

Pontificia Universidad Católica del Ecuador

Facultad De Ingeniería



TEMA:

DESARROLLO E IMPLEMENTACIÓN DE UN MODELO DE SEGMENTACIÓN EN PERFILES DE DISOLUCIÓN UTILIZANDO TÉCNICAS DE MACHINE LEARNING PARA DETERMINAR EL FACTOR DE SIMILITUD (F2) EN UN ESTUDIO DE BIOEQUIVALENCIA IN VITRO PARA METFORMINA CLORHIDRATO.

TRABAJO PREVIA A LA OBTENCIÓN DEL TÍTULO DE MAGISTER EN SISTEMAS DE INFORMACIÓN MENCION DATA SCIENCE

AUTOR:

CRIOLLO LLUMIQUINGA BRYAN SANTIAGO

Quito, Enero – 2024

DEDICATORIA

Dedico este trabajo a mis padres, cuyo amor incondicional, paciencia infinita, apoyo constante y su ejemplo inspirador han sido el pilar fundamental en mi formación académica, profesional y personal. Su presencia constante en cada paso, así como la certeza de que siempre contaré con su apoyo, han sido el motor que me impulsa a seguir cosechando logros, a ustedes les dedico con gratitud este esfuerzo.

Criollo Llumiquinga Bryan Santiago

AGRADECIMIENTO

A mis padres, gracias por su amor constante y por inculcarme valores que me han hecho crecer como persona, siempre agradecido por su apoyo incondicional, su arduo trabajo, esfuerzo y sacrificio para ayudarme a cosechar logros.

A mis hermanos por su apoyo, gracias por estar siempre ahí. Especialmente a mi hermana, agradezco sus consejos y su amistad. Su apoyo ha sido fundamental en este camino.

Criollo Llumiquinga Bryan Santiago

RESUMEN

La presente investigación se centró en la aplicación y evaluación de modelos de segmentación para clasificar los perfiles de disolución del medicamento genérico Metformina Clorhidrato 500 mg en variados niveles de pH (6,8 - 4,5 - 1,2). La técnica empleada fue el Random Forest, la cual se eligió por su capacidad para abordar eficazmente relaciones no lineales y gestionar variables predictoras sin una preparación exhaustiva de datos, lo cual es crucial en el contexto de resultados de laboratorio con limitaciones para manipulación o transformación. Los resultados obtenidos destacaron un rendimiento sólido de los modelos, alcanzando precisiones del 79.76% para pH 6,8, 82.14% para pH 4,5 y 80.95% para pH 1,2. Estos resultados demuestran la capacidad de los modelos para prever de manera acertada la disolución del medicamento en diferentes condiciones de pH, siendo el modelo para pH 4,5 el que mostró una leve inclinación hacia una mayor precisión. La evaluación de sensibilidad y especificidad reveló la eficiencia general de los modelos en la identificación de casos positivos y negativos, respectivamente. Aunque se observaron algunas variaciones en ciertos intervalos de tiempo, los modelos demostraron su confiabilidad para clasificar los perfiles de disolución de Metformina Clorhidrato 500 mg. El análisis del factor de similitud (F2) confirmó la equivalencia terapéutica entre el medicamento genérico y el innovador ya que todos los lotes cumplieron con los criterios de bioequivalencia in vitro, con valores de F2 superiores a 50, respaldando la capacidad del medicamento genérico para disolverse y liberar el principio activo de manera similar al medicamento innovador.

ABSTRACT

The present research focused on the application and evaluation of segmentation models to classify the dissolution profiles of the generic drug Metformina Hydrochloride 500 mg at various pH levels (6.8 - 4.5 - 1.2). The technique used was Random Forest, which was chosen for its ability to effectively address non-linear relationships and manage predictor variables without extensive data preparation, which is crucial in the context of laboratory results with limitations for manipulation or transformation. The results obtained highlighted a solid performance of the models, reaching accuracies of 79.76% for pH 6.8, 82.14% for pH 4.5 and 80.95% for pH 1.2. These results demonstrate the capacity of the models to accurately predict the dissolution of the drug under different pH conditions, with the model for pH 4.5 being the one that showed a slight inclination towards greater precision. The sensitivity and specificity evaluation revealed the overall efficiency of the models in identifying positive and negative cases, respectively. Although some variations were observed at certain time intervals, the models demonstrated their reliability in classifying the dissolution profiles of Metformina Hydrochloride 500 mg. The analysis of the similarity factor (F2) confirmed the therapeutic equivalence between the generic drug and the innovator since all batches met the in vitro bioequivalence criteria, with F2 values greater than 50, supporting the ability of the generic drug to dissolve and release the active ingredient in a similar way to the innovator medicine.

ÍNDICE DE CONTENIDO

ÍNDICE DE FIGURAS.....	8
ÍNDICE DE TABLAS.....	10
1. Introducción.....	11
1.1. Antecedentes.....	12
1.2. Planteamiento del problema.....	13
1.3. Justificación.....	15
1.4. Objetivos.....	16
1.4.1. Objetivo general.....	16
1.4.2. Objetivos específicos.....	16
CAPITULO II: FUNDAMENTACIÓN TEÓRICA.....	17
2. Marco teórico.....	17
2.1. Bioequivalencia de un medicamento.....	17
2.2. Biodisponibilidad de un medicamento.....	17
2.3. Medicamento genérico.....	17
2.4. Medicamento innovador.....	18
2.5. Principio activo.....	18
2.6. Tabletas de liberación inmediata.....	18
2.7. Sistema de Clasificación Biofarmacéutica (BCS).....	18
2.8. Alta solubilidad.....	19
2.9. Alta permeabilidad.....	19
2.10. Bioexención.....	20
2.11. Entidad sanitaria ARCSA.....	20
2.12. Metformina.....	21
2.13. Glucofage.....	22
2.14. Perfil de disolución.....	23
2.15. Factor de similitud (f2).....	24
2.16. Machine learning.....	25
2.16.1. Aprendizaje supervisado.....	25
2.16.2. Aprendizaje no supervisado.....	26
2.16.3. Aprendizaje por refuerzo.....	26
CAPITULO III: FUNDAMENTACIÓN METODOLÓGICA.....	28
3. Metodología.....	28
3.1. Aplicación de la metodología.....	30

3.1.1.	Comprensión del negocio y del problema	30
3.1.2.	Comprensión de los datos.....	31
3.1.2.1.	Recolección de los datos	31
3.1.2.2.	Descripción de los datos.....	31
3.1.2.3.	Exploración de los datos.....	32
3.1.3.	Preparación de los datos	33
3.1.3.1.	Selección de los datos	33
3.1.3.2.	Limpieza de datos.....	35
3.1.4.	Modelado	36
3.1.4.1.	Selección de la técnica	36
3.1.4.2.	Preparación del modelo filtrado por nivel de Ph	37
3.1.4.3.	Construcción del modelo para un nivel de pH de 6,8	37
3.1.4.4.	Construcción del modelo para un nivel de pH de 4,5	38
3.1.4.5.	Construcción del modelo para un nivel de pH de 1,2	39
3.1.5.	Evaluación	40
3.1.5.1.	Evaluación del modelo para un nivel de pH de 6,8	41
3.1.5.2.	Evaluación del modelo para un nivel de pH de 4,5	43
3.1.5.3.	Evaluación del modelo para un nivel de pH de 1,2	44
3.1.6.	Implementación	46
CAPITULO IV: RESULTADOS.....		48
4.	Resultados	48
4.1.	Resultados de los modelos obtenidos.....	48
4.2.	Resultados del factor de similitud	51
4.2.1.	Curvas de disolución de los medicamentos genérico e innovador, pH 6,8	51
4.2.2.	Curvas de disolución de los medicamentos genérico e innovador, pH 4,5	52
4.2.3.	Curvas de disolución de los medicamentos genérico e innovador, pH 1,2	53
4.2.4.	Cálculo del factor de similitud por lote para un nivel de pH de 6,8 – 4,5 y 1,2 54	
CAPITULO V: CONCLUSIONES Y RECOMENDACIONES.....		55
5.1.	Conclusiones	55
5.2.	Recomendaciones	56
Bibliografía		58

ÍNDICE DE FIGURAS

Ilustración 1. Sistema de Clasificación Biofarmacéutica.	19
Ilustración 2. Listado de principios activos que el ARCSA solicita estudios de bioequivalencia in vitro	21
Ilustración 3. Modelo de machine learning supervisado	26
Ilustración 4. Machine learning no supervisado	26
Ilustración 5. Machine learning por refuerzo	27
Ilustración 6. Metodología CRISP-DM.....	28
Ilustración 7. Estructura de base de datos del medicamento Innovador.....	32
Ilustración 8. Estructura de base de datos del medicamento Genérico.....	32
Ilustración 9. Salida de la base del medicamento Innovador para determinar valores nulos o faltantes	35
Ilustración 10. Salida de la base del medicamento Genérico para determinar valores nulos o faltantes	36
Ilustración 11. Segmentación por nivel de pH.....	37
Ilustración 12. Partición de la data para prueba y entrenamiento, pH 6,8	38
Ilustración 13. Desarrollo del modelo de segmentación para un pH de 6,8	38
Ilustración 14. Partición de la data para prueba y entrenamiento, pH 4,5	39
Ilustración 15. Desarrollo del modelo de segmentación para un pH de 4,5	39
Ilustración 16. Partición de la data para prueba y entrenamiento, pH 1,2	39
Ilustración 17. Desarrollo del modelo de segmentación para un pH de 1,2	40
Ilustración 18. Evaluación del rendimiento del modelo a un nivel de pH 6.8.....	41
Ilustración 19. Matriz de estadísticas por clase, pH 6,8	42
Ilustración 20. Evaluación del rendimiento del modelo a un nivel de pH 4,5.....	43
Ilustración 21. Matriz de estadísticas por clase, pH 4,5	44
Ilustración 22. Evaluación del rendimiento del modelo a un nivel de pH 1,2.....	45
Ilustración 23. Matriz de estadísticas por clase, pH 1,2	46
Ilustración 24. Curvas de disolución Generico vs Innovador, pH 6,8	51

Ilustración 25. Curvas de disolución Generico vs Innovador, pH 4,5	52
Ilustración 26. Curvas de disolución Generico vs Innovador, pH 1,2	53
Ilustración 27. Resultados del factor de similitud Lote vs pH.....	54

ÍNDICE DE TABLAS

Tabla 1. Variables para el desarrollo del modelo de segmentación de perfiles de disolución	33
Tabla 2. Descripción de la variable TiempDisol	34

CAPÍTULO I: INTRODUCCIÓN

1. Introducción

La bioequivalencia es un aspecto crítico en el proceso de evaluación y aprobación de medicamentos genéricos ya que mediante esta técnica se busca garantizar que estos fármacos sean igual de efectivos y seguros que sus equivalentes de referencia con el objetivo de brindar al paciente un tratamiento calidad. En este aspecto, el estudio de bioequivalencia in vitro desempeña un papel fundamental al evaluar la similitud de la liberación de un principio activo en diversas formulaciones farmacéuticas.

Uno de los principios activos ampliamente utilizados es la metformina clorhidrato, un fármaco diseñado para controlar los niveles de azúcar en la sangre, efectivo en el tratamiento de la diabetes tipo 2. La precisión en la evaluación de la bioequivalencia de formulaciones de metformina es esencial para garantizar la calidad y la seguridad de los medicamentos disponibles en el mercado, lo cual implica un análisis minucioso de los perfiles de disolución en diferentes niveles de pH y determinar si cumplen con los estándares requeridos, los cuales se representan por el factor de similitud (f_2) para posteriormente ser presentados a la Agencia Nacional de Regulación, Control y Vigilancia Sanitaria (ARCSA).

La presente investigación se enfoca en el desarrollo e implementación de un modelo de segmentación de perfiles de disolución para cada uno de los niveles de pH establecidos, aprovechando las técnicas de machine learning. Este modelo tendrá como objetivo principal segmentar los los perfiles de disolución para posteriormente determinar el factor de similitud (f_2) e identificar parones que se desvíen significativamente en estudios de bioequivalencia in vitro para Metformina Clorhidrato. Al automatizar y optimizar este proceso, se busca agilizar la evaluación de bioequivalencia, reducir los costos asociados, mejorar la precisión y objetividad en estudios de este tipo.

El impacto de este estudio busca extenderse más allá del ámbito académico, ya que la aprobación de medicamentos genéricos de alta calidad y su disponibilidad en el mercado son fundamentales para la salud pública y la accesibilidad a tratamientos

efectivos. Además, este estudio contribuye al avance científico y tecnológico al aplicar técnicas de machine learning en el análisis de perfiles de disolución, lo que potencialmente podría tener aplicaciones en otros campos de la farmacología y la ciencia de datos.

1.1. Antecedentes

En la investigación realizada sobre “Estudio de Bioequivalencia "in vitro" de tres productos genéricos (tabletas) de liberación inmediata que contienen Metformina Clorhidrato 850 mg con el medicamento innovador comercializados en Perú” realizada en el año 2017 se concluyó que: Evaluando los perfiles de disolución en tres medios con diferentes niveles de pH. Los resultados demostraron que todos los productos cumplen con los estándares de calidad en pruebas de disolución, valoración y uniformidad, pero, se encontraron diferencias en los factores de similitud en comparación con el medicamento referente. El medicamento A mostró bioequivalencia a pH 4.5 y 6.8, mientras que el B y C solo lo lograron a pH 6.8, hallazgos para evaluar la calidad y seguridad de los medicamentos genéricos (Paz, 2017).

En la publicación desarrollada acerca de “Equivalencia farmacéutica in vitro de metformina clorhidrato frente a Glucofage®” en el año 2018 se obtuvo que: Al evaluar la equivalencia farmacéutica in vitro del medicamento genérico metformina clorhidrato de 850 mg en comparación con su innovador Glucofage®. De las marcas genéricas (B, C, D, E) demostraron ser equivalentes farmacéuticos al producto de referencia según el factor de similitud (f_2), lo que respalda su calidad y semejanza en términos de liberación del principio activo. Sin embargo, se identificaron problemas con la marca F, que no cumplió con el criterio de f_2 y mostró deficiencias en aspectos físicos y de dureza. Además, la marca B no cumplió con la uniformidad de contenido (Encalada & Sánchez, 2018).

Según la investigación sobre “Estudio de bioequivalencia in vitro de comprimidos de liberación inmediata de metformina de 850 mg comercializados en Bolivia” del año 2018 se encontró que: Todos los productos pasaron con éxito las pruebas de calidad, los resultados del estudio de bioequivalencia revelaron diferencias estadísticamente

significativas entre los productos de estudio y el de referencia. Esto indicó que no se pudo demostrar la bioequivalencia in vitro de los productos B, C, D y E en comparación con el producto de referencia A, basándose en las pruebas de disolución in vitro. Estos hallazgos resaltan la importancia de una evaluación rigurosa de la bioequivalencia para garantizar la calidad y eficacia de los medicamentos genéricos (Saravia & Daza, 2020).

Según el estudio “Estudio de bioequivalencia in vitro de dos formas farmacéuticas perorales multifuente de liberación inmediata con metformina como principio activo” del año 2013 se encontró que, los resultados revelaron diferencias estadísticamente significativas entre los productos B y C en comparación con el producto de referencia A en términos de sus perfiles de disolución in vitro. Estos hallazgos sugieren que no se puede afirmar que los productos B y C sean bioequivalentes al producto de referencia en las condiciones de este estudio. Esta investigación resalta la importancia de una evaluación rigurosa de la bioequivalencia para garantizar la calidad y eficacia de los medicamentos, especialmente en el contexto de la metformina, un fármaco crucial en el tratamiento de la diabetes (Pérez, 2013).

En la investigación sobre “Estudio de la bioequivalencia in vitro de las tabletas de ciprofloxacino de 500 mg, elaboradas por dos laboratorios farmacéuticos del Ecuador, mediante el empleo de los perfiles de disolución” del año 2016 se encontró que, los resultados revelaron que ninguno de los medicamentos genéricos cumplió con los criterios de similitud establecidos, ya que los valores de F2 se encontraron por debajo del rango de aceptación. Esto indica que los productos genéricos no son intercambiables con el medicamento de referencia en términos de su perfil de disolución lo que sugiere que estos medicamentos genéricos pueden no garantizar la misma actividad farmacológica que el producto de referencia (Calle & Copete, 2016).

1.2.Planteamiento del problema

La evaluación de la bioequivalencia es fundamental en el desarrollo y registro de medicamentos genéricos, ya que garantiza su efectividad y seguridad. El ARCSA (Agencia de Regulación y Control Sanitario) establece los criterios y requisitos necesarios para demostrar la bioequivalencia y biodisponibilidad en los medicamentos de uso y

consumo humano, por lo cual, en base a la bioexención este tipo de estudios se pueden respaldar con estudios in vitro para demostrar la equivalencia terapéutica comparado el medicamento genérico con el de innovador.

En el caso específico de metformina clorhidrato como principio activo, el ARCSA dentro de su listado de principios activos solicita un estudio de bioequivalencia el cual implica analizar los perfiles de disolución utilizando el factor de similitud (f_2). Este factor se utiliza como parámetro para determinar la comparabilidad entre un medicamento genérico (Metformina clorhidrato 500 mg) y su producto de innovador (Glucofage 500 mg) para lo cual dentro de este estudio se considerarán diferentes condiciones de pH que simulan el ambiente gastrointestinal.

Para llevar a cabo este estudio, se requiere que un analista digite y procese los resultados de las pruebas de laboratorio referente a los perfiles de disolución ubicándolos en plantillas Excel para posteriormente calcular el factor de similitud y dar una conclusión. Sin embargo, debido los costos, el número de muestras y el tiempo que toma realizar este tipo de análisis, es necesario desarrollar un modelo que permita identificar, clasificar de manera automática y precisa los perfiles de disolución de la metformina en diferentes intervalos de tiempo ($t_1=10$ minutos, $t_2=15$ minutos, $t_3=20$ minutos, $t_4=30$ minutos, $t_5=45$ minutos y $t_6=60$ minutos) para los distintos niveles de pH (6,8-4,5 y 1,2)

En este sentido, se utilizarán técnicas de machine learning para llevar a cabo dicha clasificación. Sin embargo, también es necesario identificar patrones de disolución que no cumplan con los estándares del principio activo del medicamento de innovador o innovador.

La presencia de patrones de disolución fuera de los estándares establecidos implica una falta de comparabilidad potencial en términos de efectividad y seguridad entre el medicamento genérico vs el producto de innovador. Por lo tanto, es crucial investigar las posibles causas y factores influyentes que contribuyen a estas variaciones. Además, se requieren ajustes y optimizaciones en la formulación y el proceso de fabricación de los medicamentos genéricos para garantizar que cumplan con los criterios de calidad, efectividad y seguridad requeridos.

Por lo cual surge la interrogante ¿A través de la aplicación de este modelo se podrá identificar la similitud de un medicamento genérico comparado con el original e identificar patrones de disolución que se desvíen significativamente del estándar establecido en los niveles de pH (6,8-4,5 y 1,2)?

1.3.Justificación

La investigación y desarrollo del modelo de segmentación en perfiles de disolución utilizando técnicas de machine learning para determinar el factor de similitud (f_2) en un estudio de bioequivalencia in vitro para el clorhidrato de metformina es de vital importancia por las siguientes razones:

Mejora en la evaluación de la bioequivalencia: La bioequivalencia es un parámetro crucial para garantizar la comparabilidad en efectividad y seguridad entre medicamentos genéricos y sus productos de innovador. Al desarrollar un modelo de segmentación automatizado, se optimiza y agiliza el proceso de evaluación de bioequivalencia in vitro para metformina clorhidrato, lo cual facilita la toma de decisiones sobre la aprobación y registro de medicamentos genéricos.

Eficiencia y reducción de costos: La implementación de técnicas de machine learning en la clasificación automatizada de perfiles de disolución permite una mayor eficiencia en la identificación de patrones de liberación del fármaco. Esto reduce la carga de trabajo y el tiempo requerido para el análisis de los perfiles de disolución, lo que a su vez reduce los costos asociados al estudio de bioequivalencia.

Mayor precisión y objetividad: El uso de técnicas de machine learning en la segmentación de perfiles de disolución ofrece una mayor precisión y objetividad en la clasificación de estos. El modelo desarrollado puede identificar de manera automática las características relevantes de la liberación de la metformina y detectar patrones de disolución que no se ajusten a los estándares establecidos. Esto proporciona una evaluación más precisa de la bioequivalencia y garantiza que los medicamentos genéricos cumplan con los criterios de calidad, efectividad y seguridad requeridos.

Detección de desviaciones y optimización de formulaciones: El modelo de segmentación basado en machine learning permite identificar patrones de disolución que

se desvíen significativamente de los estándares establecidos. Esta información es importante para detectar posibles problemas en la formulación y el proceso de fabricación de los medicamentos genéricos, lo que a su vez permite proponer ajustes y optimizaciones que mejoren la calidad y consistencia de los productos.

Contribución al conocimiento científico: La investigación en este campo contribuye al avance científico y tecnológico al aplicar técnicas de machine learning en el análisis de perfiles de disolución. Los resultados obtenidos y el modelo desarrollado pueden servir como base para futuras investigaciones en el campo de la bioequivalencia y el uso de inteligencia artificial en la evaluación de medicamentos genéricos.

1.4.Objetivos

1.4.1. Objetivo general

Desarrollar e implementar un modelo de segmentación en perfiles de disolución utilizando técnicas de machine learning para determinar el factor de similitud (f_2) en un estudio de bioequivalencia in vitro para Metformina clorhidrato.

1.4.2. Objetivos específicos

- Investigar la relación entre la liberación de metformina clorhidrato y la bioequivalencia, identificando los factores que pueden influir en las variaciones de la liberación del fármaco.
- Desarrollar y evaluar la precisión de un modelo de segmentación basado en técnicas de machine learning que permita clasificar los perfiles de disolución de metformina clorhidrato 500 mg según el pH.
- Determinar el factor de similitud (f_2) utilizando el modelo de segmentación desarrollado evaluando la comparabilidad entre el medicamento genérico (metformina clorhidrato 500 mg) y el producto de innovador (Glucofage 500 mg) según su perfil de disolución y pH identificando patrones que se desvíen significativamente de los estándares establecidos.

CAPITULO II: FUNDAMENTACIÓN TEÓRICA

2. Marco teórico

2.1. Bioequivalencia de un medicamento

La bioequivalencia se refiere a la cualidad que demuestra que un medicamento es equivalente en términos de calidad, eficacia y seguridad al fármaco de referencia, teniendo el mismo principio activo y dosis, aunque sea producido por diferentes laboratorios farmacéuticos. Esta condición permite que ambos medicamentos tanto el original y el genérico, puedan ser intercambiables, ya que el medicamento genérico produce el mismo efecto terapéutico que el medicamento original. Las pruebas de bioequivalencia se realizan para comparar el medicamento genérico con el original o también llamado innovador y garantizar que ambos contengan la misma cantidad de principio activo y alcancen los mismos niveles en sangre en el mismo tiempo, lo que asegura que sus efectos en el organismo sean *los mismos* (Eurofarma, 2023).

2.2. Biodisponibilidad de un medicamento

La biodisponibilidad se refiere a la velocidad y medida en que un fármaco ingresa en el torrente sanguíneo y llegan a su lugar de acción. Esta característica está relacionada con las propiedades de la forma farmacéutica y su proceso de diseño y fabricación. Es esencial determinar si diferentes formulaciones del mismo medicamento son equivalentes, ya que las diferencias en la biodisponibilidad pueden ser clínicamente significativas. (Le, 2022).

2.3. Medicamento genérico

Un medicamento genérico es una alternativa diseñada para ser idéntica en todos los aspectos importantes a un medicamento de marca ya establecido en el mercado. Esto incluye la dosis, seguridad, potencia, forma de administración, calidad y efecto terapéutico. La bioequivalencia es la clave para demostrar que el medicamento genérico es efectivamente igual al medicamento de marca, lo que significa que ambos proporcionan los mismos resultados clínicos y beneficios para el paciente, por lo cual un medicamento genérico puede reemplazar sin problema al medicamento de marca,

ofreciendo el mismo nivel de eficacia y resultados positivos para el tratamiento (FDA, 2018).

2.4. Medicamento innovador

Se considera un producto innovador o de patente a aquel medicamento que ha sido desarrollado mediante un proceso de investigación y se encuentra protegido por una patente, lo que le otorga exclusividad a la empresa farmacéutica que lo creó. Estos medicamentos se identifican tanto por el nombre de su sustancia activa como por una denominación comercial o marca. Es importante destacar que, debido a su protección por patente, solo el laboratorio farmacéutico que lo desarrolló tiene el derecho exclusivo de fabricarlo y comercializarlo por un periodo determinado (Ledón & Hernández, 2020).

2.5. Principio activo

Un principio activo es una sustancia o combinación de sustancias esenciales a partir de las cuales se fabrican ciertos medicamentos para producir un efecto específico en el organismo, ya sea metabólico, inmunológico o farmacológico. Incorporado al proceso de producción, el ingrediente activo el cual se convierte en una parte esencial del funcionamiento del fármaco y este es responsable del efecto terapéutico que el medicamento produce en el organismo del paciente (Unilabs, 2023).

2.6. Tabletas de liberación inmediata

Las tabletas de liberación inmediata son una forma farmacéutica común obtenida por la compresión de polvos o gránulos con sustancias medicinales, su rápida liberación del principio activo permite una eficacia terapéutica ágil, la formulación adecuada es esencial para asegurar su eficacia y seguridad, brindando opciones terapéuticas confiables y efectivas para los pacientes (Paz, 2017).

2.7. Sistema de Clasificación Biofarmacéutica (BCS)

El Sistema de Clasificación Biofarmacéutica (BCS) clasifica los medicamentos según sus parámetros de solubilidad y permeabilidad, los cuales desempeñan un papel fundamental en el proceso de absorción. La SCB busca como objetivo sustituir las pruebas in vivo en humanos con pruebas de disolución in vitro, de acuerdo con la

clasificación obtenida para el fármaco. En la actualidad, el enfoque principal de la aplicación del BCS se centra especialmente en los estudios de bioequivalencia, que tienen la finalidad de demostrar la intercambiabilidad de los fármacos administrados por vía oral en forma sólida de liberación inmediata. De esta manera, el BCS juega un papel relevante al optimizar y agilizar la evaluación de la bioequivalencia, facilitando la aprobación y comercialización de medicamentos genéricos en el competitivo mercado farmacéutico (Sánchez, 2018).

El Sistema de Clasificación Biofarmacéutica (BCS), sistema empleado para clasificar principios activos en base a características de solubilidad acuosa y permeabilidad intestinal se categorizan según la (USP-NF, 2019) de la siguiente forma:

<i>Clase 1</i>	<i>Alta solubilidad, alta permeabilidad</i>
<i>Clase 2</i>	<i>Baja solubilidad, alta permeabilidad</i>
<i>Clase 3</i>	<i>Alta solubilidad, baja permeabilidad</i>
<i>Clase 4</i>	<i>Baja solubilidad, baja permeabilidad</i>

Ilustración 1. Sistema de Clasificación Biofarmacéutica.

Elaborado por: Criollo, 2023.

2.8. Alta solubilidad

“Un principio activo es considerado altamente soluble cuando la mayor potencia registrada en el país sea soluble en 250 mL o menos de medio acuoso en el rango de pH 1,2 – 6,8. Por ejemplo, si un principio activo se encuentra en forma de comprimidos de 500 mg, 850 mg y 1000 mg, en una FFSO-LI, este principio activo se considera altamente soluble cuando la mayor potencia (1000 mg) es soluble en 250 mL o menos, en todo el rango de pH 1,2 – 6,8” (Instituto de Salud Pública de Chile, 2018).

2.9. Alta permeabilidad

“Un principio activo es considerado altamente permeable cuando i) La extensión de la absorción en humanos es 85% o más basado en la determinación de balance de masa o en comparación con una dosis intravenosa (biodisponibilidad absoluta). ii) Cuando al menos un 85% de la droga administrada es excretada inalterada en la orina. iii) Cuando al menos un 85% de la droga administrada es recuperada en la orina como droga madre y

metabolitos con evidencia de su estabilidad en el tracto. (Instituto de Salud Pública de Chile, 2018)

2.10. Bioexención

Es la capacidad que posee la autoridad sanitaria de exonerar al fabricante o titular del medicamento de presentar los estudios de bioequivalencia *in vivo* para establecer la equivalencia terapéutica cuando esta pueda demostrarse mediante estudios *in vitro* (USP-NF, 2020).

Por lo cual la bioexención representa la autorización de comercialización para formas de dosificación orales, basadas exclusivamente en criterios de disolución, eximiendo así la necesidad de realizar estudios de bioequivalencia "*in vivo*", por tanto, estas excepciones permiten evitar ensayos *in vivo* para demostrar la bioequivalencia del fármaco, optando en cambio por utilizar ensayos de disolución *in vitro* para asegurar la intercambiabilidad y eficacia del medicamento, este enfoque se encuentra respaldado por rigurosas normas científicas las cuales agilizan y simplifican el proceso de aprobación y comercialización de medicamentos, brindando tanto a los fabricantes como a los pacientes opciones de tratamiento seguras y confiables (Subirán, 2018).

2.11. Entidad sanitaria ARCSA

La Agencia Nacional de Regulación, Control y Vigilancia Sanitaria (ARCSA), es la entidad pública adscrita al Ministerio de Salud Pública (MSP) que se encarga de controlar y vigilar las condiciones higiénico – sanitarias de los productos de uso y consumo humano, además de brindar servicios que facilitan la obtención de permisos de funcionamiento y Notificaciones Sanitarias. (Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA, s.f.).

Dentro del comunicado del ARCSA en la resolución No. ARCSA-DE-015-2018-JCGO modificada en el 2021, menciona en el artículo 4, que. “La ARCSA determinará y mantendrá actualizada la lista de principios activos que requieren estudios de bioequivalencia "*in vivo*" e "*in vitro*" y la lista de medicamentos comparadores, conforme

el Instructivo que la ARCSA elabore para el efecto (Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA, 2021).

Por lo cual, en el “Instructivo externo criterios y requisitos para demostrar bioequivalencia y biodisponibilidad, en los medicamentos de uso y consumo humano”. Anexo 1, emitido por el ARCSA se encuentra el listado de los principios activos que requieren de un estudio in vitro (Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA, 2023). Dentro de este estudio se toma al principio activo de metformina la cual se encuentra en la posición 11.



Agencia Nacional de Regulación,
Control y Vigilancia Sanitaria

INSTRUCTIVO EXTERNO CRITERIOS Y REQUISITOS PARA DEMOSTRAR BIOEQUIVALENCIA Y BIODISPONIBILIDAD, EN LOS MEDICAMENTOS DE USO Y CONSUMO HUMANO	CÓDIGO	IE-B.3.2.1- MED-02
	VERSIÓN	2
	Página 11 de 17	

4	Capecitabina	In vitro	17/03/2023
5	Ciclofosfamida	In vitro	17/03/2023
6	Clonazepam	In vitro	17/03/2023
7	Clorambucilo	In vitro	17/03/2023
8	Gabapentina	In vitro	17/03/2023
9	Lamivudina, combinaciones	In vitro	17/03/2023
10	Letrozol	In vitro	17/03/2023
11	Metformina	In vitro	17/03/2023
12	Propranolol	In vitro	17/03/2023
13	Selegilina	In vitro	17/03/2023
14	Sunitinib	In vitro	17/03/2023
15	Temozolomida	In vitro	17/03/2023
16	Zidovudina, combinaciones	In vitro	17/03/2023

Ilustración 2. Listado de principios activos que el ARCSA solicita estudios de bioequivalencia in vitro

Fuente: (Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA, 2021)

2.12. Metformina

Es un medicamento antidiabético de administración oral que tiene la función de regular los niveles de glucosa en el organismo. Por ende, es ampliamente utilizado por personas que padecen diabetes para mantener bajo control la concentración de azúcar en su sangre. Este fármaco logra este efecto al actuar en tres ubicaciones específicas del cuerpo: el hígado, el músculo y el intestino (Torres, 2023).

Las tabletas de metformina están disponibles en una formulación de liberación inmediata y este compuesto tiene una clasificación 3 (alta solubilidad y baja permeabilidad) bajo el sistema de clasificación biofarmacéutico y se considera candidato para la exención de los estudios de bioequivalencia, principalmente debido al hecho de que, para este tipo de fármaco, la absorción se produce a través de la permeabilidad a través de la membrana intestinal, en lugar de la disolución (Pérez , Orobio, & Baena, 2013).

La biodisponibilidad absoluta de metformina tras la administración de un comprimido de 500 mg en ayunas es aproximadamente de 50%-60%. Estudios que utilizaron dosis orales únicas de Glucophage® de 500 a 1500 mg y de 850 a 2550 mg indican que existe una falta de proporcionalidad de dosis debido a una absorción reducida en lugar de una alteración en la eliminación (Bristol-Myers Squibb Company, 2017); se absorbe proporcionalmente menos fármaco después de dosis más altas que con dosis más bajas, posiblemente debido a un proceso de absorción saturable (Metry , Abrahamsson, Cristofolletti, & Dressman, 2021).

2.13. Glucofage

El medicamento innovador de metformina clorhidrato se originó a partir de la planta *Galega officinalis*, que históricamente se utilizaba para tratar a pacientes diabéticos. En 1956, el diabetólogo francés Jean Sterne realizó estudios que demostraron la eficacia de la metformina como hipoglucemiante, lo que llevó a su aplicación en pacientes diabéticos en diferentes hospitales. Los resultados de estos estudios indicaron que la metformina podía reducir o eliminar la necesidad de insulina en los pacientes tratados, sin riesgo de hipoglucemia grave, y sin alteraciones en individuos no diabéticos. Sterne publicó los hallazgos científicos sobre la metformina en 1957 en una revista médica marroquí (Encalada & Sánchez, 2018).

Glucophage® es la marca comercial de metformina, un medicamento que se usa en adultos con diabetes tipo 2 para regular los niveles de azúcar en la sangre, esta tableta viene en presentaciones de liberación inmediata y liberación prolongada. Su función principal es aumentar la sensibilidad a la insulina, lo que ayuda a las células a absorber mejor la glucosa y reduce la producción de glucosa por parte del hígado. La metformina

se prescribe para lograr niveles óptimos de glucosa y también para mantener un control general de los niveles de azúcar en la sangre, este fármaco también ayuda a personas con prediabetes o síndrome de ovario poliquístico, ya que ayuda a combatir la resistencia a la insulina en estos casos (Texas Children's Hospital, 2016).

2.14. Perfil de disolución

La disolución se considera como un procedimiento para el control de calidad y una prueba de equivalencia in vitro para las formas farmacéuticas de liberación inmediata, la prueba de control de calidad consiste en evaluar la disolución del medicamento en un solo punto y en un medio específico. Por el contrario, la prueba de equivalencia in vitro implica comparar el perfil de disolución en diferentes medios con pH 1.2, 4.5 y 6.8 entre el medicamento innovador y el genérico (USP-NF, 2020)

- Buffer de ácido clorhídrico de pH 1,2 (Ambiente estomacal)
- Buffer de acetato de pH 4,5 (Ambiente intestinal, grueso)
- Buffer de fosfato de pH 6,8 (Ambiente intestinal, delgado)

Por lo cual los perfiles de disolución se convierten en una herramienta importante para evaluar la liberación de un ingrediente activo en el tiempo a través de pruebas de disolución en diversos medios de pH los cuales se preparan de acuerdo a lo establecido en la farmacopea vigente simulando así el ambiente gastrointestinal, además, estos perfiles son un indicador importante para estimar la biodisponibilidad del fármaco y para establecer correlaciones entre los parámetros evaluados in vitro y los resultados de biodisponibilidad observados en organismos (Polio, 2014).

Según (Encalada & Sánchez, 2018) afirma que “La comparación de perfiles de disolución está ganando importancia como un medio de documentación comparativa con los estudios de bioequivalencia ya que con esta prueba in vitro se pueden reemplazar los estudios in vivo”.

La segmentación de los perfiles de disolución es otro aspecto importante de la evaluación de bioequivalencia, implica dividir el perfil de disolución en intervalos de tiempo significativos para analizar la liberación del ingrediente activo en diferentes

etapas. La aplicación de técnicas de aprendizaje automático a los perfiles de disolución de segmentos es particularmente útil, ya que permite la identificación automática de patrones de liberación relevantes y significativos. Esto contribuye a una mayor precisión y objetividad de la clasificación, lo que permite una evaluación más precisa de la bioequivalencia. Asimismo, la reducción de costes y tiempo asociada a la evaluación convierte a este método en una opción muy eficaz en el proceso de evaluación de bioequivalencia in vitro.

2.15. Factor de similitud (f2)

Según la (FDA, 2018) “El factor de similitud (f2) es una transformación de raíz cuadrada recíproca logarítmica de la suma del error cuadrado y es una medición de la similitud en la disolución porcentual (%) entre las dos curvas”.

Es una medida de similitud entre las curvas de disolución obtenidas de los productos de prueba y de referencia, es decir el medicamento genérico vs el innovador, este factor se calcula a partir del perfil de disolución promedio en cada tiempo de muestreo utilizando la siguiente ecuación (Jung, Anda, Rubio, & Mayet, 2012):

$$f_2 = 50 \times \log \left\{ \left[1 + \left(\frac{1}{n} \right) \sum_{t=1}^n (R_t - P_t) \right]^{-0.5} \times 100 \right\}$$

Donde:

n= Tiempo de muestreo

R_t= Porcentaje disuelto promedio en el tiempo t del medicamento de referencia

P_t= Porcentaje disuelto promedio en el tiempo t del medicamento de prueba

El factor de similitud (f2) varía en un rango de 50 a 100, siendo 100 el valor ideal, cuando el f2 es igual o mayor a 50, se considera que el producto cumple con el criterio de similitud en comparación con el producto de referencia lo cual implica que las formulaciones son consideradas como iguales o muy similares al producto de referencia (Redondo, 2015).

Las pruebas comparativas de disolución in vitro son útiles cuando la velocidad de disolución es un factor clave que limita la absorción del fármaco. Estos estudios permiten

establecer criterios de disolución específicos en el control de calidad, lo que asegura la consistencia en el proceso de fabricación. Si existe una correlación bien probada entre los resultados in vitro e in vivo, es posible predecir el comportamiento de las drogas en los organismos utilizando los modelos desarrollados. De esta forma, los perfiles de disolución in vitro se pueden utilizar como sustitutos para evaluar la bioequivalencia de fármacos (Jung, Anda, Rubio, & Mayet, 2012).

2.16. Machine learning

El Machine Learning es una variante de la Inteligencia Artificial que permite que un sistema adquiera conocimientos a partir de datos, en lugar de depender de una programación explícita. Mediante el uso de datos de entrenamiento, el algoritmo es capaz de desarrollar modelos más precisos y fundamentados en la información recopilada, este tipo de modelos representa la información resultante del proceso de entrenamiento, cuando el algoritmo es capacitado con los datos disponibles, por lo cual, luego de suministrar datos al modelo, se obtiene una predicción basada en la información utilizada durante el entrenamiento de este (IBM, 2020).

2.16.1. Aprendizaje supervisado

El aprendizaje supervisado utiliza algoritmos que aprenden de datos para descubrir información sin necesidad de una programación detallada. Estos algoritmos construyen modelos basados en datos de entrenamiento para predecir valores con nuevos datos de entrada. Las máquinas desarrollan relaciones basadas en lo aprendido de conjuntos de datos anteriores, lo que mejora la toma de decisiones y la predicción de resultados (Inesdi Business Techschool, 2022).

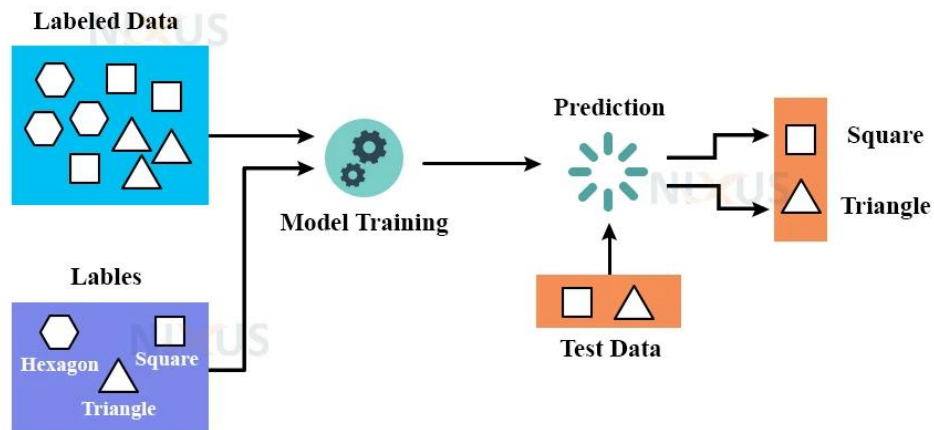


Ilustración 3. Modelo de machine learning supervisado

Fuente: (Nixus, 2023)

2.16.2. Aprendizaje no supervisado

El aprendizaje no supervisado es un método en la ciencia de datos que busca patrones y estructuras ocultas en conjuntos de datos sin etiquetas ni resultados conocidos, a diferencia del aprendizaje supervisado, cabe mencionar que suele ser menos predecible debido a la falta de datos de entrenamiento estructurados (Aprendeia, 2020).

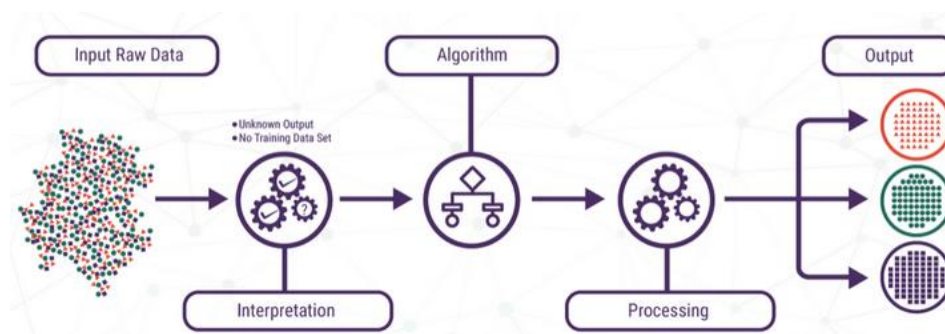


Ilustración 4. Machine learning no supervisado

Fuente: (Domingo, 2018)

2.16.3. Aprendizaje por refuerzo

El aprendizaje por refuerzo es una técnica donde las máquinas aprenden de forma autónoma al recibir recompensas por acciones correctas y sanciones por acciones incorrectas, similar a cómo los seres vivos aprenden de la experiencia. En este enfoque, las máquinas buscan constantemente tomar decisiones que generen recompensas y evitan

acciones que, según su experiencia previa, conllevan sanciones (Centro Europeo de Posgrado - CEUPE, 2020).



Ilustración 5. Machine learning por refuerzo

Fuente: (Velazquez, 2021)

CAPITULO III: FUNDAMENTACIÓN METODOLÓGICA

3. Metodología

La metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) que significa “Proceso estándar interindustrial para la minería de datos”. Cuyo objetivo es desarrollar modelos a partir del análisis de la información y los datos de una organización para predecir futuras fallas y resoluciones (Caio, 2022)

Dentro de esta metodología se desglosan seis etapas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelamiento, evaluación e implementación, cabe recalcar que el proceso no sigue una secuencia estricta, permitiendo el avance y retroceso entre fase, además cada fase guía la siguiente según su resultado.

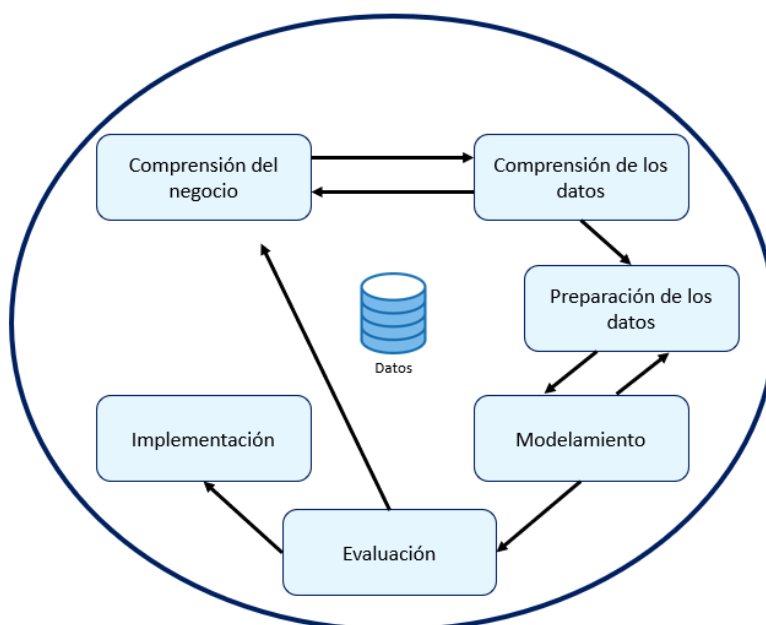


Ilustración 6. Metodología CRISP-DM

Elaborado por: Criollo, 2023.

Dentro de esta metodología se desglosan seis etapas: comprensión del negocio, comprensión de los datos, preparación de los datos, modelamiento, evaluación e implementación, cabe recalcar que el proceso no sigue una secuencia estricta, permitiendo el avance y retroceso entre fase, además cada fase guía la siguiente según su resultado.

En este trabajo se aplicará la metodología CRISP-DM la cual ofrece una estructura sólida que se ajusta a las necesidades de la investigación. Su enfoque permite abordar los resultados obtenidos del estudio in vitro de metformina clorhidrato y las técnicas de machine learning, además, esta metodología inicia con la comprensión del contexto de negocio, lo que es esencial para definir claramente los objetivos de la investigación asegurando que estos estén alineados con las necesidades de la industria farmacéutica y las regulaciones sanitarias.

Esta metodología también destaca la gestión de datos, garantizando que se trabajaran con datos de calidad y confiabilidad, dado que la precisión del modelo dependerá de los resultados obtenidos en los perfiles de disolución en los distintos niveles de pH para posteriormente determinar el del factor de similitud (f_2), lo cual, es fundamental dado que los resultados de la investigación tienen un impacto significativo en la evaluación de los medicamentos genéricos.

Por lo cual, se enlista cada una de las etapas que se deben cumplir dentro de esta metodología para alcanzar satisfactoriamente los objetivos del proyecto.

- Comprensión del negocio
- Comprensión de los datos
- Preparación de los datos
- Modelamiento
- Evaluación
- Implementación

Por lo cual, se enlista cada una de las etapas que se deben cumplir dentro de esta metodología para alcanzar satisfactoriamente los objetivos del proyecto.

3.1. Aplicación de la metodología

3.1.1. Comprensión del negocio y del problema

Esta etapa es quizás la más importante de todo el proceso. Si no se hace correctamente, el resto del proyecto puede quedar inválido en el futuro. En esta etapa se determinan los objetivos del proyecto y las necesidades de la empresa o proyecto que se analiza. Por lo tanto, todos deben estar bien informados y plenamente alineados (Caio, 2022).

Como lo menciona (Vallalta, 2023). En esta fase deberemos hay que tener la capacidad de:

- Establecer los objetivos de negocio.
- Evaluar la situación actual.
- Fijar los objetivos a nivel de minería de datos.
- Obtener un plan de proyecto.

Desde el enfoque del negocio, se realiza un análisis detallado sobre la importancia de la bioequivalencia en la industria farmacéutica y cómo la aplicación de modelos de aprendizaje automático puede aumentar la eficiencia de este importante proceso. Este enfoque se basa en un profundo conocimiento de cómo la equivalencia terapéutica entre medicamentos genéricos e innovadores es un pilar fundamental para garantizar la seguridad y eficacia de los productos farmacéuticos utilizando técnicas como la bioequivalencia la cual es un requisito regulatorio establecido por la autoridad sanitaria Agencia de Regulación y Control Sanitario (ARCSA).

Además, se exploran los perfiles de disolución y el factor de similitud (f_2) en el contexto de la evaluación de bioequivalencia. Se aborda cómo los perfiles de disolución representan patrones de liberación del fármaco a lo largo del tiempo y su relación con la eficacia del tratamiento. El f_2 , por su parte es una métrica para medir la similitud de los perfiles de disolución entre los medicamentos genéricos y su innovador de los cuales destacan los criterios y estándares establecidos para determinar la comparación entre estos perfiles, lo que a su vez influye en la aprobación y comercialización de medicamentos genéricos en el mercado (FDA, 2018)

En este contexto, la aplicación de un modelo de machine learning emerge como una solución potencial. La cual tiene la capacidad de transformar y modernizar la evaluación de bioequivalencia in vitro a través del aprendizaje automático el cual es capaz de clasificar perfiles de disolución en las condiciones de pH establecidas.

3.1.2. Comprensión de los datos

La fase de comprensión de los datos comienza con la recopilación inicial de datos y continúa con actividades que le permiten familiarizarse con los datos, identificar problemas de calidad, determinar conocimientos preliminares sobre los datos y/o descubrir subconjuntos interesantes para generar hipótesis sobre la configuración de los datos (Azevedo & Zantos, 2022).

En esta etapa se realiza un análisis exhaustivo de los datos recopilados en las pruebas de disolución in vitro tanto del medicamento innovador (Glucofage 500 mg) y el medicamento genérico (Metformina clorhidrato 500 mg) en los distintos niveles de pH (6,8 – 4,5 – 1,2) en donde se verificará el cumplimiento del factor de similitud y los patrones que sigue cada uno de ellos.

3.1.2.1.Recolección de los datos

Los resultados obtenidos de las pruebas de disolución se ingresarán dentro del paquete de Microsoft office, Excel, donde se estructura y clasifican las muestras por tiempos de disolución y pH previo al cálculo de f_2 , esto se realizará para todos los lotes manufacturados que ingresarán en el estudio, estos registros formarán parte de la base de datos final con la que se desarrollara el modelo de machine learning.

3.1.2.2.Descripción de los datos

Para desarrollar el modelo de segmentación se consideran variables tales como el peso de la tableta en (mg), la absorbancia, la concentración en (mg/mL), el factor de dilución y el porcentaje de disolución de Metformina clorhidrato, mismos que se encuentran clasificados para los seis tiempos de disolución (10, 15, 20, 30, 45 y 60) minutos para los niveles de pH establecidos, estos datos se recopilan de los análisis in vitro de los perfiles de disolución.

3.1.2.3.Exploración de los datos

Para el estudio de bioequivalencia, se utilizó Glucofage® 500 mg como producto de referencia, identificado por el lote M93224 y el registro sanitario número 22.221-1-02-05. Se recolectaron 12 muestras para cada uno de los seis tiempos de disolución, categorizado por pH (6,8 - 4,5 y 1,2), lo que resultó en un conjunto de 216 muestras analizadas para 10 variables, los cuales servirán como referencia para calcular el factor de similitud f_2 .

```
> head(Innovador)
# A tibble: 6 x 10
  Lote Tiempo TiempDisol pH Muestra Peso Absorbancia Concentracion FD Metformina
<chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 M93224 10 T_1 6.8 1 525. 0.333 0.00469 45000 42.3
2 M93225 10 T_1 6.8 2 526. 0.36 0.00504 45000 45.3
3 M93226 10 T_1 6.8 3 526. 0.35 0.00491 45000 44.2
4 M93227 10 T_1 6.8 4 533. 0.335 0.00472 45000 41.9
5 M93228 10 T_1 6.8 5 522. 0.371 0.00518 45000 46.9
6 M93229 10 T_1 6.8 6 524. 0.317 0.00448 45000 40.5
> dim(Innovador)
[1] 216 10
```

Ilustración 7. Estructura de base de datos del medicamento Innovador

Elaborado por: Criollo, 2023.

En el caso del medicamento genérico, Metformina clorhidrato 500 mg, se obtuvieron datos de seis lotes manufacturados en el año 2022 y 2023 (220908, 220909, 220910, 230101, 230402 y 230503), de los cuales se recolectaron 12 muestras para cada uno de los seis tiempos de disolución, clasificados en los mismos niveles de pH, lo cual generó 1.296 muestras analizadas de igual forma para 10 variables. Cabe mencionar que el modelo de segmentación de los perfiles de disolución se realizara con estos datos obtenidos para continuar con el cálculo del f_2 , comparando así el medicamento innovador vs el genérico.

```
> head(Generico)
# A tibble: 6 x 10
  Lote Tiempo T_Clas pH Muestra Peso Absorbancia Concentracion FD Metformina
<chr> <dbl> <chr> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1 L-220908 10 T_1 6.8 1 722 0.242 0.00316 45000 28.5
2 L-220908 10 T_1 6.8 2 728 0.298 0.00384 45000 34.3
3 L-220908 10 T_1 6.8 3 735. 0.255 0.00332 45000 29.4
4 L-220908 10 T_1 6.8 4 718. 0.19 0.00253 45000 22.9
5 L-220908 10 T_1 6.8 5 702. 0.201 0.00266 45000 24.6
6 L-220908 10 T_1 6.8 6 727. 0.243 0.00317 45000 28.4
> dim(Generico)
[1] 1296 10
```

Ilustración 8. Estructura de base de datos del medicamento Genérico

Elaborado por: Criollo, 2023.

3.1.3. Preparación de los datos

La etapa de preparación de datos engloba todas las acciones requeridas para generar el conjunto de datos definitivo el cual será empleado en las herramientas de modelamiento a partir de los datos originales. Las labores comprenden la elección de tablas, sets de datos y características, además de la modificación y depuración de los datos para adecuarlos a las herramientas de modelado (Azevedo & Zantos, 2022).

En esta fase implica analizar y procesar los resultados obtenidos de los perfiles de disolución en los distintos niveles de pH, también y tendencias. Cada paso está cuidadosamente alineado con los objetivos del proyecto, asegurando que los datos sean coherentes con los criterios de bioequivalencia in vitro para desarrollar el modelo de segmentación y el cálculo del factor de similitud (f_2) entre el medicamento genérico (Metformina clorhidrato 500 mg) y su innovador (Glucofage® 500 mg).

3.1.3.1. Selección de los datos

Dentro de las muestras recopiladas tanto para el medicamento innovador como el genérico se tomaron las 10 variables de relevancia dentro del estudio de bioequivalencia in vitro las cuales se enlistan a continuación:

Tabla 1. Variables para el desarrollo del modelo de segmentación de perfiles de disolución

ID	Variable	Grupo
1	Lote	Identificación
2	Tiempo	Factor
3	TiempDisol	Factor
4	pH	Factor
5	Muestra	Factor
6	Peso	Numérica
7	Absorbancia	Numérica
8	Concentración	Numérica
9	FD	Numérica
10	Metformina	Numérica

Elaborado por: Criollo, 2023.

Las variables de la tabla 3 surgen del estudio de bioequivalencia las cuales se dan para cada uno de los 3 niveles de pH, cuya descripción es la siguiente:

- 1) Lote: identificación de cada lote manufacturado del cual se obtuvieron las muestras para el estudio.
- 2) Tiempo: clasificación de los tiempos (10, 15, 20, 30, 45 y 60) minutos para la disolución de Metformina clorhidrato.
- 3) TiempDisol: variable identificadora de cada uno de los tiempos de disolución de Metformina clorhidrato.

Tabla 2. Descripción de la variable TiempDisol

TiempDisol	Minuto
T_1	10
T_2	15
T_3	20
T_4	30
T_5	45
T_6	60

Elaborado por: Criollo, 2023.

- 4) pH: indica los medios de disolución con los que se trabajó, en este caso un nivel de pH de 6,8 – 4,5 y 1,2.
- 5) Muestra: indica el número de muestras con el que trabajo en cada tiempo de disolución.
- 6) Peso: el peso de la tableta de Metformina Clorhidrato expresado en miligramos (mg)
- 7) Absorbancia: la absorbancia monitorea la liberación del principio activo de una formulación farmacéutica con el paso del tiempo. Dentro de los perfiles de disolución describe cómo se disuelve y se libera el principio activo del medicamento en diferentes condiciones como el pH.
- 8) Concentración: es la concentración de la cantidad de Metformina clorhidrato que se ha disuelto en una solución por unidad de volumen, expresada en miligramos por mililitro (mg/mL).
- 9) FD: representa el factor de disolución el cual es una medida para evaluar cuánto de un medicamento se ha disuelto en un tiempo determinado indicando

si este se está disolviendo y liberando gradualmente evaluando que el medicamento se disuelva adecuadamente.

- 10) Metformina: representa el porcentaje de Metformina clorhidrato que se ha liberado o disuelto en el medio de disolución en un tiempo determinado midiendo así la eficacia de la disolución lo cual es relevante para determinar la bioequivalencia del medicamento genérico en comparación con el producto de referencia.

3.1.3.2.Limpieza de datos

Dentro de este estudio no fue necesario realizar una depuración de los datos ya que los resultados obtenidos del estudio de bioequivalencia invitro provienen de distintas pruebas de laboratorio en las cuales se analizaron 12 muestras para cada tiempo de disolución y nivel de pH, por lo cual no existen valores nulos o registros en blanco.

```
> summary(Innovador)
  Lote      Tiempo  TiempDisol      pH      Muestra      Peso
Length:216  Min.   :10      Length:216  Min.   :1.200  Min.   : 1.00  Min.   :522.0
Class :character 1st Qu.:15      Class :character 1st Qu.:1.200  1st Qu.: 3.75  1st Qu.:526.0
Mode  :character Median :25      Mode  :character Median :4.500  Median : 6.50  Median :528.8
                Mean  :30                Mean  :4.167  Mean  : 6.50  Mean  :528.9
                3rd Qu.:45                3rd Qu.:6.800  3rd Qu.: 9.25  3rd Qu.:530.1
                Max.  :60                Max.  :6.800  Max.  :12.00  Max.  :548.0

Absorbancia  Concentracion      FD      Metformina
Min.   :0.0700  Min.   :0.003274  Min.   :45000  Min.   : 29.55
1st Qu.:0.1635  1st Qu.:0.005495  1st Qu.:45000  1st Qu.: 49.46
Median :0.4285  Median :0.007167  Median :45000  Median : 64.26
Mean   :0.4260  Mean   :0.007346  Mean   :45000  Mean   : 66.12
3rd Qu.:0.6310  3rd Qu.:0.009068  3rd Qu.:45000  3rd Qu.: 82.01
Max.   :0.8790  Max.   :0.011448  Max.   :45000  Max.   :103.76

> colSums(is.na(Innovador))
  Lote      Tiempo  TiempDisol      pH      Muestra      Peso      Absorbancia
0         0         0         0         0         0         0
Concentracion      FD      Metformina
0         0         0
```

Ilustración 9. Salida de la base del medicamento Innovador para determinar valores nulos o faltantes

Elaborado por: Criollo, 2023.

```

> summary(Generico)
  Lote      Tiempo      T_Clas      pH      Muestra      Peso
Length:1296  Min.   :10  Length:1296  Min.   :1.200  Min.   : 1.00  Min.   :702.4
Class :character  1st Qu.:15  Class :character  1st Qu.:1.200  1st Qu.: 3.75  1st Qu.:722.5
Mode  :character  Median :25  Mode  :character  Median :4.500  Median : 6.50  Median :726.5
                                Mean  :30                                Mean  :4.167  Mean  : 6.50  Mean  :727.0
                                3rd Qu.:45                                3rd Qu.:6.800  3rd Qu.: 9.25  3rd Qu.:730.3
                                Max.  :60                                Max.  :6.800  Max.  :12.00  Max.  :764.3

Absorbancia  Concentración      FD      Metformina
Min.   :0.0340  Min.   :0.001502  Min.   :45000  Min.   : 13.41
1st Qu.:0.1690  1st Qu.:0.004869  1st Qu.:45000  1st Qu.: 43.73
Median :0.3760  Median :0.007736  Median :45000  Median : 69.84
Mean   :0.4283  Mean   :0.007453  Mean   :45000  Mean   : 67.09
3rd Qu.:0.7288  3rd Qu.:0.010199  3rd Qu.:45000  3rd Qu.: 91.83
Max.   :0.8980  Max.   :0.012620  Max.   :45000  Max.   :109.49

> colSums(is.na(Generico))
  Lote      Tiempo      T_Clas      pH      Muestra      Peso      Absorbancia
0         0         0         0         0         0         0
Concentración      FD      Metformina
0         0         0

```

Ilustración 10. Salida de la base del medicamento Genérico para determinar valores nulos o faltantes

Elaborado por: Criollo, 2023.

3.1.4. Modelado

3.1.4.1. Selección de la técnica

Debido a que el modelo random forest es conocido por su capacidad para abordar conjuntos de datos complejos y no lineales, además, por la naturaleza de las variables del estudio de bioequivalencia in vitro los cuales pueden presentar interacciones y relaciones no lineales, se seleccionó esta técnica para el desarrollo de los modelos de segmentación en los distintos niveles de pH (6,8 - 4,5 y 1,2) la resistencia al sobreajuste del bosque aleatorio es especialmente relevante para garantizar la generalización del modelo buscando que estos sean capaces de realizar predicciones acertadas para estudios posteriores

Dentro del machine learning es importante comprender la distinción entre los datos de entrenamiento y los datos de prueba. Los datos de entrenamiento son esenciales para instruir a un algoritmo de aprendizaje automático, mientras que los datos de prueba se utilizan para evaluar y validar la eficacia del modelo después del entrenamiento. Esta separación implica dividir el conjunto de datos en dos porciones: una dedicada al proceso de enseñanza y la otra reservada para comprobar qué tan bien ha asimilado el algoritmo (Khan Academy, 2020).

Para la selección de los datos de entrenamiento y los datos de prueba se realizará una partición del 80 % y el 20 % respectivamente.

3.1.4.2. Preparación del modelo filtrado por nivel de Ph

Dentro de la base de datos de medicamentos “Genérico” es necesario hacer un filtro para separarlos por niveles de pH ya que se va a realizar un modelo de segmentación para cada uno de ellos debido a que cada pH simula distintas secciones del ambiente gastrointestinal:

pH 6.8: yeyuno-íleon (intestino delgado)

pH 4.5: duodeno (intestino delgado)

pH 1.2: estómago

```
# Filtro por nivel de pH
Generico_6.8 <- subset(Generico, pH == 6.8)
Generico_4.5 <- subset(Generico, pH == 4.5)
Generico_1.2 <- subset(Generico, pH == 1.2)
```

Ilustración 11. Segmentación por nivel de pH

Elaborado por: Criollo, 2023.

Por lo cual el porcentaje de disolución del Metformina Clorhidrato varía según el nivel de pH en el que se encuentre. Para comprender estas diferencias, es necesario dividir y analizar los datos en subconjuntos específicos en este caso por nivel de pH ya que este enfoque permitirá entender cómo el medicamento se comporta en distintas condiciones, lo que es decisivo al momento de garantizar la eficacia del fármaco.

Además de este análisis por niveles de pH, es necesario evaluar cómo cambian los tiempos de disolución en estas condiciones ya que ayudara a descubrir tendencias entre el pH y el tiempo de disolución para determinar el factor de similitud.

3.1.4.3. Construcción del modelo para un nivel de pH de 6,8

Con el dataframe creado para un nivel de pH de 6,8 se procede a realizar la partición de la data para prueba y entrenamiento del modelo a desarrollar utilizando el software R studio utilizando el criterio 80 % y 20 % debido a que los datos con los que se cuentan no requieren de un tratamiento previo ya que son resultados obtenidos de

análisis de laboratorio y contrastados con un patrón de referencia. El subconjunto de entrenamiento se utiliza para estimar los parámetros del modelo, mientras que el conjunto de prueba verifica cómo se comporta el modelo estimado esto con el objetivo de evitar un sobreajuste en el desarrollo del modelo (Gil, 2022)

```
##### MODELO para pH 6.8
set.seed(123) # Para reproducibilidad
train_6.8<- createDataPartition(Generico_6.8$Tiempo, p=0.8, list = F)

data_train6.8 <- Generico_6.8[train_6.8, ]
data_test6.8 <- Generico_6.8[-train_6.8, ]
```

Ilustración 12. Partición de la data para prueba y entrenamiento, pH 6,8

Elaborado por: Criollo, 2023.

Una vez obtenidos los datos de prueba y entrenamiento se desarrolla el modelo random forest tomando al “Tiempo” como variable de respuesta y al “Peso”, “Absorbancia”, “Concentración” y “Metformina” como variables explicativas dentro del modelo la variable que hace referencia al factor de dilución no resulto relevante ya que este resultado es un valor de referencia para todos los niveles de pH.

Adicional a esto se ejecutan los códigos correspondientes para obtener las predicciones en función de los datos de prueba para evaluar el rendimiento del modelo a un nivel de pH de 6,8.

```
# Crear el modelo Random Forest
modelo_6.8 <- randomForest(Tiempo ~ Peso + Absorbancia + Concentración + Metformina, data = data_train6.8)

# Hacer predicciones en el conjunto de prueba
predicciones_6.8 <- predict(modelo_6.8, data_test6.8)

# Calcular la matriz de confusión y otras métricas de evaluación
precision_6.8 <- confusionMatrix(predicciones_6.8, data_test6.8$Tiempo)
print(precision_6.8)
```

Ilustración 13. Desarrollo del modelo de segmentación para un pH de 6,8

Elaborado por: Criollo, 2023.

3.1.4.4. Construcción del modelo para un nivel de pH de 4,5

De igual forma utilizando el mismo criterio del modelo a un nivel de pH de 6,8, se realiza la partición de la información en datos de 80 % para pruebas y 20 % para entrenamiento.

```
##### MODELO para pH 4.5
set.seed(123) # Para reproducibilidad
train_4.5<- createDataPartition(Generico_4.5$Tiempo, p=0.8, list = F)

data_train4.5 <- Generico_4.5[train_4.5, ]
data_test4.5 <- Generico_4.5[-train_4.5, ]
```

Ilustración 14. Partición de la data para prueba y entrenamiento, pH 4,5

Elaborado por: Criollo, 2023.

Una vez obtenidos los datos de prueba y entrenamiento se desarrolla el modelo random forest tomando al “Tiempo” como variable de respuesta y al “Peso”, “Absorbancia”, “Concentración” y “Metformina” como variables explicativas dentro del modelo la variable que hace referencia al factor de dilución no resulto relevante ya que este resultado es un valor de referencia para todos los niveles de pH.

Adicional a esto se ejecutan los códigos correspondientes para obtener las predicciones en función de los datos de prueba para evaluar el rendimiento del modelo a un nivel de pH de 4,5.

```
# Crear el modelo Random Forest
modelo_4.5 <- randomForest(Tiempo ~ Peso + Absorbancia + Concentracion + Metformina, data = data_train4.5)

# Hacer predicciones en el conjunto de prueba
predicciones_4.5 <- predict(modelo_4.5, data_test4.5)

# Calcular la matriz de confusión y otras métricas de evaluación
precision_4.5 <- confusionMatrix(predicciones_4.5, data_test4.5$Tiempo)
print(precision_4.5)
```

Ilustración 15. Desarrollo del modelo de segmentación para un pH de 4,5

Elaborado por: Criollo, 2023.

3.1.4.5. Construcción del modelo para un nivel de pH de 1,2

De igual forma utilizando el mismo criterio del modelo a un nivel de pH de 6,8 – 4,5 y 1,2 se realiza la partición de la información en datos de 80 % para pruebas y 20 % para entrenamiento.

```
##### MODELO para pH 1.2
set.seed(123) # Para reproducibilidad
train_1.2<- createDataPartition(Generico_1.2$Tiempo, p=0.8, list = F)

data_train1.2 <- Generico_1.2[train_1.2, ]
data_test1.2 <- Generico_1.2[-train_1.2, ]
```

Ilustración 16. Partición de la data para prueba y entrenamiento, pH 1,2

Elaborado por: Criollo, 2023.

Una vez obtenidos los datos de prueba y entrenamiento se desarrolla el modelo random forest tomando al “Tiempo” como variable de respuesta y al “Peso”, “Absorbancia”, “Concentración” y “Metformina” como variables explicativas dentro del modelo la variable que hace referencia al factor de dilución no resulto relevante ya que este resultado es un valor de referencia para todos los niveles de pH.

Adicional a esto se ejecutan los códigos correspondientes para obtener las predicciones en función de los datos de prueba para evaluar el rendimiento del modelo a un nivel de pH de 1,2.

```
# Crear el modelo Random Forest
modelo_1.2 <- randomForest(Tiempo ~ Peso + Absorbancia + Concentracion + Metformina, data = data_train1.2)

# Hacer predicciones en el conjunto de prueba
predicciones_1.2 <- predict(modelo_1.2, data_test1.2)

# Calcular la matriz de confusión y otras métricas de evaluación
precision_1.2 <- confusionMatrix(predicciones_1.2, data_test1.2$Tiempo)
print(precision_1.2)
```

Ilustración 17. Desarrollo del modelo de segmentación para un pH de 1,2

Elaborado por: Criollo, 2023.

3.1.5. Evaluación

Dentro de CRISP-DM es esencial validar la efectividad de los modelos desarrollados, en este caso se evalúa cada uno de los modelos de segmentación creados para los niveles de pH (6,8 - 4,5 - 1,2), esta fase implica una interpretación detallada de los resultados en relación con los objetivos del proyecto, validando así los resultados obtenidos.

En esta fase, se analiza cómo los modelos respondieron a los datos de prueba evaluando su capacidad para generar resultados coherentes con el estudio de bioequivalencia in vitro. En el caso de que alguno de los modelos no alcance las expectativas establecidas, se procederá a identificarlos campos de mejora que sean necesario, para lo cual es importante ejecutar cambios tales como, la eliminación de atributos estadísticamente insignificantes, corrección en la entrada de datos o ajustes en el tratamiento de los atributos, entre otras medidas (Caio, 2022). En última instancia, esta etapa no solo consolida la robustez de los modelos, sino que también proporciona una base sólida para su aplicación práctica y significativa en el campo farmacéutico.

3.1.5.1. Evaluación del modelo para un nivel de pH de 6,8

La matriz de confusión muestra cómo se comparan las predicciones del modelo con los valores reales. Las filas representan las predicciones del modelo, y las columnas representan los valores reales.

```
> print(precision_6.8)
Confusion Matrix and Statistics

      Reference
Prediction 10 15 20 30 45 60
10      14   1   0   0   0   0
15       0   9   0   0   0   0
20       0   4  14   2   0   0
30       0   0   0  12   2   0
45       0   0   0   0   6   2
60       0   0   0   0   6  12

Overall Statistics

      Accuracy : 0.7976
      95% CI   : (0.6959, 0.8775)
      No Information Rate : 0.1667
      P-Value [Acc > NIR] : < 2.2e-16

      Kappa   : 0.7571

      McNemar's Test P-Value : NA
```

Ilustración 18. Evaluación del rendimiento del modelo a un nivel de pH 6.8

Elaborado por: Criollo, 2023.

Una vez realizada la evaluación, el modelo demostró una alta precisión en los distintos intervalos de tiempo, destacando especialmente en el tiempo de disolución a 15 minutos, donde acertó en todas las predicciones. Sin embargo, se presentaron desafíos en las predicciones de los tiempos de 10, 20, 30, 45 y 60 minutos, con algunas clasificaciones erróneas en estas categorías.

Específicamente, el modelo cometió errores notables al clasificar incorrectamente algunas instancias de 20 minutos como 15 minutos y otras como 30 minutos. Además, se observaron confusiones en la predicción de tiempos de 30 minutos, asignando erróneamente dos casos a la categoría de 45 minutos, de igual forma en el minuto 45 el modelo tiende a confundirlo en una categoría de 60 minutos y finalmente al minuto 60 el modelo asignó equivocadamente predicciones en el minuto 45.

A pesar de estos desafíos, la precisión general del modelo se mantiene en un nivel aceptable del 79,76 %. Este valor indica que aproximadamente el 79,76 % de las predicciones del modelo son correctas. Además, el índice Kappa de 0,7571 sugiere una buena concordancia entre las predicciones del modelo y las clases reales, corrigiendo la posibilidad de coincidencias por azar. Estos resultados reflejan un modelo que, aunque presenta ciertas dificultades, tiene un rendimiento generalmente confiable en la predicción de los tiempos de disolución.

La matriz de estadísticas por clase permite comprender cómo se está comportando el modelo para cada clase y proporciona una visión más detallada del comportamiento del modelo.

```

McNemar's Test P-Value : NA

Statistics by Class:

```

	Class: 10	Class: 15	Class: 20	Class: 30	Class: 45	Class: 60
Sensitivity	1.0000	0.6429	1.0000	0.8571	0.42857	0.8571
Specificity	0.9857	1.0000	0.9143	0.9714	0.97143	0.9143
Pos Pred Value	0.9333	1.0000	0.7000	0.8571	0.75000	0.6667
Neg Pred Value	1.0000	0.9333	1.0000	0.9714	0.89474	0.9697
Prevalence	0.1667	0.1667	0.1667	0.1667	0.16667	0.1667
Detection Rate	0.1667	0.1071	0.1667	0.1429	0.07143	0.1429
Detection Prevalence	0.1786	0.1071	0.2381	0.1667	0.09524	0.2143
Balanced Accuracy	0.9929	0.8214	0.9571	0.9143	0.70000	0.8857

Ilustración 19. Matriz de estadísticas por clase, pH 6,8

Elaborado por: Criollo, 2023.

La evaluación de la Sensibilidad y Especificidad proporciona una visión completa de cómo el modelo aborda las distintas clases. La alta Sensibilidad para la clase perteneciente al tiempo de disolución al minuto 10 y al 20 indica que el modelo es altamente competente para identificar correctamente los casos reales de 10 y 20. Sin embargo, la Sensibilidad más baja para el tiempo de disolución del minuto 45 con un valor de 0,42857 el cual nos indica que el modelo tiene una dificultad para predecir con precisión los valores del tiempo de disolución al minuto 45.

En cuanto a la especificidad, se destaca la habilidad del modelo para identificar correctamente los verdaderos negativos. Las altas Especificidades en todos los tiempos de disolución indican que el modelo es efectivo al discriminar los casos que no pertenecen

a estas clases. Esto sugiere una capacidad robusta del modelo para clasificar adecuadamente los valores que no son de estas clases.

3.1.5.2. Evaluación del modelo para un nivel de pH de 4,5

La matriz de confusión para el modelo de segmentación a un nivel de pH 4,5. Donde las filas representan las predicciones del modelo, y las columnas representan los valores reales.

```
> print(precision_4.5)
Confusion Matrix and Statistics

      Reference
Prediction 10 15 20 30 45 60
10 13  1  0  0  0  0
15  0 12  1  0  0  0
20  1  1 12  0  0  0
30  0  0  1 12  3  0
45  0  0  0  2  7  1
60  0  0  0  0  0  4 13

Overall Statistics

      Accuracy : 0.8214
      95% CI   : (0.7226, 0.8965)
No Information Rate : 0.1667
P-Value [Acc > NIR] : < 2.2e-16

      Kappa : 0.7857

McNemar's Test P-Value : NA
```

Ilustración 20. Evaluación del rendimiento del modelo a un nivel de pH 4,5

Elaborado por: Criollo, 2023.

En la evaluación global del modelo para predecir los tiempos de disolución, se evidencia un rendimiento sólido donde el modelo presentó resultados que destacan en diversos intervalos de tiempo de disolución, especialmente en los tiempos de 10 minutos y 15 minutos, donde las predicciones fueron acertadas a excepción de uno en cada segmento, sin embargo el modelo enfrentó a ciertos desafíos en las predicciones de los tiempos de 20, 30, 45 y 60 minutos, con algunas clasificaciones incorrectas en estas categorías.

A pesar de estos desafíos, la precisión general del modelo se mantiene en un nivel aceptable del 82 %. Este valor indica que aproximadamente el 82 % de las predicciones del modelo son correctas. Además, el índice Kappa de 0,7857 sugiere una buena

concordancia entre las predicciones del modelo y las clases reales, corrigiendo la posibilidad de coincidencias por azar.

La matriz de estadísticas por clase para el modelo de segmentación de un nivel de pH del 4,5.

Statistics by Class:						
	Class: 10	Class: 15	Class: 20	Class: 30	Class: 45	Class: 60
Sensitivity	0.9286	0.8571	0.8571	0.8571	0.50000	0.9286
Specificity	0.9857	0.9857	0.9714	0.9429	0.95714	0.9429
Pos Pred Value	0.9286	0.9231	0.8571	0.7500	0.70000	0.7647
Neg Pred Value	0.9857	0.9718	0.9714	0.9706	0.90541	0.9851
Prevalence	0.1667	0.1667	0.1667	0.1667	0.16667	0.1667
Detection Rate	0.1548	0.1429	0.1429	0.1429	0.08333	0.1548
Detection Prevalence	0.1667	0.1548	0.1667	0.1905	0.11905	0.2024
Balanced Accuracy	0.9571	0.9214	0.9143	0.9000	0.72857	0.9357

Ilustración 21. Matriz de estadísticas por clase, pH 4,5

Elaborado por: Criollo, 2023.

La evaluación de la sensibilidad y especificidad de este modelo destacada la sensibilidad para el tiempo de disolución al minuto 10 y 60 (92,86 %) indica que el modelo es altamente competente para identificar con precisión los casos reales de esta clase, sin embargo la sensibilidad más baja para la clase es al minuto 45 (50 %) revela ciertas dificultades del modelo al predecir con exactitud esta categoría temporal específica, sugiriendo posibles áreas de mejora en la capacidad predictiva del modelo para esta clase.

En cuanto a la especificidad, se subraya la habilidad del modelo para identificar correctamente los verdaderos negativos. Las elevadas Especificidades en todos los tiempos de disolución indican que el modelo es eficaz al discriminar los casos que no pertenecen a estas clases, fortaleciendo la confianza en las predicciones negativas. Sin embargo, se observa una especificidad ligeramente menor para el tiempo de disolución al minuto 30 (94,29 %).

3.1.5.3. Evaluación del modelo para un nivel de pH de 1,2

La matriz de confusión para el modelo de segmentación a un nivel de pH 1,2. Donde las filas representan las predicciones del modelo, y las columnas representan los valores reales.

```
> print(precision_1.2)
Confusion Matrix and Statistics

      Reference
Prediction 10 15 20 30 45 60
 10 13  1  0  0  0  0
 15  1 13  0  0  0  0
 20  0  0 12  2  0  0
 30  0  0  2 11  5  0
 45  0  0  0  0  7  2
 60  0  0  0  0  1  2 12

Overall Statistics

      Accuracy : 0.8095
      95% CI   : (0.7092, 0.887)
  No Information Rate : 0.1667
  P-value [Acc > NIR] : < 2.2e-16

      Kappa : 0.7714

McNemar's Test P-value : NA
```

Ilustración 22. Evaluación del rendimiento del modelo a un nivel de pH 1,2

Elaborado por: Criollo, 2023.

En la evaluación del modelo para clasificar los tiempos de disolución se destaca un rendimiento sólido ya que se logró altos niveles de precisión en varios intervalos de tiempo, particularmente en los momentos de 10 minutos y 15 minutos, donde todas las predicciones fueron acertadas. Sin embargo, se identificaron desafíos en la predicción de los tiempos de 20, 30, 45 y 60 minutos.

Por otro lado, el modelo cometió errores notables al clasificar algunas instancias de 20 minutos como 30 minutos, mostrando una confusión en estos tiempos de disolución, en el caso de los tiempos de 45 minutos, el modelo no acertó en dos predicciones finalmente en el caso de la disolución al minuto 60 el modelo presento fallos al clasificar erróneamente en el minuto 30 y 45 respectivamente.

A pesar de estos desafíos, la precisión general del modelo alcanzo un nivel del 80,95 %. Este valor indica que aproximadamente el 80,95% de las predicciones del modelo son correctas. El índice Kappa de 0,7714 también sugiere una buena concordancia entre las predicciones del modelo y las clases reales, corrigiendo la posibilidad de coincidencias por azar. Estos resultados confirman que el modelo, aunque enfrenta ciertas complejidades en la predicción de los tiempos de disolución, sigue siendo confiable en su rendimiento general.

La matriz de estadísticas por clase para el modelo de segmentación de un nivel de pH del 1,2.

Statistics by class:						
	Class: 10	Class: 15	Class: 20	Class: 30	Class: 45	Class: 60
Sensitivity	0.9286	0.9286	0.8571	0.7857	0.50000	0.8571
Specificity	0.9857	0.9857	0.9714	0.9000	0.97143	0.9571
Pos Pred Value	0.9286	0.9286	0.8571	0.6111	0.77778	0.8000
Neg Pred Value	0.9857	0.9857	0.9714	0.9545	0.90667	0.9710
Prevalence	0.1667	0.1667	0.1667	0.1667	0.16667	0.1667
Detection Rate	0.1548	0.1548	0.1429	0.1310	0.08333	0.1429
Detection Prevalence	0.1667	0.1667	0.1667	0.2143	0.10714	0.1786
Balanced Accuracy	0.9571	0.9571	0.9143	0.8429	0.73571	0.9071

Ilustración 23. Matriz de estadísticas por clase, pH 1,2

Elaborado por: Criollo, 2023.

La sensibilidad que evalúa la capacidad del modelo para identificar correctamente los casos positivos revela un rendimiento sólido, especialmente para los tiempos de disolución 10, 15, y 60, todas superando el 85%. Este resultado subraya la eficiencia del modelo al predecir con precisión los eventos temporales asociados con estas clases, destacando su capacidad en situaciones positivas.

En términos de especificidad, que indica la capacidad del modelo para identificar correctamente los casos negativos, se observa una alta especificidad en general. El modelo demuestra una destacada habilidad para discernir con precisión los valores que no pertenecen a las clases "10", "15", "20", y "45".

3.1.6. Implementación

La implementación de estos modelos se realizará después de tener la aprobación de la jefatura de Aseguramiento de Calidad y tienen como objetivo fundamental la optimización de los estudios de bioequivalencia in vitro, especialmente en la comparación entre medicamentos genéricos y de referencia. Con lo cual se busca una reducción en el número de muestras, la cantidad de materiales utilizados y el tiempo que toma realizar en laboratorio análisis de este tipo. Este enfoque no solo permitirá calcular de manera precisa e inmediata el factor de similitud, sino también facilitará la emisión oportuna de criterios y decisiones para evaluar la calidad de los medicamentos. Además, posibilitará la repetición de análisis en etapas específicas, la toma de nuevas muestras y la verificación de resultados, incluso entre diferentes analistas.

El estudio piloto se centra en la comparación entre Metformina Clorhidrato 500 mg y Glucophage® 500 mg, sirviendo como referencia para futuros análisis que involucren diferentes principios activos y formas farmacéuticas. Las áreas clave de impacto de este proyecto abarcan Producción, Control de Calidad, Validaciones e Investigación y Desarrollo. Estos departamentos desempeñan un papel crucial al generar análisis y llevar a cabo estudios para garantizar la calidad de los medicamentos, utilizando la información generada a partir de los estudios de bioequivalencia.

El manejo y desarrollo de los modelos estarán a cargo del área de estadística, los cuales se encargarán del análisis, desarrollo e implementación de los mismos. Se prevé la replicación del estudio para diversos medicamentos solicitados por ARCSA, utilizando algoritmos de R Studio para calcular de manera automática y rápida el factor de similitud, cabe recalcar que es importante continuar con procesos de validación a medida que se desarrollan e implementan más modelos, asegurando la precisión de los resultados garantizando así la calidad de los medicamentos genéricos.

Una vez generados los modelos es indispensable la comunicación y capacitación del personal involucrado para garantizar una interacción adecuada con los resultados del modelo, desde la comprensión de los datos utilizados hasta la interpretación de los resultados obtenidos, mismos que servirán para futuras referencias y auditorías, fortaleciendo la confianza y robustez del sistema implementado.

CAPITULO IV: RESULTADOS

4. Resultados

4.1. Resultados de los modelos obtenidos

De los modelos obtenidos se llevó a cabo un análisis exhaustivo enfocado en su rendimiento y capacidad predictiva mediante la técnica Random Forest para cada nivel de pH (6,8 – 4,5 – 1,2), esta técnica se utilizó debido a su idoneidad para obtener de manera eficaz las relaciones no lineales y manejar variables predictoras sin necesidad de una preparación exhaustiva de datos lo cual es importante en resultados de laboratorio los cuales tienen ciertas limitaciones para su manipulación o transformación.

Tabla 3. Resultados del accuracy

pH	Accuracy
6,8	0,7976
4,5	0,8214
1,2	0,8095

Elaborado por: Criollo, 2023.

Los resultados obtenidos para cada uno de los niveles de pH reflejan la precisión de los modelos demostrando la proporción de predicciones correctas.

En el modelo pH 6,8, este alcanzó una precisión del 79.76%, indicando que alrededor del 79.76% de las predicciones fueron correctas en este nivel de pH.

A un nivel de pH 4,5, la precisión aumentó a 82.14%, sugiriendo un rendimiento ligeramente superior en comparación con pH 6,8.

En pH 1,2, la precisión fue del 80.95%, demostrando un desempeño sólido en la predicción de resultados en este pH.

En vista de los resultados obtenidos se puede decir que los modelos tienen un buen desempeño en la predicción de resultados para los tres niveles de pH, con una leve inclinación hacia una mayor precisión en modelo de pH 4,5, seguido por el modelo de pH 1,2 y finalmente por el de 6,8.

Tabla 4. Resultados de la sensibilidad y especificidad

pH	Tiempo (min)	Sensitivity	specificity
6,8	10	1,00	0,99
	15	0,64	1,00
	20	1,00	0,91
	30	0,86	0,97
	45	0,43	0,97
	60	0,86	0,91
4,5	10	0,93	0,99
	15	0,86	0,99
	20	0,86	0,97
	30	0,86	0,94
	45	0,50	0,96
	60	0,93	0,94
1,2	10	0,93	0,99
	15	0,93	0,99
	20	0,86	0,97
	30	0,79	0,90
	45	0,50	0,97
	60	0,86	0,96

Elaborado por: Criollo, 2023.

La tabla 4 muestra los resultados sobre la sensibilidad y especificidad de los modelos en función de diferentes niveles de pH y tiempos de disolución. Estos parámetros son esenciales para evaluar el rendimiento de los modelos en la identificación de resultados positivos y negativos, respectivamente.

Para el pH 6,8, se observa una alta sensibilidad alcanzando el valor de 1 en la mayoría de los tiempos de exposición, indicando la capacidad del modelo para detectar positivos verdaderos. Sin embargo, a los 15 minutos, la sensibilidad disminuye significativamente a 0,64, lo que puede señalar una disminución en la capacidad del modelo para identificar resultados positivos a este tiempo específico. La especificidad,

por otro lado, se mantiene generalmente alta, indicando una capacidad constante para evitar falsos positivos.

En el pH 4,5, se destaca una sensibilidad consistente en todos los tiempos, alcanzando un máximo de 0,93; sugiriendo una buena capacidad del modelo para identificar resultados positivos. La especificidad también es alta, indicando una habilidad para evitar falsos positivos.

En el pH 1,2, la sensibilidad es relativamente estable en la mayoría de los tiempos, alcanzando un máximo de 0,93; con algunas disminuciones notables a los 30 y 45 minutos. La especificidad muestra una tendencia a disminuir a medida que aumenta el tiempo de exposición.

Los modelos desarrollados, basados en la técnica Random Forest y aplicados a diferentes niveles de pH y tiempos de disolución, han demostrado un rendimiento general positivo en la predicción de resultados. La elección de este modelo se revela acertada debido a su capacidad para gestionar la complejidad de los datos, capturar relaciones no lineales y proporcionar resultados robustos.

En términos de precisión (accuracy), se observa que el modelo muestra un desempeño consistente en los tres niveles de pH evaluados, con una ligera tendencia a ser más preciso en el pH 4,5. Esto sugiere que los modelos son efectivos en la predicción de resultados a diferentes condiciones de pH, en el análisis de sensibilidad y especificidad revela que, en general, los modelos son capaces de identificar positivos verdaderos de manera efectiva. Sin embargo, se destacan algunas limitaciones, como la disminución de la sensibilidad a los 15 minutos en el pH 6,8 y las fluctuaciones en la sensibilidad y especificidad a medida que aumenta el tiempo de exposición en el pH 1,2, resultados que proporcionan una base sólida para futuras investigaciones y aplicaciones prácticas en el contexto de estudios de bioequivalencia in vitro.

4.2. Resultados del factor de similitud

El factor de similitud resulto ser un indicador esencial para garantizar la equivalencia terapéutica, ya que refleja la capacidad del medicamento genérico para disolverse y liberar el principio activo de manera similar al medicamento innovador, las curvas de disolución también son un indicador clave para determinar la bioequivalencia de un medicamento genérico, asegurando que este se disuelva de manera adecuada y libere el principio activo de manera comparable al medicamento innovador.

4.2.1. Curvas de disolución de los medicamentos genérico e innovador, pH 6,8

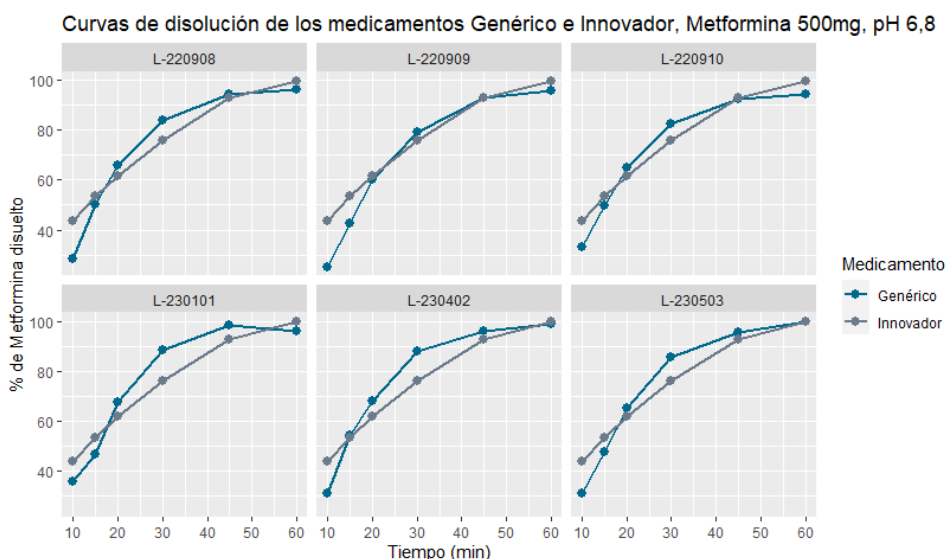


Ilustración 24. Curvas de disolución Genérico vs Innovador, pH 6,8

Elaborado por: Criollo, 2023.

En la gráfica se puede evidenciar que las curvas del medicamento genérico a un nivel de pH de 6,8 son muy similares en comparación a la curva del medicamento innovador o de referencia en cada uno de los lotes de los cuales se analizaron las muestras tanto en tiempo y en porcentaje de disolución por lo cual se podría decir que dentro de un ambiente ligeramente alcalino existe una disolución adecuada del medicamento dentro del tiempo establecido.

4.2.2. Curvas de disolución de los medicamentos genérico e innovador, pH 4,5

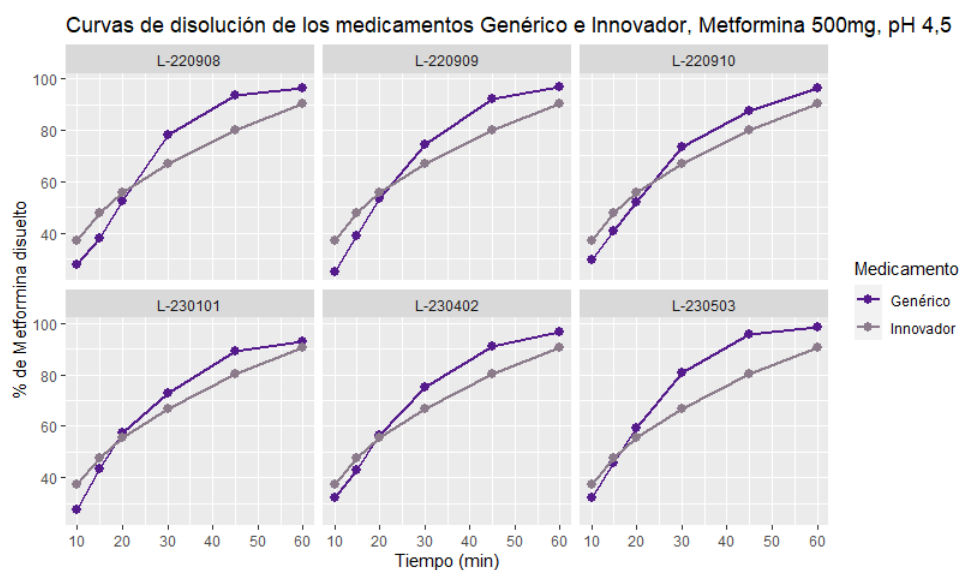


Ilustración 25. Curvas de disolución Genérico vs Innovador, pH 4,5

Elaborado por: Criollo, 2023.

En la gráfica se puede evidenciar que las curvas del medicamento genérico a un nivel de pH de 4,5 son muy similares en comparación a la curva del medicamento innovador o de referencia en cada uno de los lotes de los cuales se analizaron las muestras tanto en tiempo y en porcentaje de disolución, además se puede observar que a partir del minuto 30 la disolución va en aumento en relación a la curva del innovador, por tanto se podría decir que a este nivel de pH ligeramente ácido permitiendo una rápida y eficiente disolución dentro del tiempo establecido.

4.2.3. Curvas de disolución de los medicamentos genérico e innovador, pH 1,2

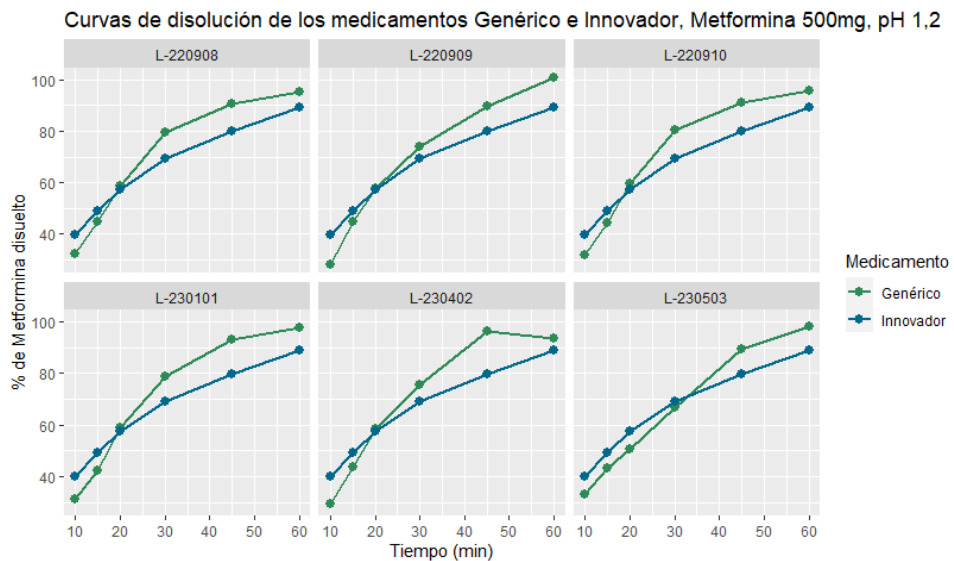


Ilustración 26. Curvas de disolución Genérico vs Innovador, pH 1,2

Elaborado por: Criollo, 2023.

En la gráfica se puede evidenciar que las curvas del medicamento genérico a un nivel de pH de 1,2 son muy similares en comparación a la curva del medicamento innovador o de referencia en cada uno de los lotes de los cuales se analizaron las muestras tanto en tiempo y en porcentaje de disolución, además se puede observar que a partir del minuto 30 la disolución va en aumento en relación a la curva del innovador, por tanto se podría decir que a este nivel de pH ácido el porcentaje de disolución el medicamento genérico tiene una velocidad de disolución relativamente mayor en relación con el medicamento de referencia por tanto, este es altamente soluble, lo que permite una rápida y eficiente disolución dentro de los tiempo establecidos.

4.2.4. Cálculo del factor de similitud por lote para un nivel de pH de 6,8 – 4,5 y 1,2

	Lote	F2_6.8	F2_4.5	F2_1.2
1	L-220908	55.99	51.09	56.10
2	L-220909	52.20	52.72	54.06
3	L-220910	61.40	59.59	54.59
4	L-230101	55.86	59.92	52.69
5	L-230402	55.47	58.33	52.32
6	L-230503	56.97	51.01	56.98

Ilustración 27. Resultados del factor de similitud Lote vs pH

Elaborado por: Criollo, 2023.

Tras calcular el factor de similitud (F2) para cada nivel de pH, se observa que todos los lotes cumplieron con el criterio establecido para la bioequivalencia in vitro en donde, el medicamento genérico Metformina 500 mg tiene una similitud y comparabilidad significativas con su innovador, Glucofage®, ya que todos los resultados superan el valor de 50.

Es importante destacar que, aunque todos los lotes cumplen con el F2, existen variaciones entre ellos. Por ejemplo, en los lotes 220908 y 230503, a un pH de 4.5, se obtuvieron resultados más bajos, con valores de 51,09 y 51,01 respectivamente. Estos valores, aunque ligeramente más bajos, siguen indicando una similitud moderada entre las curvas de disolución del medicamento genérico y del innovador en este pH.

CAPITULO V: CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Dentro de la investigación para determinar modelos de segmentación adecuados para clasificar los perfiles de disolución de Metformina Clorhidrato 500 mg se utilizó la técnica Random Forest la cual se evaluó para demostrar la aplicabilidad de los modelos, además se realizó el cálculo del factor de similitud concluyendo que:

- Las variables, "Peso", "Absorbancia", "Concentración" y "Porcentaje de liberación (Metformina)" fueron de relevancia para predecir el "Tiempo" de disolución, mostrando un enfoque adecuado para el desarrollo de los modelos de segmentación utilizando la técnica Random Forest, para clasificar los perfiles de disolución de Metformina Clorhidrato 500 mg en los distintos niveles de pH (6,8, 4,5, 1,2).
- Los modelos de segmentación para distintos niveles de pH demostraron una exactitud aceptable en los modelos para, pH 6,8 (79,76%), pH 4,5 (82,14%), y pH 1,2 (80,95%). Aunque enfrentaron a desafíos específicos en la predicción de ciertos tiempos de disolución, como los momentos de 20 y 60 minutos en pH 6,8, o los tiempos de 45 y 60 minutos en pH 4,5, estos modelos demuestran su capacidad para segmentar eficazmente los perfiles de disolución de Metformina Clorhidrato 500 mg en los distintos niveles de pH.
- La evaluación de los modelos reveló un rendimiento sólido en cuanto a sensibilidad y especificidad demostrando eficiencia en identificar correctamente los casos positivos, destacando los tiempos de 10, 15 y 60 minutos, aunque se observan dificultades en la sensibilidad para el tiempo de 45 minutos, estas métricas demuestran la confiabilidad general de los modelos en la clasificación de los perfiles de disolución.

- Los resultados del factor de similitud indican que todos los lotes cumplen con el criterio de evaluación establecido para la bioequivalencia in vitro, (F2) mayor o igual a 50, demostrando así la equivalencia terapéutica del medicamento genérico Metformina Clorhidrato 500 mg con Glucofage®, validando su capacidad para disolverse y liberar el principio activo de manera similar en diversas condiciones de pH y tiempos.
- Las curvas de disolución de Metformina Clorhidrato 500 mg mostraron una semejanza notable con las del medicamento innovador, demostrando una disolución efectiva en el tiempo requerido sin embargo a un de pH 4,5, se observa un aumento en el porcentaje de disolución después del minuto 30 en el genérico en comparación con el innovador y para el pH 1,2, el genérico demuestra una velocidad de disolución relativamente superior después del minuto 30.

5.2.Recomendaciones

Con los resultados obtenidos dentro del proyecto de investigación se pueden emitir las siguientes recomendaciones:

- Considerar la posibilidad de realizar ajustes en los modelos de segmentación para mejorar la precisión en momentos específicos de disolución, como los observados en los tiempos de 20 y 60 minutos en pH 6,8, y los tiempos de 45 y 60 minutos en pH 4,5.
- Explorar la posibilidad para la inclusión de variables adicionales en los modelos de segmentación, considerando factores como temperatura, humedad o características específicas de los lotes, para evaluar su impacto en la precisión de los modelos y su capacidad para generalizar a diversas condiciones experimentales.

- Implementar medidas adicionales para garantizar la consistencia en la toma de muestras, el análisis y la interpretación de resultados estableciendo así protocolos específicos para la recolección de muestras en estudios de bioequivalencia in vitro, asegurando que los análisis sigan procedimientos estandarizados.
- Implementar un sistema de monitoreo continuo para evaluar el rendimiento de los modelos en el tiempo y en diferentes condiciones experimentales permitiendo identificar cualquier cambio en la eficacia de los modelos y realizar ajustes de ser necesario.
- En virtud de lo observado en las curvas de disolución se recomienda realizar revisiones en la formulación de Metformina Clorhidrato 500 mg para evaluar el porcentaje de disolución en los tiempos donde la curva se aleja de la tendencia que sigue el medicamento innovador.

Bibliografía

- Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA. (2021). *Normativa Sanitaria Bioequivalencia en Medicamentos Consumo Humano*. Obtenido de https://www.gob.ec/sites/default/files/regulations/2021-10/Documento_ARCSA-DE-015-2018-JCGO_Normativa_t%C3%A9cnica_sanitaria_criterios_requisitos_para_demostrar_bioequivalencia_biodisponibilidad_medicamentos_uso_humano.pdf
- Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA. (2023). *INSTRUCTIVO EXTERNO: CRITERIOS Y REQUISITOS PARA DEMOSTRAR BIOEQUIVALENCIA Y BIODISPONIBILIDAD, EN LOS MEDICAMENTOS DE USO Y CONSUMO HUMANO*. Obtenido de <https://www.controlsanitario.gob.ec/wp-content/uploads/downloads/2023/06/Instructivo-Externo-IE-B.3.2.1-MED-02-BIOEQUIVALENCIA.pdf>
- Agencia Nacional de Regulación, Control y Vigilancia Sanitaria - ARCSA. (s.f.). *Arcsa, una institución que mejora continuamente su servicio de atención al usuario*. Obtenido de <https://www.controlsanitario.gob.ec/arcsa-una-institucion-que-mejora-continuamente-su-servicio-de-atencion-al-usuario/>
- Aprendeia. (2020). *Aprendizaje no Supervisado*. Obtenido de <https://aprendeia.com/aprendizaje-no-supervisado-machine-learning/>
- Azevedo, A., & Zantos, M. (2022). *Metología CRIPS-DM*. Obtenido de <https://www.sngular.com/es/data-science-crisp-dm-metodologia/>
- Bristol-Myers Squibb Company. (2017). *Glucophage*. Obtenido de https://www.accessdata.fda.gov/drugsatfda_docs/label/2017/020357s037s039,021202s021s023lbl.pdf
- Caio, R. (2022). Obtenido de <https://blog.mbauspesalq.com/es/2022/05/31/crisp-dm-las-6-etapas-de-la-metodologia-del-futuro/>
- Calle, J., & Copete, C. (2016). *Estudio de la bioequivalencia in vitro de las tabletas de ciprofloxacino de 500 mg, elaboradas por dos laboratorios farmacéuticos del Ecuador, mediante el empleo de los perfiles de disolución. (Tesis de grado, Universidad de Guayaquil)*. Obtenido de <https://repositorio.ug.edu.ec/items/40313799-d758-4700-b50a-5b5eb1cc1218>

- Centro Europeo de Posgrado - CEUPE. (2020). *Aprendizaje por refuerzo: Concepto, características y ejemplo ceupe*. Obtenido de <https://www.ceupe.com/blog/aprendizaje-por-refuerzo.html>
- Domingo, J. (2018). *Los algoritmos supervisados nos llevan al personalized learning y a sus interfaces (construcción)*. Obtenido de <https://juandomingofarnos.wordpress.com/2018/11/03/los-algoritmos-supervisados-nos-llevan-al-personalized-learning-y-a-sus-interfaces-construccion/>
- Encalada, R., & Sánchez, S. (2018). *Equivalencia farmacéutica in vitro de metformina clorhidrato frente a Glucofage (Tesis de grado, Universidad de Cuenca)*. Obtenido de <http://dspace.ucuenca.edu.ec/handle/123456789/31443>
- Eurofarma. (2023). *¿Qué es la bioequivalencia en medicamentos?* Obtenido de <https://www.eurofarma.com.ar/articulos/que-es-la-bioequivalencia-en-medicamentos>
- FDA. (2018). *Medicamentos Genéricos: Preguntas y Respuestas*. Obtenido de <https://www.fda.gov/drugs/generic-drugs/medicamentos-genericos-preguntas-y-respuestas>
- FDA. (2018). *Pruebas de disolución de formas de dosificación oral sólidas de liberación inmediata*. Obtenido de [https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guia-para-la-industria-pruebas-de-disolucion-de-formas-de-dosificacion-oral-solidas-de-liberacion#:~:text=El%20factor%20de%20similitud%20\(f2,\(%25\)%20entre%20las%20dos%20curvas.](https://www.fda.gov/regulatory-information/search-fda-guidance-documents/guia-para-la-industria-pruebas-de-disolucion-de-formas-de-dosificacion-oral-solidas-de-liberacion#:~:text=El%20factor%20de%20similitud%20(f2,(%25)%20entre%20las%20dos%20curvas.)
- Gil, R. (2022). *Modelos de machine learning para clasificar la cartera en un fondo de pensiones*. (Tesis de grado, Universidad Santo Tomas). Obtenido de <https://repository.usta.edu.co/handle/11634/47294>
- IBM. (2020). *¿Qué es machine learning?* Obtenido de <https://www.ibm.com/mx-es/analytics/machine-learning>
- Inesdi Business Techschool. (2022). *¿Qué es el Aprendizaje supervisado en Machine Learning?* Obtenido de <https://www.inesdi.com/blog/aprendizaje-supervisado-machine-learning/>
- Instituto de Salud Pública de Chile. (2018). *Guía para optar a bioexención de estudios de biodisponibilidad comparativa*. Obtenido de <https://www.ispch.cl/sites/default/files/G-Biof%2002%20-%20Gu%C3%ADa%20para%20optar%20a%20bioexenci%C3%B3n%20de%20estudios%20de%20biodisponibilidad%20comparativa.pdf>

- Jung, H., Anda, G. d., Rubio, K., & Mayet, L. (2012). Impacto de los criterios de diferentes agencias regulatorias en el cálculo de f2. *Revista mexicana de ciencias farmacéuticas*, 43(3), 67-71. Obtenido de https://www.scielo.org.mx/scielo.php?pid=S1870-01952012000300007&script=sci_abstract
- Khan Academy. (2020). *Algoritmos de aprendizaje automático*. Obtenido de <https://ciberseguridad.com/guias/nuevas-tecnologias/machine-learning/datos-entrenamiento/>
- Le, J. (06 de 2022). *Metabolismo de los fármacos*. Obtenido de Manual MSD: <https://www.msmanuals.com/es-ec/professional/farmacolog%C3%ADa-cl%C3%ADnica/farmacocin%C3%A9tica/biodisponibilidad-de-los-f%C3%A1rmacos>
- Ledón, P. L., & Hernández, C. A. (2020). *Medicamentos de patente, genéricos intercambiables y similares*. Obtenido de ACCESS Medicina: <https://accessmedicina.mhmedical.com/content.aspx?bookid=1489§ionid=96952814>
- Metry, M., Abrahamsson, B., Cristofolletti, R., & Dressman, J. (2021). Biowaiver Monographs for Immediate Release Solid Oral Dosage Forms: Metformin Hydrochloride. *Journal of Pharmaceutical Sciences*, 110(4), 1513 - 1526. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S0022354921000241>
- Nixus. (2023). *Tipos, algoritmos y aplicaciones de aprendizaje supervisado*. Obtenido de <https://nixustechnologies.com/supervised-machine-learning/>
- Paz, K. M. (2017). *Estudio de Bioequivalencia "in vitro" de tres productos genéricos (tabletas) de liberación inmediata que contienen Metformina Clorhidrato 850 mg con el medicamento innovador comercializados en Perú. (Tesis de grado, Universidad Peruana Cayetano Heredia)*. Obtenido de <https://repositorio.upch.edu.pe/handle/20.500.12866/1023>
- Peréz, M., Orobio, Y., & Baena, Y. (2013). Estudio comparativo de la liberación in vitro de metformina, a partir de dos productos multifuente de liberación inmediata, comercializados en Colombia. *Revista Colombiana de Ciencias Químico-Farmacéuticas*, 42(2), 169-189.
- Pérez, M. (2013). *Estudio de bioequivalencia in vitro de dos formas farmacéuticas perorales multifuente de liberación inmediata con metformina como principio activo. (Tesis de maestría, Universidad Nacional de Colombia)*. Obtenido de <https://repositorio.unal.edu.co/handle/unal/20555>

- Polio, C. (2014). *Perfiles de Disolución*. Obtenido de <https://prezi.com/jis8mub8jqas/perfiles-de-disolucion/>
- Redondo, L. (2015). *Estudios de equivalencia in vitro de formulaciones genéricas. (Tesis de grado, Universidad Complutense Madrid)*. Obtenido de <https://docta.ucm.es/rest/api/core/bitstreams/56786698-b065-45b3-94ab-c5f32ea54786/content>
- Sánchez, S. (2018). *Sistema de clasificación biofarmacéutica en la solicitud de una bioexención. (Tesis de grado, Universidad Complutense Madrid)*. Obtenido de <https://docta.ucm.es/entities/publication/9de5678e-d1fd-4c9e-b708-94e396df82e5>
- Saravia, G., & Daza, M. (2020). Estudio de Bioequivalencia in vitro de comprimidos de liberación inmediata de Metformina de 850 mg comercializados en Bolivia. *Revista CON-CIENCIA*, 8(2), 1-14. Obtenido de http://www.scielo.org/bo/scielo.php?script=sci_arttext&pid=S2310-02652020000200006
- Subirán, R. (2018). *Bioxenciones. (Tesis de grado, Universidad Complutense de Madrid)*. Obtenido de <http://147.96.70.122/Web/TFG/TFG/Memoria/RAFAEL%20SUBIRAN%20ADRADO S.pdf>
- Texas Children's Hospital. (2016). *Metformina (Glucophage)*. Obtenido de <https://www.texaschildrens.org/sites/default/files/uploads/documents/diabetes/transition/Metformin%20-%205-31-2016%20-%20SP.pdf>
- Torres, A. (2023). *Metformina: qué es, para qué sirve, dosis y efectos secundarios*. Obtenido de Azsalud: <https://azsalud.com/medicamentos/metformina>
- Unilabs. (2023). *Principio Activo*. Obtenido de <https://www.unilabs.es/glosario/principio-activo>
- USP-NF. (2020). *Evaluación del Desempeño e Intercambiabilidad de Medicamentos Sólidos Orales—Biodisponibilidad, Bioequivalencia y Disolución*. doi:https://doi.org/10.31003/USPNF_M99809_04_02
- Vallalta, J. (2023). *CRISP-DM: una metodología para minería de datos en salud*. Obtenido de <https://healthdataminer.com/data-mining/crisp-dm-una-metodologia-para-mineria-de-datos-en-salud/>
- Velazquez, N. (2021). *Aprendizaje por Refuerzo*. Obtenido de <https://trustnet.com.mx/aprendizaje-por-refuerzo/>

