióticos del tipo *Bifidobacterium* se requiere que después de los tanques de maduración el paso de la base de yogurt hacia saborización se haga sin la inclusión de oxígeno dado que esta bacteria ácido-láctica es anaerobia estricta y el ingreso de oxígeno al sistema provoca su muerte [11]. Cuando se elaboran derivados cárnicos es natural que la materia prima contenga *Salmonella* o *Listeria monocytogenes* si esta es de origen avícola o bovina respectivamente [12]. La elaboración del derivado cárnico implica la validación de tiempos y temperaturas de cocción al igual que tiempos y temperatura de enfriamiento para su empaque para la eliminación de estos dos microorganismos patógenos. Generalmente los sistemas están diseñados para que el tratamiento térmico alcance los 75°C en cámara y se mantenga. Se debe a que *Salmonella* logra una reducción logarítmica a 70°C cada un minuto y este género cuenta con una resistencia térmica de 4,5°C y la temperatura debe reducirse a 4°C en menos de 2 horas dado que tiene crecimiento óptimo entre los 30 a 45°C [13]. No siempre se miden bacterias patógenas en los procesos por un tema de costos y

se hace uso de microorganismos índice, un incremento en la línea de proceso de *Escherichia coli* se correlaciona con un aumento en la probabilidad de aparición de *Salmonella*, evolutivamente *Escherichia coli* tiene un antecesor común con *Salmonella* y conserva genes de patogenicidad con este género lo que la convierte en un excelente indicador para este patógeno [13]–[15].

Aunque el concepto de modelo predictivo se introdujo a partir de la década de 1920 con el modelo de Bigelow, que describe la cinética de inactivación de los microorganismos o, en su defecto, la resistencia al calor de los mismos, se tardó unos 60 años en desarrollar modelos de crecimiento y muerte microbiana. Desde entonces, la microbiología predictiva se ha centrado en el desarrollo de modelos primarios deterministas para predecir el comportamiento microbiano en los alimentos en función del tiempo de almacenamiento (crecimiento y supervivencia) y del tiempo de tratamiento (inactivación) con la ayuda de modelos secundarios que describen el efecto de los factores intrínsecos, extrínsecos o de procesamiento sobre las variables cinéticas. A pesar de los avances en este ámbito, la mayoría de los modelos desarrollados en los años 90 se basaban en enfoques deterministas sin tener en cuenta la variabilidad de los factores que afectan a las respuestas microbianas. Los modelos deterministas de crecimiento o inactivación que producen estimaciones únicas "óptimas" no suelen ser óptimos para gestionar satisfactoriamente la seguridad de los alimentos. Si, por ejemplo, los efectos posteriores de los niveles insatisfactorios de un patógeno superviviente en los alimentos después del procesado son graves, es poco probable que la información sobre la disminución media de la población sea una base suficiente para el diseño del procesado, o si no se conocen los límites de crecimiento de un patógeno transmitido por los alimentos a nivel unicelular, la formulación del producto para la producción de alimentos más seguros sería una cuestión difícil. Además, el uso de un enfoque basado en el peor de los casos en el procesamiento de alimentos conduce a estimaciones poco realistas con un impacto negativo en la calidad de los alimentos [5] [16]–[18]

El desafío actual es cómo implementar estas tecnologías en la cadena industrial “de la granja a la mesa”, puesto que la principal limitación de los métodos dependientes de cultivo es que

solo permiten el estudio de microorganismos cultivables, por lo tanto, se corre el riesgo de no tener una visión de la biodiversidad de la población microbiana presente en el alimento [19]. Sin embargo, el uso de datos abiertos obtenidos a través de diferentes ómicas y analizados bajo modelos, algoritmos o tecnologías de predicción o aprendizaje automatizado permite la extracción la información cifrada del microbioma e identificar si hay microorganismos patógenos y/o alterantes. Esta información brinda nuevas oportunidades para explorar la diversidad y la funcionalidad de los microorganismos en los alimentos, así como en los entornos de procesamiento y producción de alimentos [19] [20] [21] [22]

Si partimos que la investigación del microbioma es el análisis de la composición, función y las interacciones bióticas y abióticas de los microorganismos en comunidades complejas [23]El análisis de microbioma actual generalmente consta de dos componentes importantes: perfiles de la comunidad aguas arriba (¿cuál es la abundancia de todos los microorganismos en cada muestra?) y análisis de alto nivel aguas abajo (análisis de diversidad alfa/beta, análisis de abundancia diferencial). En los últimos años, los métodos de análisis de datos, visualización y aprendizaje automático en evolución se han aplicado gradualmente al desarrollo de muchas herramientas de software y servidores web para el análisis de datos de microbiomas que cubren estos dos componentes[24]

Al tomar todo lo anterior lo que obtenemos es la observación de los factores ambientales, la integración de los datos en modelos estadísticos y la predicción del comportamiento bacteriano en los alimentos que nos permiten la identificación de patrones ocultos, dando lugar a implementación de intervenciones basadas en datos. La industria de alimentos necesita tecnologías adecuadas de pruebas de calidad y seguridad para monitorear y evaluar los indicadores en tiempo real antes de que la cadena de suministro llegue al consumidor, de manera que el personal de cada eslabón pueda realizar los ajustes oportunos para garantizar la inocuidad de los alimentos y reducir las pérdidas económicas [25]

La creciente sensibilización sobre la necesidad de reforzar la seguridad microbiológica de los alimentos es uno de los principales retos del sector alimentario. La gestión de riesgos de la seguridad alimentaria ha sido adoptada por las industrias alimentarias como parte de su

sistema de control oficial a lo largo de la cadena de producción, dado que los patógenos emergentes causan un número cada vez mayor de brotes y se consideran el principal riesgo para la salud pública, por lo que se requiere la evaluación exacta del potencial patógeno para predecir y gestionar su riesgo para la salud de manera anticipada. Se han utilizado métodos convencionales como la serotipificación y las combinaciones de genes de virulencia para evaluar los potenciales patógenos bacterianos. O las pruebas de productos finales basadas en un enfoque en el peligro ya han sido catalogadas como ineficientes ya que no proporcionan en muchos casos información cuantitativa sobre las condiciones sanitarias de un determinado alimento. El paso a un enfoque (cuantitativo) basado en el riesgo (probabilidad de que un peligro esté presente en un alimento) requiere la evaluación de los conocimientos generados (métodos estadísticos y computacionales) y la interpretación de los resultados mediante el desarrollo y aplicación de recursos (bases de datos y herramientas informáticas de software) disponibles para su uso por parte de la seguridad alimentaria, incluyendo evaluadores y gestores de riesgos, operadores alimentarios o instituciones de investigación [26], [27]

2. Microbiología predictiva

La microbiología predictiva se ha definido como el conocimiento de las respuestas microbianas a las condiciones ambientales, sintetizadas en modelos matemáticos, que permite una evaluación objetiva de las operaciones de procesamiento, distribución y almacenamiento sobre la seguridad microbiológica y la calidad de los alimentos. La microbiología predictiva que se basó inicialmente en tres principios fundamentales: en primer lugar, el crecimiento, la supervivencia y la inactivación de los microorganismos se consideran respuestas reproducibles; en segundo lugar, el comportamiento microbiano está dictado principalmente por un número limitado de factores ambientales y tercero, al cuantificar el efecto combinado de estos factores es posible predecir las respuestas de los microorganismos [27]–[29].

Los modelos predictivos son las herramientas matemáticas más efectivas para estimar los cambios en los niveles microbianos en los alimentos a medida que el producto avanza a través de la cadena de la granja a la mesa, y han sido reconocidos como componentes importantes en el proceso de evaluación de riesgos en seguridad y calidad de los alimentos[30], [31]. Las

bases de datos de las respuestas microbianas en los alimentos afectadas por factores intrínsecos y extrínsecos de los alimentos, como la temperatura, el pH y la actividad del agua (*aw*), son muy útiles para el desarrollo y la validación de modelos predictivos y análisis de riesgos. Actualmente se encuentra disponible una amplia gama de herramientas de software de seguridad alimentaria. Hoy en día, las herramientas de software de inocuidad de los alimentos ofrecen cada vez más una gama más amplia de soluciones a través de la creación de entornos computacionales dinámicos que facilitan la apertura y el informe sistemático del proceso de análisis de riesgos. Entre 2016 y 2023 se han lanzado los siguientes recursos y herramientas de software de microbiología predictiva (Tabla 1) [5] [31] [32].

Software	Escribe	Descripción	Disponible
<i>BiOED</i>	Paquete R para diseño experimental óptimo	Incluye algoritmos matemáticos para análisis y diseño experimental óptimo para la caracterización óptima de la inactivación microbiana	https://CRAN.R-project.org/package=bioOED
Patógenos en los alimentos	Base de datos de respuestas microbianas en alimentos	Contiene más de 6000 registros de comportamiento microbiano y facilita el acceso y recuperación de datos de acuerdo a diferentes criterios de selección, incluyendo el tipo de microorganismos, categoría de alimento y etapa de la cadena de producción	https://vcadavez.shinyapps.io/MeatProducts/
<i>base de datos D</i>	Base de datos de inactivación microbiana en alimentos	Proporciona a los científicos información para construir modelos de inactivación. Esta base de datos abarca más de 5000 registros de parámetros de inactivación microbiana en alimentos	https://foodmicrowur.shinyapps.io/Ddatabase/
<i>FSMR abierto</i>	Repositorio de modelos	Motor de búsqueda impulsada por la comunidad para modelos microbiológicos predictivos	https://knime.fmr.berlin/opensfmr/
<i>ComBase Premium</i>	Herramienta de predicción y repositorio de modelos	Portal web de modelos predictivos y herramientas de gestión de riesgos para ayudar a mejorar la seguridad y calidad de los alimentos. Los modelos CB Premium se centran en cómo crecen y mueren los microorganismos en los alimentos comerciales	https://www.cbpremium.org/
<i>Modelos predictivos DMRI para carne</i>	Herramienta de predicción	Incluye modelos predictivos de seguridad y vida útil para varios microorganismos en diferentes productos cárnicos, validados en condiciones estáticas y dinámicas	http://dmripredict.dk/

<i>ListeriaCardenalModelo</i>	Herramienta de predicción	Esta herramienta permite simular un modelo de parámetro cardinal metanalizado de <i>L.monocytogenes</i> desarrollado utilizando datos derivados de experimentos de crecimiento de 2009	https://vcadavez.shinyapps.io/ListeriaCardinalModel/
<i>praedicere possumus</i>	Herramienta de predicción	Incluye modelos predictivos para evaluar la probabilidad de crecimiento y la densidad de diversas bacterias patógenas en los alimentos	http://praedicer.e.uniud.it/
<i>MicroHibro</i>	Herramienta de microbiología predictiva y evaluación de riesgos en alimentos	Es un software de microbiología predictiva on-line de acceso gratuito. Se basa en una completa base de datos de modelos predictivos. Los usuarios pueden ingresar y compartir modelos utilizando un sistema estandarizado. Se pueden realizar modelos de evaluación de riesgos microbianos de estimación puntual y probabilística.	http://www.microhibro.com
<i>Food Spoilage and Safety Predictor (FSSP)</i>	Crecimiento de <i>Listeria monocytogenes</i> en queso	Software integrado a modelos predictivos desarrollados en quesos	http://fssp.food.dtu.dk/
<i>Dairy Product Safety Predictor</i>	Crecimiento de <i>Listeria monocytogenes</i> , <i>Salmonella spp</i> , <i>E.coli</i> y <i>Staphylococcus aureus</i> en queso azul, queso prensado, queso fresco	Software integrado a modelos predictivos desarrollados en quesos	www.aqr.maisondulait.fr
<i>GroPIN</i>	Varios microorganismos en quesos para evaluar crecimiento o supervivencia	Software integrado a modelos predictivos desarrollados en quesos	www.aua.gr/p/somas/gropin/
<i>Combase Premium</i>	Crecimiento de <i>E.coli</i> y <i>Listeria monocytogenes</i> en	Software integrado a modelos predictivos desarrollados en quesos	www.cbpremiun.org

	Queso Camembert, Queso Brie, Mozarella, Cheddar con leche sin pasteurizar y pasteurizada		
--	---	--	--

Tabla 1: *Modelos de predicción entre 2016 y 2023.*

Estos modelos son clasificados en 3 tipos: primario, secundario y terciario. Los modelos primarios se ocupan de la descripción de la evolución microbiana (crecimiento, supervivencia o inactivación) en función del tiempo (todos los demás parámetros son conocidos y fijos). Las cantidades modeladas pueden incluir unidades formadoras de colonias, biomasa, mediciones de absorbancia, así como niveles de sustrato o compuestos metabólicos producidos. Los modelos secundarios caracterizan diferentes parámetros que aparecen en los modelos primarios en función de las condiciones ambientales como la temperatura, el pH, la actividad del agua, entre otros. Por último, los modelos terciarios representan un modelo predictivo con una interfaz fácil de usar que combina modelos primarios y secundarios con bases de datos que recogen los parámetros de entrada del modelo [29]. Los tres tipos de modelos, primarios, secundarios y terciarios son necesarios para construir una herramienta completa de microbiología predictiva que pueda ayudar a predecir el crecimiento microbiano [33].

La integración de modelos predictivos de microbiología y evaluación de riesgos en herramientas informáticas de inocuidad de los alimentos puede dimensionarse según la perspectiva de los usuarios, clasificándose en tres componentes principales: matemático, de predicción y de aplicabilidad. [34]

La dimensión matemática se corresponde con la detección y selección de las funciones matemáticas más adecuadas que describen las observaciones microbianas utilizando diferentes enfoques matemáticos (por ejemplo, análisis de regresión, enfoque bayesiano, aprendizaje automático). Al analizar la función matemática, se puede obtener información

sobre el sistema, es decir, la definición de variables independientes específicas y la relación entre ellas y la variable respuesta (es decir, modelos deterministas). Además de la variable respuesta, como el crecimiento, la tasa de mortalidad, otras variables ambientales pueden depender de otras y este hecho debe describirse matemáticamente. Un ejemplo de esta relación de dependencia podría ser el pH o el Aw de los embutidos como factores dependientes del tiempo de maduración o almacenamiento.[35]

El componente de predicción del modelo representa el uso de modelos predictivos para prever cómo se comportarán los microorganismos bajo ciertas condiciones. La capacidad de predicción de los modelos generalmente se evalúa comparando los modelos con las observaciones obtenidas en experimentos independientes con los utilizados para el desarrollo del modelo. Este procedimiento normalmente se denomina validación y puede ser respaldado por el cálculo de índices de bondad de ajuste. La predicción a partir de modelos validados se puede utilizar para obtener información por adelantado que se puede utilizar para desarrollar medidas preventivas.[34], [35]

Los dos componentes anteriores son importantes desde un punto de vista científico, mientras que el componente de aplicación es relevante a nivel operativo o normativo a lo largo de la cadena alimentaria. La aplicación de microbiología predictiva y modelos de análisis de riesgo microbiológicos se basa en la capacidad de hacerlos disponibles una vez que se publican y validan. La integración de elementos de ingeniería de software en el modelado es crucial para proporcionar una dimensión de aplicabilidad a los modelos microbiológicos predictivos. [5], [36]

Ahora bien, el rápido progreso en la disciplina de la microbiología es inseparable de la actualización de los métodos o técnicas de observación en el mismo período. Con el advenimiento de la era de Big Data, las preguntas apremiantes para los investigadores han evolucionado gradualmente hacia cómo filtrar/condensar rápida y eficientemente este crecimiento exponencial de información para obtener datos de calidad generalizados y cómo transformar los datos masivos del microbiota en datos fácilmente comprensibles y visualizables. En comparación con la investigación tradicional con datos insuficientes o

técnicas que causan sesgo cognitivo, la baja reproducibilidad y lapsos de tiempo prolongados, es más probable que el proceso moderno de investigación en microbiología incorpore nuevas tecnología y métodos de macrodatos para hacerlo mejor y correcto [37].

3. Aprendizaje automático.

El aprendizaje automático (ML), propuesto por primera vez por Arthur Samuel en 1959, un subcampo central de la inteligencia artificial (IA), es una técnica comprobada que genera impulso en el dominio de la microbiología. Se ha utilizado hasta hora en problemas computacionalmente intensos, como la predicción de objetivos farmacológicos y vacunas candidatas, el diagnóstico de microorganismos que causan enfermedades infecciosas, la clasificación de la resistencia a los medicamentos contra los medicamentos antimicrobianos, la predicción de brotes de enfermedades y la exploración de interacciones microbianas. El aprendizaje automático se ha aplicado a todas las áreas de investigación en microbiología, incluidas virología, parasitología [37]–[39].

Antes del aprendizaje automático, los intentos de realizar tareas complejas como el reconocimiento de letras e imágenes implicaban programar computadoras con sistemas basados en reglas. Estos sistemas tendían a fallar fuera de los escenarios de prueba con datos no vistos. El aprendizaje automático ha cambiado el paradigma de programar computadoras sobre cómo realizar una tarea a uno en el que las computadoras aprenden cómo realizar la tarea sin ser programadas y ha ganado popularidad en la era de los grandes datos en la comunidad científica. Mientras que el modelado mecanicista se basa principalmente en formulaciones matemáticas simplificadas para resolver problemas relacionados con conjuntos de datos complejos, los algoritmos de ML reconocen patrones y/o estructuras de datos subyacentes y hacen predicciones basadas en los patrones aprendidos en dichos conjuntos de datos[40]. El aprendizaje automático es un campo de estudio extenso que se superpone y hereda ideas de campos relacionados como la estadística, la informática e inteligencia artificial, dedicado a explorar para simular o implementar comportamientos de aprendizaje humano, reorganizar las estructuras de conocimiento existentes y mejorar continuamente su rendimiento para predecir o clasificar otros datos desconocidos [39]. Y se

refiere a un tema emergente en el que los algoritmos pueden aprender y desarrollarse por sí mismos sin necesidad de una codificación explícita [25], [41].

El aprendizaje automático tiene dos modos de aprendizaje principales: supervisado o predictivo para hacer predicciones futuras a partir de datos de entrenamiento, y no supervisado o descriptivo, que es de naturaleza exploratoria sin datos de entrenamiento, objetivo definido o resultado. El objetivo del aprendizaje automático supervisado es encontrar un modelo que relacione lo mejor posible los datos de entrada y de salida. Por ejemplo una imagen de una placa de Petri que contiene un medio de cultivo podría considerarse como “crecimiento presente” o “crecimiento ausente”. Un ordenador desarrolla un algoritmo para clasificar cada imagen y compara la precisión del algoritmo con la “verdad supervisada”; normalmente, los resultados iniciales no proporcionan un rendimiento adecuado y el algoritmo ajustará iterativamente sus parámetros por ensayo y error para optimizar la precisión [42]. Otro ejemplo de aprendizaje automático supervisado es su eficacia para atribuir las fuentes de alimentos a los casos clínicos de enfermedades transmitidos por alimentos como la listeriosis a partir de los datos de secuenciación. La inferencia de información genética a partir de los genotipos patógenos a menudo resulta crucial para la inferencia biológica. La atribución del origen de las infecciones por *Listeria monocytogenes* permite a los profesionales de la industria alimentaria, gestores de datos, epidemiológicos, microbiólogos y bioinformáticas adaptar sus prácticas para prevenir la propagación de patógenos transmitidos por los alimentos. También permite utilizar de forma más eficiente los recursos para contener la supervivencia y la proliferación de patógenos en su origen [43].

Los modelos de entrenamiento de ML suelen tener muestras divididas aleatoriamente en tres partes: conjunto de entrenamiento, conjunto de validación y conjunto de prueba. Entre ellos, el conjunto de entrenamiento se usa para establecer el modelo, el conjunto de validación se usa para ajustar los parámetros que controlan la complejidad del modelo y monitorear si el modelo parece estar sobreajustado, y el conjunto de prueba verifica la capacidad de generalización del modelo óptimo[25], [39]–[41] . ML se aplica típicamente en situaciones donde hay grandes conjuntos de datos disponibles y se pueden relacionar con resultados de

interés conocidos [40], [41]. Los principales algoritmos tradicionales de aprendizaje automático utilizados en la investigación alimentaria, son la regresión de mínimos cuadrados parciales (PLSR), la regresión lineal múltiple (MLR), vecinos más cercanos K (KNN), el análisis discriminante de mínimos cuadrados parciales (PLS-DA), árboles de decisión, el bosque aleatorio (RF), el modelado de independencia suave de analogía de clase (SIMCA), red neuronal (NN) y máquina de vectores de soporte lineal y radial (SVM) [25], [40], [42].

Las redes neuronales artificiales (ANN) se han utilizado para evaluar el comportamiento microbiano. Las ANN se han utilizado para adaptarse a patrones complicados en los datos. Las ANN se han examinado esporádicamente como un enfoque de modelado predictivo para la inactivación bacteriana, para el crecimiento bacteriano, para el comportamiento de crecimiento e inactivación y crecimiento /no crecimiento. El modelo basado en ANN mejora la precisión en comparación con los modelos estadísticos clásicos. También se han aplicado modelos de máquinas de soporte vectorial y Random forest a la seguridad alimentaria y agricultura [33].

El uso de algoritmos de aprendizaje automático en la seguridad alimentaria desde el 2020 se ha incrementado en respuesta a la complejidad de cuantificar algunas características y sus efectos en la dinámica de la población microbiana [44]. En la tabla 2 se discriminan los estudios entre el 2019 a 2023 publicados en el dominio de seguridad alimentaria usando aprendizaje automático.

Tanui, Karanth, et al., 2022. utilizaron algoritmos de aprendizaje automático para predecir la presencia de *Salmonella* en carne de pollo molida. Utilizando genotipos conocidos como características de entrada, el Random forest semisupervisado mostró la exactitud general más alta de 94% (intervalo de confianza del 95%: 85%-99%) y un valor Kappa de 82% y un 87% de predicción. El modelo predijo genes asociados con la virulencia que podrían usarse como características en los esfuerzos de modelado predictivo en el futuro, demostrando su utilidad en diferentes áreas de la seguridad alimentaria, incluida la identificación de fuentes de patógenos, la predicción de la resistencia a los antibióticos y la evaluación de los patógenos transmitidos por los alimentos.

Im et al., 2021 desarrollaron varios modelos de aprendizaje automático para predecir la patogenicidad de *Escherichia coli productora de toxina Shiga* (STEC) utilizando datos de secuenciación completa de genoma. El conjunto de datos de entrada para los modelos de aprendizaje automático se reveló utilizando distintos repertorios de genes de grupos de control positivos (patógenos) y negativos (no patógenos) en las que cada STEC aislada se designó en función de la atribución de la fuente, el potencial de riesgo relativo a las fuentes de aislamiento. Entre los modelos de aprendizaje automático examinados, el modelo Máquina de soporte vectorial (SVM), discriminó entre los dos grupos con mayor precisión. El modelo SVM predijo con éxito la patogenicidad de los aislamientos de las principales fuentes de brotes de STEC, sugiriendo que el modelo SVM es un método confiable y ampliamente aplicable para evaluar el potencial patógeno de los aislamientos de STEC en comparación con los métodos convencionales.

Huang et al., 2022 realizaron un estudio para definir el límite de crecimiento y no crecimiento de *Listeria monocytogenes* bajo el uso de diferentes conservantes usados en la industria cárnica y pH como factores de control. Desarrollaron un modelo de regresión logística multifactorial para calcular la probabilidad de crecimiento y luego definir el límite de crecimiento utilizando 2 umbrales. Con el umbral 1 ($p=0,0104$), la exactitud de clasificación para eventos de crecimiento fue de 68,6%, con una tasa de verdaderos positivos de 77,6% y una tasa de verdaderos negativos de 45,5% para los crecimientos obtenidos en agar PALCAM. Con el umbral 2 ($p=0,04$) la exactitud se convierte en 82,6% con una tasa de verdaderos positivos de 95.5% y una tasa de verdaderos negativos del 80%. Esta información se puede usar para formular productos listos para consumo que pueden prevenir el crecimiento de *Listeria monocytogenes* si se almacenan a la temperatura óptima.

Aplicación	Algoritmo Aprendizaje Automatizado	Exactitud	Tipo de datos	Referencia
Seguridad Alimentaria y Calidad	Máquina de soporte vectorial	100%	Imagen	[41].
	Vecinos más cercanos	90%		
	Naive Bayes	80%		
Peligros específicos Seguridad Alimentaria. Peligros Biológicos	Árbol aleatorio (Random Forest)	99.3%	Datos abiertos (genómicos)	[40].
	Logit boost (LB)	100%		
	Potenciación del gradiente (GB)	99.3%		
	Máquina de soporte vectorial con kernel lineal	98.7%		
	Máquina de soporte vectorial con kernel radial	97.4%		
Peligros específicos Seguridad Alimentaria. Peligros Biológicos Seguridad alimentaria General	Máquina de soporte vectorial	84%	Datos de monitoreo genómicos	[26].
	Árbol aleatorio (Random Forest)	84%		
	DTs	84%		
	Naive Bayes	84%		
	Árbol aleatorio (Random Forest)	80,20%		
	Naive Bayes	85%		
	Regresión logística	79,70%		
	Potenciación del gradiente (GB)	81,60%		
	Red Neuronal	80,30%		
	Máquina de soporte vectorial	86,95		
Adaboost(AB)	79.5% %	Texto	[46].	

	Bidirectional encoder representations from transformers (BERT)	85,50%		
Peligros específicos Seguridad Alimentaria. Peligros Biológicos	Regresión logística multifactorial	68,60%	Datos de monitoreo	[45]
Seguridad alimentaria General	Red Neuronal	82,50%	Datos de monitoreo	[47]
	Regresión logística kernel			
Peligros específicos Seguridad Alimentaria. Peligros Biológicos	Árbol aleatorio (Random Forest)	78,30%	Datos abiertos(genómicos)	[48]
	Logit boost (LB)	93,30%		
Seguridad alimentaria y Calidad	Adaboost(AB)	92,41%	Imagen	[49]
Peligros específicos Seguridad Alimentaria. Peligros Biológicos	Máquina de soporte vectorial con kernel lineal	89%	Datos abiertos(genómicos)	[50]
	Árbol aleatorio (Random Forest)	89%		
	Máquina de soporte vectorial con kernel radial	70%		
	Potenciación del gradiente (GB)	88%		
	Logit boost (LB)	86%		
Peligros específicos Seguridad Alimentaria. Peligros Biológicos	Regresión logística	76%	Datos abiertos(genómicos)	[51]
	Árbol aleatorio (Random Forest)	72%		
	Máquina de soporte vectorial con kernel lineal	70% %		
	Adaboost(AB)	72%		
	Árbol de decisión	63%	Datos abiertos	[52]

Peligros específicos Seguridad Alimentaria. Peligros Biológicos	Árbol aleatorio (Random Forest)	64%		
	Potenciación del gradiente (GB)	69%		
	Adaboost(AB)	67%		
Seguridad alimentaria General	Red Neuronal	96,25%	Imagen	[25]
	Máquina de soporte vectorial	90%		
Agua para agricultura. Riesgo Biológico	Árbol aleatorio (Random Forest)	83.2%	Datos de monitoreo	[53]
	Red Neuronal	86.1%		
	Máquina de soporte vectorial	77.3%		
	Naive Bayes	78.7%		
Seguridad alimentaria. Riesgo Biológico	Árbol aleatorio (Random Forest)	72.2%	Datos abiertos y Datos de monitoreo	[43]
	Logit boost (LB)	73.2%		
	Gradiente estocástico	70.1%		
	Máquina de soporte vectorial	61.4%		

Tabla 2: Estudios entre el 2019 a 2023 publicados en el dominio de seguridad alimentaria usando aprendizaje automático.

Los diferentes algoritmos tienen sus propias ventajas y desventajas, y no hay superioridad ni inferioridad. Lo que debe hacerse es interpretar completamente los datos de entrada en función de diferentes escenarios de demanda y luego construir modelos adecuados para ajustarse continuamente para lograr el mejor rendimiento. Además, la creencia de que “mientras se utilice el modelo más avanzado y complejo, el problema científico se resolverá” no es objetivo. En esencia, la tecnología informática solo ayuda a las personas a tomar decisiones o automatiza el proceso de toma de decisiones y mejora la eficiencia. Por lo tanto, la elección del modelo debe ser el más adecuado, en lugar de optar por el más complejo. Hay cuatro criterios utilizados para juzgar los méritos de los algoritmos de aprendizaje automático [54] (1). Corrección, el criterio más importante para juzgar los méritos de un algoritmo. (2) Robustez, es decir, tolerancia a fallas, que representa la capacidad del algoritmo para responder y abordar la entrada de datos ilegibles. (3) La legibilidad, los algoritmos fáciles de entender significan un proceso de depuración, modificación y expansión que requiere menos tiempo. (4) La temporalidad, es decir, la complejidad del tiempo y la complejidad del espacio, representan el esfuerzo computacional y el espacio de memoria necesarios para ejecutar el algoritmo, respectivamente [37].

Los avances y aplicaciones de nuevos métodos experimentales y computacionales impulsan la integración de la investigación en ecología microbiana con tecnologías de punta en biología computacional y otros campos. Simplificar el problema científico al predecir las relaciones entre las especies o su entorno (diversidad natural, ciclos de vida, interacciones y coevolución) alterando las estrategias de predicción basadas en las características de las especies.

La investigación del aprendizaje automático está evolucionando rápidamente, con arquitecturas, combinaciones de algoritmos y estrategias computacionales que cambian rápidamente. El objetivo final no es solo predecir la precisión de la tarea, sino también descubrir los procesos biológicos que subyacen a los problemas científicos. El desarrollo de herramientas analíticas efectivas, incluido el aprendizaje automático, garantiza la validez de los datos, la anotación adecuada y el intercambio abierto, y que la mayoría de los estudios

nacidos de la intersección de la microbiología y el aprendizaje automático muestren resultados.

4. Modelos de aprendizaje automatizado

4.1. Regresión logística

Es una técnica que pertenece a lo que se denomina modelos lineales generalizados. La principal característica de este modelo es que puede predecir una variable cualitativa a partir de varias predictivas, analiza la relación entre múltiples variables independientes y una variable dependiente categórica. [55]

Es un algoritmo de clasificación que modela la probabilidad de que un valor X pertenezca a una clase K . La probabilidad se mantiene en el rango 0-1.

La forma general del modelo es:

$$\log\left(\frac{p(X)}{1-p(X)}\right) = \beta_0 + \beta_1 X$$

Para mantener los valores en el rango 0-1, se utiliza la función de enlace:

$$p(x) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

Esto es llamado “*odds*”, toman valores entre 0 - ∞ .

El concepto de “*odds*” puede entenderse como la razón (*probabilidad*) que un evento ocurra. Valores cercanos a 0 indican baja probabilidad [46].

El modelo de regresión logística se utiliza ampliamente como enfoque estadístico estándar para estudiar la influencia de los determinantes en la presencia o ausencia de un evento peligroso en cualquier organismo o situación particular de seguridad de alimentos para humanos o animales. Un problema estadístico común que se produce en la aplicación de

modelos logísticos para analizar datos de respuesta binaria (ausencia/presencia), es el problema de la separación, que, si no se aborda adecuadamente, puede comprometer el proceso de inferencia y llevar a conclusiones sesgadas, por ejemplo, respecto a la relevancia de un factor concreto sobre la presencia o ausencia de un efecto en un organismo (por ejemplo, las conclusiones pueden indicar que no hay efecto mientras que el factor influye fuertemente en el resultado), produciendo inferencias y conclusiones erróneas. Por lo tanto, es necesario prestar más atención a los conjuntos de datos dispersos analizados mediante modelos de regresión logística. [56]. La regresión logística es útil cuando la varianza de la variable respuesta no es constante o no se distribuye normalmente o cuando está limitada en forma binaria. [57], [58]

4.2.Árbol de decisión para clasificación

Un árbol de decisión es una especie de modelo predictivo para clasificación y predicción. Representa una relación de mapeo entre los atributos del objeto y los valores del objeto. Cada nodo del árbol representa un objeto y cada ruta representa un posible valor de atributo. Cada hoja corresponde al valor del objeto representado por la ruta desde la raíz hasta el nodo hoja. El árbol de decisión es una técnica de uso frecuente en la minería de datos, que puede usarse para analizar datos y hacer predicciones [59].

Este modelo consiste en predecir la variable objetivo aprendiendo reglas de decisión simples inferidas de las características de los datos. Está definido como un clasificador que utiliza valores de los atributos para crear una partición recursiva basada en los datos de entrada. Construye diagramas lógicos en forma de árboles jerárquicos. Representan la categorización de los datos bajo una serie de condiciones aplicadas en forma de árboles sucesivos. En el árbol, cada nodo representa una prueba o decisión sobre un atributo, cada rama es el resultado de la prueba y cada nodo termina en una etiqueta de clase [55].

Son conceptualmente simples y fáciles de interpretar, pero su poder predictivo no es el mejor en términos de precisión. Realizan una segmentación de los datos predictores en regiones más simples, en las cuales se ajusta un modelo simple en cada una. Estos ajustes nos llevan a clasificar de acuerdo a si la condición es verdadera o falsa.

El algoritmo decide automáticamente las variables y puntos de corte, así como la forma del árbol.

$$f(x) = \sum_{m=1}^M C_m I(x \in R_m)$$

M= total de regiones

R= Región i

En los árboles de clasificación se predice que cada observación pertenece a la región según la ocurrencia más común de la categoría a predecir.

Existen diferentes medidas para realizar el proceso de segmentación de un árbol:

Error de clasificación (classification error rate)

$$E = 1 - \max_k(P_{mk})$$

Índice de Gini (Gini index)

$$G = \sum_{k=1}^K P_{mk}(1 - P_{mk})$$

Toma valores pequeños que indican que cada nodo (región) contiene principalmente observaciones de una única clase.

Entropía (Entropy)

$$D = - \sum_{k=1}^K P_{mk} \log p_{mk}$$

4.2.1. Ventajas y desventajas de los árboles de decisión

Fáciles de explicar e interpretar.

Se pueden observar de forma gráfica, lo cual facilita la interpretación de los resultados

Pueden usarse para obtener respuestas cualitativas o cuantitativas

No son el mejor algoritmo debido a su poder predictivo

Muy sensibles a cambios mínimos en el conjunto de datos

4.3.Árbol aleatorio (Random Forest)

Es un método de ensamble, que busca combinar muchos modelos más simples (*weak learners*) para obtener un único y potente modelo. Es un conjunto que consta de una colección de árboles de clasificación donde cada árbol es un vector aleatorio independiente distribuido idénticamente y estos árboles de clasificación votan por la clase más popular [60]. La técnica consiste en combinarlos y promediar los modelos mejorando los resultados. En comparación con los árboles de decisión, esta técnica suele reducir el sobreajuste, que es un problema de ellos [55].

Disminuye la varianza del modelo reduciendo las correlaciones entre los árboles

Este modelo mejora la capacidad predictiva a costa de la interpretabilidad que ofrecen los árboles de decisión. Realiza un muestreo aleatorio de las variables predictoras cada que realiza una división en el árbol. De esta forma no siempre se van a considerar los mismos predictores y particiones en cada árbol construido (árboles descorrelacionados) [46]

$$f_{rf}^B(x) = \frac{\sum_{b=1}^B T_b(x)}{B}$$

Para clasificación se tiene que

$$C_{rf}^B(x) = \text{mayoríadevotos}\{C_b(x)\}_1^B$$

Se siguen unos pasos básicos:

1. Seleccionar m variables aleatorias de las p variables posibles
2. Seleccionar el mejor punto de corte a través de los m variables
3. Fraccionar los nodos en 2 nodos hijos

Normalmente el valor de m es \sqrt{p}

Random Forest hace uso de la metodología de *out of the bag (OOB)*. Esta consiste en muestreo con reemplazo, en este caso los datos para el OOB son los datos escogidos durante el proceso de muestreo [46] [61]. Las estimaciones del error de OOB son casi idénticas a las obtenidas por validación cruzada. Una vez el error se estabiliza el entrenamiento del modelo puede terminar.

En resumen, es un algoritmo de aprendizaje que promedia los resultados de varios árboles de decisión. La combinación de múltiples árboles de decisión puede abordar problemas de sobreajuste, donde los clasificadores individuales a menudo aprenden patrones muy irregulares basados en datos de entrenamiento, lo que genera un sesgo bajo y una varianza extremadamente alta, lo que limita su aplicación más allá del conjunto de entrenamiento [57], [58].

4.4. Máquina de soporte vectorial (Support Vector Machine SVM)

Se basa en la separación por hiperplanos, de esta forma se logra separar lo mejor posible las categorías de una clase. Es de gran utilidad cuando existen relaciones no lineales en los datos usados para predecir una categoría [46] [55].

En SVM se hace uso de kernels.

- *Kernel*: la idea del kernel es usar funciones para transformar los datos con relaciones o superficies no lineales, a ecuaciones lineales, esto se hace aumentando las dimensiones del espacio. Con este kernel cuantifica la similitud entre las observaciones.
- *Lineal*

$$K(X_i, X_{i'}) = \sum_{j=1}^p X_{ij} X_{i'j}$$

Es un clasificador binario no probabilístico que tiene como objetivo encontrar el hiperplano óptimo entre dos clases en un espacio 2D. Los nuevos puntos de datos se asignan a una de dos clases dependiendo de en qué lado del hiperplano caen [57], [58].

- *Polinomial*

$$K(X_i, X_{i'}) = (1 + \sum_{j=1}^p X_{ij} X_{i'j})^d$$

- *Radial*

$$K(X_i, X_{i'}) = \exp(-\gamma \sum_{j=1}^p X_{ij} X_{i'j}^2)$$

Además de realizar una clasificación lineal, la máquina de vectores también puede realizar una clasificación no lineal aplicando un truco de kernel. Mediante el cual los predictores del

modelo se reemplazan con funciones de núcleo. Esto permite que el algoritmo opere en un espacio de características implícitas de alta dimensión; por ejemplo, un espacio 3D. Las coordenadas de los puntos de datos en el espacio recién transformado no se calculan explícitamente, lo que significa que este enfoque es más eficiente desde el punto de vista computacional que otros. En cambio, se calcula entre pares de datos [7], [57], [58]

Red Neuronal (single layer neural network)

Es el tópico más conocido como *deep learning*. Utilizado para problemas tales como clasificación de imágenes y vídeo, modelación de texto y audios.

La idea de las redes es tomar p variables y construir funciones no lineales para predecir la respuesta. Esto se logra haciendo activaciones diferentes entre las variables predictoras usando funciones de activación no-lineal.

La forma general de la red neuronal de una sola capa es la siguiente:

$$f(x) = \beta_0 + \sum_{k=1}^K \beta_k A_k$$

Donde A_k

$$A_k = h_k(X) = g(\omega_{k0} + \sum_{j=1}^p \omega_{kj} X_j)$$

- $g(z)$ es una función de activación no-lineal.
- A_k puede ser entendido como una transformación de las características (variables) originales.

Las métricas de calificación de los modelos fueron:

- **Exactitud (Accuracy):** No confundir este término con el de precisión. El accuracy hace referencia a que tan cerca los resultados de las mediciones de los valores reales. Mientras precisión hace referencia al grado de dispersión de dichas medidas y si este valor es bajo se relaciona con reproducibilidad de los resultados

$$Accuracy = \frac{\text{predicciones correctas}}{\text{todas las clasificaciones}}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Sensibilidad:** También llamado “rata de positivos verdaderos”. Como su nombre indica hace referencia a la probabilidad de predecir correctamente los valores positivos, es decir, devuelve el porcentaje de la categoría que cumple con la condición de ser la categoría deseada.

$$TPR = \frac{TP}{TP + FN}$$

- **Especificidad:** También llamado “rata de negativos verdaderos”. Contrario a la sensibilidad, busca la probabilidad de predecir correctamente los valores negativos.

$$TNR = \frac{TN}{TN + FP}$$

- **Área sobre la curva (ROC_AUC):** Es una representación gráfica de la sensibilidad vs la especificidad de un modelo de clasificación binario. Su cálculo parte de los conceptos anteriores. Se genera una curva con cada predicción realizada para la matriz de confusión, se toma cada punto y se genera la curva teniendo los valores de porcentaje de verdadero positivo vs porcentaje de falso positivo.

$$TPR(T) = \int_T^\infty f_1(x)dx \text{ y } FPR(T) = \int_T^\infty f_0(x)dx$$

El valor de la curva ROC puede interpretarse así:

Valor	Interpretación
0.5	Es el peor valor a obtener, equivale a lanzar una moneda
0.5-0.6	Malo
0.6-0.75	Regular
0.75-0.90	Bueno

- **Puntaje FI (f_means):** Es una medida del accuracy. Combina los test de precisión y recall

- **Precisión:** número de verdaderos positivos dividido todos los positivos resultados positivos
- **Recall:** número de verdaderos positivos dividido el número real de lo que debió ser predicho como positivo.

$$F_1Score = \frac{2TP}{2TP + FP + FN}$$

- **Matriz de confusión:** Es una forma de visualización del desempeño de un algoritmo. Se genera una matriz de filas y columnas donde el elemento representa los valores predichos y reales. La matriz de confusión es el punto de partida de muchas otras medidas de desempeño, entre ellas, exactitud, especificidad, sensibilidad, puntaje F1 y otras.

Todos los ajustes se hicieron partiendo de la metodología propuesta e implementada en *tidymodels*. La metodología consiste en diversos pasos (no todos tienen que ser implementados, solo es una sugerencia)

5. Condiciones de crecimiento de patógenos de interés en la industria de alimentos

En la industria de alimentos existen varios microorganismos patógenos que pueden afectar la salud pública, sin embargo, en este caso nos concentraremos en *Listeria monocytogenes* y *Salmonella* spp.

Listeria monocytogenes es una bacteria que puede causar una enfermedad intestinal leve no invasiva o una enfermedad grave, a veces potencialmente mortal denominada listeriosis invasiva [62].

Es psicotrópico, capaz de multiplicarse por debajo de 0°C. Puede crecer en pH entre 4.6 y 9.5 y a una actividad de agua (aw) por debajo de 0,92. Encuentra condiciones favorables de crecimiento en pisos, drenajes y equipos en la industria de alimentos. Las cepas de *Listeria monocytogenes* son frecuentemente encontradas en superficies de la industria de alimentos. No es claro aún, pero los hallazgos sugieren que en los lugares donde *Listeria monocytogenes* persiste, no existen otras especies, es decir, no se mezcla con otros niveles de microcolonias

y coincide con los reportes como un contaminante de “plantas limpias” o el “síndrome de las plantas limpias”

Por tanto la presencia de células bacterianas de una cepa específica en un lugar de “refugio” tras el proceso de limpieza y desinfección depende de: 1) la eficacia de la limpieza, 2) la eficacia de la desinfección y 3) el número de células antes del proceso de limpieza y desinfección. Se debe considerar el caso en el que el número inicial de células bacterianas en el lugar de acogida es lo suficientemente grande como para que algunas células sobrevivan después de varios ciclos de limpieza y desinfección diarios. Las bacterias supervivientes son capaces de adaptarse a bajas concentraciones residuales de los productos químicos utilizados para los procesos de saneamiento. Así, sea cual sea el número de células supervivientes adheridas, si se cumplen las condiciones de crecimiento, en términos de temperatura, agua, pH y nutrientes, el crecimiento se reanuda entre los ciclos de saneamiento. Al cabo de un tiempo si la reducción del número de células debida al proceso de limpieza y desinfección, la cepa persiste en el lugar de refugio. Por lo tanto, la persistencia de las cepas de *Listeria monocytogenes* se debe en gran medida a la incapacidad para desalojarlas de los lugares donde se alojan o matarlas allí, y a su propia capacidad de crecer a bajas temperaturas. [63], [64]

Salmonella spp es una bacteria que causa una enfermedad denominada salmonelosis. La mayoría de las personas sanas se recuperan, sin embargo la enfermedad puede diseminarse por el sistema sanguíneo y a otras partes del cuerpo provocando infecciones fatales. La dosis infectiva es poca de 15 a 20 células. La tasa de letalidad de la mayoría de las formas de salmonelosis es inferior al 1%, aunque suele ser superior en casos de fiebre tifoidea. Sin embargo, varias cepas pueden causar una enfermedad grave, por ejemplo, la tasa de letalidad de *Salmonella* Dublin es del 15% en caso de septicemia en ancianos, y la tasa de letalidad de *Salmonella* Enteritidis es de aproximadamente el 3,6% en brotes hospitalarios o en residencias de ancianos [65].

5.1. Características intrínsecas para la proliferación de microorganismos.

En algunos casos, la presencia de enfermedades transmitidas por alimentos, no presentan un riesgo a los consumidores a menos que ellos estén expuestos a un elevado número del microorganismo como por ejemplo esporas de *Bacillus cereus*. En otros casos, la presencia de un patógeno transmitido por los alimentos puede suponer un riesgo importante, incluso cuando están expuestos a un número bajo del organismo, como es el caso de Salmonella en un alimento listo para consumo [62].

A continuación se analizan los principales factores intrínsecos y extrínsecos que pueden influir en el crecimiento de patógenos bacterianos.

5.1.1. Impacto de la actividad de agua (aw)

La actividad de un alimento es un factor intrínseco clave que afecta al crecimiento de patógenos transmitidos por los alimentos. El término se refiere a la cantidad de agua no ligada que un microorganismo necesita para crecer. A medida que se elimina la humedad de un alimento o se liga a solutos como la sal o el azúcar, la aw disminuye [62].

La aw varía de 0 a 1, donde 1 representa agua pura, donde el crecimiento microbiano no es posible. Así, la adición de nutrientes reduce gradualmente el valor aw de los alimentos. Para considerar que un alimentos tiene aw bajo, debe tener valores inferiores a 0,83. Sin embargo, los estudios muestran que Salmonella puede sobrevivir en alimentos con una aw $\geq 0,94$, lo que refuerza su capacidad de causar enfermedades incluso si está presente en un número bajo de células. El número de brotes de Salmonella relacionados con alimentos con bajo contenido de aw ha aumentado cada año, y los principales alimentos involucrados son productos con alto contenido de azúcar, como el chocolate y los cereales de miel, los productos a base de semillas de sésamo, la mantequilla de maní y las almendras [66].

5.1.2. Impacto del pH

El pH es un factor intrínseco. La mayoría de las bacterias crecen mejor a un pH cercano a la neutralidad. Valores bajos de pH inhiben el crecimiento de bacterias patógenas transmitidas por alimentos y en algunos casos este factor intrínseco permite el control de patógenos dando

lugar a su muerte. Aunque muchas especies de patógenos transmitidos por alimentos mueren en condiciones de pH bajo, otras especies como *Escherichia coli* O157:H7 y *Salmonella* pueden sobrevivir en condiciones de pH bajo durante mucho tiempo, aunque su crecimiento pueda inhibirse. Por lo tanto la eficacia del pH como medida de intervención para eliminar o impedir el crecimiento de patógenos bacterianos transmitidos por alimentos es variable. Tales medidas de intervención requieren un control estricto durante la fabricación. La falta de dicho control puede dar lugar a la supervivencia y el crecimiento de patógenos transmitidos por alimentos provocando consecuencias para la salud o muerte [62].

Salmonella puede lograr la homeostasis del pH, que es cuando el pH intracelular se mantiene en comparación con el pH ambiental. La homeostasis se ve facilitada por las bombas de protones celulares y los sistemas antiporter de potasio/protón y sodio/protón. La capacidad de *Salmonella* para disminuir la extrusión de protones y la conductancia de protones de la membrana permite proteger a la célula contra el estrés ácido. Además *Salmonella typhimurium* tiene una respuesta regulada para proteger aún más del estrés ácido, lo que se denomina respuesta de tolerancia al ácido (ATR). El ATR protege a *Salmonella* spp. a niveles de pH de 3,0 a 4,0 pero se activa cuando los valores de pH ambiental están entre 6,0 y 5,5 y cuando falla la homeostasis del pH. Estas condiciones de pH se denominan etapa post-shock y etapa previa al shock, respectivamente. Durante la etapa postshock, se produce la estimulación de 43 proteínas de choque ácido para prevenir y reparar el daño causado a las macromoléculas por los ácidos [67].

5.1.3. Impacto de la temperatura

La temperatura es un factor extrínseco que afecta las condiciones de crecimiento de los patógenos transmitidos por alimentos. Todos los microorganismos tienen una condición de temperatura en la que ocurre su crecimiento. Algunos microorganismos patógenos no crecen o crecen de forma lenta en condiciones de refrigeración (ejemplo algunas especies de *Salmonella*), mientras que otros patógenos como *Listeria monocytogenes* crece en temperaturas de refrigeración.

Una aplicación inadecuada de temperaturas (bajas o altas temperaturas) puede dar lugar a la supervivencia y el crecimiento de patógenos transmitidos por los alimentos, lo que puede tener graves consecuencias para la salud o provocar la muerte [62] [65].

COVER LETTER

Manuscript Title: Predictive microbiology and machine learning by optimization
productive process: Metanalysis.

I/We certify that I/we have participated and contributed sufficiently for the completion of the manuscript and have agreed to have my/our name listed as a contributor.

I/We believe that the manuscript represents valid and credible work. Neither the content of this manuscript nor any other unified content with substantially similar or comparable substance under my/our creation or authorship has been published/is constantly considered for publication elsewhere. I/We certify that complete data regarding the study is proclaimed in this manuscript and no data from the study have been/will be published solely or separately. I/we attest that, I/we will cooperate in providing any required data/information related to the study during peer review by the editor or their assignees. Any direct or indirect financial interests or conflicts that exist or may be perceived to exist have been disclosed in the cover letter. Sources of external support for this work are mentioned in the cover letter.

I/We hereby exchange all copyright proprietorship, including all rights incidental thereto, solely to Acta Scientific, in the issue that such work is published by Acta Scientific. Acta Scientific grants the rights to the corresponding author to incorporate any necessary changes and he/she will act as the guarantor or surety for the manuscript on our behalf.

All persons who have made substantial and significant contributions to this work, but are not contributors, are mentioned in the acknowledgement with the written permission of the contributors to include their name. If there is no acknowledgement part in the manuscript, that means I/we have not received any contributions and also, no contributor has been omitted.

I/We have no conflicts of interest to disclose.

If you feel that the manuscript is appropriate for your journal, I/We suggest the following reviewers:

Name	Email	Affiliation	Country
------	-------	-------------	---------

Corresponding Author details

Name: Verónica Yepes Medina

Department: Antioquia


University/Organization: Universidad EAFIT

Country: Colombia

Phone Number: +57 312 8937254

Email: yepesmedinaveronica@gmail.com

List of Authors

S. No.	Name	Email	Signature
	Verónica Yepes Medina	yepesmedinaveronica@gmail.com	

I certify that the above information is correct, and I would like to submit manuscript towards the journal for publication.



Verónica Yepes Medina

Corresponding Author

(Name, Signature)

Capítulo 2

6. Métodos y Materiales

6.1. *Evaluar modelos de aprendizaje automático supervisado para predecir la presencia de microorganismos patógenos en la elaboración de alimentos cárnicos y lácteos.*

En concordancia con el estado del arte los modelos de aprendizaje automático supervisado usados principalmente en la industria son: regresión logística, Random Forest, Red Neuronal, árbol de decisión y máquina de soporte vectorial radial por lo que fueron incluidos como los modelos para evaluar y predecir la presencia de microorganismos patógenos, específicamente *Salmonella* spp y *Listeria monocytogenes* en la elaboración de alimentos cárnicos y lácteos.

6.2. *Recolección de datos*

Los datos para la evaluación de los modelos de aprendizaje automático supervisado fueron adquiridos del monitoreo microbiológico y fisicoquímico en producto terminado realizado en la industria de alimentos cárnicos y lácteos en una industria colombiana. Los datos recolectados correspondieron a la evaluación realizada entre 2011 a 2022. Para garantizar la confiabilidad de los datos entregados se tuvieron en cuenta los siguientes criterios en la revisión de los mismos:

- Industria con reconocimiento nacional como marca en derivados cárnicos y en derivados lácteos.
- Industria con certificación en HACCP bajo el decreto 60:2002
- Industria con laboratorio interno de control de calidad para la ejecución de las pruebas, con estándares de buenas prácticas de laboratorio bajo Resolución 1619:2015 y requisitos técnicos y de gestión de acuerdo con la ISO/IEC 17025:2017..
- Métodos analíticos normalizados para los análisis microbiológicos y fisicoquímicos.

La base de datos incluyó información sobre el tipo de producto analizado, resultados asociados con parámetros microbiológicos como: recuento de aerobios mesófilos, recuento de coliformes totales, recuento de *Escherichia coli* biotipo 1, recuento de esporas *Clostridium* sulfito reductor, recuento de estafilococo coagulasa positivo, recuento de mohos y levaduras,

detección de *Salmonella* spp y detección de *Listeria monocytogenes*. Y como variables fisicoquímicas: pH, porcentaje de grasa, porcentaje de humedad y porcentaje de proteína.

En la evaluación de los modelos de aprendizaje automático supervisado las variables respuesta fueron *Listeria monocytogenes* y *Salmonella* spp y las variables para predicción incluyeron parámetros microbiológicos conocidos como indicadores de calidad en los que se incluyen microorganismos como: aerobios mesófilos, coliformes totales, *Escherichia coli* biotipo 1 (no patógena) esporas *Clostridium* sulfito reductor (no patógeno), estafilococo coagulasa positivo (patógeno con dosis infectiva baja) y mohos y levaduras. Y predictores fisicoquímicos como pH, grasa, humedad y proteína.

El objetivo de la evaluación de los modelos de aprendizaje automático supervisado para las variables respuesta (*Salmonella* spp y *Listeria monocytogenes*) fue poder identificar si hay una correlación directa o significativa con alguna de las variables predictoras microbiológicas o fisicoquímicas y de este modo entregar un modelo de aprendizaje automático supervisado a la industria que complementen las acciones de análisis de peligros y determinación de riesgos en sus procesos que conlleven a una identificación de las variables que afectan la presencia de estos microorganismos patógenos de importancia en salud pública.

6.3. Realizar metanálisis de modelos de aprendizaje automático supervisado (Regresión logística, Random Forest, Red Neuronal, Árbol de decisión y Máquina de soporte vectorial radial) aplicados en la industria de alimentos cárnicos y lácteos.

Como análisis estadístico de una gran colección de resultados de estudios publicados, el metanálisis tiene como objetivo integrar e interpretar los hallazgos para lograr conclusiones integrales que los estudios individuales por sí solos no demostrarían claramente. Se realizó una búsqueda de literatura electrónica en las bases de datos Web of Science y PubMed para encontrar artículos e informes oficiales publicados desde 2018 hasta 2023 que resumieran la predicción de microorganismos patógenos, principalmente *Listeria monocytogenes* y *Salmonella* spp a través de modelos de aprendizaje automático supervisado en cualquier

etapa de la producción de alimentos a nivel global. La búsqueda se realizó de forma sistemática en inglés y tuvo como objetivo encontrar estudios de calidad validados por la comunidad científica.

Las revisiones bibliográficas se realizaron mediante una sintaxis que combinaba términos sobre la existencia (prevalencia, incidencia, ocurrencia, calidad, inocuidad, seguridad alimentaria) de patógenos (*Salmonella spp*, *Listeria monocytogenes*) en los alimentos y su predicción a través de inteligencia artificial (aprendizaje automático, minería de datos, modelos, algoritmos, predicción), mientras que se excluyeron otros estudios en el dominio clínico, sobre insectos y/o parásitos, donde los productos fueron inoculados artificialmente (inóculos, enriquecidos) y publicados antes del 2018. Todos los términos se combinaron aplicando correctamente los conectores lógicos AND, y AND NOT.

En la valoración inicial se recuperaron 331 publicaciones en la base de datos Web of Science y 194 en la base de datos PubMed, tras la revisión de título se identificó que los 194 artículos de PubMed se encontraban incluidos en las 331 recuperaciones realizadas en Web of Science, por lo que se fueron eliminados como duplicados. En la figura 1, se describe la revisión sistemática para la obtención final de los artículos que fueron utilizados en el metanálisis.

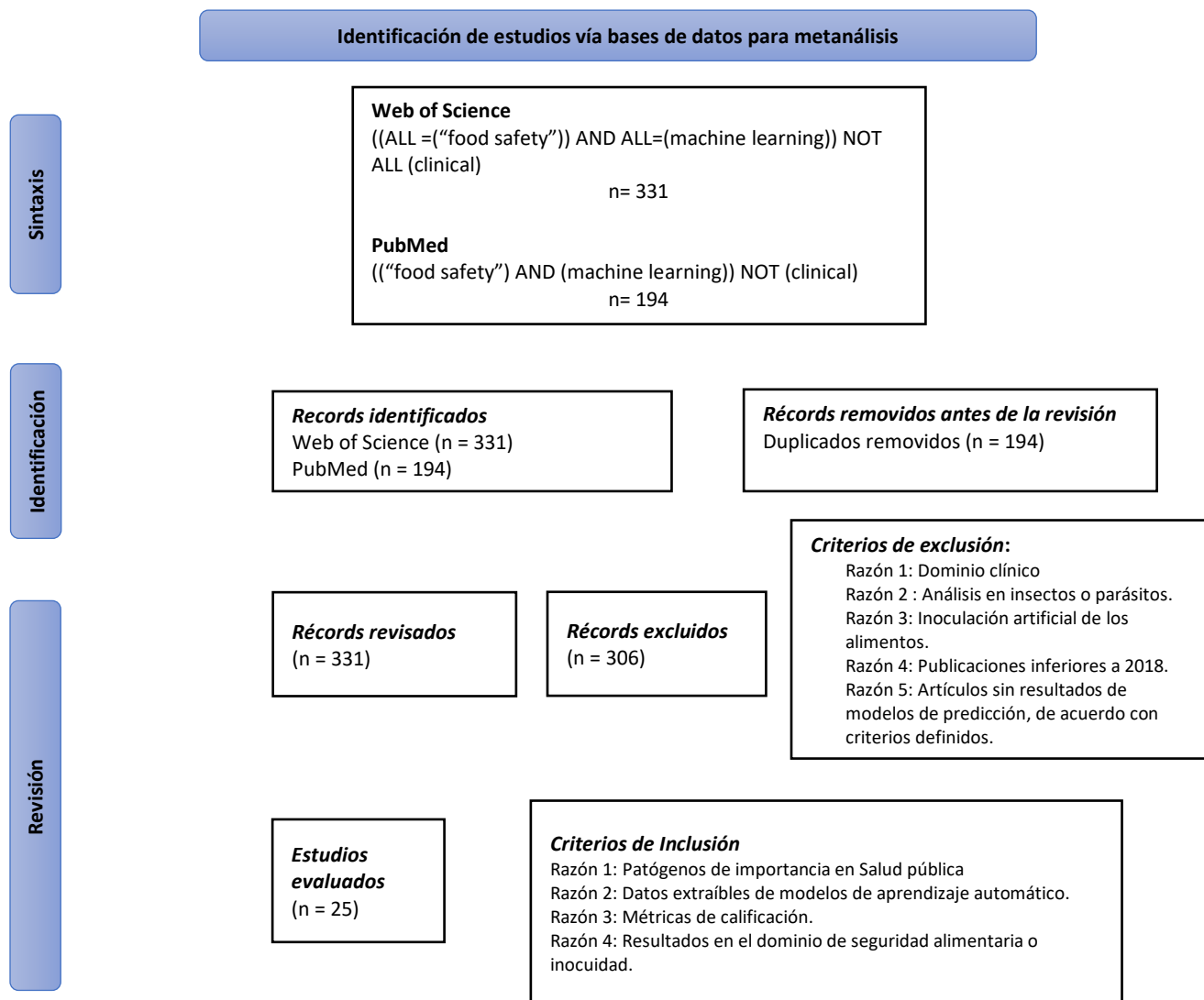


Figura 1: *Revisión sistemática para metanálisis.*

Tras valorar toda la información de las publicaciones recuperadas, veinticinco (25) estudios primarios publicados desde 2019 hasta 2023 se consideraron apropiados para su inclusión ya que cumplían los siguientes criterios: (1) presencia de microorganismos patógenos de importancia en salud pública (*es importante aclarar que los dos microorganismos patógenos de interés son Listeria monocytogenes y Salmonella spp, pero no se excluyeron microorganismos como Escherichia coli enteropatógena o Campylobacter spp dado que también son de interés en la industria de alimentos y en salud pública*); (2) presentar datos extraíbles; (3) resultados sobre las métricas de calificación de modelos de aprendizaje automático supervisado: exactitud, puntaje F1, área sobre la curva (ROC-AUC), sensibilidad,

especificidad y/o matriz de confusión y (4) resultados en cualquier etapa de producción de alimentos a nivel global.

El tratamiento inicial de los 25 artículos para la ejecución del metanálisis se realizó con la creación de una matriz en Microsoft Excel 2010 cuyas variables principales fueron: Referencia, algoritmos, métricas de calificación del modelo (exactitud, puntaje F1, sensibilidad, especificidad, área bajo la curva y coeficiente Kappa de Cohen, además del tipo de datos y aplicación en seguridad alimentaria. (Ver figura 2: representación esquemática de análisis de información para meta análisis).

El metanálisis fue ejecutado en R versión 4.2.1 (2022-06-23 ucrt) usando la librería *metafor* y las librerías complementarias de *tidyverse*, *readxl* y *janitor*.

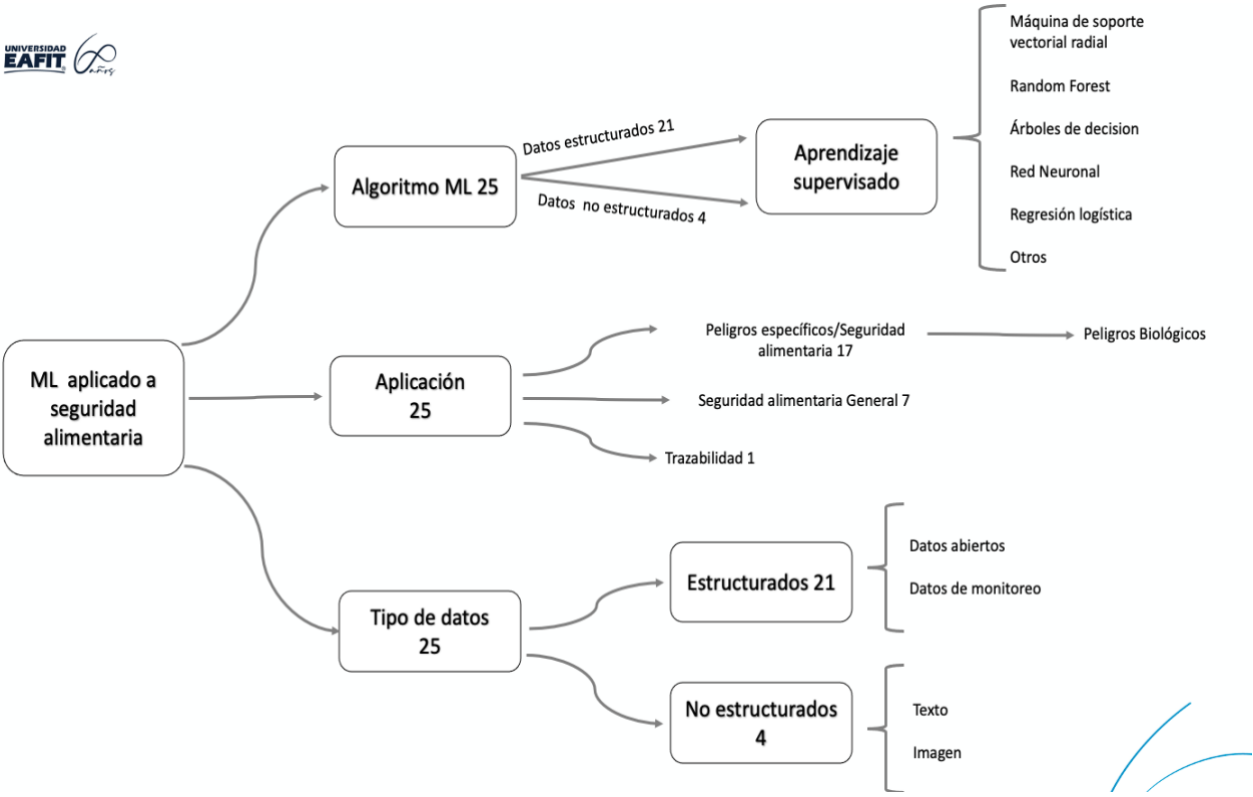


Figura 2: Representación esquemática de análisis de información para metanálisis.

6.4. Identificar las variables intrínsecas e extrínsecas que impactan el crecimiento microbiológico de Salmonella spp y Listeria monocytogenes en alimentos cárnicos y lácteos.

La identificación de las variables intrínsecas e extrínsecas que impactan el crecimiento microbiológico de *Salmonella* spp y *Listeria monocytogenes* se realizó a través de revisión bibliográfica. Esta información fue útil para indicar cuáles son las variables de proceso que deben ser consideradas para evitar su desarrollo y para la interpretación de la evaluación de los modelos con las variables predictoras microbiológicas y fisicoquímicas usadas a partir de verificación a través de métodos normalizados en producto terminado.

La información entregada por este objetivo complementó el objetivo general de predecir la presencia de *Salmonella* spp y *Listeria monocytogenes* y entregar datos significativos para contar con soporte técnico para el análisis de peligros y riesgos de los procesos a través del control de las variables intrínsecas e extrínsecas que favorecen el crecimiento de estos dos microorganismos patógenos.

6.5. Entrenar modelos de aprendizaje automático supervisado para predecir la presencia o ausencia de Salmonella spp y Listeria monocytogenes en alimentos cárnicos y lácteos.

Se realizó entrenamiento de los modelos de aprendizaje automático seleccionados en el objetivo general correspondientes a: Regresión logística, Random Forest, Red Neuronal, Árbol de decisión y Máquina de soporte vectorial radial.

Antes del entrenamiento se llevó a cabo exploración de los datos, ajuste de la información para evitar errores durante la fase de entrenamiento

6.5.1. Exploración datos

Usando R versión 4.2.1 y las librerías *tidyverse*, *janitor*, *readxl*, *skimr* y *GGally*) se realizó revisión, depuración y evaluación estadística descriptiva de los datos originales para su posterior modelación. En la revisión y depuración se aseguró que cada variable estuviera ingresada correctamente como categoría o como valor numérico, además de ajustar los valores para permitir los cálculos, es decir, valores como “<10” “<100” que corresponden al modelo de reporte en microbiología de unidades formadoras de colonia (ufc) por gramo o mililitro cuando no se obtiene ningún tipo de recuento sobre la base de las diluciones en base 10 y el límite de detección del método microbiológico usado se transformaron al número entero “0” y para valores con el signo mayor se incluyó el número 1 al final del resultado obtenido. Esto se realizó para todos los parámetros microbiológicos evaluados. Se ejecutó agrupamiento de los productos teniendo en cuenta las similitudes de producción. Y se definieron dos grandes categorías para el entrenamiento y prueba de los modelos: cárnico y lácteo.

6.5.2. Análisis exploratorio de datos

Se realizó análisis exploratorio de datos o EDA por sus siglas en inglés para examinar los datos sin ninguna idea preconcebida con el fin de descubrir qué contienen los datos. Se revisó el porcentaje de positividad para las variables respuesta: *Salmonella spp* y *Listeria monocytogenes*, la correlación con los parámetros fisicoquímicos y la distribución de positividad dependiendo de las dos categorías: cárnica y láctea.

El análisis exploratorio de datos (EDA) es un paso esencial en cualquier análisis de investigación. El objetivo principal del análisis exploratorio fue examinar los datos en busca de distribución, valores atípicos y anomalías para dirigir pruebas específicas de su hipótesis. También proporcionó herramientas para la generación de hipótesis al visualizar y comprender los datos, generalmente a través de representaciones gráficas. EDA tiene el propósito de ayudar al reconocimiento de patrones naturales del analista y es un primer paso fundamental después de la recopilación de datos y el reprocesamiento, donde los datos

simplemente se visualizan, grafican, manipulan, sin suposiciones, para ayudar a evaluar la calidad de los datos . Con el análisis exploratorio de datos se buscó:

- Maximizar la percepción de la base de los datos/Comprender la estructura de la base de datos.
- Visualizar las posibles relaciones entre la exposición y las variables de resultado.
- Detectar valores atípicos y anomalías
- Desarrollar modelos predictivos o explicativos que funcionan con la menor cantidad posible de variables de exposición o selección preliminar de modelos apropiados.
- Extraer y crear variables relevantes.

6.5.3. Entrenamiento y prueba de modelos de aprendizaje automático

Para el entrenamiento y prueba de los modelos se usó el programa R versión 4.2.1, plataforma Windows x86_64-w64-mingw32/x64 (64 bit). Windows 10x 64.

Todos los ajustes se hicieron partiendo de la metodología propuesta e implementada en *tidymodels*. La metodología consiste en diversos pasos (no todos tienen que ser implementados, solo es una sugerencia)

Entre los pasos de la metodología se encuentran:

1. Construir un modelo: implica establecer qué modelos se usaran, todo esto después de un análisis exploratorio previo.
2. Preprocesar los datos (construcción de la “receta”). Se realiza el preprocesamiento de los datos, estos se resumen como balanceo de clases, imputación de datos (de ser necesario), normalización de datos (de ser necesario).
3. Evaluación de los modelos con las diferentes muestras de los datos con la consecuente creación de muestras con validación cruzada para evaluar los modelos.
4. Evaluar y comparar las métricas de los modelos

5. Ajustar (Tunear) los hiperparámetros del modelo. Esta etapa es cuando por defecto se permite al algoritmo de programación elegir los mejores hiperparámetros para cada modelo, sin la elección previa por parte del evaluador.

La base de datos con 4033 datos y 12 variables fue particionada en una proporción 80/20 con 5 iteraciones para validación cruzada. En los modelos para el entrenamiento se discriminó por categoría y variable respuesta, para *Listeria monocytogenes* y la categoría láctea cada submuestra quedó compuesta por 1829 datos para entrenamiento y 457 datos para prueba, cada submuestra contenía en total 2286 datos. Para *Listeria monocytogenes* y la categoría cárnica cada submuestra quedó compuesta por 591 datos para entrenamiento y 147 datos para prueba, cada submuestra contenía 738 datos en total. Para *Salmonella* spp y la categoría cárnica cada submuestra quedó compuesta por 591 datos para entrenamiento y 147 datos para prueba, cada submuestra contenía 738 datos en total. No se realizó entrenamiento de los modelos con la categoría láctea para *Salmonella* spp porque en los datos utilizados no se contaba con resultados de presencia para esta variable respuesta.

Los hiperparámetros usados fueron:

Random forest: 500 árboles

Árboles de decisión y clasificación: profundidad: 5 ramas

Máquina de soporte vectorial: Sigma: 0,2.

Red Neuronal: penalidad: 0,1.

Se realizó estratificación de los datos para asegurar que el mismo porcentaje de positividad de la variable respuesta (*Salmonella* spp y *Listeria monocytogenes*) se mantuviera en todas las iteraciones, además de aplicar SMOTE (*Synthetic Minority Over-Sampling Technique*) como técnica estadística de sobremuestreo de minorías sintéticas para aumentar el número de casos de un conjunto de datos de forma equilibrada, puesto que a partir del análisis exploratorio se encontró que sólo hubo un 1,07% y 1.19% de muestras positivas para *Salmonella* spp y *Listeria monocytogenes* respectivamente.

Se ejecutó el entrenamiento y prueba de los 5 modelos de aprendizaje automático de forma independiente para cada variable respuesta con 10 predictores que combinaron información microbiológica y fisicoquímica. Al final se obtuvieron los resultados en tabla y gráficas comparativas para todos los criterios de calificación seleccionados para los modelos evaluados.

7. Resultados

7.1. Metanálisis

Se realizó la síntesis de los resultados de los diferentes estudios previos sobre el uso de algoritmos de aprendizaje automatizado sobre la predicción de microorganismos patógenos como riesgo biológico dentro de la seguridad alimentaria para las métricas: precisión, sensibilidad, especificidad, área bajo la curva, coeficiente Kappa de Cohen y puntaje F1.

Bajo una estimación de máxima verosimilitud restringida a partir del n de cada artículo y de la media de cada métrica para los diferentes algoritmos de aprendizaje automatizado, se estima que la precisión obtenida para predecir la presencia de un microorganismo patógeno de importancia en seguridad alimentaria es de 81,10% con un intervalo entre 77,07-85,13%. Para la sensibilidad de 74,74% con un intervalo entre 68,88-80,59%. Para la especificidad de 67,09% con un intervalo entre 58,87 – 75,31%. Estimo para el área bajo la curva un valor de 70,75% (intervalo 66,28- 75,21%) y para el puntaje F1 76,89% (intervalo 64,72%- 89,06%) (Ver figura 3 a figura 7).

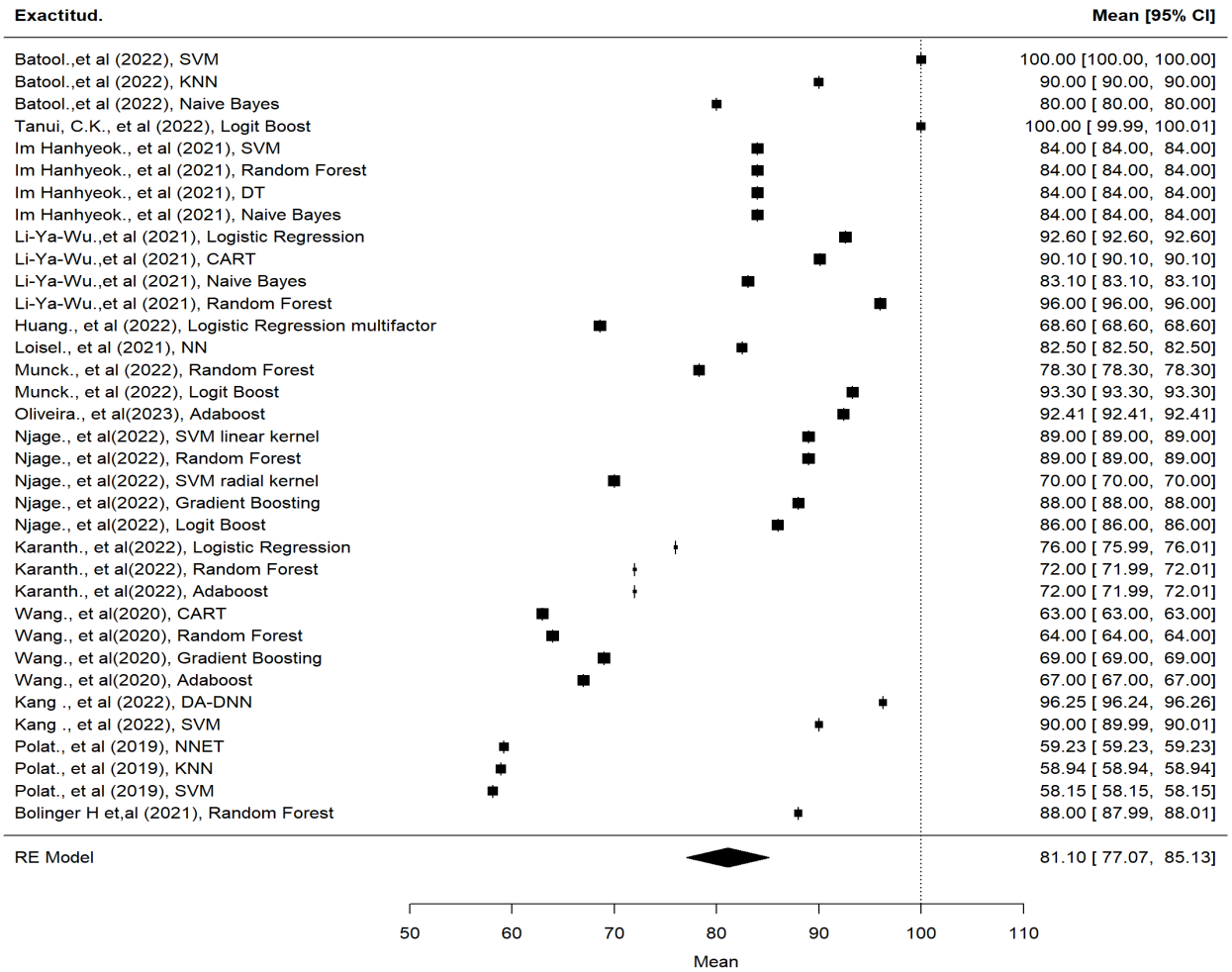


Figura 3: Comparación de estudios en términos de exactitud.

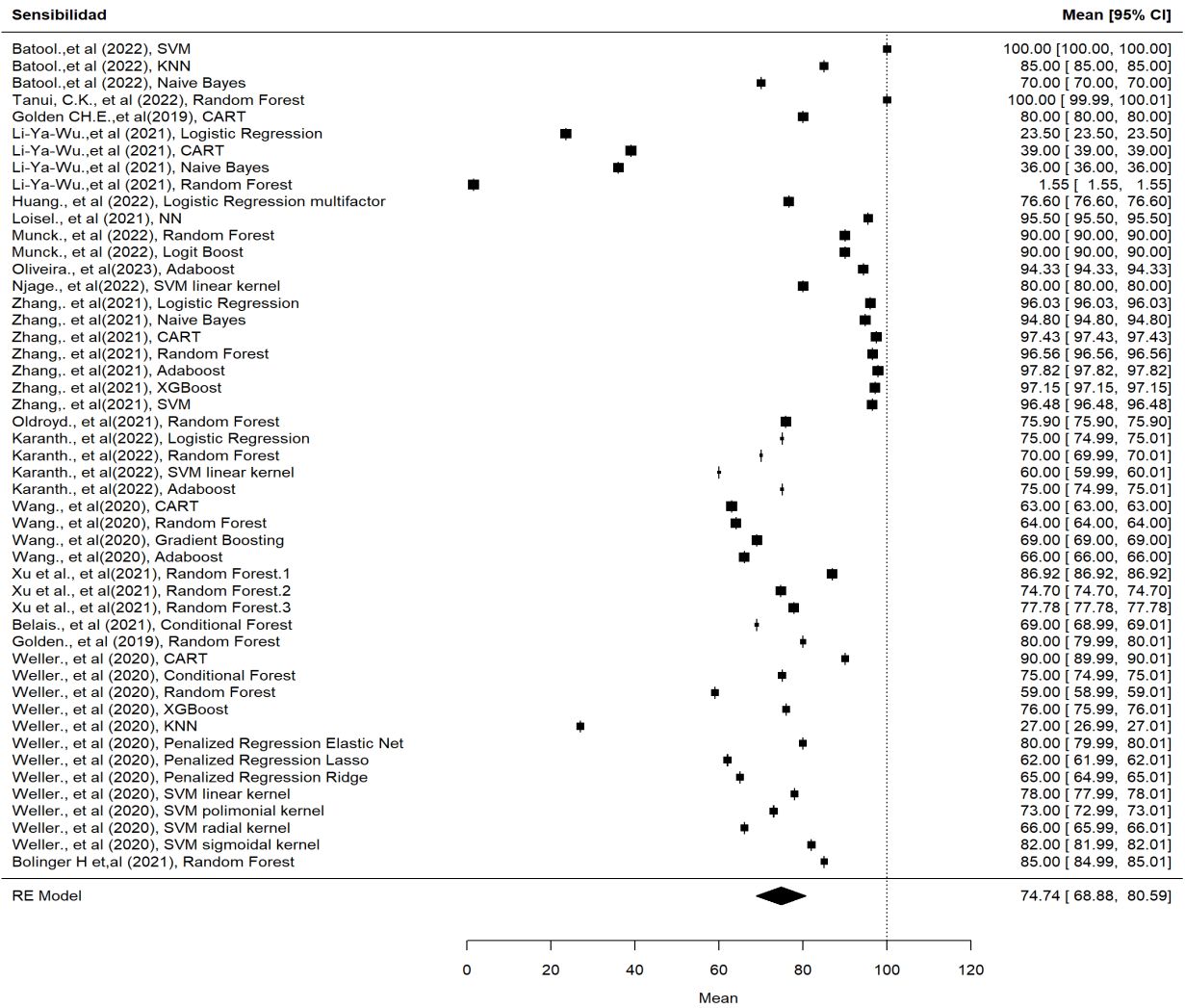


Figura 4: Comparación de estudios en términos de sensibilidad.

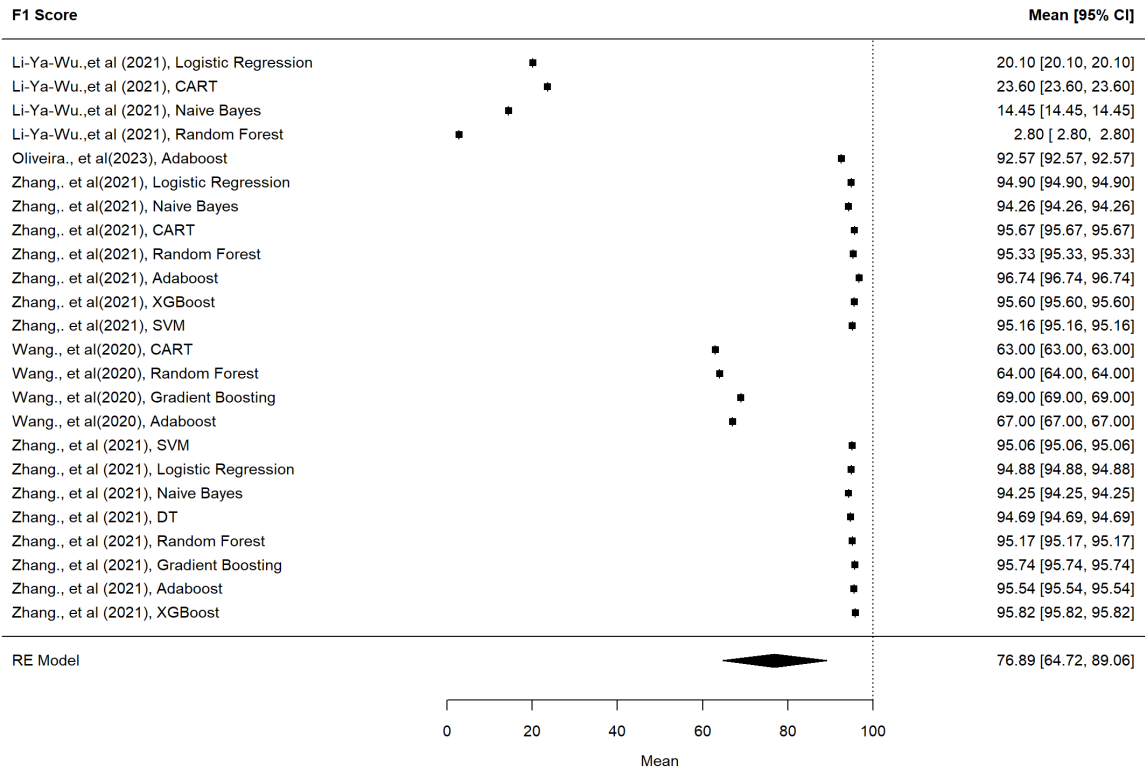


Figura 5: Comparación de estudios en términos de F1 Score.

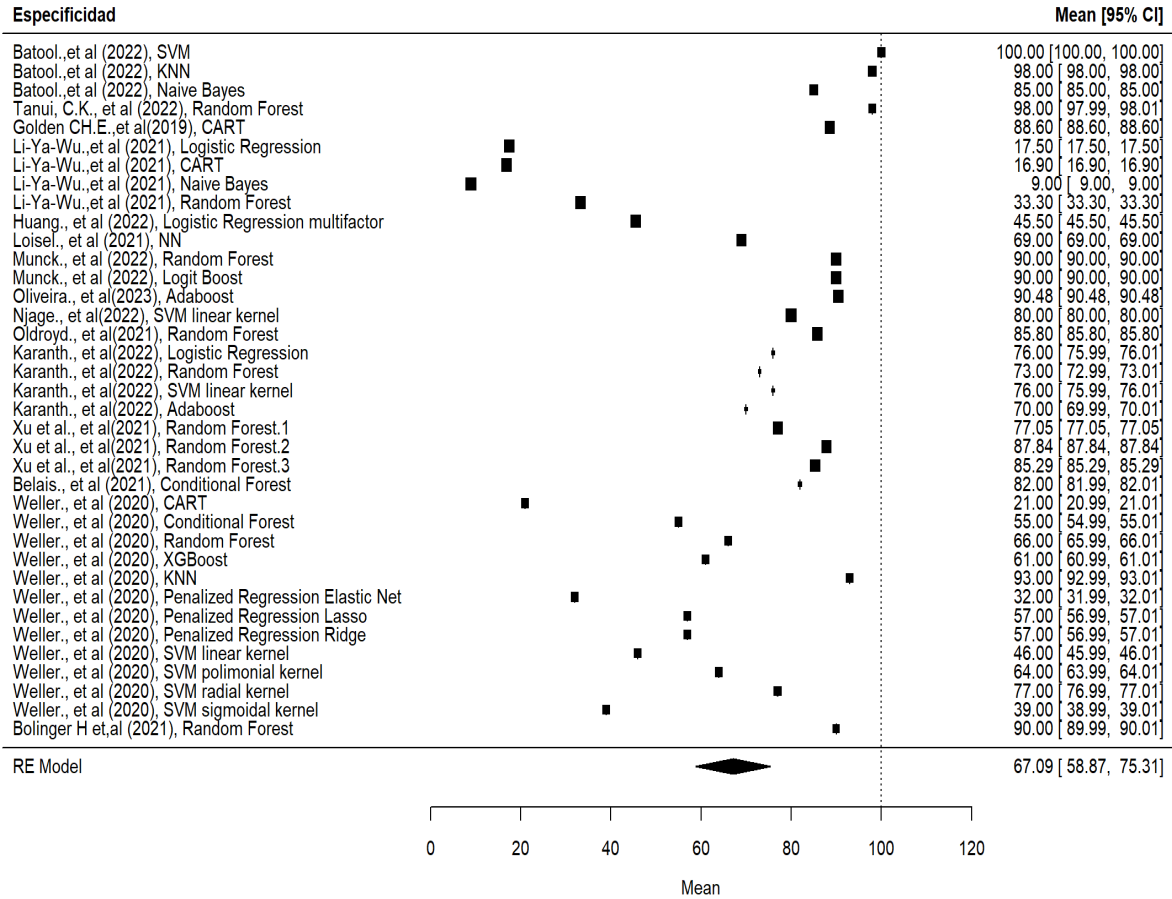


Figura 6: Comparación estudios en términos de especificidad.

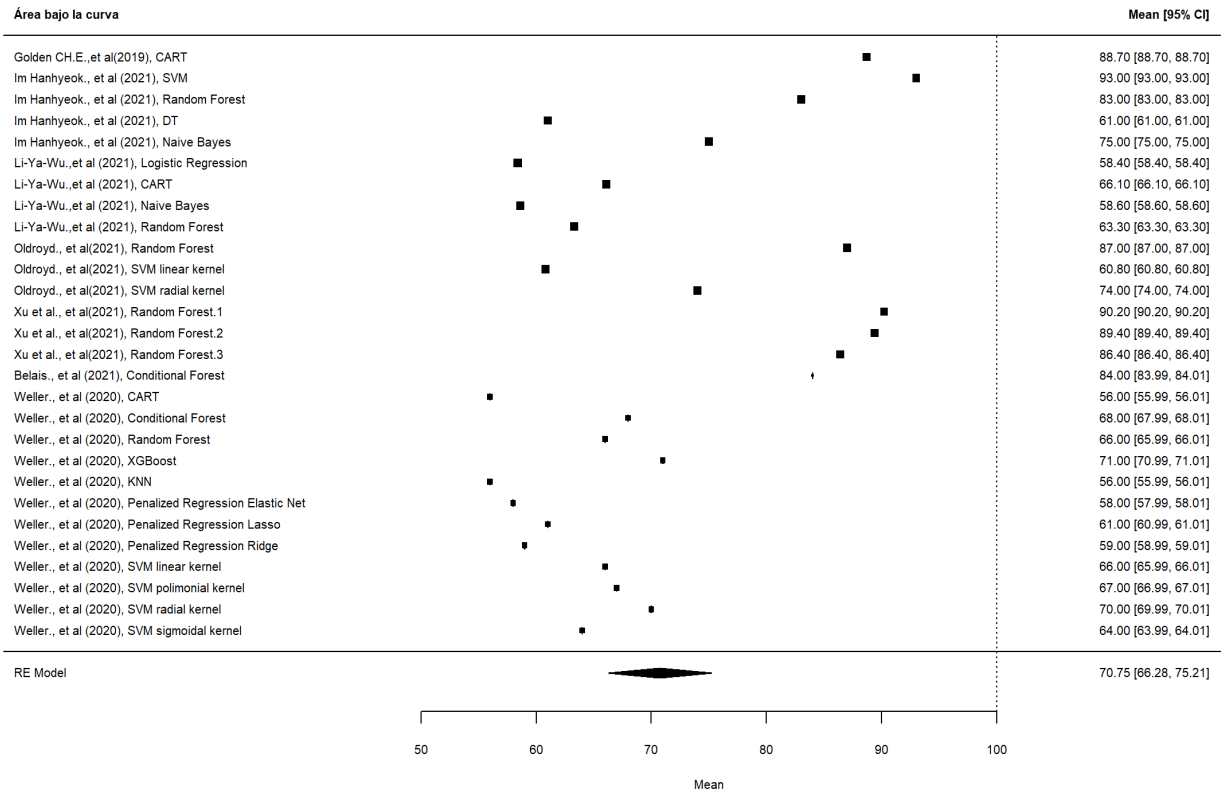


Figura 7: Comparación estudios en términos de área bajo la curva

7.2. Análisis descriptivo exploratorio

Se procesaron un total de 4033 datos, de los cuales 3049 fueron para productos lácteos y 984 para productos cárnicos. El 98,93% de los resultados no reportaron presencia para el microorganismo patógeno *Salmonella* spp y el 1,07% de positividad. El comportamiento para *Listeria monocytogenes* fue similar, 98,81% fueron resultados negativos y el 1,19% resultados positivos.

Se realizó evaluación de correlación entre la variable respuesta *Salmonella* spp y *Listeria monocytogenes* con los predictores fisicoquímicos (pH, grasa, humedad, proteína). No es posible determinar correlaciones significativas entre las características fisicoquímicas del alimento y la presencia de *Salmonella* spp o *Listeria monocytogenes* dado que el número de resultados positivos es significativamente bajo, no supera el 1% para ambas categorías de alimento: cárnico y lácteo.

- Correlación de *Salmonella* spp con predictores fisicoquímicos

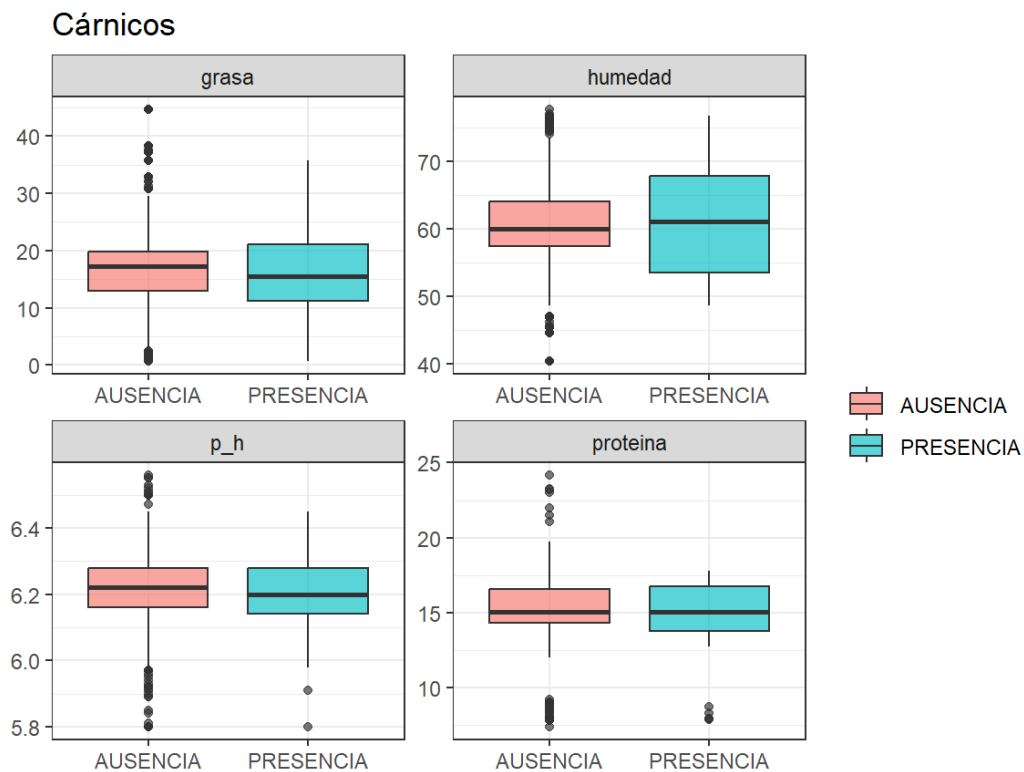


Figura 8: Correlación predictores fisicoquímicos con la variable respuesta *Salmonella* spp.

- **Correlación de *Listeria monocytogenes* con predictores fisicoquímicos**

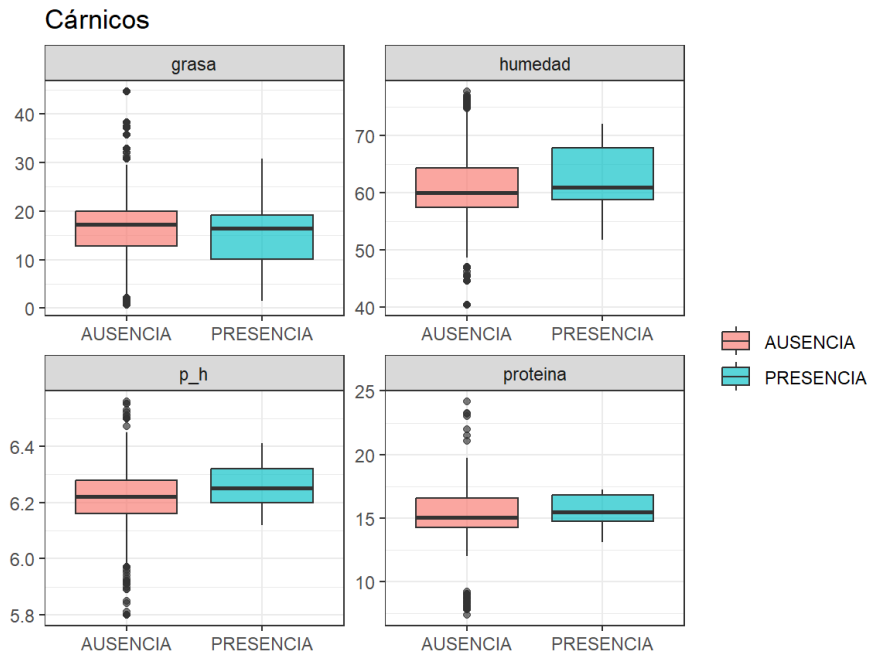


Figura 9: Correlación predictores fisicoquímicos con la variable respuesta *Listeria monocytogenes*

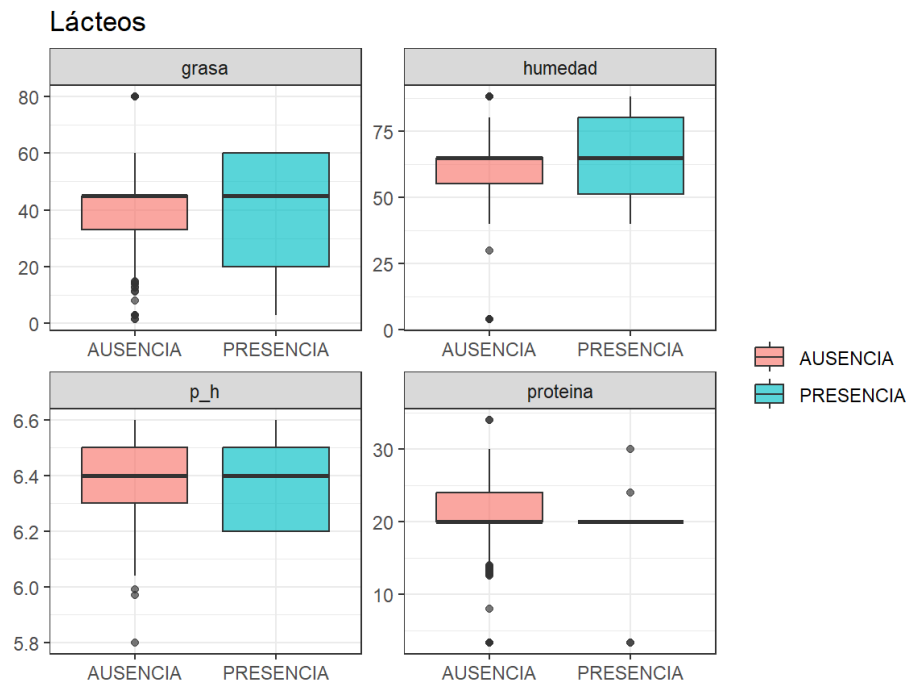


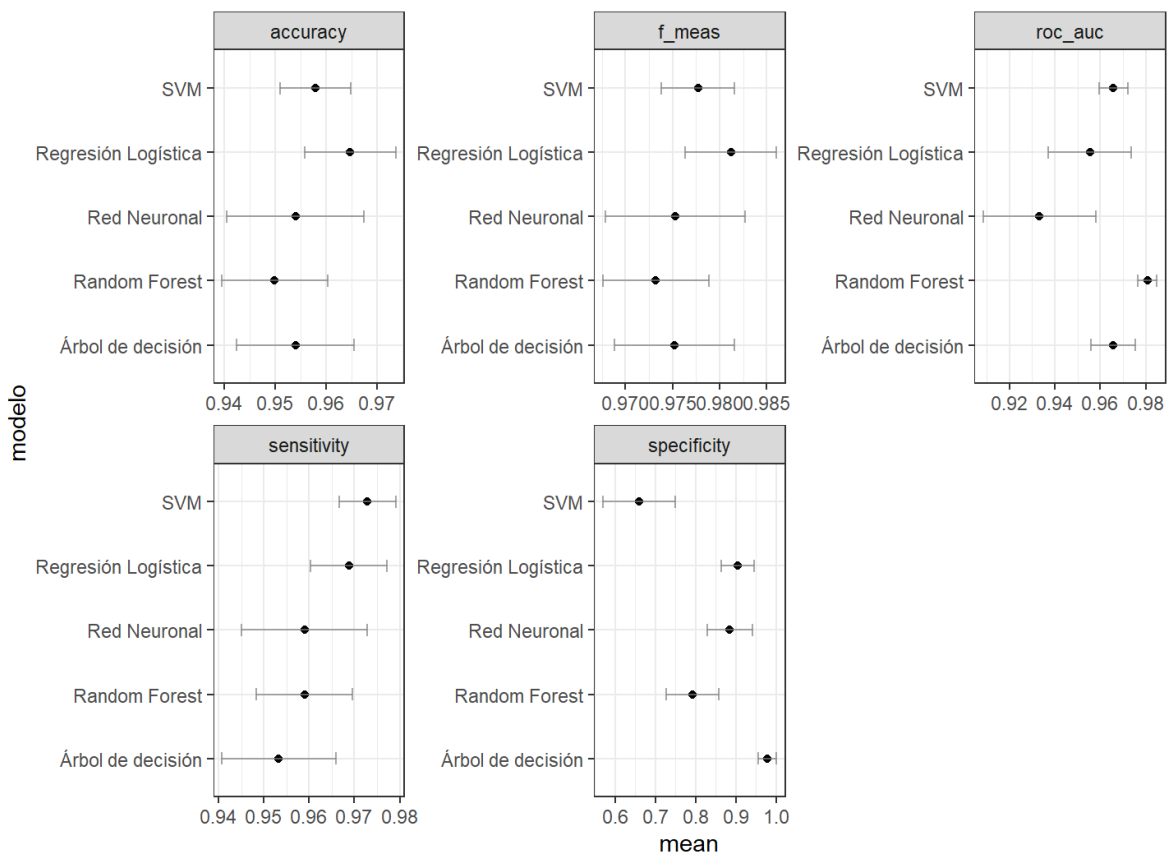
Figura 10: Correlación predictores fisicoquímicos con la variable respuesta *Listeria monocytogenes*

7.3. Modelación de algoritmos

Tras la aplicación de la metodología de *tidymodels* los resultados obtenidos de la modelación de los 5 algoritmos bajo las métricas de exactitud, sensibilidad, especificidad, puntaje F1, área bajo la curva y matriz de confusión para los datos de las categorías cárnica y láctea, el comportamiento fue el siguiente:

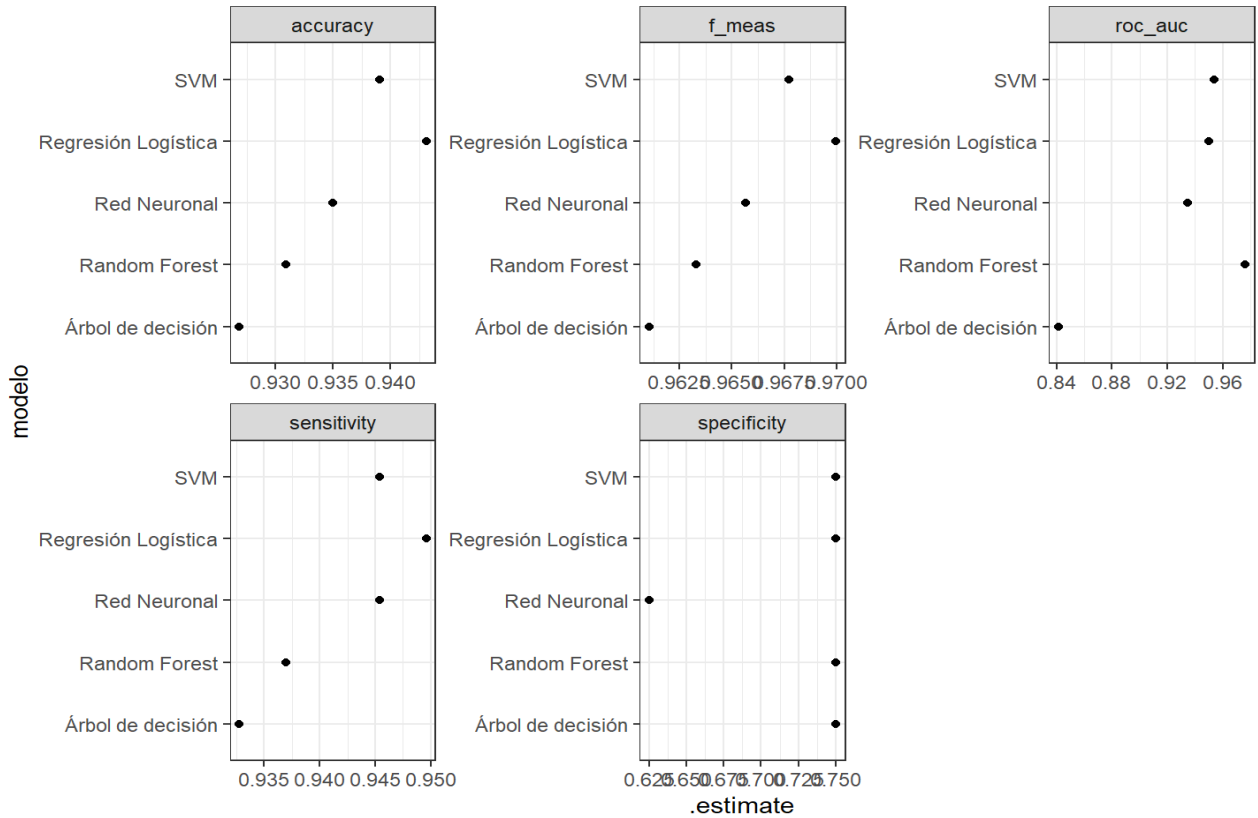
- **Comparación de resultados de entrenamiento para *Salmonella* spp categoría cárnica.**

En los resultados de entrenamiento en la categoría cárnica para el microorganismo patógeno *Salmonella* spp, el algoritmo con mejor especificidad es el árbol de decisión, seguido de la regresión logística, el cual también cuenta con más del 96% en exactitud con respecto a los demás algoritmos. El algoritmo para los criterios de calificación evaluados correspondió a la regresión logística.



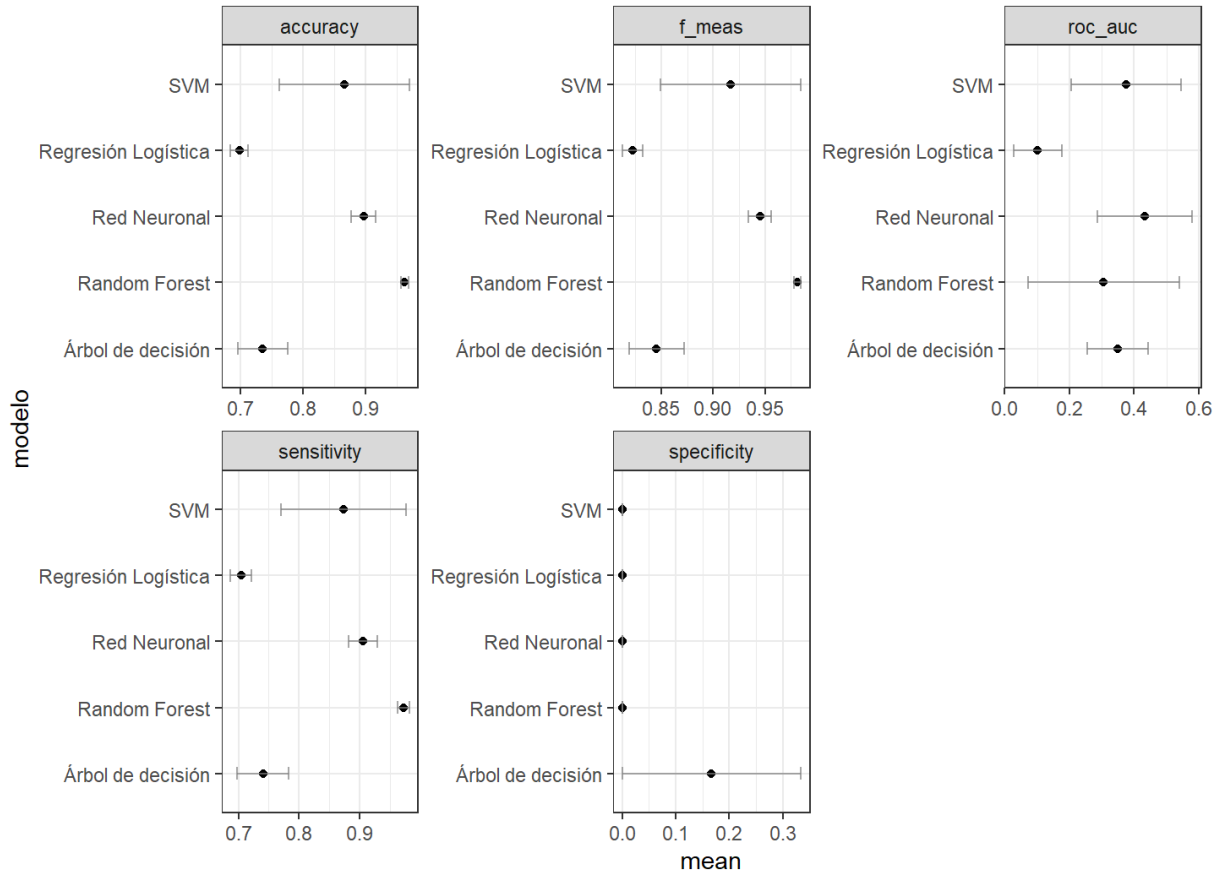
- **Comparación de resultados de prueba para *Salmonella* spp categoría cárnica**

Con base a los resultados de prueba el mejor algoritmo para determinar la presencia de *Salmonella* spp en productos cárnicos es la Regresión logística.



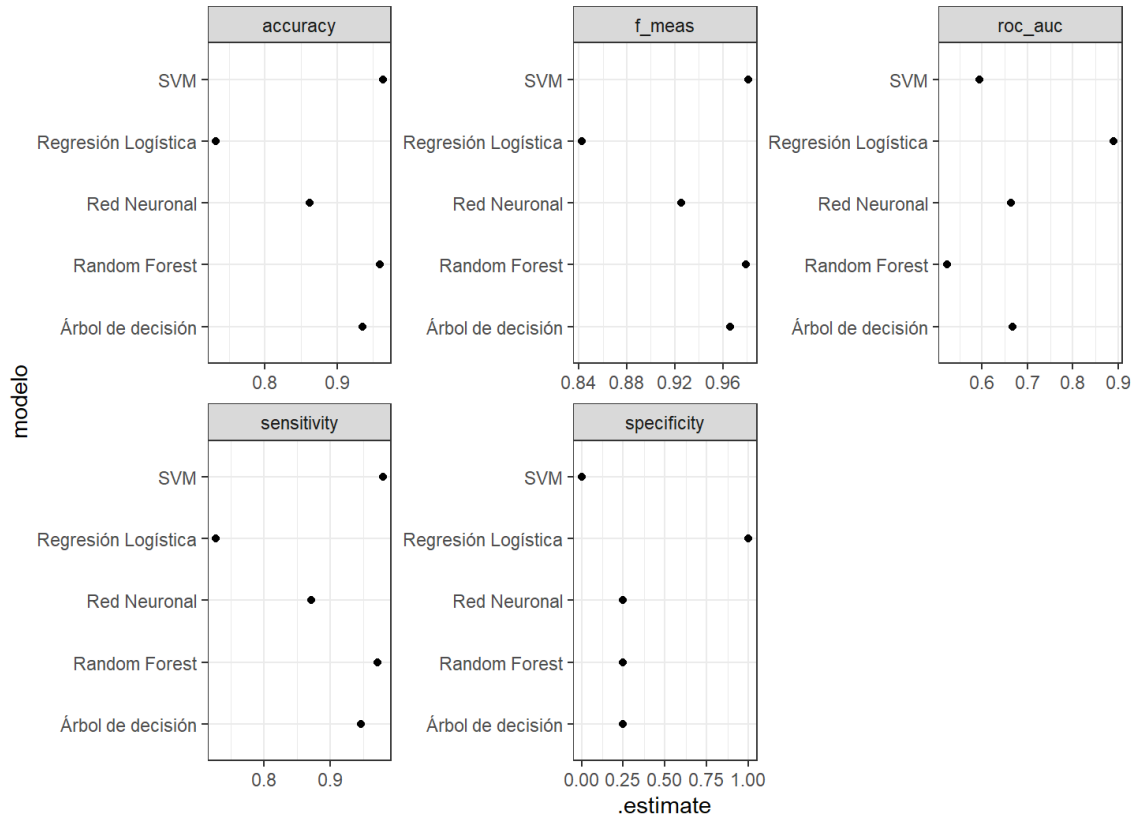
-Comparación de resultados de entrenamiento para *Listeria monocytogenes* categoría cárnica.

De acuerdo con los resultados obtenidos ningún algoritmo podría ser usado para la predicción de *Listeria monocytogenes* dada la alta variabilidad de todos en los criterios de calificación usados.



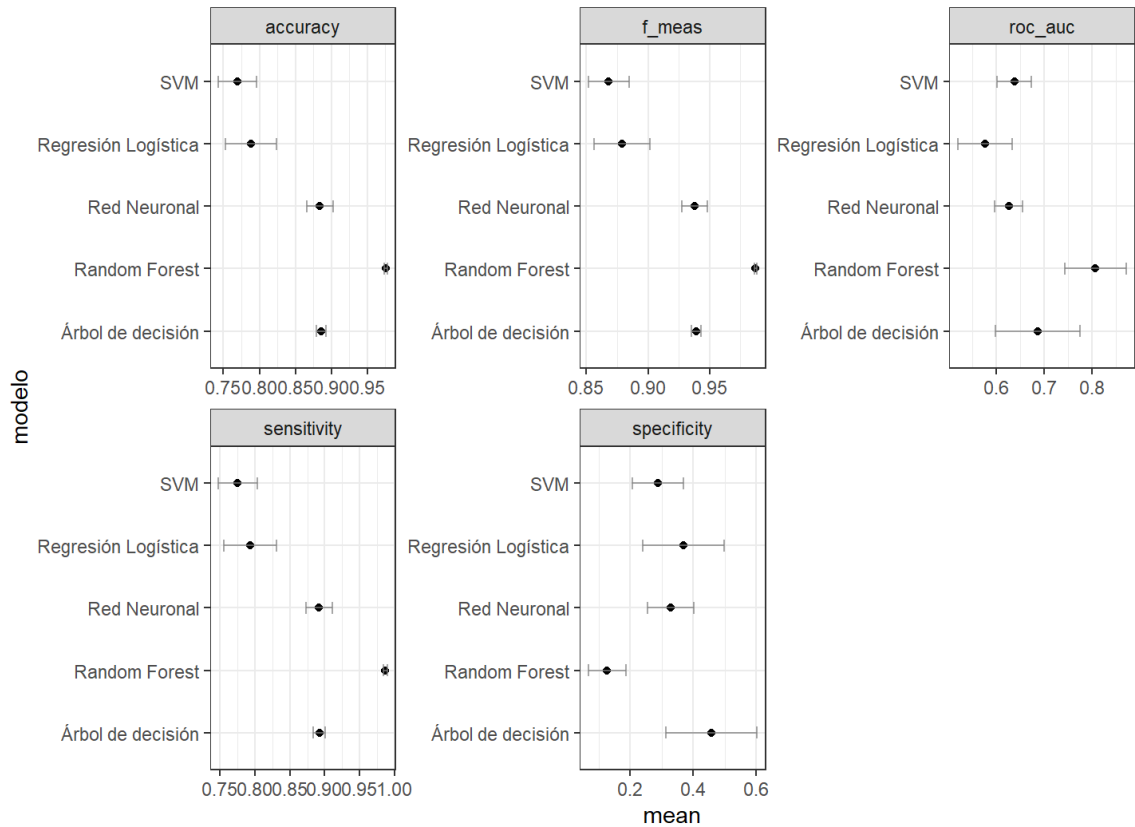
-Comparación de resultados de prueba para *Listeria monocytogenes* categoría cárnica

En los datos de prueba se confirman los resultados obtenidos en el entrenamiento de los algoritmos, la alta variabilidad de los algoritmos en los porcentajes bajos obtenidos en cada uno de los criterios de calificación.



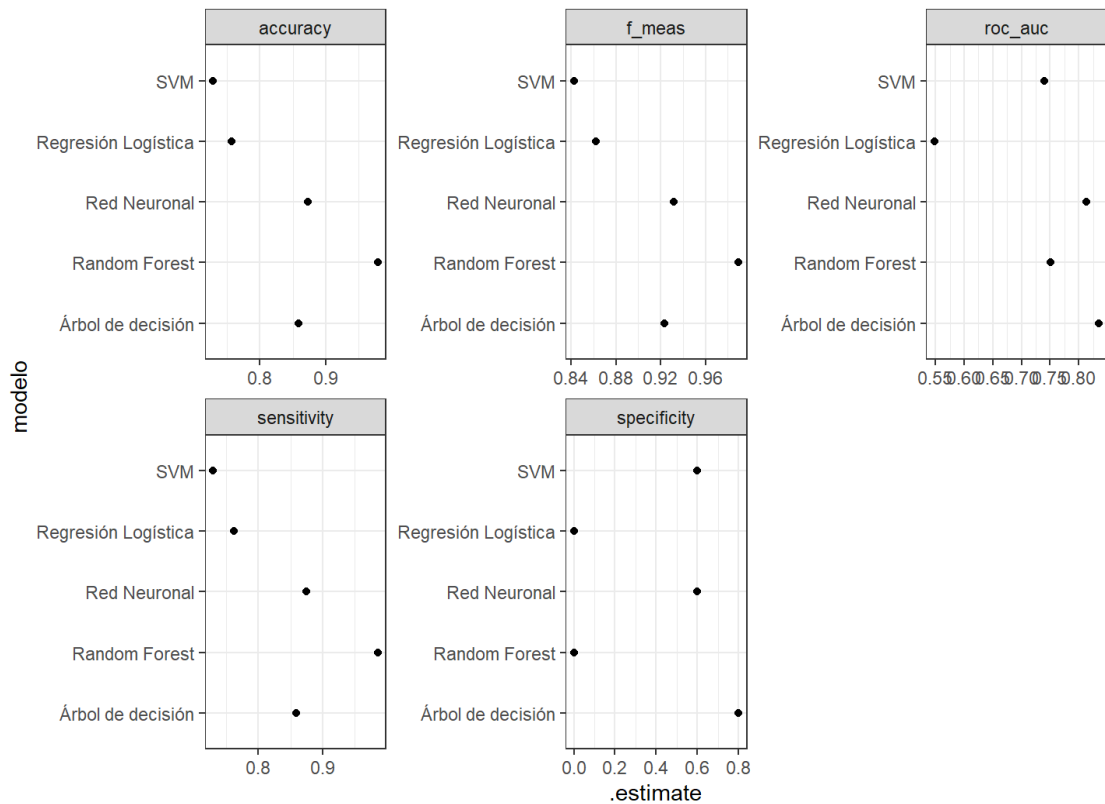
- **Comparación de resultados de entrenamiento para *Listeria monocytogenes* categoría láctea**

El algoritmo que reportó mejores resultados durante el entrenamiento fue el Random Forest excepto para el criterio de calificación de la especificidad. Los 4 algoritmos entrenados mostraron resultados muy variables.



- **Comparación resultados datos de prueba para *Listeria monocytogenes* categoría láctea.**

Se ratifican los resultados del entrenamiento en los resultados de prueba, en los que el algoritmo que logró los mejores porcentajes es el Random Forest.



Discusión

Los algoritmos de aprendizaje automatizado son una herramienta de alto impacto en la seguridad alimentaria y con mayor inclusión de datos se puede entregar el mejor modelo para predecir comportamientos microbiológicos de patógenos de importancia en salud pública sin las limitaciones de la microbiología convencional dependiente de cultivo. Dado que las limitaciones del enfoque de aprendizaje automático dependen del tamaño y la calidad del conjunto de datos, es necesario contar con un número mayor de datos para evaluar si la presencia del microorganismo es afectada o no por las características intrínsecas y extrínsecas del producto elaborado.

No se encontró una correlación entre los predictores usados microbiológicos y fisicoquímicos y las variables respuesta: presencia de *Salmonella* spp y *Listeria monocytogenes* a partir de los datos usados. No fue posible determinar para las dos categorías de alimento: cárnica y láctea y las dos variables respuesta el mejor algoritmo de predicción, hecho que es coherente con el porcentaje significativamente bajo de resultados positivos para estos dos microorganismos patógenos.

Se requiere de una evaluación posterior considerando resultados de proceso asociadas con monitoreo ambiental y variables de control de proceso, dado que existe una conciencia significativa de los fabricantes en garantizar la seguridad de los alimentos, por lo que no introducirían una herramienta de predicción a menos que sea “efectiva” porque se entiende que si hay una falla en la elaboración del alimento y no se detecta la presencia de un microorganismo patógeno se pueden ocasionar problemas fatales en la salud pública. El uso de un algoritmo de inteligencia artificial se usaría como complemento de los métodos actuales y sólo después de un tiempo prolongado de validación donde se obtengan resultados comparables y confiables podría ponerse en uso en los entornos productivos.

El algoritmo puede ser una herramienta de detección de variables para el control de los procesos siempre y cuando se garantice al menos lo siguiente:

- Datos: **acceso** a datos que alimenten el modelo y permitan su entrenamiento de manera permanente. A mayor cantidad de datos mayor precisión en las predicciones. **Calidad:** los datos deben ser obtenidos con métodos analíticos validados o verificados en las condiciones de operación de los laboratorios de las plantas de proceso en las que funcionan.
- Matriz de parámetros de proceso donde estén identificadas las variables que pueden favorecer la presencia de microorganismos patógenos. Esta debería incluir resultados de monitoreo ambiental, parámetros bromatológicos, condiciones ambientales de proceso, condiciones de operación: temperatura, tiempos, características de las materias primas.
- Claridad sobre las condiciones de crecimiento de los microorganismos y sus interacciones con el alimento.

- **Confiabilidad:** los resultados que el modelo de predicción entregue deben tener una precisión entre el 95% y 99% para que pueda ser complementario con los análisis microbiológicos realizados en proceso y en producto.
- **Simplicidad:** debe permitir ser usado con facilidad que requiera un conocimiento mínimo en programación. Mejor aún que el modelo se entregue de tal forma que los usuarios solo deben alimentarlo con los datos y se entrene de manera permanente para que en línea se está mostrando si el proceso favorece o no la aparición de un microorganismo patógeno.

Conclusiones

El desarrollo de herramientas que permitan una fácil implementación de la evaluación de la exposición a lo largo de la cadena alimentaria es importante para armonizar las prácticas y apoyar a los evaluadores de riesgos y a los tomadores de decisiones sobre las recomendaciones de la futura producción comercial de alimentos.

La capacidad predictiva de un modelo de ML depende en gran medida de la calidad y cantidad de los datos de entrenamiento. Por lo tanto, el primer paso debe ser la adquisición y el procesamiento adecuado de los datos sin procesar para obtener un conjunto de datos de entrenamiento adecuado.

Referencias

- [1] F. Borges *et al.*, "Contribution of omics to biopreservation: Toward food microbiome engineering," *Frontiers in Microbiology*, vol. 13. Frontiers Media S.A., Aug. 02, 2022. doi: 10.3389/fmicb.2022.951182.
- [2] S. N. Garcia, B. I. Osburn, and J. S. Cullor, "A one health perspective on dairy production and dairy food safety," *One Health*, vol. 7. Elsevier B.V., Jun. 01, 2019. doi: 10.1016/j.onehlt.2019.100086.
- [3] Committee on the Review of the Use of Scientific Criteria and Performance Standards for Safe Food. and Institute of Medicine (U.S.), *Scientific criteria to ensure safe food*. National Academies Press, 2003.
- [4] Z. Liu, A. N. Mutukumira, and H. Chen, "Food safety governance in China: From supervision to coregulation," *Food Sci Nutr*, vol. 7, no. 12, pp. 4127–4139, Dec. 2019, doi: 10.1002/fsn3.1281.

- [5] S. C. González *et al.*, “‘MicroHibro’: A software tool for predictive microbiology and microbial risk assessment in foods,” *Int J Food Microbiol*, vol. 290, pp. 226–236, Feb. 2019, doi: 10.1016/j.ijfoodmicro.2018.10.007.
- [6] K. L. Beck *et al.*, “Monitoring the microbiome for food safety and quality using deep shotgun sequencing,” *NPJ Sci Food*, vol. 5, no. 1, Dec. 2021, doi: 10.1038/s41538-020-00083-y.
- [7] X. Zhang, Y. Sun, and Y. Sun, “Application of Intelligent Taste Analysis Based on Random Forest Algorithm in Food Quality Inspection,” *Comput Intell Neurosci*, vol. 2022, 2022, doi: 10.1155/2022/6901184.
- [8] L. Pérez-Lavalle, E. Carrasco, and A. Valero, “Microbiological criteria: Inestablishment principles food quality for their and applications safety,” *Ital J Food Saf*, vol. 9, no. 1, 2020, doi: 10.4081/ijfs.2020.8543.
- [9] I. H. Sarker, “Data Science and Analytics: An Overview from Data-Driven Smart Computing, Decision-Making and Applications Perspective,” *SN Computer Science*, vol. 2, no. 5. Springer, Sep. 01, 2021. doi: 10.1007/s42979-021-00765-8.
- [10] M. E. Hibbing, C. Fuqua, M. R. Parsek, and S. B. Peterson, “Bacterial competition: Surviving and thriving in the microbial jungle,” *Nature Reviews Microbiology*, vol. 8, no. 1. pp. 15–25, Jan. 2010. doi: 10.1038/nrmicro2259.
- [11] P. M. Fratamico, A. K. Bhunia, and J. L. (James L.) Smith, *Foodborne pathogens : microbiology and molecular biology*. Caister Academic Press, 2005.
- [12] A. Gonçalves-Tenório, B. Nunes Silva, V. Rodrigues, V. Cadavez, and U. Gonzales-Barron, “Prevalence of pathogens in poultry meat: A meta-analysis of European published surveys,” *Foods*, vol. 7, no. 5, May 2018, doi: 10.3390/foods7050069.
- [13] L. I. Ma, J. L. Kornacki, G. Zhang, C.-M. Lin, and M. P. Doyle, “Development of Thermal Surrogate Microorganisms in Ground Beef for In-Plant Critical Control Point Validation Studies,” 2007. [Online]. Available: http://meridian.allenpress.com/jfp/article-pdf/70/4/952/1678375/0362-028x-70_4_952.pdf
- [14] N. M. Vega, K. R. Allison, A. N. Samuels, M. S. Klempner, and J. J. Collins, “Salmonella typhimurium intercepts Escherichia coli signaling to enhance antibiotic tolerance,” *Proc Natl Acad Sci U S A*, vol. 110, no. 35, pp. 14420–14425, Aug. 2013, doi: 10.1073/pnas.1308085110.
- [15] K. Wessels, D. Rip, and P. Gouws, “Salmonella in chicken meat: Consumption, outbreaks, characteristics, current control methods and the potential of bacteriophage use,” *Foods*, vol. 10, no. 8. MDPI AG, Aug. 01, 2021. doi: 10.3390/foods10081742.
- [16] S. Koseki, “Predictive Modeling for Estimation of Bacterial Behavior from Farm to Table,” *Food Safety*, vol. 4, no. 2, pp. 33–44, 2016, doi: 10.14252/foodsafetyfscj.2016006.
- [17] E. Stavropoulou and E. Bezirtzoglou, “Predictive modeling of microbial behavior in food,” *Foods*, vol. 8, no. 12. MDPI Multidisciplinary Digital Publishing Institute, Dec. 06, 2019. doi: 10.3390/foods8120654.

- [18] T. A. Mcmeekin, L. A. Mellefont, and T. Ross, "Predictive microbiology: Past, present and future," in *Modelling Microorganisms in Food*, Elsevier Ltd, 2007, pp. 7–21. doi: 10.1533/9781845692940.1.7.
- [19] C. Sabater, J. F. Cobo-Díaz, A. Álvarez-Ordóñez, P. Ruas-Madiedo, L. Ruiz, and A. Margolles, "Novel methods of microbiome analysis in the food industry," *International Microbiology*, vol. 24, no. 4. Springer Science and Business Media Deutschland GmbH, pp. 593–605, Nov. 01, 2021. doi: 10.1007/s10123-021-00215-8.
- [20] K. Melnyk, S. Klus, G. Montavon, and T. O. F. Conrad, "GraphKKE: graph Kernel Koopman embedding for human microbiome analysis," *Appl Netw Sci*, vol. 5, no. 1, Dec. 2020, doi: 10.1007/s41109-020-00339-2.
- [21] G. Cammarota *et al.*, "Gut microbiome, big data and machine learning to promote precision medicine for cancer," *Nat Rev Gastroenterol Hepatol*, vol. 17, no. 10, pp. 635–648, Oct. 2020, doi: 10.1038/s41575-020-0327-3.
- [22] T. W. Whon, N.-R. Shin, J. Y. Kim, and S. W. Roh, "Omics in gut microbiome analysis," *JOURNAL OF MICROBIOLOGY*, vol. 59, no. 3, pp. 292–297, Mar. 2021, doi: 10.1007/s12275-021-1004-0.
- [23] E. Sakowski *et al.*, "Current State of and Future Opportunities for Prediction in Microbiome Research: Report from the Mid-Atlantic Microbiome Meet-up in Baltimore on 9 January 2019," *mSystems*, vol. 4, no. 5, Sep. 2019, doi: 10.1128/mSystems.00392-19.
- [24] X. Zhang, Z. Chen, J. Ma, and P. Pandey, "Predicting Escherichia coli levels in manure using machine learning in weeping wall and mechanical liquid solid separation systems."
- [25] Z. Kang, Y. Zhao, L. Chen, Y. Guo, Q. Mu, and S. Wang, "Advances in Machine Learning and Hyperspectral Imaging in the Food Supply Chain," *Food Engineering Reviews*. Springer, Dec. 01, 2022. doi: 10.1007/s12393-022-09322-2.
- [26] H. Im, S.-H. Hwang, B. S. Kim, and S. H. Choi, "Pathogenic potential assessment of the Shiga toxin-producing Escherichia coli by a source attribution-considered machine learning model," 2021, doi: 10.1073/pnas.2018877118/-/DCSupplemental.
- [27] S. C. González *et al.*, "'MicroHibro': A software tool for predictive microbiology and microbial risk assessment in foods," *Int J Food Microbiol*, vol. 290, pp. 226–236, Feb. 2019, doi: 10.1016/j.ijfoodmicro.2018.10.007.
- [28] S. Koseki, K. Koyama, and H. Abe, "Recent advances in predictive microbiology: theory and application of conversion from population dynamics to individual cell heterogeneity during inactivation process," *Current Opinion in Food Science*, vol. 39. Elsevier Ltd, pp. 60–67, Jun. 01, 2021. doi: 10.1016/j.cofs.2020.12.019.
- [29] E. Chaix, O. Couvert, C. Guillaume, N. Gontard, and V. Guillard, "Predictive Microbiology Coupled with Gas (O₂/CO₂) Transfer in Food/Packaging Systems: How to Develop an Efficient Decision Support Tool for Food Packaging Dimensioning," *Compr Rev Food Sci Food Saf*, vol. 14, no. 1, pp. 1–21, Jan. 2015, doi: 10.1111/1541-4337.12117.
- [30] K. P. Koutsoumanis, A. Lianou, and M. Gougouli, "Last developments in foodborne pathogens modeling," *Current Opinion in Food Science*, vol. 8. Elsevier Ltd, pp. 89–98, Apr. 01, 2016. doi: 10.1016/j.cofs.2016.04.006.

- [31] A. Possas, A. Valero, and F. Pérez-Rodríguez, "New software solutions for microbiological food safety assessment and management," *Current Opinion in Food Science*, vol. 44. Elsevier Ltd, Apr. 01, 2022. doi: 10.1016/j.cofs.2022.100814.
- [32] A. Possas, A. Valero, and F. Pérez-Rodríguez, "New software solutions for microbiological food safety assessment and management," *Current Opinion in Food Science*, vol. 44. Elsevier Ltd, Apr. 01, 2022. doi: 10.1016/j.cofs.2022.100814.
- [33] K. Koyama, K. Kubo, S. Hiura, and S. Koseki, "Is skipping the definition of primary and secondary models possible? Prediction of *Escherichia coli* O157 growth by machine learning," *J Microbiol Methods*, vol. 192, Jan. 2022, doi: 10.1016/j.mimet.2021.106366.
- [34] E. Stavropoulou and E. Bezirtzoglou, "Predictive modeling of microbial behavior in food," *Foods*, vol. 8, no. 12. MDPI Multidisciplinary Digital Publishing Institute, Dec. 06, 2019. doi: 10.3390/foods8120654.
- [35] J. B. Xavier *et al.*, "iScience Mathematical models to study the biology of pathogens and the infectious diseases they cause," *iScience*, vol. 25, p. 104079, 2022, doi: 10.1016/j.isci.
- [36] F. Pérez-Rodríguez, E. Carrasco, A. K. Pradhan, A. S. Sant'Ana, V. P. Valdramidis, and A. Valero, "Special issue on 10th international conference of predictive modelling in foods: Towards a new paradigm in predictive microbiology," *International Journal of Food Microbiology*, vol. 291. Elsevier B.V., pp. 65–66, Feb. 16, 2019. doi: 10.1016/j.ijfoodmicro.2018.11.010.
- [37] Y. Jiang, J. Luo, D. Huang, Y. Liu, and D. D. Li, "Machine Learning Advances in Microbiology: A Review of Methods and Applications," *Frontiers in Microbiology*, vol. 13. Frontiers Media S.A., May 26, 2022. doi: 10.3389/fmicb.2022.925454.
- [38] C. E. Golden, M. J. Rothrock, and A. Mishra, "Using Farm Practice Variables as Predictors of *Listeria* spp. Prevalence in Pastured Poultry Farms," *Front Sustain Food Syst*, vol. 3, Mar. 2019, doi: 10.3389/fsufs.2019.00015.
- [39] S. J. Goodswen, J. L. N. Barratt, P. J. Kennedy, A. Kaufer, L. Calarco, and J. T. Ellis, "Machine learning and applications in microbiology," *FEMS Microbiology Reviews*, vol. 45, no. 5. Oxford University Press, Sep. 01, 2021. doi: 10.1093/femsre/fuab015.
- [40] C. K. Tanui, S. Karanth, P. M. K. Njage, J. Meng, and A. K. Pradhan, "Machine learning-based predictive modeling to identify genotypic traits associated with *Salmonella enterica* disease endpoints in isolates from ground chicken," *LWT*, vol. 154, Jan. 2022, doi: 10.1016/j.lwt.2021.112701.
- [41] A. Batool *et al.*, "An IoT and Machine Learning-Based Model to Monitor Perishable Food towards Improving Food Safety and Quality," *J Food Qual*, vol. 2022, 2022, doi: 10.1155/2022/6302331.
- [42] E. M. Mateo, J. V. Gómez, A. Tarazona, M. Á. García-Esparza, and F. Mateo, "Comparative analysis of machine learning methods to predict growth of *Sporotrichioides* and production of t-2 and ht-2 toxins in treatments with ethylene-vinyl alcohol films containing pure components of essential oils," *Toxins (Basel)*, vol. 13, no. 8, Aug. 2021, doi: 10.3390/toxins13080545.

- [43] C. K. Tanui, E. O. Benefo, S. Karanth, and A. K. Pradhan, "A Machine Learning Model for Food Source Attribution of *Listeria monocytogenes*," *Pathogens*, vol. 11, no. 6, Jun. 2022, doi: 10.3390/pathogens11060691.
- [44] S. Hiura, S. Koseki, and K. Koyama, "Prediction of population behavior of *Listeria monocytogenes* in food using machine learning and a microbial growth and survival database," *Sci Rep*, vol. 11, no. 1, Dec. 2021, doi: 10.1038/s41598-021-90164-z.
- [45] L. Huang, Z. Jia, and C. A. Hwang, "Growth and No-Growth boundary of *Listeria monocytogenes* in beef – A logistic modeling," *Food Research International*, vol. 152, Feb. 2022, doi: 10.1016/j.foodres.2021.110919.
- [46] L. M. van den Bulk, Y. Bouzembrak, A. Gavai, N. Liu, L. J. van den Heuvel, and H. J. P. Marvin, "Automatic classification of literature in systematic reviews on food safety using machine learning," *Curr Res Food Sci*, vol. 5, pp. 84–95, Jan. 2022, doi: 10.1016/j.crfs.2021.12.010.
- [47] J. Loisel *et al.*, "Cold chain break detection and analysis: Can machine learning help?," *Trends in Food Science and Technology*, vol. 112. Elsevier Ltd, pp. 391–399, Jun. 01, 2021. doi: 10.1016/j.tifs.2021.03.052.
- [48] N. Munck, P. M. K. Njage, P. Leekitcharoenphon, E. Litrup, and T. Hald, "Application of Whole-Genome Sequences and Machine Learning in Source Attribution of *Salmonella* Typhimurium," *Risk Analysis*, vol. 40, no. 9, pp. 1693–1705, Sep. 2020, doi: 10.1111/risa.13510.
- [49] A. N. de Oliveira *et al.*, "Tomato classification using mass spectrometry-machine learning technique: A food safety-enhancing platform," *Food Chem*, vol. 398, Jan. 2022, doi: 10.1016/j.foodchem.2022.133870.
- [50] P. M. K. Njage *et al.*, "Quantitative microbial risk assessment based on whole genome sequencing data: Case of *Listeria monocytogenes*," *Microorganisms*, vol. 8, no. 11, pp. 1–24, Nov. 2020, doi: 10.3390/microorganisms8111772.
- [51] S. Karanth, C. K. Tanui, J. Meng, and A. K. Pradhan, "Exploring the predictive capability of advanced machine learning in identifying severe disease phenotype in *Salmonella enterica*," *Food Research International*, vol. 151, Jan. 2022, doi: 10.1016/j.foodres.2021.110817.
- [52] Z. Lin *et al.*, "Machine learning approach for predicting single cell lag time of *Salmonella* Enteritidis after heat and chlorine treatment," *Food Research International*, vol. 156, Jun. 2022, doi: 10.1016/j.foodres.2022.111132.
- [53] S. Buyrukoğlu, "New hybrid data mining model for prediction of *Salmonella* presence in agricultural waters based on ensemble feature selection and machine learning algorithms," *J Food Saf*, vol. 41, no. 4, Aug. 2021, doi: 10.1111/jfs.12903.
- [54] J. G. Greener, S. M. Kandathil, L. Moffat, and D. T. Jones, "A guide to machine learning for biologists," *Nature Reviews Molecular Cell Biology*, vol. 23, no. 1. Nature Research, pp. 40–55, Jan. 01, 2022. doi: 10.1038/s41580-021-00407-0.
- [55] A. Nogales, R. Díaz-Morón, and Á. J. García-Tejedor, "A comparison of neural and non-neural machine learning models for food safety risk prediction with European Union RASFF data," *Food Control*, vol. 134, Apr. 2022, doi: 10.1016/j.foodcont.2021.108697.

- [56] C. Ensoy, T. Rakhmawati, C. Faes, and M. Aerts, "Separation Issues and Possible Solutions: Part I – Systematic Literature Review on Logistic Models - Part II – Comparison of different methods for separation in logistic regression," *EFSA Supporting Publications*, vol. 12, no. 9, Mar. 2017, doi: 10.2903/sp.efsa.2015.en-869.
- [57] R. A. Oldroyd, M. A. Morris, and M. Birkin, "Predicting food safety compliance for informed food outlet inspections: A machine learning approach," *Int J Environ Res Public Health*, vol. 18, no. 23, Dec. 2021, doi: 10.3390/ijerph182312635.
- [58] R. A. Oldroyd, M. A. Morris, and M. Birkin, "Food safety vulnerability: Neighbourhood determinants of non-compliant establishments in England and Wales," *Health Place*, vol. 63, May 2020, doi: 10.1016/j.healthplace.2020.102325.
- [59] H. H. Ku, P. Lin, and M. P. Ling, "Assessment of potential human health risks in aquatic products based on the heavy metal hazard decision tree," *BMC Bioinformatics*, vol. 22, Nov. 2021, doi: 10.1186/s12859-022-04603-3.
- [60] X. Xu, M. J. Rothrock, A. Mohan, G. D. Kumar, and A. Mishra, "Using farm management practices to predict *Campylobacter* prevalence in pastured poultry farms," *Poult Sci*, vol. 100, no. 6, Jun. 2021, doi: 10.1016/j.psj.2021.101122.
- [61] C. E. Golden, M. J. Rothrock, and A. Mishra, "Using Farm Practice Variables as Predictors of *Listeria* spp. Prevalence in Pastured Poultry Farms," *Front Sustain Food Syst*, vol. 3, Mar. 2019, doi: 10.3389/fsufs.2019.00015.
- [62] "Qualitative Risk Assessment: Risk of Activity/Food Combinations for Activities (Outside the Farm Definition) Conducted in a Facility Co-Located on a Farm," 2015.
- [63] B. Carpentier and O. Cerf, "Review - Persistence of *Listeria monocytogenes* in food industry equipment and premises," *International Journal of Food Microbiology*, vol. 145, no. 1, pp. 1–8, Jan. 31, 2011. doi: 10.1016/j.ijfoodmicro.2011.01.005.
- [64] R. E. Nilsson, T. Ross, and J. P. Bowman, "Variability in biofilm production by *Listeria monocytogenes* correlated to strain origin and growth conditions," *Int J Food Microbiol*, vol. 150, no. 1, pp. 14–24, Oct. 2011, doi: 10.1016/j.ijfoodmicro.2011.07.012.
- [65] T. Milkiewicz, V. Badia, V. B. Souza, D. A. Longhi, A. C. Galvão, and W. da Silva Robazza, "Development of a general model to describe *Salmonella* spp. growth in chicken meat subjected to different temperature profiles," *Food Control*, vol. 112, Jun. 2020, doi: 10.1016/j.foodcont.2020.107151.
- [66] R. M. Morasi, V. L. M. Rall, S. T. A. Dantas, V. P. P. Alonso, and N. C. C. Silva, "Salmonella spp. in low water activity food: Occurrence, survival mechanisms, and thermoresistance," *J Food Sci*, vol. 87, no. 6, pp. 2310–2323, Jun. 2022, doi: 10.1111/1750-3841.16152.
- [67] T. P. Keerthirathne, K. Ross, H. Fallowfield, and H. Whiley, "A review of temperature, pH, and other factors that influence the survival of *Salmonella* in mayonnaise and other raw egg products," *Pathogens*, vol. 5, no. 4. MDPI AG, Dec. 01, 2016. doi: 10.3390/pathogens5040063.

Agradecimientos

Gracias a mi hija María José por su paciencia mientras hago una de las cosas que más me apasiona en la vida: la microbiología. Gracias a mi asesor al Doctor Nicolás Pinel por su guía en la mejora permanente de mi tesis y al profesor Doctor Geysson Fernández, por sus enseñanzas en aprendizaje automático y su integración con los sistemas biológicos.

Financiación

No se tuvo ninguna financiación, la ejecución del proyecto fue con recursos propios.

Conflicto de intereses

El autor declara que no tiene conflicto de intereses.