



UNIVERSIDAD JOSÉ ANTONIO PÁEZ

**APLICACIÓN WEB PARA DETERMINAR POSIBLES
CANDIDATOS A FÁRMACO UTILIZANDO REDES
NEURONALES**

Autor
Salas, Mariángeles

Urb. Yuma II, Calle N° 3, Municipio San Diego
Teléfono: (0241) 8714240 (Máster) - Fax: (0241) 8712394



**REPÚBLICA BOLIVARIANA DE VENEZUELA
UNIVERSIDAD JOSÉ ANTONIO PÁEZ
FACULTAD DE INGENIERÍA
ESCUELA DE COMPUTACIÓN**

**APLICACIÓN WEB PARA DETERMINAR POSIBLES CANDIDATOS A
FÁRMACO UTILIZANDO REDES NEURONALES**

Trabajo de grado presentado como requisito para optar al título de
INGENIERO EN COMPUTACIÓN

Autor:
Salas, Mariángeles
C.I. 25.754.059
Tutor Académico:
Ing. Maldonado, Mayerlin

San Diego, abril 2021



FI-C-001-2020-3CR (TG)

Valencia, 26 de marzo de 2021

Ciudadana:
Salas López, Mariangeles Constanza.
CI. 25.754.059
Presente-

Cumplo con informarle que la Comisión de Trabajo de Grado y Pasantías de la Facultad de Ingeniería en su reunión N° **05-2021** de fecha **22-01-2021** aprobó el proyecto de trabajo de grado titulado **APLICACIÓN WEB PARA DETERMINAR POSIBLES CANDIDATOS A FÁRMACO UTILIZANDO REDES NEURONALES** presentado por usted (es) como requisito para optar al título de Ingeniero en Computación.

Se ratifica la designación de la Ing. Mayerlin Maldonado C.I: 11.810.356 como Tutora Académica que la asesorara en el desarrollo de este proyecto.

Atentamente,

Dr. Francisco Gelanzé Sevilla.
Decano

c.c. Coordinación de Pasantías y Trabajo de Grado (1).

GF/aa



REPÚBLICA BOLIVARIANA DE VENEZUELA
UNIVERSIDAD JOSÉ ANTONIO PÁEZ
FACULTAD DE INGENIERÍA
ESCUELA DE COMPUTACIÓN

ACEPTACIÓN DEL TUTOR

Quien suscribe, Ing. Mayerlin Maldonado, portador(a) de la cédula de identidad N° 11.810.356, en mi carácter de tutor del trabajo de grado presentado por la ciudadana Salas López, Mariángeles Constanza, portadora de la cédula de identidad N° 25.754.059, titulado **“APLICACIÓN WEB PARA DETERMINAR POSIBLES CANDIDATOS A FÁRMACO UTILIZANDO REDES NEURONALES.”**, presentado como requisito parcial para optar al título Ingeniero de Computación, considero que dicho trabajo reúne los requisitos y méritos suficientes para ser sometido a la presentación pública y evaluación por parte del jurado examinador que se designe.

En San Diego, a los 02 días del mes de abril de dos mil veinte.

Ing. Mayerlin Maldonado

C.I.: 11.8y10.356

DEDICATORIA

El presente trabajo está dedicado a Dios, por darme la inspiración, la sabiduría y la fuerza para culminar esta etapa.

A mi mamá, por su amor y su entrega, por estar para mí en todo lo que he necesitado y apoyarme en cada paso que he dado para llegar hasta aquí.

A mi abuela Flor, por su amor hacia mí, porque sé que donde estés estas muy orgullosa de mí.

AGRADECIMIENTOS

A Dios, por guiarme y bendecirme en todo momento, ser la fortaleza en los momentos difíciles y darme la sabiduría para alcanzar este logro tan deseado.

A mi mamá, porque sin ella no hubiese sido posible esto, por confiar en mí y ser un apoyo incondicional en todo el camino recorrido para culminar la carrera.

A mis compañeros de estudios, por ser acompañantes en esta travesía, ser apoyo y guías en esta experiencia.

A la Ing. Mayerlin Maldonado, por ser una profesora dedicada y aceptar que yo realizara este trabajo bajo su dirección, orientándome no solo en las aulas si no también en el proceso de realización de este proyecto.

A la Universidad José Antonio Páez, por ser mi centro de estudio y permitirme formarme entre sus aulas, a pesar de todas las dificultades vividas.

A mi familia, por apoyarme y ayudarme en los momentos difíciles y por siempre estar para mí durante este viaje.

A mis amigos, por acompañarme y motivarme a lo largo de la carrera y en el desarrollo de este trabajo.

ÍNDICE GENERAL

Contenido

pp.

| | |
|---|------------|
| RESUMEN | xiv |
| INTRODUCCIÓN | 1 |
| CAPÍTULO I | 3 |
| EL PROBLEMA | 3 |
| 1.1. Planteamiento del problema | 3 |
| 1.2. Formulación del problema..... | 6 |
| 1.3. Objetivos de la investigación..... | 6 |
| 1.3.1. Objetivo general..... | 6 |
| 1.3.2. Objetivos específicos | 6 |
| 1.4. Justificación de la investigación..... | 7 |
| 1.5. Alcance..... | 8 |
| CAPÍTULO II | 9 |
| MARCO TEÓRICO | 9 |
| 2.1. Antecedentes de la investigación..... | 9 |
| 2.2. Bases Teóricas | 14 |
| 2.2.1. Alzheimer..... | 14 |
| 2.2.2. Análisis predictivo | 14 |
| 2.2.3. Aprendizaje automático | 15 |
| 2.2.4. Aplicación Web..... | 16 |
| 2.2.5. Ciencia de datos | 16 |
| 2.2.6. Coeficiente de reparto octanol-agua..... | 17 |
| 2.2.7. Concentración inhibitoria media máxima | 17 |
| 2.2.8. Dengue | 18 |
| 2.2.9. Diabetes..... | 18 |
| 2.2.10. Diana Terapéutica | 19 |
| 2.2.11. Hipertensión | 19 |
| 2.2.12. Manejador de base de datos | 20 |
| 2.2.13. Reglas de Lipinski..... | 20 |

| | |
|--|-----------|
| 2.2.14. SMILES | 21 |
| 2.3. Bases Legales | 24 |
| 2.4. Definición de términos básicos | 26 |
| CAPÍTULO III..... | 28 |
| MARCO METODOLÓGICO..... | 28 |
| 3.1. Tipo de investigación | 28 |
| 3.2. Diseño de la investigación..... | 28 |
| 3.3. Nivel de investigación | 29 |
| 3.4. Población y muestra | 29 |
| 3.5. Técnicas e instrumentos de recolección de datos | 30 |
| 3.6. Fases de la investigación | 31 |
| CAPÍTULO IV | 33 |
| RESULTADOS..... | 33 |
| 4.1. Fase I: Diagnostico de la situación actual de las plataformas de determinación de posibles candidatos a fármaco, aplicando las técnicas respectivas de recolección de datos..... | 33 |
| 4.1.1 Actividad I: Investigación de sistemas con similitudes a la plataforma que se planteó desarrollar. | 34 |
| 4.1.2 Actividad II: Elaboración del guion para la entrevista su aplicación | 34 |
| 4.2. Fase II: Determinación de los requerimientos funcionales y no funcionales para una aplicación web basada en redes neuronales, en la determinación posibles candidatos a fármaco. | 44 |
| 4.2.1. Actividad I: Análisis y definición de los requisitos funcionales y no funcionales de la plataforma colaborativa. | 44 |
| 4.3. Fase III: Diseño de una aplicación web para la determinación de posibles candidatos a fármaco usando metodología Extreme Programming (XP) | 45 |
| 4.3.1. Actividad I: Diseño de casos de uso. | 45 |
| 4.3.2. Actividad II: Descripción de casos de uso. | 47 |
| 4.3.3. Actividad III: Declaración de módulos y de estados en la plataforma. | 51 |
| 4.3.4. Actividad IV: Modelado de base de datos. | 52 |
| 4.3.5. Actividad V: Descripción de la arquitectura del sistema. | 53 |

| | | |
|--|---|-----------|
| 4.3.6. | Actividad VI: Servicio de redes neuronales..... | 54 |
| 4.3.7. | Actividad VII: Diseño de interfaces..... | 58 |
| 4.3.8. | Actividad VIII: Estructuración de las secciones de la plataforma. | 69 |
| 4.4. | Fase IV: Construcción de una aplicación web que prediga propiedades químicas y farmacéuticas para la determinación de posibles candidatos a fármaco utilizando redes neuronales | 70 |
| 4.5. | Fase V: Ejecución de un plan de pruebas de software para la verificación de el correcto funcionamiento de la aplicación web. | 71 |
| CAPÍTULO V | | 78 |
| CONCLUSIONES Y RECOMENDACIONES..... | | 78 |
| 5.1. | Conclusiones | 78 |
| 5.2. | Recomendaciones | 80 |
| REFERENCIAS BIBLIOGRÁFICAS..... | | 81 |
| | Electrónicas: | 81 |
| | Impresas: | 83 |

ÍNDICE DE TABLAS

| Tabla | Contenido | pp. |
|-------|---|-----|
| 1. | Respuesta pregunta n° 1 entrevista. | 34 |
| 2. | Respuesta pregunta n° 2 entrevista. | 35 |
| 3. | Respuesta pregunta n° 3 entrevista. | 37 |
| 4. | Respuesta pregunta n° 4 entrevista. | 38 |
| 5. | Respuesta pregunta n° 5 entrevista. | 38 |
| 6. | Respuesta pregunta n° 6 entrevista. | 39 |
| 7. | Respuesta pregunta n° 7 entrevista. | 40 |
| 8. | Respuesta pregunta n° 8 entrevista. | 41 |
| 9. | Respuesta pregunta n° 9 entrevista. | 42 |
| 10. | Respuesta pregunta n° 10 entrevista. | 42 |
| 11. | Caso de Uso (Iniciar Sesión). | 47 |
| 12. | Caso de Uso (Registrar Usuario). | 47 |
| 13. | Caso de Uso (Modificar Usuario). | 48 |
| 14. | Caso de Uso (Predicción de Fármacos). | 49 |
| 15. | Caso de Uso (Historial de predicciones general). | 50 |
| 16. | Caso de Uso (Historial de predicciones por diana terapéutica). | 50 |
| 17. | Caso de Uso (Historial de predicciones por investigador). | 50 |
| 18. | Módulos de la plataforma. | 51 |
| 19. | Datos enviados y recibidos del modelo. | 56 |
| 20. | Inicio de sesión. | 71 |
| 21. | Registro de usuario. | 71 |
| 22. | Dibujo de las moléculas a ingresar. | 72 |
| 23. | Ingreso de diana terapéutica para la predicción. | 73 |
| 24. | Historial general y por diana terapéutica. | 73 |
| 25. | Historial según investigador. | 74 |
| 26. | Privacidad de información. | 75 |

| | | |
|-----|--|----|
| 27. | Confirmación de la contraseña..... | 75 |
| 28. | Redireccionamiento por falta de datos de sesión..... | 76 |
| 29. | Predicción de candidatos a fármaco..... | 76 |

ÍNDICE DE GRAFICOS

| Gráfico | Contenido | pp. |
|---------|---|-----|
| 1. | Diagrama de caso de uso (Administrador)..... | 45 |
| 2. | Diagrama de caso de uso (Usuario). | 46 |
| 3. | Diagrama de base de datos. | 52 |
| 4. | Diagrama de la arquitectura del sistema. | 53 |
| 5. | Arquitectura del servicio de redes neuronales. | 54 |
| 6. | Esquema general de la transformación a descriptores moleculares. | 56 |
| 7. | Arquitectura del modelo de redes neuronales | 57 |
| 8. | Error Absoluto medio obtenido del modelo..... | 58 |
| 9. | Esquema básico del diseño de la interfaz..... | 58 |
| 10. | Paleta de colores..... | 60 |
| 11. | Captura de pantalla (Pantalla de inicio) | 60 |
| 12. | Captura de pantalla (Información, Introducción) | 61 |
| 13. | Captura de pantalla (Información, Dianas terapéuticas) | 61 |
| 14. | Captura de pantalla (Inicio de sesión)..... | 62 |
| 15. | Captura de pantalla (Inicio post-ingreso-administrador) | 62 |
| 16. | Captura de pantalla (Inicio post-ingreso-usuario)..... | 63 |
| 17. | Captura de pantalla (Usuarios-administrador) | 63 |
| 18. | Captura de pantalla (Ingresar usuario-administrador)..... | 64 |
| 19. | Captura de pantalla (Editar usuario-administrador)..... | 64 |
| 20. | Captura de pantalla (Ingreso de moléculas-Administrador) | 65 |
| 21. | Captura de pantalla (Ingreso de moléculas-usuario)..... | 65 |
| 22. | Captura de pantalla (Predicción-administrador) | 66 |
| 23. | Captura de pantalla (Predicción-usuario)..... | 66 |
| 24. | Captura de pantalla (Historial por investigador-administrador) | 67 |
| 25. | Captura de pantalla (Historial general-administrador)..... | 67 |
| 26. | Captura de pantalla (Historial general-usuario)..... | 68 |

| | | |
|-----|--|----|
| 27. | Captura de pantalla (Historial por diana terapéutica-administrador)..... | 68 |
| 28. | Captura de pantalla (Historial por diana terapéutica-usuario) | 69 |
| 29. | Carta estructurada (Mapa de navegación)..... | 70 |



REPÚBLICA BOLIVARIANA DE VENEZUELA
UNIVERSIDAD JOSÉ ANTONIO PÁEZ
FACULTAD DE INGENIERÍA
ESCUELA DE COMPUTACIÓN

**APLICACIÓN WEB PARA DETERMINAR POSIBLES CANDIDATOS A
FÁRMACO UTILIZANDO REDES NEURONALES**

Autor: Mariángeles Salas

Tutor: Ing. Mayerlin Maldonado

Fecha: abril 2020

RESUMEN

La presente investigación tuvo como objetivo desarrollar una plataforma para determinar si un compuesto químico ingresado por un usuario es un posible candidato a fármaco para las enfermedades Alzheimer, Dengue, Diabetes e Hipertensión, implementando algoritmos de redes neuronales para la predicción de ciertos parámetros fisicoquímicos y farmacodinámicos tales como el IC₅₀ y las reglas de Lipinski, para así mejorar los tiempos y costos del desarrollo de nuevos medicamentos para las enfermedades anteriormente mencionadas. Para ello se realizó la presente investigación enmarcada en el modelo de proyecto especial, con un diseño de campo para recolección de información, paralelamente se hizo uso de la metodología de desarrollo XP. Aunado a ello, se realizó enfocado en la línea de investigación, gestión de proyectos de tecnologías de información y comunicación. ya que el proyecto genero una aplicación web que permite un avance en las investigaciones para el desarrollo de nuevos medicamentos al no tener que recurrir a ensayos experimentales. Para finalizar, se desarrolló un sistema que permite el manejo de datos ingresados organizados de manera tal que se puedan distinguir por su actividad química o por la diana terapéutica.

Descriptor: Candidatos a fármacos, redes neuronales, desarrollo de software, inteligencia artificial, análisis predictivo, química computacional

INTRODUCCIÓN

El proceso de llevar un nuevo medicamento al mercado es extremadamente desafiante, cuesta un promedio de \$2.6 mil millones y toma hasta 12 años. Por lo tanto, es importante que las compañías farmacéuticas optimicen el proceso de descubrimiento y desarrollo de medicamentos, para maximizar la vida útil del medicamento antes de que expiren las patentes.

En general, el proceso puede dividirse en tres fases principales: Descubrimiento del fármaco, o fase en la que se eligen las moléculas candidatas en función de sus propiedades farmacológicas, desarrollo preclínico, o fase en la que se realiza un amplio abanico de estudios en seres no humanos (p. ej., pruebas de toxicidad, análisis farmacocinético/farmacodinámico y formulación) y el desarrollo clínico, o fase en la que se prueba la eficacia, los efectos secundarios y los peligros potenciales que el fármaco seleccionado provoca en voluntarios sanos y en pacientes, El Sevier (2020).

En cada uno de estos pasos existe un riesgo importante de que el compuesto no cumpla los requisitos necesarios para seguir adelante con el proceso. En tal caso, es posible que se deba abandonar el compuesto en el que se ha realizado una inversión importante.

La caracterización cuantitativa o cualitativa de las características similares a la de los fármacos conocidos puede ayudar a los químicos médicos y computacionales a seleccionar pistas de fármacos de mayor calidad a partir de una gran cantidad de compuestos y mejorar la eficiencia de las líneas de diseño de fármacos.

Los algoritmos de aprendizaje automático se desarrollaron originalmente en campos alejados de la investigación farmacéutica, sin embargo, empresas como AstraZeneca, Merck y Pfizer han comenzado a utilizar herramientas de aprendizaje automatizado para identificar rápidamente candidatos prometedores de medicamentos, validándolos a través de ensayos preclínicos y ensayos clínicos de manera eficiente, a diferencia de los métodos tradicionales de descubrimiento y desarrollo ayudando así a

los investigadores a disminuir su tiempo de investigación y ahorrar fondos para más estudios a largo plazo.

Los beneficios potenciales y a largo plazo de la Inteligencia Artificial (IA) en el descubrimiento de fármacos son un tiempo de comercialización más rápido y una mayor eficiencia en todo el proceso de descubrimiento de fármacos, por lo cual se desarrolló una plataforma web la cual mediante redes neuronales predice parámetros fisicoquímicos y farmacodinámicos para determinar si una estructura química ingresada por un usuario puede ser un candidato a fármaco para cuatro enfermedades de gran impacto en Venezuela como lo son dengue, hipertensión, diabetes y alzhéimer.

De esta forma, la presente investigación está compuesta de cinco (5) capítulos, definidos en orden a continuación:

Capítulo I, El Problema, enmarca una de las partes más importantes del trabajo de investigación y se refiere a la descripción general del problema, justificación, objetivos de la investigación, incluyendo objetivos específicos y el alcance de la investigación.

Capítulo II, Marco Teórico, comprende los antecedentes más influyentes sobre la investigación, las bases teóricas que le brindan al investigador el apoyo inicial para el conocimiento del objeto de estudio, las bases legales y los términos básicos que sustentan dicha investigación.

Capítulo III, Marco Metodológico, el cual comprende la descripción de las metodologías utilizadas para dar solución al problema planteado. Aquí se incluyen necesariamente métodos, técnicas y/o procedimientos empleados en el estudio, sustentados con autores.

Capítulo IV, Describe los resultados obtenidos en la realización del proyecto investigativo, especificado por fases.

Capítulo V, Conlleva las conclusiones y recomendaciones que se han definido en la realización de todo el proceso investigativo.

CAPÍTULO I

EL PROBLEMA

1.1. Planteamiento del problema

Durante la historia, el descubrimiento de muchos medicamentos ha sido por casualidad o serendipia, dichos remedios típicamente naturales lograban curar una enfermedad a través del ensayo y error, sin embargo, actualmente esto ha cambiado y el descubrimiento y desarrollo de nuevos fármacos se ha convertido en un proceso largo y complejo.

El proceso de desarrollo de un fármaco es largo y costoso ya que se tiene que demostrar, a lo largo de las distintas fases, que el nuevo medicamento reúne los requisitos de eficacia, seguridad y calidad exigidos para su comercialización, este proceso también incluye numerosas actividades y disciplinas dirigidas a demostrar una relación costo-eficacia optima, lo que implica un gran esfuerzo intelectual y un coste económico enorme.

Desde que se comienza a investigar una molécula innovadora hasta que el medicamento llega al mercado pueden transcurrir entre 10 a 15 años, este tiempo se divide en las diferentes fases del proceso de investigación de un fármaco que comprenden la fase de descubrimiento, la fase preclínica, la fase clínica y la fase de aprobación y registro (Merck Sharp and Dohme, 2016).

La fase de descubrimiento es el primer paso en el proceso de descubrimiento de un nuevo fármaco, en esta etapa miles de compuestos pueden ser estudiados, sin embargo, solo muy pocos pasan a las siguientes etapas, normalmente los investigadores descubren nuevos fármacos a través de varias formas, una de ellas es el conocimiento sobre el proceso de una enfermedad que permite a los investigadores diseñar un producto para detener o revertir los efectos de la enfermedad, otra es a través de muchas pruebas de compuestos moleculares para encontrar posibles efectos beneficiosos frente

a una gran cantidad de enfermedades, la tercera manera son los tratamientos existentes que tienen efectos inesperados y por último las tecnologías que proporcionan nuevas formas de orientar los productos médicos a sitios específicos del cuerpo o manipular el material genético.

La fase de descubrimiento más utilizada industrialmente se divide en cuatro sub-fases; la primera es la identificación de la diana terapéutica, estas son sustancias químicas asociadas a células o genes las cuales pueden ser el origen de ciertas enfermedades, actualmente la mayoría de las dianas suelen ser proteínas, la segunda sub-fase es la validación de ella, donde una vez que los científicos identifican una posible diana deben validarla hasta demostrar que verdaderamente interviene en la enfermedad, una vez demostrado se debe investigar si puede realizarse un fármaco contra dicha enfermedad el cual sea seguro y eficaz.

Por otro lado la tercera sub-fase corresponde a la identificación de un compuesto líder, un compuesto líder es aquel que puede tener potencial para tratar la enfermedad, este puede ser una estructura química, un compuesto natural, un péptido o un anticuerpo que se une a la diana terapéutica y tiene un efecto sobre ella, al obtener un compuesto líder se pueden desarrollar otras moléculas hasta tener los candidatos que se trabajan en las fases de estudio en humanos, para finalizar se lleva a cabo la validación del compuesto líder en esta fase se comparan esos compuestos potencialmente buenos para tratar la enfermedad con la intención de obtener la información sobre cuales podrán ser medicamentos fiables y competentes.

Teniendo en cuenta que el número de moléculas orgánicas que son sintéticamente factibles esta entre 10^{20} y 10^{24} , es evidente que su análisis sería muy complejo sin el uso de técnicas computacionales por lo cual la industria farmacéutica no se ha quedado ajena a los sistemas informáticos.

El Machine Learning (ML) o aprendizaje automatizado es un subconjunto de la inteligencia artificial (IA) en la informática y designa cualquier método donde los resultados de acciones, decisiones u observaciones pasadas, se utilizan para mejorar las predicciones o la toma de decisiones futuras. Las técnicas de aprendizaje automatizado

son ahora extremadamente populares en el desarrollo de fármacos, ya que permiten la automatización de análisis de datos biológicos ruidosos y de gran dimensión.

Por otra parte, las redes neuronales se están convirtiendo en un método muy popular para resolver problemas de aprendizaje automático e inteligencia artificial. Se utilizan ampliamente en los campos de la biología de sistemas y el descubrimiento de fármacos para resolver la complejidad asociada con modelos matemáticos, cribado virtual de compuestos, descifrar relaciones cuantitativas estructura-actividad, estimación de propiedades farmacocinéticas y farmacodinámicas y durante el desarrollo de formulaciones, las redes neuronales son excelentes para descifrar relaciones no lineales entre las variables y predecir el resultado mucho antes en el proceso de descubrimiento de fármacos.

Los métodos computacionales no diseñan fármacos por sí solos, ellos forman parte de un esfuerzo multidisciplinario para alcanzar un objetivo extremadamente complejo, los especialistas reconocen que diseñar la estructura de nuevas moléculas observando su complementariedad con la estructura de la proteína con la que quieren interaccionar, es algo muy útil.

Conociendo que la etapa de descubrimiento de un fármaco es la más larga, en promedio 50.2 meses, con una tasa de éxito muy baja, ya que solo 250 de cada 10000 moléculas de la investigación básica pasan a la fase preclínica se planteó desarrollar una aplicación web para predecir parámetros como el IC50 y las reglas de Lipinski utilizando redes neuronales, las cuales fueron entrenadas con información obtenida de ChEMBL una base de datos open source, la cual posee los datos experimentales de compuestos distintos según las dianas terapéuticas que para efectos de este proyecto se utilizaron cuatro (4), DENV (Dengue), DPP4 (Diabetes), Acetilcolinesterasa (Alzheimer) y Angiotensina II tipo 1 (Hipertensión), dichas dianas terapéuticas fueron tomadas por su gran impacto en la sociedad venezolana.

La data utilizada será dividió en dos, donde 75% de la data se utilizó para entrenar el modelo y el otro 15% se utilizó para realizar pruebas y así verificar el funcionamiento

de las redes neuronales y que este cumpliera con los requerimientos, por otro lado, se determinó el margen de error mediante pruebas estadísticas.

Con este proyecto se logra disminuir costos al no necesitar laboratorios ni reactivos, y también se reduce tiempo, ya que no se tienen que realizar pruebas experimentales a todos los compuestos, si no solamente a los que presenten las mejores características.

1.2. Formulación del problema

A partir de lo anteriormente planteado se formula la siguiente interrogante. ¿Cómo se puede optimizar el descubrimiento de candidatos a fármacos mediante técnicas computacionales para disminuir los tiempos y los costos del desarrollo de medicamentos?

1.3. Objetivos de la investigación

1.3.1. Objetivo general

Desarrollar una aplicación web de determinación de posibles candidatos a fármaco utilizando redes neuronales para la disminución de tiempos y costos en el desarrollo farmacéutico.

1.3.2. Objetivos específicos

- Diagnosticar la situación actual de las plataformas de determinación de posibles candidatos a fármaco, aplicando las técnicas respectivas de recolección de datos
- Determinar los requerimientos funcionales y no funcionales para una aplicación web basada en redes neuronales, en la determinación de posibles candidatos fármaco
- Diseñar una aplicación web de para la determinación de posibles candidatos a fármaco usando metodología Extreme Programming (XP).
- Construir una aplicación web que prediga propiedades químicas y farmacéuticas para la determinación de posibles candidatos a fármaco utilizando redes neuronales

- Ejecutar un plan de pruebas de software para la verificación de el correcto funcionamiento de la aplicación web.

1.4. Justificación de la investigación

El presente trabajo de investigación tuvo como finalidad la necesidad de crear una alternativa computacional a los métodos experimentales tradicionales de descubrimiento y desarrollo de nuevos medicamentos.

Esto es debido a que la fase de descubrimiento de nuevos medicamentos o compuestos líderes es la fase que lleva más tiempo de ejecución y con los costos más elevados, en tal sentido esta plataforma supone un gran impacto para la industria farmacéutica, ya que al lograr predecir el IC_{50} y las reglas de lipinski, se puede determinar si un compuesto puede ser un buen candidato a fármaco, lo que conlleva a una disminución de tiempos y de costos al no tener que realizar las determinaciones de estos parámetros de manera experimental.

Para efectos de este proyecto se utilizaron cuatro enfermedades las cuales fueron tomadas por su gran impacto en la sociedad venezolana, siendo estas, Dengue una enfermedad vírica que en su última epidemia entre finales del 2019 y comienzos del 2020 ha dejado más de 14.701 casos confirmados y 22 muertos según la organización mundial de la salud (OMS) , el Alzheimer por su parte afecta entre 7% y 9% de la población mayor de 65 años en nuestro país, según la Organización Mundial de la Salud (ONU) el Alzheimer tiene un crecimiento interanual de 1.3% en el mundo, sin embargo, la proyección del país es alarmante en el largo plazo, pues encabeza la tasa de crecimiento proyectado a 2050 con 491%: la cual irá a un ritmo de 14,87% anual según El Nacional (2017) y por último, la hipertensión y la diabetes debido a que al menos 7 millones de venezolanos (34,1 %) sufren de hipertensión arterial, este es el factor de riesgo más prevalente en la población, junto con la diabetes (12,3 %) de acuerdo con el Estudio Venezolano de Salud Cardio metabólica EVESCAM (2017).

Este proyecto también aporta información al área académica como soporte en próximas investigaciones, ya que al ingresar las moléculas y obtener los parámetros, el

sistema las clasifica como activa, inactiva o intermedia, y la guardara en la base de datos donde tanto el administrador como el usuario podrán verificarlas en otro momento si lo consideran necesario, además de utilizar redes neuronales para la predicción que puede ser de gran utilidad para otros proyectos.

1.5. Alcance

Considerando lo anterior expuesto, la aplicación para la determinación de posibles candidatos a fármaco tuvo como funcionalidad agilizar los tiempos de la investigación farmacéutica, para ello la plataforma evalúa compuestos químicos ingresados por el usuario y determinara mediante la predicción de ciertos parámetros químicos y farmacéuticos si pueden ser o no candidatos a fármaco para la diana terapéutica escogida.

Esta aplicación se dividió en dos partes fundamentales, en primer lugar, para la red neuronal, se utilizó como lenguaje de programación Python y en segundo lugar el desarrollo de la aplicación web donde se utilizaron lenguajes de programación como JavaScript y el microframework Flask, por otro lado, para el diseño de la interfaz del usuario: HTML, CSS y Bootstrap, como gestor de base de datos se utilizó Firebase, la cual es una base de datos NoSQL o no relacional.

La metodología aplicada fue Extreme Programming (XP), la cual se centra en procedimientos rápidos y ágiles que permitió aplicar de mejor manera las técnicas de desarrollo de software durante el ciclo de desarrollo de la aplicación, además hace énfasis en la adaptabilidad obteniendo mejores resultados.

El tiempo que se dispuso para el desarrollo del proyecto fue de 32 semanas.

CAPÍTULO II

MARCO TEÓRICO

2.1. Antecedentes de la investigación

En todo trabajo de investigación es fundamental realizar la búsqueda de otras publicaciones y trabajos para así comprobar su relación y afinidad con nuestra investigación. Esta sección de la investigación permite realizar dicha comparación y así como también justificar la factibilidad del desarrollo del proyecto.

Qureshi, Tandon y Kumar (2015). Elaboraron un trabajo de investigación en el Instituto de Tecnología Microbiológica en India, el cual fue publicado en la revista PeptideScience titulado **“AVP-IC50Pred: Multiple Machine Learning Techniques-Based Prediction of Peptide Antiviral Activity in Terms of Half Maximal Inhibitory Concentration (IC50)”**. Ellos desarrollaron un servidor web llamado “AVP-IC₅₀Pred”, utilizando algoritmos basados en regresión para predecir la actividad antiviral en términos de valores de IC₅₀ (μM). Para ello dividieron un total de 759 péptidos no redundantes de AVPdb y HIPdb en un conjunto de entrenamiento/prueba que tenía 683 péptidos (T683) y un conjunto de validación con 76 péptidos independientes (V76) para evaluación. Emplearon cuatro técnicas diferentes de aprendizaje automático (ML), siendo estas, Support vector machine, Random Forest, Instance-based classifier y K-Star. Los resultados obtenidos fueron determinados mediante correlación de Pearson (PCC) tanto para la el aprendizaje como para la validación obtenido máximos mayores de 0.6 en la mejor combinación de conjuntos de características.

Se tomó en consideración esta investigación, dado que para el desarrollo de la aplicación se utilizaron múltiples herramientas de aprendizaje automatizado para construir modelos de predicción completos que explotan características importantes de la secuencia de péptidos, como composiciones de aminoácidos y dipéptidos, perfil binario de residuos N8-C8 y propiedades fisicoquímicas seleccionada, ayudando así a

que los investigadores que trabajan en terapias antivirales puedan disminuir el costo y el tiempo que implica la validación experimental.

Por otro lado, Bjerrum (2017). Como parte del Wildcard Pharmaceuticals Consulting en Dinamarca publicó un proyecto en el archivo en línea para las prepublicaciones de artículos científicos “arXiv” titulado **“SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules”**. En este trabajo el evaluó el sistema de especificación de introducción lineal molecular simplificada (SMILES) el cual es una representación de texto en una sola línea de una molécula única.

Debido a que una molécula puede tener múltiples cadenas SMILES, razón por la cual se han definido SMILES canónicas, lo que asegura una correspondencia uno a uno entre la cadena SMILES y la molécula. Sin embargo, el hecho de que múltiples SMILES representen la misma molécula hizo que el explorara esto como una técnica para el aumento de datos de un conjunto de datos QSAR molecular modelado por una red neuronal basada en células de memoria a largo y corto plazo (LSTM).

El conjunto de datos aumentado era 130 veces más grande que el original. La red entrenada con el conjunto de datos aumentados muestra un mejor rendimiento en un conjunto de prueba en comparación con un modelo construido con solo una cadena SMILES canónica por molécula. El coeficiente de correlación R^2 en el conjunto de prueba mejoró de 0,56 a 0,66 cuando se utilizó la enumeración SMILES, y el error cuadrático medio (RMS) también descendió de 0,62 a 0,55. La técnica también funciona en la fase de predicción. Al tomar el promedio por molécula de las predicciones para las SMILES enumeradas, se encontró una mejora adicional a un coeficiente de correlación de 0,68 y un RMS de 0,52. sobre los datos de SMILES.

Este proyecto se consideró pertinente ya que en él se determina que la enumeración SMILES proporciona modelos QSAR más robustos, siendo esta mejor cuando no se utilizan las SMILES canónicas, si no el conjunto de SMILES que puede tener una molécula, ya que al aumentar la data mejoran las predicciones realizadas a través de la red neuronal.

En paralelo Brown, Sherer, Johnson, Holloway, Sherborne (2017), Realizaron una publicación desde el departamento de Química Estructural adscrito al laboratorio de investigación de materiales de la universidad de Illinois en la revista Journal of Computer-Aided Molecular Design, titulada “**The evolution of drug design at Merck Research Laboratories**” En este trabajo los autores hablan sobre la evolución de Merck que con más de 40 años de inversión, han estado en el negocio del diseño de fármacos más tiempo que la mayoría de las empresas. Ellos describen que, a lo largo de su historia, el grupo de diseño de fármacos de Merck ha tenido varios nombres, pero siempre ha estado en el negocio del "diseño", con el objetivo final de proporcionar una hipótesis procesable que pudiera probarse experimentalmente. A menudo, el resultado era una molécula pequeña, pero también podría ser un péptido, biológico, modelo predictivo, reacción, proceso, etc.

Según ellos en la actualidad, el grupo de diseño de Merck ha incluido la química computacional, la determinación de la estructura de proteínas y química informática, lo que ha dado como resultado una expansión del 80% en el tamaño del departamento, lo que refleja el aumento en el impacto debido a un énfasis significativo en toda la organización para "diseñar primero" a lo largo de toda la ruta de descubrimiento de fármacos desde la identificación de componentes potenciales (LID) por sus siglas en inglés, a la primera dosis en humanos (FIH) por sus siglas en inglés.

Entre las herramientas que han incluido en sus estudios están los servicios de aplicaciones como Biovia, Openeye y servicios de nube, además de Spotfire, Excel, ChemDraw para realizar los dibujos de las moléculas, OneNote, aplicaciones web y de escritorio para simulaciones y modelaje computacional, Machine Learning y entre otras tecnologías utilizando lenguajes de programación como Java, C#, .NET, JavaScript, Python, Perl y C++, para realizar así predicciones a lo largo de todo el proceso de diseño de fármacos.

Al analizar a profundidad esta publicación, fue de significativa importancia para el presente estudio ya que demuestra que las industrias farmacéuticas están llevando a

cabo cambios en sus estructuras, aplicando cada vez más herramientas computacionales en el descubrimiento y desarrollo de fármacos.

Por otra parte, Patankar, Jurs (2018) realizaron un trabajo desde el departamento de Química en la Universidad de Pensilvania el cual fue publicado en la revista Journal of Chemical Information and Computer Sciences, titulado **“Prediction of IC₅₀ Values for ACAT Inhibitors from Molecular Structure”**. En dicha publicación realizaron un estudio cuantitativo de estructura-actividad en varias series de compuestos derivados del isocianato de N-clorosulfonilo para desarrollar modelos que relacionen sus estructuras con la actividad IC₅₀ para la inhibición de la acil-CoA: colesterol O-aciltransferasa (ACAT).

Los descriptores numéricos se utilizaron para codificar información topológica, electrónica y geométrica de las estructuras moleculares de los inhibidores. Se utilizaron un conjunto de datos de 157 compuestos que muestran actividad reductora de triglicéridos y colesterol para desarrollar modelos de regresión lineal exitosos y modelos de redes neuronales computacionales no lineales. Los modelos se validaron mediante un conjunto de predicciones externo. Se han desarrollado modelos utilizando la metodología QSAR que relacionan con éxito la estructura molecular con las concentraciones inhibitorias de inhibidores de ACAT. Los modelos informados aquí son los primeros con la capacidad de predecir los valores de IC₅₀ a partir de información estructural sola para los inhibidores de ACAT derivados del isocianato de N-clorosulfonilo.

La información estructural de los compuestos que reducen el colesterol y los triglicéridos se codificó numéricamente como descriptores moleculares que incluyen descriptores de enlace de hidrógeno y descriptores de índice de estado electrotopológico. La selección de características lineales condujo al desarrollo de modelos lineales tipo I y modelos CNN no lineales mejorados tipo II. La selección de características no lineales condujo al desarrollo de modelos CNN de tipo III totalmente no lineales.

En conclusión, el modelo tipo III tuvo los mejores errores RMS de todos los modelos que se generaron. Por lo cual este trabajo tuvo gran importancia ya que demostró que los valores de IC₅₀ pueden predecirse basándose únicamente en la estructura molecular, sin la inclusión de datos obtenidos experimentalmente, como los coeficientes de partición.

Para finalizar, Mishra, Singh, Lahiri, Mista (2019) Realizaron una publicación desde Instituto de información tecnológica en india, en la revista Biomedical Informatics titulada **“A comparative study on the molecular descriptors for predicting drug-likeness of small molecules”**. En dicho trabajo, utilizaron un conjunto de nueve descriptores moleculares elegidos heurísticamente, incluidos cuatro de la regla de Lipinski, como parámetro de clasificación para el cribado de moléculas "similares a fármacos". La robustez de la clasificación se comparó con cuatro descriptores fundamentales de Lipinski.

El clasificador basado en redes neuronales se aplicó en una base de datos de 60000 moléculas para la clasificación de moléculas "similares a fármacos" y "no similares a fármacos". El resultado de la clasificación con nueve descriptores mostró una alta precisión de clasificación del 96,1% en comparación con el resultado de la clasificación con cuatro descriptores de Lipinski que arrojaron una precisión del 82,48%. También se produjo una disminución significativa de falsos positivos al utilizar nueve descriptores, lo que provocó un fuerte aumento del 18% en la especificidad de la clasificación.

Se tomó en consideración esta investigación, ya que de este estudio se desprende que los descriptores de Lipinski, forman la base para la identificación de moléculas similares a fármacos que pueden mejorarse sustancialmente añadiendo más descriptores que representan las propiedades farmacodinámicas de las moléculas, demostrando que el enfoque basado en reglas para la predicción de la similitud con las drogas tiene margen de mejora. Además, también señala que el enfoque de aprendizaje automático muestra una buena eficiencia para la predicción de la similitud con las drogas y que la elección adecuada y el número de descriptores es de suma importancia

para la aplicación de enfoques de aprendizaje automático para la predicción de la similitud con las drogas.

2.2. Bases Teóricas

Según Arias (2006), “Las bases teóricas implican un desarrollo amplio de los conceptos y proposiciones que conforman el punto de vista o enfoque adoptado, para sustentar o explicar el problema planteado”.

Por lo cual a continuación se expondrán diversos conceptos para el desarrollo y comprensión de la problemática.

2.2.1. Alzheimer

Según Romano, Nissen, Paredes (2007), la enfermedad de Alzheimer es un trastorno neurológico que provoca la muerte de las células nerviosas del cerebro. Por lo general, la Enfermedad de Alzheimer comienza paulatinamente y sus primeros síntomas pueden atribuirse a la vejez o al olvido común. A medida en que avanza la enfermedad, se van deteriorando las capacidades cognitivas, entre ellas la capacidad para tomar decisiones y llevar a cabo las tareas cotidianas, y pueden surgir modificaciones de la personalidad, así como conductas problemáticas. En sus etapas avanzadas, la Enfermedad de Alzheimer conduce a la demencia y finalmente a la muerte.

2.2.2. Análisis predictivo

La Universidad de la Rioja (2020), lo define como modelos matemáticos que predicen el comportamiento de una variable en función de un conjunto de otras variables. Cuanto más relacionadas estén el conjunto de variables predictoras con la variable a predecir (correlación), más exactas serán las predicciones.

Para la construcción de estos modelos, una vez desarrollado el algoritmo predictivo, es necesario disponer de un conjunto de datos históricos. Para ello se dividen estos en dos conjuntos: uno de datos de entrenamiento y otro de prueba. Al modelo se le pasan como entrada los datos de entrenamiento para calibrar la predicción y, posteriormente, los datos de prueba. Después se compara el resultado de la predicción con los valores reales (históricos) para comprobar su exactitud.

El proceso de un análisis predictivo consiste en:

- **Recopilar datos:** El desarrollo de un modelo de análisis predictivo comienza con la obtención de los datos en función de los que se va a realizar las predicciones. Los datos podrán proceder de diferentes fuentes, tales como archivos, BBDD, sensores, etc. y se exploran para conocer su naturaleza, estructura, la calidad de los mismos...
- **Ordenar y adaptar los datos:** El siguiente paso consiste en realizar un procesado inicial de los mismos para ordenarlos, transformarlos y adecuarlos a las necesidades del modelo. En este punto se suelen eliminar los valores extremos que distorsionan el funcionamiento del modelo y se crea una única estructura con los datos procesados de las distintas fuentes.
- **Analizar los datos:** Una vez creada la estructura de datos se analizan los mismos para identificar sus características, detectar los patrones y tendencias en sus valores y obtener información relevante para el desarrollo del algoritmo predictivo, que es la base del modelo.
- **Elegir el modelo matemático:** Para el desarrollo del algoritmo se utilizará la información obtenida del análisis de los datos y, en función de ella, se decidirán las técnicas matemáticas que permitirán procesar los datos de entrada al modelo y realizar las predicciones (salidas).
- **Optimizar los parámetros:** Finalmente, se optimizarán los parámetros del algoritmo mediante un conjunto de datos de entrenamiento (datos reales de las variables de entrada y de salida del modelo) y se verificará su exactitud con un conjunto de datos de prueba (datos reales de las variables de entrada y predicciones del modelo que se comparan con los datos reales con los que deberían coincidir las predicciones).

2.2.3. Aprendizaje automático

Según Moreno, Benjar y Armengol (2015), El Aprendizaje Automático es una rama de la Inteligencia Artificial que abarca diferentes técnicas, las cuales permiten

dotar a los computadores de la capacidad de "aprender" modelos tales que, de forma automática, pueden ser usados, por un lado, para resolver problemas nuevos o, por otro lado, para mejorar el rendimiento en problemas ya vistos.

2.2.4. Aplicación Web

Molina (2007), describe una aplicación Web como “un sistema informático que los usuarios utilizan accediendo a un servidor Web a través de internet o de una intranet. En otras palabras, es un programa que se codifica en un lenguaje interpretable por los navegadores web en la que se confía la ejecución al navegador. Las aplicaciones Web son populares debido a la practicidad del navegador Web como cliente ligero. La facilidad para actualizar y mantener aplicaciones Web sin distribuir e instalar software en miles de potenciales clientes es otra razón de su popularidad”

2.2.5. Ciencia de datos

Por su parte Henderson (2007), describen que la ciencia de datos fusiona minería de datos, análisis estadístico y aprendizaje de maquina junto con capacidades de modelamiento de datos e integración de datos, para construir modelos predictivos que exploran patrones de contenido de datos. El desarrollo de modelos predictivos es llamado algunas veces Ciencia de Datos porque el analista de datos, o científico de datos, utiliza el método científico para desarrollar y evaluar un modelo.

El desarrollo de soluciones de ciencia de datos implica la inclusión iterativa de orígenes de datos en modelos que producen información. La ciencia de datos depende de:

- **Fuentes de datos enriquecidos:** Los datos con la posibilidad de mostrar de otra manera patrones invisibles en el comportamiento organizacional o del cliente.
- **Alineación y análisis de información:** Técnicas para comprender el contenido de los datos y combinar conjuntos de datos para probar y formular hipótesis de patrones significativos.

- **Entrega de información:** Ejecutar modelos y algoritmos matemáticos con los datos y producir visualizaciones y otra salida para obtener información del comportamiento
- **Presentación de resultados e información de los datos:** El análisis y presentación de los resultados de manera que la información pueda ser compartida

2.2.6. Coeficiente de reparto octanol-agua

Para Figueruel, Dávila (2004), El coeficiente de reparto octanol-agua (K_{ow}) de una sustancia, también llamado coeficiente de partición (P_{ow}), es el cociente o razón entre las concentraciones de esa sustancia en una mezcla bifásica formada por dos disolventes inmiscibles en equilibrio: n-octanol y agua. Se utiliza el n-octanol por ser un compuesto orgánico que simula bien el material lipídico de la materia viva.

Ese coeficiente mide, pues, la solubilidad diferencial de un soluto en esos dos disolventes. Este coeficiente simula bien el carácter hidrófobo de una sustancia o la afinidad hacia los lípidos de una sustancia disuelta en agua. Es comúnmente utilizado en su forma logarítmica $\log P$.

2.2.7. Concentración inhibitoria media máxima

Según Kalliokoski, Kramer, Vulpetty, Gedeck (2013), "la concentración inhibitoria media máxima (IC_{50}) es una medida de la eficacia de una sustancia para inhibir una función biológica o bioquímica específica".

Esta medida cuantitativa indica qué cantidad de un fármaco en particular u otra sustancia (inhibidor) se necesita para inhibir un proceso biológico determinado (o componente de un proceso, es decir, una enzima, célula, receptor celular o microorganismo) a la mitad. Se usa comúnmente como una medida de la potencia de un fármaco antagonista en la investigación farmacológica. Según la FDA, IC_{50} representa la concentración de un fármaco que se requiere para una inhibición del 50% *in vitro*.

Esta es la métrica más utilizada para la actividad en el objetivo en la optimización de fármacos potenciales.

2.2.8. Dengue

Para Torres (2008), El dengue es una enfermedad viral, de carácter endémico-epidémico, transmitida por mosquitos del género *Aedes*, principalmente por *Aedes aegypti*, que constituye hoy la arbovirosis más importante a nivel mundial en términos de morbilidad, mortalidad y afectación económica que tiene diversas formas de expresión clínica: desde fiebre indiferenciada (frecuente en niños) y fiebre con cefalea, gran malestar general, dolores osteomioarticulares, con o sin exantema, leucopenia y algún tipo de sangrado hasta formas graves que – habiendo comenzado con lo anterior – presenta choque hipovolémico por extravasación de plasma, con trombocitopenia moderada o intensa y con grandes hemorragias en aparato digestivo y otras localizaciones. También el dengue es capaz de expresarse mediante las llamadas formas "atípicas" que son relativamente infrecuentes y resultan de la afectación particularmente intensa de un órgano o sistema: encefalopatía, miocardiopatía o hepatopatía por dengue, entre otras

2.2.9. Diabetes

Según la American Diabetes Association (2005), la diabetes mellitus es un grupo de enfermedades metabólicas caracterizadas por hiperglucemia resultante de defectos en la secreción de insulina, la acción de la insulina o ambos. La hiperglucemia crónica de la diabetes se asocia con daño, disfunción e insuficiencia a largo plazo de varios órganos, especialmente ojos, riñones, nervios, corazón y vasos sanguíneos.

La gran mayoría de los casos de diabetes se clasifican en dos amplias categorías etiopatogénicas, la diabetes tipo 1, la cual es una deficiencia absoluta de secreción de insulina. Los individuos con mayor riesgo de desarrollar este tipo de diabetes a menudo pueden identificarse mediante la evidencia serológica de un proceso patológico autoinmune que ocurre en los islotes pancreáticos y mediante marcadores genéticos. En la otra categoría, mucho más prevalente, la diabetes tipo 2, su causa es una combinación de resistencia a la acción de la insulina y una respuesta secretora de

insulina compensatoria inadecuada. En la última categoría, un grado de hiperglucemia suficiente para causar cambios patológicos y funcionales puede estar presente durante un largo período de tiempo antes de que se detecte la diabetes. Durante este período asintomático, es posible demostrar una anomalía en el metabolismo de los carbohidratos mediante la medición de la glucosa plasmática en ayunas o después de una provocación con una carga de glucosa oral.

2.2.10. Diana Terapéutica

Respecto a las dianas terapéuticas Indacea (2018), describe que “una diana terapéutica es una molécula o proteína con una forma específica sobre la que el fármaco actúa”. La identificación de una diana terapéutica es el primer paso para el desarrollo de un fármaco y se calcula que hay unas 8.000 dianas que podrían ser base para nuevos fármacos. Algunas se encuentran en exceso (con lo que habría que bloquear su actividad) y otras en defecto (con lo que habría que incrementarla o reemplazarla).

2.2.11. Hipertensión

La hipertensión, también conocida como tensión arterial alta o elevada, es un trastorno en el que los vasos sanguíneos tienen una tensión persistentemente alta, lo que puede dañarlos. Cada vez que el corazón late, bombea sangre a los vasos, que llevan la sangre a todas las partes del cuerpo. La tensión arterial es la fuerza que ejerce la sangre contra las paredes de los vasos (arterias) al ser bombeada por el corazón. Cuanto más alta es la tensión, más esfuerzo tiene que realizar el corazón para bombear.

La tensión arterial normal en adultos es de 120 mm Hg¹ cuando el corazón late (tensión sistólica) y de 80 mm Hg cuando el corazón se relaja (tensión diastólica). Cuando la tensión sistólica es igual o superior a 140 mm Hg y/o la tensión diastólica es igual o superior a 90 mm Hg, la tensión arterial se considera alta o elevada.

La mayoría de las personas con hipertensión no muestra ningún síntoma; por ello se le conoce como el "asesino silencioso". En ocasiones, la hipertensión causa síntomas como dolor de cabeza, dificultad respiratoria, vértigos, dolor torácico, palpitaciones del corazón y hemorragias nasales, pero no siempre Organización Mundial de la Salud (2016)

2.2.12. Manejador de base de datos

Para Osorio (2008), una base de datos es aquella que almacena datos y los conecta en una unidad lógica junto a los metadatos necesarios para su procesamiento, con el objetivo de presentar el contenido de forma adecuada y clara.

Un manejador de base de datos (en inglés DBMS, Data Base Management System) es un módulo de programa que constituye la interfaz entre los datos de bajo nivel almacenados en la base de datos y los programas de aplicaciones y las consultas hechas al sistema. Sus tareas básicas son:

- **Interacción con el manejador de archivos.** Es quien se encarga realmente del almacenamiento, recuperación y actualización de los datos en la base de datos
- **Implantación de la integridad.** Realiza las operaciones que limitan la consistencia de los datos.
- **Puesta en práctica de la seguridad.** Hace cumplir los requisitos de seguridad para tener acceso a los datos
- **Respaldo y recuperación.** Es responsable de detectar fallas y restaurar la base de datos al estado que existía antes presentarse esta.
- **Control de concurrencia.** Vela por la consistencia de la información cuando varios usuarios actualizan a la vez la base de datos

2.2.13. Reglas de Lipinski

Según Voet, Voet (2006) Son unas reglas basadas en datos empíricos, formuladas por Christopher Lipinski y conocidas como “reglas del cinco” de Lipinski, estas establecen que es probable que un compuesto exhiba mala absorción o penetración si:

- Su masa molecular es mayor que 500Da
- Tiene más de 5 donantes de puentes de hidrógenos (expresados como la suma de grupos OH y NH)
- Presenta más de 10 aceptores de puente de hidrogeno (expresados como la suma de sus átomos de N y O)

- Su valor del log P es mayor que 5

Es probable que los fármacos candidatos que no cumplen la regla 1 tengan solubilidades bajas y solo atraviesen las membranas celulares con dificultad, es posible que los que no cumplen las reglas 2, 3 o ambas sean demasiados polares para atravesar las membranas celulares; y es factible que los que no cumplen la regla 3 sean poco solubles en solución acuosa e incapaces de ganar acceso a las superficies de la membrana.

2.2.14. SMILES

Según Hornos (2017), SMILES es un acrónimo que significa Simplified Molecular Input Line Entry Specification, es una especificación para describir sin ambigüedades la estructura de una molécula usando cadenas ASCII cortas. Suele considerarse que SMILES tiene la ventaja de ser más inteligible a la lectura humana además también tiene una base más amplia de soporte de software fundamentados en teoría de grafos.

Generalmente se puede escribir un número de cadenas SMILES igualmente válidas para una molécula. Por ejemplo, CCO, OCC y C(O)C especifican por igual al etanol. Se han desarrollado algoritmos para asegurar que se genera la misma cadena SMILES para una molécula, sin importar el orden de los átomos en la estructura. Esta cadena SMILES es única para cada estructura, aunque es dependiente del algoritmo de "canonicalización" usado para generarlo, y se denomina SMILES canónico.

Para describir estructuras con SMILES:

- **Átomos.** La forma de indicar un elemento es la siguiente: átomo: [<masa> símbolo<quiralidad> <hidrógenos> <signo<carga>>], como, por ejemplo: Carbono: [C], Carbono 14: [14C], Cu²⁺: [Cu⁺⁺] o [Cu²⁺]
- **Hidrógenos.** Como regla general los átomos de hidrógeno no se especifican. De esta forma, en nomenclatura SMILES, en realidad C no representa a un átomo de carbono por sí solo, sino a la molécula de metano. Si se desea,

pueden agregarse los hidrógenos de forma explícita dentro de los corchetes, en cuyo caso el metano sería [CH₄].

- **Enlaces.** En el caso de un enlace simple, podemos directamente omitir el símbolo del enlace y colocar los símbolos atómicos uno junto al otro, aunque si lo deseáramos podríamos mostrarlo de forma explícita usando un guion medio. Para el doble enlace usamos el símbolo “=” y para el enlace triple “#”. Existen otros símbolos como “:” para un enlace aromático y \$ para un enlace cuádruple, pero los casos comunes se cubren con los primeros 3.
- **Ramificaciones.** Las ramificaciones se indican entre paréntesis. A la hora de describir la molécula se elige una cadena lineal como la principal y luego se escribe entre paréntesis todo aquello que se encuentre unido a los átomos de la cadena.

Las ramificaciones abarcan no sólo cadenas carbonadas sino también grupos funcionales, por ejemplo, los grupos carbonilo se expresan como C(=O) ya que el átomo de oxígeno se encuentra por fuera de la cadena y se considera una ramificación. El enlace de la ramificación, cuando debe mostrarse de forma explícita, se coloca dentro de los paréntesis. La elección de cuál es la cadena principal y cuáles son las ramificaciones es indistinta.

Las ramificaciones a su vez pueden tener sus propias subramificaciones, las cuales se expresan dentro un nuevo par de paréntesis anidados dentro de la rama madre. Este proceso se puede repetir indefinidamente, agregando tantas ramificaciones anidadas como sea necesario.

- **Anillos.** El truco para describir un anillo en una sola línea consiste en hacer un corte en el mismo de forma tal de convertirlo en una cadena lineal, agregando un indicador que permita reconstruir el anillo uniendo los extremos del enlace cortado. Esto se hace numerando los cortes y etiquetando

a los átomos involucrados con dicho número, El sentido de la enumeración es indistinto, puede ser horario u antihorario.

El caso de los anillos pone de manifiesto algo que también ocurre en otras estructuras: una única molécula puede ser asociada a varias cadenas SMILES igualmente válidas.

- **Aromaticidad.** En SMILES no es necesario explicitar el carácter aromático de un enlace. Los sistemas que trabajan con SMILES cuentan con un algoritmo de detección de aromaticidad basado en la regla de $4N+2$ de Hückel. Por lo tanto, un anillo aromático puede ser escrito de dos formas: como una estructura híbrida o bien como una de sus estructuras de resonancia, denominadas representaciones de Kekulé.

Cuando se desea expresar aromaticidad, los átomos involucrados se escriben en minúsculas. No cualquier elemento puede ser aromático, estos se limitan a: C, N, O, P, S, As, Se, y *, Usar una u otra notación queda a criterio del usuario, las tres cadenas que se muestran en la figura de arriba son igualmente válidas. Sin embargo, la nomenclatura aromática con letras minúsculas representa mejor la estructura y es más sencilla, por lo que suele ser la preferida.

- **Isomería en dobles enlaces.** En SMILES se utilizan los símbolos “/” y “\” para representar los enlaces simples que unen a los grupos que definen la isomería cis/E o trans/Z en un doble enlace. Por ejemplo, el dibromoeteno: Br/C=C\Br representa al cis-1,2-dibromoeteno, mientras que Br/C=C/Br representa el isómero trans.
- **Isomería óptica en carbonos sp³.** En SMILES este tipo de estereoquímica se define localmente para cada centro quiral, sin la necesidad de establecer reglas de prioridades para los sustituyentes. En SMILES el sentido en el que se describirán los sustituyentes se indica con los símbolos “@” y “@@”. Si se mira el símbolo @ podemos notar que parece una espiral que gira, desde

el centro hacia afuera, en sentido antihorario. De esta forma, cuando el carbono quiral viene acompañado de un @ significa que los sustituyentes que se enumeran a continuación deben ser ubicados en sentido antihorario.

- **Reacciones.** En la nomenclatura SMILES puede expandirse para representar una reacción completa. Esto es muy poderoso ya que no sólo podemos crear bases de datos de sustancias sino también de reacciones.

Las reacciones se especifican con el símbolo >>, el cual representa una flecha de reacción. A la izquierda del símbolo se ubican los reactivos, los cuales se separan con un punto “.” (en SMILES nunca se usan espacios), y a la derecha los productos siguiendo la misma notación. Aquellas sustancias involucradas en la reacción que no son ni reactivos ni productos (solventes, catalizadores, etc.) se denominan “agentes” y se ubican entre los dos símbolos “>” que representan la flecha de reacción.

2.3. Bases Legales

Según Villafranca D. (2002) “Las bases legales no son más que las leyes que sustentan de forma legal el desarrollo del proyecto” explica que las bases legales “son leyes, reglamentos y normas necesarias en algunas investigaciones cuyo tema así lo amerite”.

Las bases legales de esta investigación se encuentran representadas, en primer lugar, en la **Constitución de la República Bolivariana de Venezuela (1999)**.

Artículo 98. La creación cultural es libre. Esta libertad comprende el derecho a la inversión, producción y divulgación de la obra creativa, científica, tecnológica y humanística, incluyendo la protección legal de los derechos del autor o de la autora sobre sus obras. El Estado reconocerá y protegerá la propiedad intelectual sobre las obras científicas, literarias y artísticas, invenciones, innovaciones, denominaciones, patentes, marcas y lemas de acuerdo con las condiciones y excepciones que establezcan la ley y los tratados internacionales suscritos y ratificados por la República en esta materia.

Artículo 110. El Estado reconocerá el interés público de la ciencia, la tecnología, el conocimiento, la innovación y sus aplicaciones y los servicios de información necesarios por ser instrumentos fundamentales para el desarrollo económico, social y político del país, así como para la seguridad y soberanía nacional. Para el fomento y desarrollo de esas actividades, el Estado destinará recursos suficientes y creará el sistema nacional de ciencia y tecnología de acuerdo con la ley. El sector privado deberá aportar recursos para los mismos. El Estado garantizará el cumplimiento de los principios éticos y legales que deben regir las actividades de investigación científica, humanística y tecnológica. La ley determinará los modos y medios para dar cumplimiento a esta garantía.

Ley de Medicamentos (2000)

Artículo 21. El Ministerio de Salud y Desarrollo Social exigirá el establecimiento de un sistema de aseguramiento que garantice la calidad de los medicamentos, ya sean nacionales o importados en sus etapas de producción y elaboración, distribución, tenencia, dispensación y expendio en todo el Territorio Nacional.

Artículo 59. El Ministerio de Salud y Desarrollo Social, establecerá los programas necesarios para fortalecer los servicios farmacéuticos en el ámbito institucional y de la comunidad.

Ley Orgánica de Ciencia, Tecnología e Innovación (2014)

Artículo 2. Las actividades científicas, tecnológicas, de innovación y sus aplicaciones son de interés público para el ejercicio de la soberanía nacional en todos los ámbitos de la sociedad y la cultura.

Artículo 21. La autoridad nacional con competencia en materia de ciencia, tecnología, innovación y sus aplicaciones creará mecanismos de apoyo, promoción y difusión de invenciones e innovaciones populares, que generen bienestar a la población o logren un impacto económico o social en la Nación.

2.4. Definición de términos básicos

Acetilcolinesterasa: Es una enzima humana de la familia de colinesterasas que se encuentra en los tejidos nerviosos y los glóbulos rojos, cuya función principal es hidrolizar al neurotransmisor acetilcolina. Modular la actividad de la acetilcolinesterasa es una de las maniobras terapéuticas fundamentadas en aliviar las funciones cognitivas en pacientes con la enfermedad de Alzheimer.

Angiotensina II tipo 1: Es un receptor de angiotensina. Tiene efectos vasopresores y regula la secreción de aldosterona. Es un efector importante en el control de la presión y volumen sanguíneo en el sistema cardiovascular. Los antagonistas de los receptores de angiotensina II son fármacos indicados para la hipertensión, nefropatía diabética, e insuficiencia cardíaca.

API: Significa interfaz de programación de aplicaciones. Es un conjunto de definiciones y protocolos que se utiliza para desarrollar e integrar el software de las aplicaciones, permiten que productos y servicios se comuniquen con otros, sin necesidad de saber cómo están implementados. Esto simplifica el desarrollo de las aplicaciones y permite ahorrar tiempo y dinero.

DENV: Acrónimo del virus del dengue, es el agente causal de la enfermedad conocida como dengue, que es la principal enfermedad viral transmitida por artrópodos en el mundo. Es un flavivirus que ingresa por endocitosis y se replica en el citoplasma de la célula infectada, originando tres proteínas estructurales y siete proteínas no estructurales.

DPP4: La dipeptidil peptidasa-4 es una proteína que, en los humanos, se encuentra codificada por el gen DPP4, se ha desarrollado y están siendo utilizados una nueva clase de medicamentos basados en la inhibición de la DPP-4 para el tratamiento de la diabetes

Fármaco: Un fármaco se define como cualquier sustancia (diferente de un alimento o de un dispositivo) que se utiliza para el diagnóstico, el tratamiento, la curación o la prevención de una enfermedad, o para tratar afecciones que repercutan en la estructura o el funcionamiento del organismo.

Firestore: Es una base de datos alojada en la nube. Los datos se almacenan en formato JSON (JavaScript Object Notation, Notación de Objeto de JavaScript) y se sincronizan en tiempo real con cada cliente conectado. Cuando compilas apps multiplataforma con el SDK (Software Development Kit, Kit de Desarrollo de Software) de iOS, Android y JavaScript, los clientes comparten una instancia de Realtime Database (Base de dato en tiempo real) y reciben actualizaciones automáticamente con los datos más recientes.

Ligando: Es una sustancia que usualmente es una molécula pequeña la cual forma un complejo con una biomolécula. En un sentido más estricto, es una molécula que envía una señal al unirse al centro activo de una proteína.

Multiplataforma: Es un término usado para referirse a los programas, sistemas operativos, lenguajes de programación, u otra clase de software, que puedan funcionar en diversas plataformas.

QSAR: Los modelos de relación cuantitativa entre estructura y actividad (QSAR), son modelos matemáticos que pueden utilizarse para predecir las propiedades de destino fisicoquímicas, biológicas y ambientales de los compuestos a partir del conocimiento de su estructura química.

Red Neuronal: Una red neuronal es un modelo simplificado que emula el modo en que el cerebro humano procesa la información. Las unidades de procesamiento se organizan en capas. Hay tres partes normalmente en una red neuronal: una capa de entrada, con unidades que representan los campos de entrada; una o varias capas ocultas; y una capa de salida, con una unidad o unidades que representa el campo o los campos de destino. Las unidades se conectan con fuerzas de conexión variables (o ponderaciones). Los datos de entrada se presentan en la primera capa, y los valores se propagan desde cada neurona hasta cada neurona de la capa siguiente. al final, se envía un resultado desde la capa de salida

CAPÍTULO III

MARCO METODOLÓGICO

3.1. Tipo de investigación

Para Rojas (2003), Es necesario definir o ubicar tipos de investigación o esquemas por seguir adecuados -según la naturaleza del problema por investigar-, para ello existen diferentes clasificaciones basadas en distintos criterios.

Teniendo en cuenta que la presente investigación tiene como objetivo principal el desarrollo de un software que prediga candidatos a fármaco, se determinó que el tipo de investigación del presente trabajo es “Proyecto Especial”, el cual esta descrito según la Mijares H, Garcia E (2007), como Creaciones tangibles, susceptibles de ser utilizadas como soluciones a problemas demostrados, o que respondan a necesidades e intereses de tipo cultural. Se incluyen en esta categoría los trabajos de elaboración de libros de textos y de materiales de apoyo educativo, el desarrollo de software y hardware, prototipos y de productos tecnológicos en general.

3.2. Diseño de la investigación

Para Arias (2006), en este punto se especifica el tipo de investigación según el diseño o estrategia adoptada para responder al problema planteado. Según el diseño, la investigación puede ser documental, de campo o experimental. Así mismo, cada tipo se puede realizar a nivel exploratorio, descriptivo o explicativo.

De igual manera Arias (2006) también describe la investigación de campo como “aquella que consiste en la recolección de datos directamente de los sujetos investigados, o de la realidad donde ocurren los hechos (datos primarios), sin manipular o controlar variable alguna, es decir, el investigador obtiene la información, pero no altera las condiciones existentes”.

Teniendo estos conceptos presentes se definió que esta investigación tiene una tendencia a ser de campo, debido a que los datos necesarios para el desarrollo de la

plataforma son recolectados de una base de datos de acceso libre y no se ve alterada por parte de los investigadores como ocurre en las investigaciones experimentales.

3.3. Nivel de investigación

Según Arias (2006), el nivel de investigación se refiere al grado de profundidad con que se aborda un fenómeno u objeto de estudio. Además de eso para Arias (2006) la investigación descriptiva consiste en “la caracterización de un hecho, fenómeno, individuo o grupo, con el fin de establecer su estructura o comportamiento. Los resultados de este tipo de investigación se ubican en un nivel intermedio en cuanto a la profundidad de los conocimientos se refiere”.

La presente investigación es de tipo descriptivo, ya que comprende la descripción, registro y análisis de todos los parámetros que son necesarios para predecir si una molécula puede o no, ser candidato a fármaco para las enfermedades Dengue, Alzheimer, Diabetes e Hipertensión.

3.4. Población y muestra

“La población, o en términos más precisos población objetivo, es un conjunto finito o infinito de elementos con características comunes para los cuales serán extensivas las conclusiones de la investigación. Ésta queda delimitada por el problema y por los objetivos del estudio” Arias (2006).

“La muestra es un subconjunto representativo y finito que se extrae de la población accesible” Arias (2006)

Para esta investigación se utilizó una muestra intencionada, debido a que solo centros especializados pueden realizar investigaciones sobre diseño y funcionalidades de los medicamentos, de esta manera se tomó como población a un grupo de investigadores del área de farmacología, adscrito a la Facultad de Ciencias y Tecnología en la Universidad de Carabobo, los cuales se encargan de estudios diversos temas de la quimio-informática y estudios con y sobre fármacos, con una muestra de 4 personas de los diferentes escalafones, entre ellos doctores en ciencias químicas y química médica, asistentes de laboratorio e investigadores, para así lograr obtener la mayor información.

3.5. Técnicas e instrumentos de recolección de datos

En cuanto a las técnicas e instrumentos de recolección de datos, Arias (2006) define que las técnicas de recolección de datos son las distintas formas o maneras de obtener la información. Son ejemplos de técnicas; la observación directa, la encuesta en sus dos modalidades: oral o escrita (cuestionario), la entrevista, el análisis documental, análisis de contenido, etc. Por otra parte, los instrumentos son los medios materiales que se emplear para recoger y almacenar la información.

El primer lugar, se utilizó la recopilación documental la cual Torrealba & Rodríguez (2009) definen como “un instrumento o técnica de investigación general cuya finalidad es obtener datos e información a partir de fuentes documentales con el fin de ser utilizados dentro de los límites de una investigación en concreto”. Lo cual fue la revisión bibliográfica de diversas fuentes cuya información fue necesaria y se aplicó en la investigación, en este caso se centró en diseño y desarrollo de medicamentos, así como también en el área metodológica de ingeniería del software.

Por otro lado, se utilizó la técnica de observación directa, los autores Hernández, Fernández, Baptista (2010) expresan que es: “la observación directa consiste en el registro sistemático, válido y confiable de comportamientos o conducta manifiesta”. A través de esta técnica el investigador puede observar y recoger datos mediante su propia observación. sin modificar o alterar las condiciones ambientales, además de la no intervención del procedimiento de recolección de información ya que esto ocasionaría la invalidez de la misma.

También se dispuso de otra técnica para obtener los datos, la entrevista, la cual está definida por Arias (2006), como “más que un simple interrogatorio, es una técnica basada en un diálogo o conversación “cara a cara”, entre el entrevistador y el entrevistado acerca de un tema previamente determinado, de tal manera que el entrevistador pueda obtener la información requerida”.

Cuando el entrevistado está relacionado con el tema y la información, esta ofrece una gran ventaja ya que este puede aportar información base para la investigación, esta fue aplicada de forma no estructurada en la cual se obtuvieron algunos de los

requerimientos y características que la aplicación web planteada necesitaba para elevar su eficiencia en la predicción de buenos candidatos a medicamentos.

3.6. Fases de la investigación

El desarrollo de esta aplicación se llevó a cabo mediante la implementación de la metodología XP, está definida según Perez A (2011) como “una metodología de desarrollo de software basada en los métodos ágiles, que evidencia principios tales como el desarrollo incremental, la participación activa del cliente, el interés en las personas y no en los procesos como elemento principal, aceptando el cambio y la simplicidad”.

Esta metodología fue seleccionada por su sencillez, proceso continuo y robusto, además de su diseño sencillo permitió aplicar términos de requisitos a las funciones internas de la plataforma para que tome forma y evolucione hasta llegar a los resultados esperados de la misma.

La metodología XP tiene un conjunto importante de reglas y prácticas. Esta metodología se divide en las siguientes fases:

- **Fase I: Diagnostico de la situación actual de las plataformas de determinación de posibles candidatos a fármaco, aplicando las técnicas respectivas de recolección de datos.**

Para lograr el objetivo de diagnosticar la situación actual de las plataformas de determinación de posibles candidatos a fármaco, se propuso aplicar las técnicas de recolección de información mediante la utilización de un instrumento como es la entrevista, en una población de 4 personas de centros investigativos adscritos a la Facultad de ciencias y tecnología, la cual se aplicó de forma no estructurada y de la cual obtuvimos información para la siguiente fase.

- **Fase II: Determinación de los requerimientos funcionales y no funcionales para una aplicación web basada en redes neuronales, en la determinación posibles candidatos a fármaco**

En esta fase se establecieron los requerimientos funcionales y no funcionales para la plataforma de determinación de posibles candidatos a fármacos para las enfermedades Dengue, Alzheimer, Diabetes e Hipertensión, basada en algoritmos de redes neuronales para mejorar los tiempos y los costos en el desarrollo de medicamentos, en función al diagnóstico de la situación actual.

- **Fase III: Diseño de una aplicación web para la determinación de posibles candidatos a fármaco usando metodología Extreme Programming (XP)**

En esta fase se elaboró el diseño de la arquitectura y las funciones de la plataforma para la determinación de candidatos a fármacos mediante la aplicación de la metodología XP.

- **Fase IV: Construcción de una aplicación web que prediga propiedades químicas y farmacéuticas para la determinación de posibles candidatos a fármaco utilizando redes neuronales**

Se realizó la codificación de la plataforma para la determinación de posibles candidatos a fármacos, para la realización de la red neuronal se utilizó el lenguaje de programación Python y para el desarrollo de la aplicación web se utilizará lenguajes de programación como JavaScript y el microframework Flask, para el diseño de la interfaz del usuario: HTML, CSS y Bootstrap y como gestor de la base de datos: Firebase.

- **Fase V: Ejecución de un plan de pruebas de software para la verificación de el correcto funcionamiento de la aplicación web.**

En esta fase se ejecutó un plan de pruebas de software, para verificar el correcto funcionamiento de la plataforma y en caso de que algún modulo presentase algún error, se corrigió antes de implementarse.

CAPÍTULO IV

RESULTADOS

Para el desarrollo de sistemas es fundamental seguir especificaciones que permitan a los desarrolladores tener una estructura que asegure que todas las etapas de construcción del sistema, desde la fase inicial de recolección de los requerimientos hasta las pruebas finales del sistema, logrando así que estas sean no solo coherentes si no también tener un mayor control de los recursos destinados para el trabajo de grado.

Esto se logra, utilizando ciertas metodologías, las cuales nos permiten establecer un entorno que facilite y aumente la productividad del equipo que realiza el proyecto, al mismo tiempo impulsa la creación de mejores productos, ya que se definen parámetros para cada tipo de proyecto. Teniendo esto en cuenta, se realizó un estudio de las diversas metodologías, para de esta manera obtener la más factible para desarrollar una aplicación con estas características, concluyendo que la más oportuna es la metodología XP, ya que es una metodología ágil y flexible que se centra en ciertas variables y prácticas que fortalecen la productividad a la hora del desarrollo de softwares, por lo cual, se logró dividir el desarrollo de la aplicación en 5 fases tal como dicta la metodología, siendo estas: diagnostico, determinación, diseño, desarrollo y pruebas, así alcanzar el desarrollo óptimo de la aplicación.

4.1. Fase I: Diagnostico de la situación actual de las plataformas de determinación de posibles candidatos a fármaco, aplicando las técnicas respectivas de recolección de datos.

Mediante la elaboración de un guion para la entrevista, así como la aplicación de los métodos y técnicas mencionados con anterioridad, se logró obtener una visión general de cómo se maneja la investigación en el descubrimiento de fármacos en el país, y como es la utilización de softwares de predicción en este campo, aunado a esto, con esta información se extrajeron todos

los datos necesarios para definir los requisitos funcionales y no funcionales de la plataforma en cuestión.

4.1.1 Actividad I: Investigación de sistemas con similitudes a la plataforma que se planteó desarrollar.

Se investigaron y se probaron diferentes herramientas que tuvieran características similares, como predicciones de IC₅₀ o de bioactividad de las moléculas, tales como, Molinspiration, AVP-IC₅₀, etc. Las pruebas que se realizaron se centraron en entender las funcionalidades, para así adaptar algunas características a nuestra aplicación.

4.1.2 Actividad II: Elaboración del guion para la entrevista y su aplicación.

En esta actividad se elaboró el guion para la entrevista según la investigación previamente realizada, esto con la finalidad de poder analizar las mejoras de características previamente planteadas y poder ajustarlo más a las necesidades de los centros de investigación, además también poder comprender con mayor claridad la etapa inicial del desarrollo de fármacos y así mediante algoritmos de aprendizaje automatizado llegar a la automatización de este proceso. En cuanto a los resultados de esta entrevista tenemos que:

1. Desde su punto de vista, ¿Cómo evalúa usted el desarrollo de fármacos en Venezuela?, ¿Por qué considera que ha disminuido la investigación y desarrollo de estos en el país?

Tabla 1. Respuesta pregunta n° 1 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--|
| 1 | Muy pobre. Porque se le da mucho más valor a lo investigado en el exterior |
| 2 | El desarrollo en el país actualmente es muy bajo, la investigación ha disminuido enormemente por las deficiencias presupuestarias a las que están sometidos los centros de |

| | |
|---|---|
| | investigación del país y el desarrollo también ha disminuido significativamente por la falta de políticas gubernamentales que incentiven la inversión de capital, aunado a políticas económicas erróneas perpetuadas durante años. |
| 3 | Considero que está atrasado respecto a distintos países, ya que se hace uso ahora de herramientas computacionales para el ahorro de tiempo y costos, en el país se hace difícil competir con estas nuevas técnicas a nivel internacional usando el método tradicional que es la investigación dentro del laboratorio. |
| 4 | El desarrollo de fármacos ha disminuido considerablemente en el país debido a factores económicos, estos derivan de la situación ya conocida que atravesamos y que no contamos con el presupuesto para investigaciones. |

Fuente: Salas (2021).

Breve análisis: Con esta pregunta, se evidencia como ha mermado la investigación y desarrollo de fármacos en el país, esto debido a diferentes factores, tanto económicos como sociales, esto debido a que las condiciones económicas los principales investigadores han decidido emigrar del país, las empresas farmacéuticas también han cesado operaciones, logrando así que las investigaciones queden en manos de las universidades, las cuales no cuentan con los recursos necesarios para llevar grandes proyectos, sin embargo, algunas de estas llevan a cabo ciertas investigaciones, que muchas veces no se les da el valor que deberían.

2. ¿Cómo hacen el discernimiento de un compuesto químico para que sea buen candidato a fármaco? ¿Bajo qué criterios lo hacen?

Tabla 2. Respuesta pregunta n° 2 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--|
| 1 | Eso lleva un proceso en cuatro etapas. Asociado a la formulación del producto. Primero se evalúa la estructura molecular. Luego su interacción con el medio y luego su actividad química y biológica. Finalmente, sus efectos secundarios |
| 2 | Evaluando su actividad biológica <i>in vitro</i> ante distintos ensayos clínicos. |
| 3 | Las fases de creación de medicamentos, se dividen en el descubrimiento de diana y diseño de fármacos, la fase de confirmación o desarrollo del fármaco y el registro y post-lanzamiento. |
| 4 | En primer lugar se identifican las dianas terapéuticas, que son las proteínas asociadas a la enfermedad, después de este análisis se realizan diferentes pruebas para ver que compuestos tienen efectos deseados sobre la diana seleccionada siguiendo las reglas de lipinski y su capacidad de inhibir la proteína que presenta la enfermedad, una vez realizado todo este análisis, se identifican los compuestos potenciales, que son los que posiblemente puedan ser desarrollados como futuros medicamentos eficaces. |

Fuente: Salas (2021).

Breve análisis: Con esta pregunta, se pudo identificar con mejor claridad el desarrollo de medicamentos y como se realiza el discernimiento de si un compuesto es buen candidato a farmaco, obteniendo así la información de la importancia de evaluar las reglas de lipinski y el IC₅₀, los cuales determinan la actividad biológica que tiene el compuesto químico con respecto a las dianas terapéuticas seleccionadas.

3. Enumere las causas por las cual la automatización del proceso seria de ayuda en el desarrollo de nuevos candidatos a fármaco

Tabla 3. Respuesta pregunta n° 3 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--|
| 1 | Facilita la cantidad de experimentos a desarrollar y el gasto en materiales e insumos. |
| 2 | Se pueden evaluar muchos más candidatos a la vez y así acelerar enormemente el descubrimiento de nuevos candidatos, además que los recursos pueden enfocarse a candidatos con mayor potencial y probabilidad de funcionar |
| 3 | 1. Ahorro de tiempo. 2. Disminución de costos. 3. Al disminuir el trabajo en el laboratorio implica la menor utilización de reactivos que pudieran ser dañinos para el medio ambiente. 4. Desarrollo a mayor velocidad de fármacos más efectivos. |
| 4 | Un proceso automatizado disminuiría tiempo y recursos que son valiosos. También disminuiría el error humano y nos brinda la oportunidad de trabajar con herramientas actualizadas para mejorar en el campo de la investigación. |

Fuente: Salas (2021).

Breve análisis: Teniendo estos datos, logramos evidenciar la importancia y el impacto que tendría la aplicación ya que con ella se logra ahorrar tiempo, disminuir costos y trabajo de laboratorio, mejorando la productividad del centro investigativo y generando menos huella en el medio ambiente al dejar de utilizar ciertos compuestos químicos que son altamente tóxicos.

4. En su experiencia personal y profesional, ¿Cómo evalúa usted el uso de softwares de predicción en investigaciones dentro de Venezuela?

Tabla 4. Respuesta pregunta n° 4 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--|
| 1 | Son de gran utilidad para explorar a nivel molecular los fenómenos químicos |
| 2 | Son de gran utilidad, como actividad preliminar a las investigaciones experimentales permiten enfocar mejor los recursos y los esfuerzos, permitiendo mejorar los resultados de las investigaciones a largo plazo. |
| 3 | A pesar de que hay investigadores que hacen uso de este tipo de softwares en el país, considero que está muy poco difundido y que queda un largo trayecto para que se afiance su uso en Venezuela. |
| 4 | El uso de software de predicción es una excelente herramienta para diversos análisis. Sin embargo, su uso no es tan común todavía |

Fuente: Salas (2021).

Breve análisis: Aquí se evidencia como el uso de softwares son de gran utilidad como actividad preliminar en investigaciones experimentales, ya que estos dan información necesaria y valiosa, en menos tiempo y con bajo margen de error, sin embargo, en Venezuela falta un largo camino por recorrer, ya que no están acostumbrados a reliazar ensayos *in silico*, sin embargo, cada vez más estas se están incorporando estas técnicas a los laboratorios de investigación.

5. ¿Cómo se evaluaría la calidad de un software de predicción para la determinación de candidatos a fármaco?

Tabla 5. Respuesta pregunta n° 5 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|---|
| 1 | Teniendo en cuenta el error presentado entre el la data de entrenamiento del modelo y la data de prueba, así tener la certeza que los parámetros recibidos por la aplicación estén correctos. |
| 2 | Comparando que los candidatos obtenidos en efecto cumplan con los parámetros de actividad biológica requeridos para el fármaco |
| 3 | Verificando la exactitud de los valores obtenidos mediante la plataforma y los obtenidos mediante las pruebas experimentales. |
| 4 | Teniendo en cuenta, que las pruebas realizadas con la aplicación diseñada concuerden con las pruebas realizadas experimentalmente. |

Fuente: Salas (2021).

Breve análisis: En esta pregunta, los entrevistados concuerdan que la aplicación debe proporcionar una fiabilidad de los resultados, por lo cual nuestra data efectivamente debe ser dividida en dos partes, así poder verificar el error obtenido entre la data de entrenamiento y los parámetros que se predicen con la data de pruebas, así ofrecerles a los investigadores los resultados y ganar la confiabilidad en la plataforma.

6. En promedio, ¿Cuánto sería el tiempo ahorrado al realizar el descarte de compuestos no óptimos para ser candidatos a fármaco?

Tabla 6. Respuesta pregunta n° 6 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--------------------------|
| 1 | Puede ir de meses a años |

| | |
|---|---|
| 2 | Pudieran ahorrarse incluso años |
| 3 | Puede tomarse décadas sin la utilización de software de predicción. |
| 4 | Un software de predicción ahorraría muchos meses de estudios. Aproximadamente, un intervalo de 6 meses a 1 año. |

Fuente: Salas (2021).

Breve análisis: Con esta respuesta, podemos observar el beneficio tan grande que se obtendría al hacer de manera computacional la primera fase del desarrollo de fármacos, ya que con ello se podrían adelantar meses e incluso años en la investigación farmacológica.

7. En su entorno laboral, ¿Aplican herramientas computacionales para agilizar procesos en las investigaciones que realizan?

Tabla 7. Respuesta pregunta n° 7 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|-------------------|
| 1 | Si muy poco |
| 2 | Muy poco |
| 3 | No, no utilizamos |
| 4 | Muy poco |

Fuente: Salas (2021).

Breve análisis: Aquí observamos que algunos centros utilizan herramientas computacionales en algunos de sus procesos, por lo cual ya algunos están relacionados con las aplicaciones computacionales, de igual manera es muy poco común la utilización de estos softwares en las áreas de investigación.

8. Ante una eventual falta de reactivos y/o fondos para investigación ¿Cree usted que el desarrollo de herramientas computacionales de predicción se adapta y ayudan a remediar estas?

Tabla 8. Respuesta pregunta n° 8 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--|
| 1 | Si, además de ser positivo para el medio ambiente esta práctica. |
| 2 | Si, ya que las herramientas computacionales son más económicas, permiten mayor cantidad de ensayos a un menor costo, y la capacidad de análisis puede llegar a ser muy competitiva a estándares internacionales sin tantos recursos, en comparación a los ensayos experimentales que requieren gran cantidad de reactivos y equipos costosos. |
| 3 | Si, además que estas herramientas de predicción brindan un gran conjunto de ventajas y pocas desventajas como lo pueden ser la necesidad de equipos de cómputo potentes, la posibilidad de descubrir un conjunto de nuevos fármacos de manera rápida y sin necesidad de ir a ciegas en un laboratorio, es de gran ayuda en esta industria de desarrollo. |
| 4 | Absolutamente. Este tipo de herramientas son una buena alternativa al momento de escasez de reactivo u otros insumos. |

Fuente: Salas (2021).

Breve análisis: Tener en cuenta las respuestas de esta preguntas es importante, ya que debido a la crisis venezolana, la implementación de técnicas computacionales es la mejor herramienta para superar la crisis y seguir con las investigaciones, además se puede medir la receptividad de los investigadores, que se encuentran en disposición de probar estas herramientas.

9. ¿Cuál cree usted que sería el principal inconveniente a la hora de implementar un software de predicción de candidatos a fármaco a nivel industrial?

Tabla 9. Respuesta pregunta n° 9 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------|--|
| 1 | La plataforma de sistemas operativos |
| 2 | Convencer a los empresarios del potencial de la técnica y sus beneficios, ya que estos no serían inmediatos, dificultando así la inversión de capital en el área. |
| 3 | La poca cantidad de personal capacitado para el uso de este tipo de software y el costo de los equipos de cómputo potentes. |
| 4 | Creo que el principal inconveniente sería la familiarización con este método, primeramente, en el sector industrial, y la disposición de un personal capacitado para la ejecución del mismo. |

Fuente: Salas (2021).

Breve análisis: Una de las incertidumbres de los entrevistados era la plataforma de sistemas operativos, por lo cual entre los requerimientos es hacerlo como aplicación web, ya que así no es necesario instalarlo, si no simplemente tener una computadora con acceso a internet y con un navegador actualizado, esto también abarata costos, y por último se debe realizar una aplicación intuitiva y de fácil acceso para que la capacitación del personal sea más sencilla y esto beneficie a los potenciales empresarios.

10. ¿Qué funciones cree que son necesarias en una plataforma para la determinación de candidatos a fármaco?

Tabla 10. Respuesta pregunta n° 10 entrevista.

| Informante Clave N° | Respuestas |
|--------------------------------------|---|
| 1 | <ul style="list-style-type: none"> · Determinar solubilidad en agua, · Distribución en o/w. · Actividad química y actividad biológica · Variación de los grupos funcionales activos. · Variación de la estructura molecular |
| 2 | Enfermedad a tratar, proteínas involucradas, etc. |
| 3 | Es importante que tenga una forma de ingresar las moléculas, que sea fácil y eficiente, puede ser con SMILES o Inchi, ya que esto garantiza que se están ingresando las moléculas que se desean estudiar, según proteínas ligadas a la enfermedad y sus actividades farmacológicas |
| 4 | Debe poder predecir tanto propiedades fisicoquímicas como farmacológicas, las cuales deben estar de acuerdo a las dianas terapéuticas elegidas, además de eso se deben poder verificar las predicciones previamente hechas, ya que esto es importante para futuras investigaciones. |

Fuente: Salas (2021).

Breve análisis: Para culminar se pregunto cuales deberian ser las funcionalidades prioritarias en la plataforma, de esta manera se determino que la plataforma debia determinar la solubilidad en agua y la distribucion o/w, las cuales se presentan en las reglas de Lipinski, para la actividad quimica y biologica se determina con el IC₅₀, Ademas tambien se definio que la plataforma debe tener una forma de ingresar las moleculas que sea intuitiva y ademas que quede registro de las predicciones para que puedan ser verificados posteriormente.

4.2. Fase II: Determinación de los requerimientos funcionales y no funcionales para una aplicación web basada en redes neuronales, en la determinación posibles candidatos a fármaco.

En esta fase, luego de los análisis correspondientes a los resultados de las preguntas realizadas en la entrevista y tomando en consideración las observaciones, comentarios e ideas realizadas, se procedió a definir los requerimientos funcionales y no funcionales para la aplicación.

4.2.1. Actividad I: Análisis y definición de los requisitos funcionales y no funcionales de la plataforma colaborativa.

Se realizó el análisis requerido de la información obtenida por las respectivas herramientas de recolección de datos, las cuales fueron descritas anteriormente, adicionalmente se tomó en cuenta cada una de las observaciones realizadas en el proceso de la entrevista, con el fin de definir lo mejor posible los requisitos funcionales y no funcionales.

A continuación, se presentan los requerimientos:

· Requerimientos funcionales:

- § Registrar usuarios
- § Gestionar y administrar perfiles
- § Ingreso de moléculas en formatos moleculares o SMILES
- § Predicción de parámetros fisicoquímicos y farmacodinámicos
- § Almacenamiento de las predicciones realizadas
- § Bajo error entre las predicciones realizadas con la data de entrenamiento y la data de prueba
- § Manejo de información en la plataforma

· Requerimientos no funcionales:

- § Proteger el sistema con los estándares básicos de seguridad.
- § Mantener un diseño uniforme en la interfaz en toda la plataforma.

- § Realizar todas las validaciones necesarias para garantizar la integridad de los datos ingresados en la plataforma.
- § Garantizar la escalabilidad del sistema mediante una adecuada documentación, diseño y desarrollo.

4.3. Fase III: Diseño de una aplicación web para la determinación de posibles candidatos a fármaco usando metodología Extreme Programming (XP)

Continuando con las fases de la metodología XP y luego de completar la fase de análisis y haber determinado los requerimientos funcionales y no funcionales del sistema, se diseñaron todos los diagramas para definir la estructura de la aplicación, la estructura de navegación, base de datos, casos de uso y el diseño en general de la plataforma para la predicción de candidatos a fármaco.

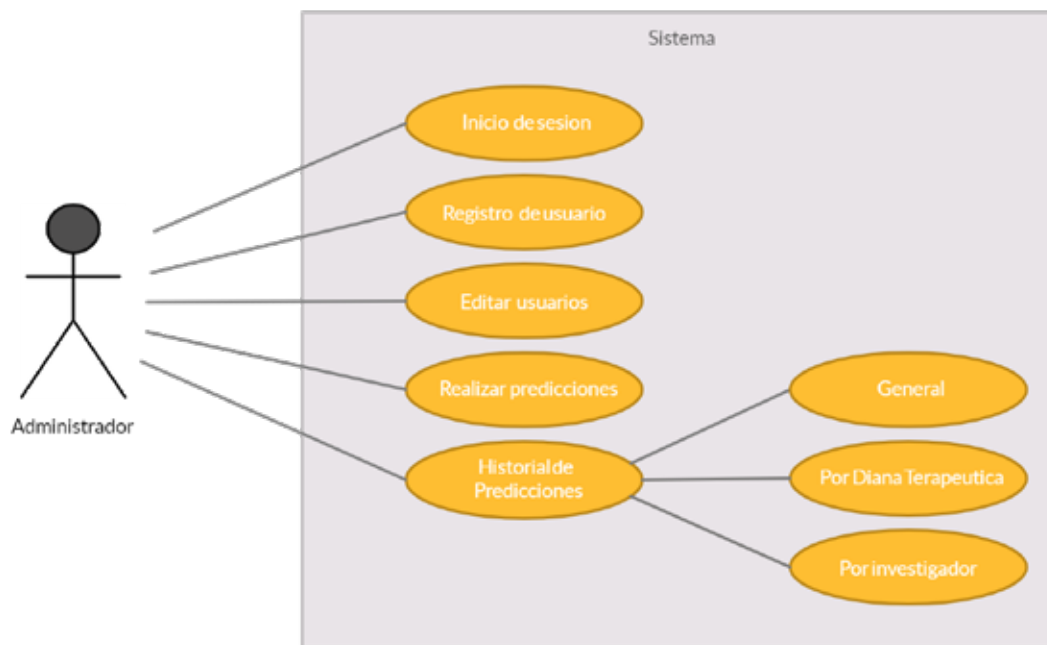
4.3.1. Actividad I: Diseño de casos de uso.

Los casos de uso son parte del conjunto incluido en el lenguaje de Modelado Unificado (UML), esta herramienta es utilizada para facilitar la visualización de las actividades que podrá realizar cada usuario según su nivel en el sistema

A continuación, se presentan los diagramas de uso de cada actor con sus descripciones correspondientes.

- **Administradores:** Estos son los jefes de laboratorio y están encargados de gestionar los perfiles de los usuarios, se ramifican en diferentes grados que son definidos para limitar y representar de forma más adecuada el control que se lleva en una organización, estos también tienen la opción de realizar predicciones y visualizar y gestionar las predicciones realizadas.

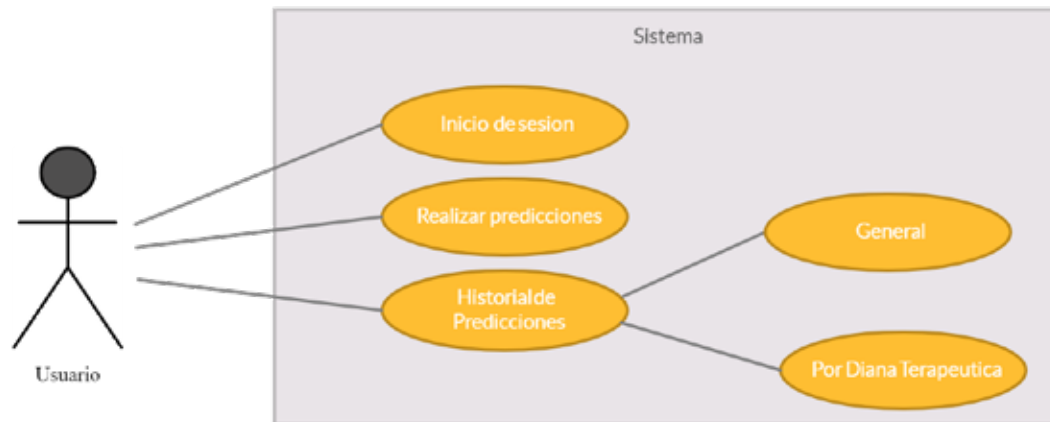
Gráfico 1. Diagrama de caso de uso (Administrador)



Fuente: Salas (2021).

- Usuarios: Los usuarios son todos aquellos que están en los escalafones inferiores al jefe de laboratorio, estos tienen menores funciones disponibles en el sistema, pero son los más indispensables en el mismo. Estos dentro de la plataforma podrán ingresar las moléculas que quieran estudiar, elegir de acuerdo a que enfermedad van a hacer la predicción y también tendrán la opción de gestionar todas las predicciones realizadas según ciertas características.

Gráfico 2. Diagrama de caso de uso (Usuario).



Fuente: Salas (2021).

4.3.2. Actividad II: Descripción de casos de uso.

Tabla 11. Caso de Uso (Iniciar Sesión).

| Iniciar Sesión | |
|--|--|
| Actor: Administrador, Usuario | |
| Objetivo: Ingresar a la aplicación | |
| Precondición: Estar registrado en el sistema | |
| Flujo normal: <ol style="list-style-type: none"> 1. Entrar al inicio de la aplicación. 2. Seleccionar Iniciar sesión. 3. Llenar el formulario con los datos requeridos. 4. Presionar el botón para iniciar sesión. 5. Entrada a la aplicación. | Flujo alterno: <ol style="list-style-type: none"> 1. Los datos ingresados son inválidos. 2. Se genera un mensaje de alerta debajo del campo erróneo e indicando el error. 3. No permite iniciar sesión hasta ingresar datos correctos. |
| Postcondición: Se accede a la aplicación | |

Fuente: Salas (2021).

Tabla 12. Caso de Uso (Registrar Usuario).

| Registrar Usuario | |
|--|---|
| Actor: Administrador | |
| Objetivo: Registrar un usuario | |
| Precondición: El administrador debe haber ingresado al sistema | |
| Flujo normal: <ol style="list-style-type: none"> 1. Entrar al inicio de la aplicación. 2. Iniciar sesión con cuenta de administrador. 3. Presionar botón de registro de nuevo usuario. 4. Llenar el formulario con los datos requeridos. 5. Presionar el botón para registrar. 6. Se borran los campos del formulario | Flujo alterno: <ol style="list-style-type: none"> 1. Los datos ingresados son inválidos. 2. Se genera un mensaje de alerta debajo del campo erróneo e indicando el error. 3. No permite el registro hasta ingresar datos correctos. |
| Postcondición: Se registra el nuevo usuario | |

Fuente: Salas (2021).

Tabla 13. Caso de Uso (Modificar Usuario).

| Modificar Usuario | |
|--|--|
| Actor: Administrador | |
| Objetivo: Modificar un usuario | |
| Precondición: El administrador debe haber ingresado al sistema | |
| Flujo normal: <ol style="list-style-type: none"> 1. Entrar al inicio de la aplicación. 2. Iniciar sesión con cuenta de administrador. | Flujo alterno: <ol style="list-style-type: none"> 1. Los datos ingresados son inválidos. |

| | |
|--|---|
| <ol style="list-style-type: none"> 3. Presionar botón de modificar usuario. 4. Llenar el formulario con los datos requeridos. 5. Presionar el botón para modificar. 6. Se borran los campos del formulario | <ol style="list-style-type: none"> 2. Se genera un mensaje de alerta debajo del campo erróneo e indicando el error. 3. No permite la modificación hasta ingresar datos correctos. |
| <p>Postcondición: Se modifica el usuario</p> | |

Fuente: Salas (2021).

Tabla 14. Caso de Uso (Predicción de Fármacos).

| Predicción de Fármacos | |
|---|---|
| Actor: Administrador, Usuario | |
| Objetivo: Predicción de fármacos | |
| Precondición: El administrador o usuario debe haber ingresado al sistema | |
| <p>Flujo normal:</p> <ol style="list-style-type: none"> 1. Iniciar sesión. 2. Presionar botón de predicción de fármacos. 3. Seleccionar la enfermedad a la cual se realizará la predicción. 4. Dibujar la estructura molecular o ingresar la estructura en formato SMILES. 5. Presionar el botón de realizar predicción | <p>Flujo alterno:</p> <ol style="list-style-type: none"> 1. Los datos ingresados son inválidos. 2. Se genera un mensaje de alerta indicando el error. 3. No se realiza la predicción de los parámetros. |
| <p>Postcondición: Se realiza la predicción de la molécula</p> | |

Fuente: Salas (2021).

Tabla 15. Caso de Uso (Historial de predicciones general).

| Historial de Predicciones general | |
|---|--|
| Actor: Administrador, Usuario | |
| Objetivo: Mostrar historial de predicciones general | |
| Precondición: El administrador o usuario debe haber ingresado al sistema | |
| Flujo normal: <ol style="list-style-type: none"> 1. Iniciar sesión. 2. Presionar botón de Historial. 3. Seleccione opción general. 4. (Opcional) Ordenar por actividad | |
| Postcondición: Se muestra el historial de predicciones general | |

Fuente: Salas (2021).

Tabla 16. Caso de Uso (Historial de predicciones por diana terapéutica).

| Historial de Predicciones por diana terapéutica | |
|--|--|
| Actor: Administrador, Usuario | |
| Objetivo: Mostrar historial de predicciones general | |
| Precondición: El administrador o usuario debe haber ingresado al sistema | |
| Flujo normal: <ol style="list-style-type: none"> 1. Iniciar sesión. 2. Presionar botón de Historial. 3. Seleccione opción diana terapéutica. | |
| Postcondición: Se muestra el historial de predicciones | |

Fuente: Salas (2021).

Tabla 17. Caso de Uso (Historial de predicciones por investigador).

| Historial de Predicciones por investigador |
|---|
| |

| | |
|--|--|
| Actor: Administrador | |
| Objetivo: Mostrar historial de predicciones por investigador | |
| Precondición: El administrador debe haber ingresado al sistema | |
| Flujo normal: | |
| <ol style="list-style-type: none"> 1. Iniciar sesión. 2. Presionar botón de Historial. 3. Seleccione opción de investigador. 4. Ingrese el nombre del investigador. 5. (Opcional) Ordenar por actividad | |
| Postcondición: Se muestra el historial de predicciones | |

Fuente: Salas (2021).

4.3.3. Actividad III: Declaración de módulos y de estados en la plataforma.

Luego de haber definido los casos de usos, se procedió a definir y dividir sus acciones en diferentes módulos para poder hacer uso de interfaces cargadas información no relevante e innecesaria, así lograr una fácil navegación y visualización, garantizando la comodidad para el usuario.

Tabla 18. Módulos de la plataforma.

| Módulos | |
|---------|---|
| Nombre | Descripción |
| Inicio | Esta comprende la página de inicio, que contiene información básica, sobre la aplicación de IA en el descubrimiento de fármacos y sobre las enfermedades estudiadas en la plataforma. |
| Ingreso | Este contiene un formulario para el inicio de sesión. |

| | |
|--|--|
| Inicio (Post ingreso) | Este módulo contiene los botones que permitirán realizar predicciones, verificar el historial y si es administrador, el botón para registrar y modificar usuarios. |
| Registro/Modificaciones (Post ingreso) | Contiene los formularios para registrar nuevos usuarios o modificar los ya existentes. |
| Predicción (Post ingreso) | Contiene el panel donde dibujarán las moléculas, un input donde podrán escribir la molécula en formato SMILES, y los botones para realizar la predicción. |
| Historial (Post ingreso) | Contiene una tabla con las predicciones realizadas, los usuarios podrán acceder al historial general o por diana terapéutica y los administradores también tendrán la opción de discriminar por investigador, además todos estos podrán ser ordenados por actividad. |
| Usuarios (Post ingreso) | Contiene una tabla en el cual están descritos los usuarios con sus datos. |

4.3.4. Actividad IV: Modelado de base de datos.

Seguido de haber determinado los actores y sus casos de uso, se diseñó el modelo de base de datos, en este caso un modelo NoSQL, el cual es una representación gráfica de como estarán almacenados los datos en la aplicación web, y en base a este modelo, se implementó la base de datos.

Gráfico 3. Diagrama de base de datos.

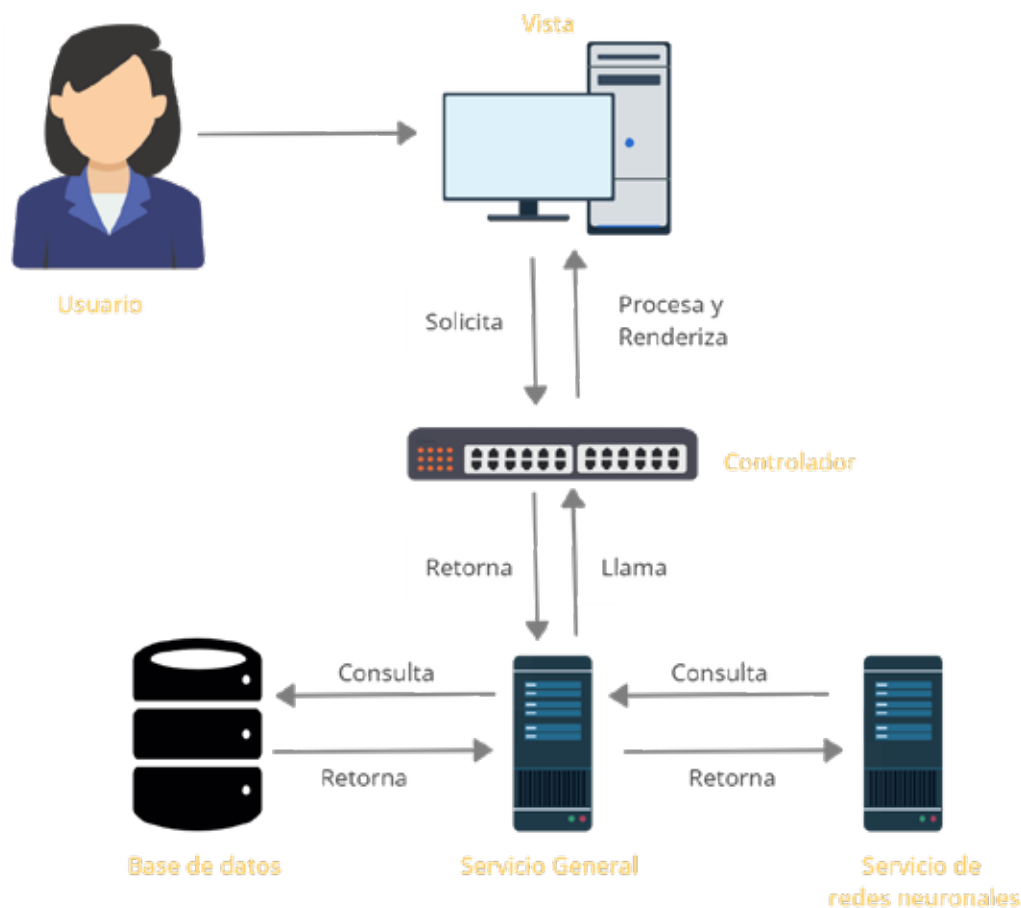


Fuente: Salas (2021).

4.3.5. Actividad V: Descripción de la arquitectura del sistema.

La arquitectura del sistema es un conjunto de patrones que proporcionan un marco que dictan y facilitan el desarrollo del sistema. Para este proyecto se utilizó el patrón Modelo-Vista-Controlador, este plantea separar los componentes del sistema en capas, las cuales son: la interfaz con el usuario, lógica de programación y la interfaz con los datos, de igual manera la lógica se separó en dos servicios separados, los cuales son, un servicio que conlleva las acciones o funciones básicas, que se encarga de realizar las consultas al otro servicio, el cual se encarga de las operaciones que realizan las redes neuronales.

Gráfico 4. Diagrama de la arquitectura del sistema.



Fuente: Salas (2021).

4.3.6. Actividad VI: Servicio de redes neuronales.

El servicio de redes neuronales se presenta como una API, la cual está definida como un conjunto de servicios y funciones que se puedan integrar en aplicaciones externas para el envío y recepción de datos procesados.

Dicha API posee una red neuronal artificial que permite predecir los parámetros fisicoquímicos y farmacodinámicos de los compuestos ingresados, la arquitectura de la misma, se describe a continuación:

Gráfico 5. Arquitectura del servicio de redes neuronales.



Fuente: Salas (2021).

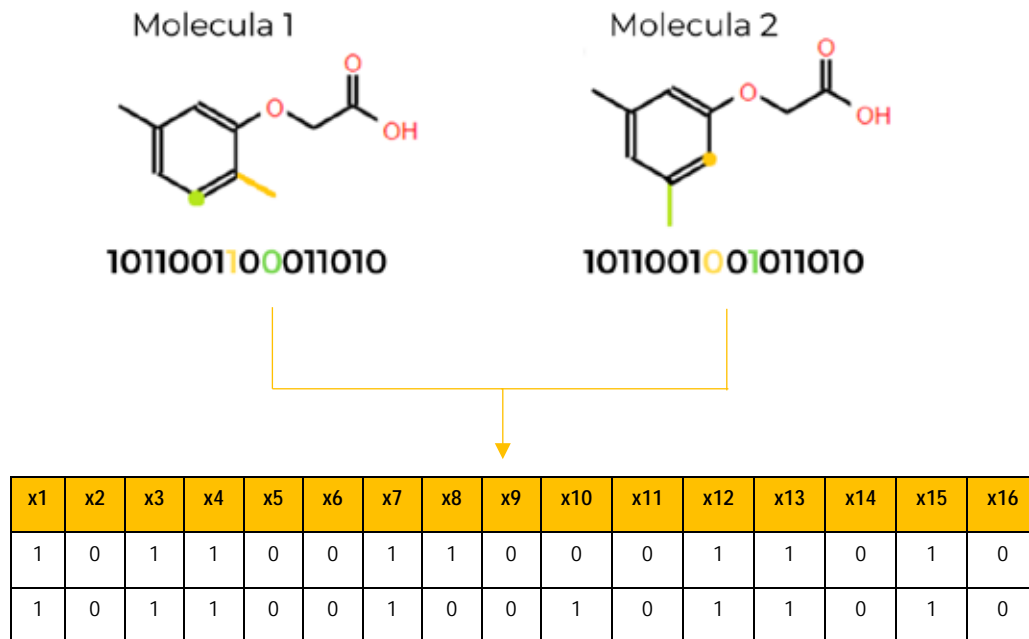
Para la realización de este modelo, se tomaron los datos de la base de datos ChEMBL, la cual es una base de datos open source de moléculas bioactivas con propiedades de tipo farmacéuticas, se obtuvieron las datos de cada target o diana terapéutica las cuales fueron Acetilcolinesterasa (Alzheimer) con 7927 compuestos bioactivos, DENV (Dengue) con 1263 compuestos bioactivos, DPP4 (Diabetes) con 5089 y por último Angiotensina II tipo I (Hipertensión) con 1607 compuestos. Teniendo esto, se procedió a limpiar la data, para así obtener solo los datos que serían útiles para nuestro proyecto, seguido de esto, se le agregó una columna extra a cada una de nuestras tablas con el nombre de la diana terapéutica, con la finalidad de poder diferenciarlas al momento de unir las.

Al tener nuestra data limpia, es decir, con los parámetros de SMILES, IC_{50} y $\log P$, se procedió a evaluar los SMILES, de tal manera de solo dejar los canónicos, esto para evitar que las moléculas se repitieran. Se unió la data y se separó la columna de las dianas terapéuticas a la cual se le aplicó el algoritmo de “one hot encoding”, el cual separó nuestra columna de dianas terapéuticas en cuatro columnas binarias para cada una de ellas, esto fue realizado para que el modelo pudiese interpretar sobre cual diana terapéutica estaba aprendiendo.

Debido a que los modelos matemáticos utilizados para las redes neuronales no pueden interpretar la data en formato SMILES, se procedió a determinar los descriptores moleculares que no son más, que el resultado final de la transformación de la información química codificada dentro de una representación simbólica de una molécula en un número útil, para que así el modelo pudiera entenderlo, por lo cual se

implementó un algoritmo capaz de determinar los sitios activos de las moléculas y eso pasarlo a una matriz binaria, donde cero es inactivo y uno es activo.

Gráfico 6. Esquema general de la transformación a descriptores moleculares.



Fuente: Salas (2021).

Seguido de eso se procedió a determinar los componentes principales de nuestra data, para ello se utilizaron los algoritmos de Feature Selector, PCA y un análisis de varianza, la cual nos proporcionó una data más limpia, para comenzar a implementar el modelo de redes neuronales, de tal manera de evitar que los datos repetidos hicieran un sobreajuste en nuestro modelo.

Los datos enviados y recibidos del modelo se pueden observar a continuación.

Tabla 19. Datos enviados y recibidos del modelo

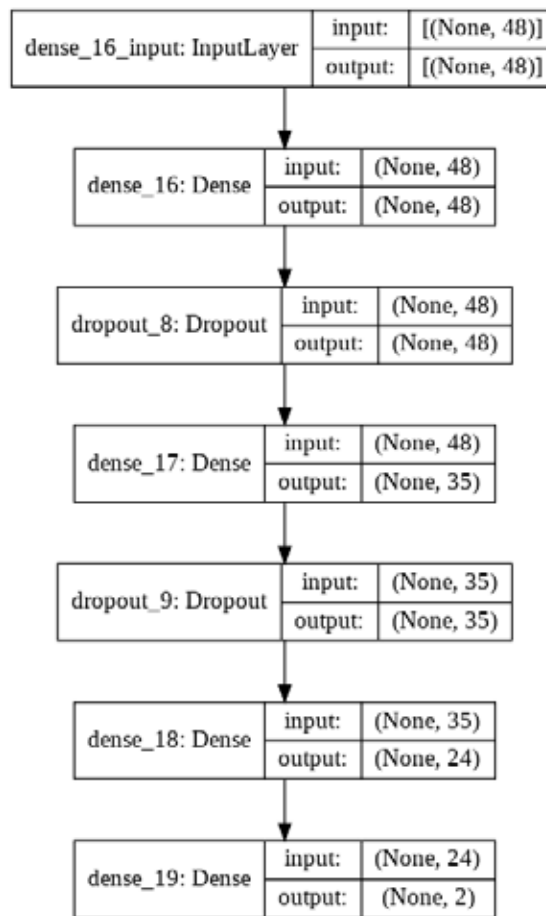
| Parámetros | Descripción | Categoría |
|-------------------------------------|--------------------------|-----------|
| Envío de características | | |
| Arg 1-45 | Descriptores moleculares | Numérico |
| Recepción de características | | |

| | | |
|--------------------|---|----------|
| Resultado 1 | logP (Coeficiente de reparto) | Numérico |
| Resultado 2 | pIC ₅₀ (Concentración inhibitoria media máxima) | |

Fuente: Salas (2021).

El esquema utilizado para el modelo de redes neuronales fue el siguiente:

Gráfico 7. Arquitectura del modelo de redes neuronales



Fuente: Salas (2021).

Para medir el rendimiento del modelo se utilizó el error absoluto medio descrito según la ecuación (1):

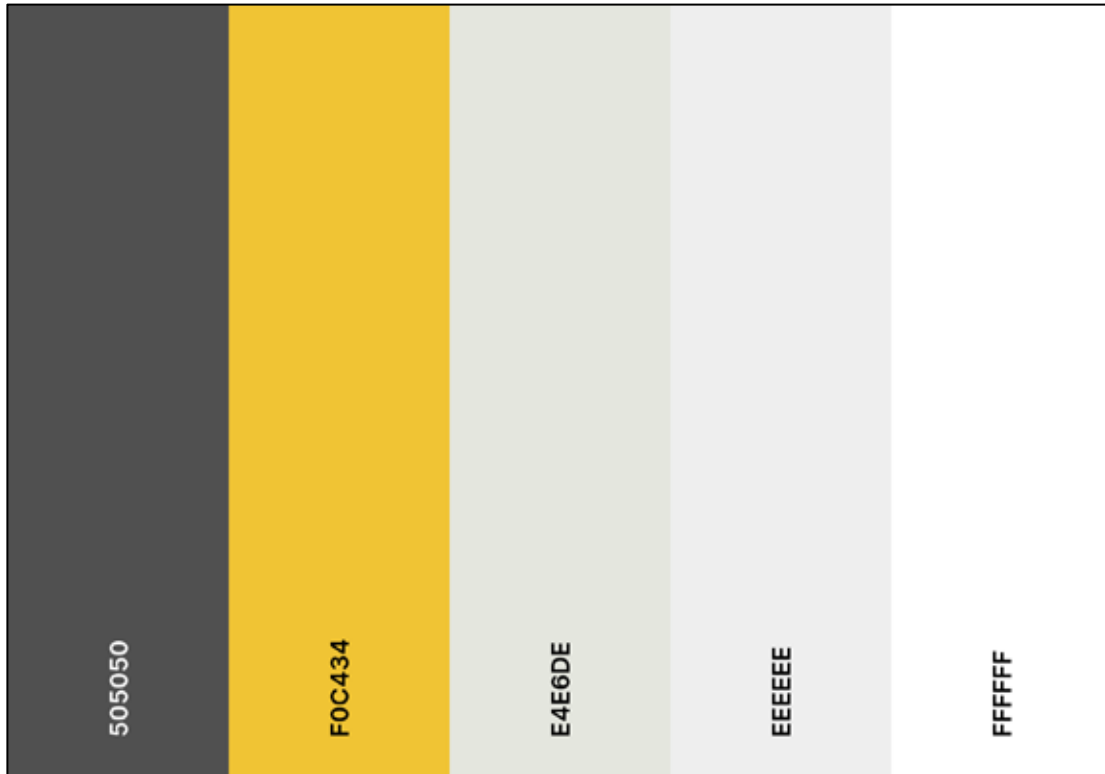
$$MAE = \frac{1}{n}$$

Menu de navegación

Contenido del modulo

Fuente: Salas (2021).

Gráfico 10. Paleta de colores



Fuente: Salas (2021).

A continuación, se presentarán las capturas de la aplicación para determinar candidatos a fármaco utilizando redes neuronales, Los datos personales como los correos, nombres, numeros de telefono fueron ocultados por la privacidad de los usuarios en las capturas:

Gráfico 11. Captura de pantalla (Pantalla de inicio)



Fuente: Salas (2021).

Gráfico 12. Captura de pantalla (Información, Introducción)



Fuente: Salas (2021).

Gráfico 13. Captura de pantalla (Información, Dianas terapéuticas)



Fuente: Salas (2021).

Gráfico 14. Captura de pantalla (Inicio de sesión)



Fuente: Salas (2021).

Gráfico 15. Captura de pantalla (Inicio post-ingreso-administrador)



Fuente: Salas (2021).

Gráfico 16. Captura de pantalla (Inicio post-ingreso-usuario)



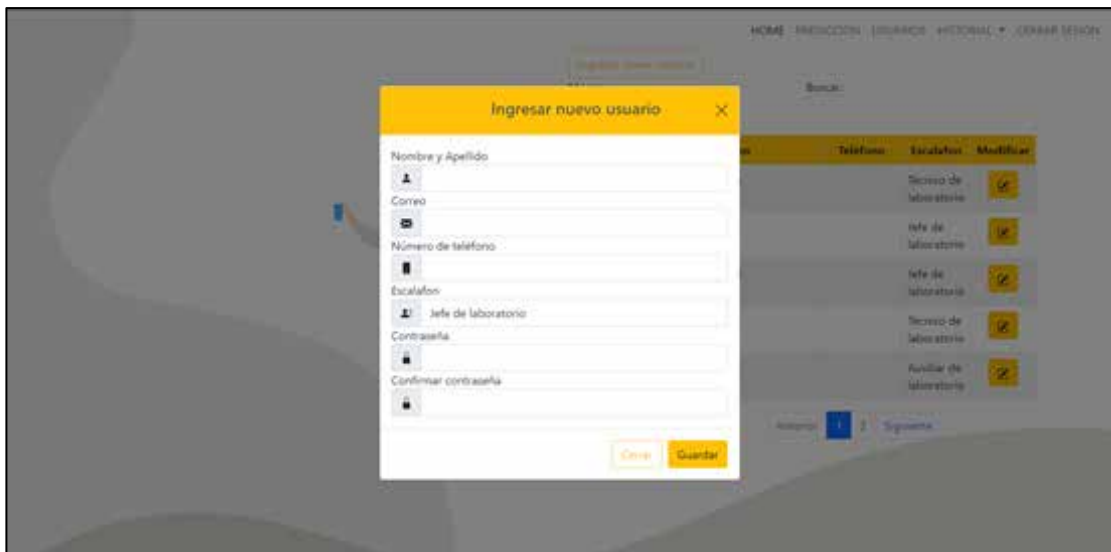
Fuente: Salas (2021).

Gráfico 17. Captura de pantalla (Usuarios-administrador)



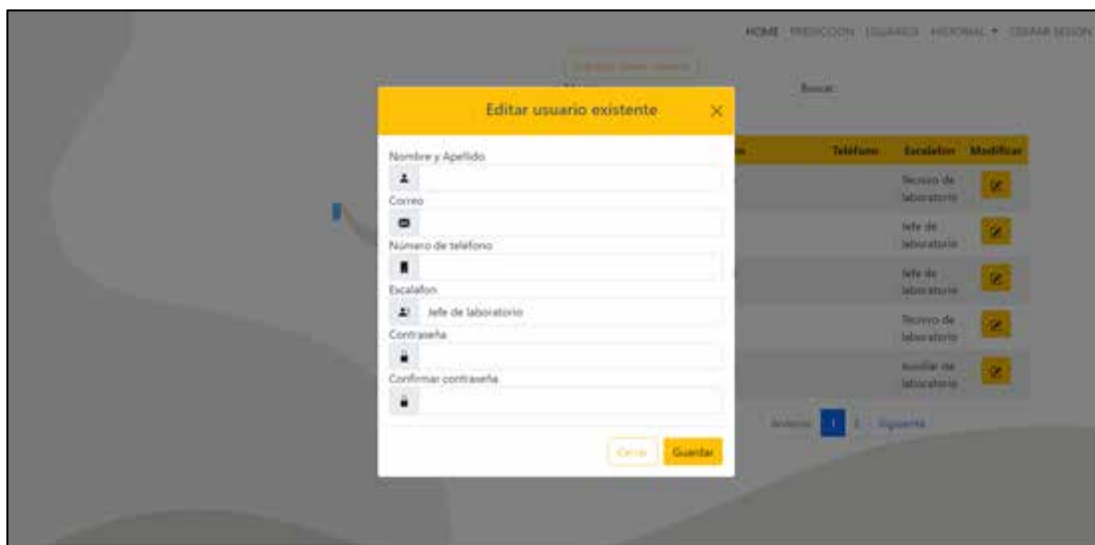
Fuente: Salas (2021).

Gráfico 18. Captura de pantalla (Ingresar usuario-administrador)



Fuente: Salas (2021).

Gráfico 19. Captura de pantalla (Editar usuario-administrador).



Fuente: Salas (2021).

Gráfico 20. Captura de pantalla (Ingreso de moléculas-Administrador)



Fuente: Salas (2021).

Gráfico 21. Captura de pantalla (Ingreso de moléculas-usuario)



Fuente: Salas (2021).

Gráfico 22. Captura de pantalla (Predicción-administrador)

| SMILES | Peso Molecular (g/mol) | Numero de Átomos | H Donadores | H Aceptores | LogP | pIC50 |
|------------------------------------|------------------------|------------------|-------------|-------------|------|-------|
| <chem>CCCCCCCCCCCC(C(=O)O)O</chem> | 284.44 | 20 | 1 | 3 | 4.50 | 3.3 |

Cumple con las reglas de Lipinski

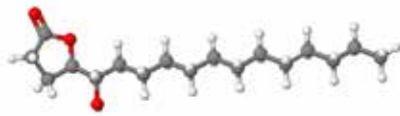
Tiene baja actividad con respecto a la diana terapéutica

Fuente: Salas (2021).

Gráfico 23. Captura de pantalla (Predicción-usuario)

HOME PREDICCIÓN HISTORIAL CERRAR SESIÓN

Predicción



| SMILES | Peso Molecular (g/mol) | Numero de átomos | H Donadores | H Aceptores | LogP | pK50 |
|---------------------------------------|------------------------|------------------|-------------|-------------|------|------|
| <chem>CCCCCCCCCCCC(C)CCCC(=O)O</chem> | 284.44 | 20 | 1 | 3 | 4.50 | 3.5 |

Cumple con las reglas de Lipinski

Tiene baja actividad con respecto a la diana terapéutica

[Inicio](#)

Fuente: Salas (2021).

Gráfico 24. Captura de pantalla (Historial por investigador-administrador)

HOME PREDICCIÓN USUARIOS HISTORIAL CERRAR SESIÓN

Mostrar 5 registros

Buscar:

| Investigador | Diana terapéutica | SMILES | Peso Molecular (g/mol) | Numero de átomos | H Donadores | H Aceptores | LogP | pK50 | Actividad |
|------------------------|-------------------|---------------------------------------|------------------------|------------------|-------------|-------------|-------|------|-----------|
| DPP4 | | <chem>CC(C)CCCCC=O</chem> | 156.26 | 11 | 1 | 0 | 4.08 | 6.50 | Inactivo |
| Acetilcolinesterasa | | <chem>CCCCCCCCCCCC(C)CCCC(=O)O</chem> | 284.44 | 20 | 1 | 1 | 4.50 | 3.50 | Inactivo |
| Angiotensina II tipo I | | <chem>CC(C)C(C)P(=O)(O)O</chem> | 138.10 | 9 | 1 | 2 | 0.24 | 8.47 | Activo |
| DENV | | <chem>CC(N)C(C)C(C)C(C)C</chem> | 130.23 | 9 | 1 | 0 | -1.13 | 4.23 | Inactivo |
| Acetilcolinesterasa | | <chem>Cl=C(C)C(C)C(=O)O(C)C</chem> | 175.40 | 8 | 2 | 1 | 1.34 | 9.50 | Activo |

Mostrando registros del 1 al 5 de un total de 6 registros

Anterior 1 2 Siguiente

Fuente: Salas (2021).

Gráfico 25. Captura de pantalla (Historial general-administrador)

HOME | PREDICCIÓN | USUARIO | HISTORIAL | CERRAR SESIÓN

Mostrar 5 registros

Buscar:

| Diana terapéutica | SMILES | Peso Molecular (g/mol) | Numero de átomos | H Donadores | H Aceptores | LogP | pIC50 | Actividad |
|------------------------|---------------------------------|------------------------|------------------|-------------|-------------|-------|-------|-----------|
| Acetilcolinesterasa | <chem>CCCCCCCCCCCC(CCC)O</chem> | 284.44 | 29 | 3 | 1 | 4.50 | 3.50 | Inactivo |
| Acetilcolinesterasa | <chem>C1=CC(C)C=C1O</chem> | 175.40 | 8 | 2 | 1 | 1.34 | 9.50 | Activo |
| Angiotensina II tipo I | <chem>CC(C)C(=O)O</chem> | 138.10 | 8 | 3 | 2 | 0.24 | 2.35 | Inactivo |
| DENV | <chem>CC(N)C(=O)C</chem> | 130.25 | 9 | 1 | 0 | -2.13 | 9.20 | Activo |
| DENV | <chem>CC(C)=C=C=C(C)C=C</chem> | 178.23 | 13 | 2 | 0 | 2.41 | 1.50 | Inactivo |

Mostrando registros del 1 al 5 de un total de 6 registros

Anterior 1 2 Siguiente

Fuente: Salas (2021).

Gráfico 26. Captura de pantalla (Historial general-usuario)

HOME | PREDICCIÓN | HISTORIAL | CERRAR SESIÓN

Mostrar 5 registros

Buscar:

| Diana terapéutica | SMILES | Peso Molecular (g/mol) | Numero de átomos | H Donadores | H Aceptores | LogP | pIC50 | Actividad |
|------------------------|--------------------------------|------------------------|------------------|-------------|-------------|-------|-------|-----------|
| DPP4 | <chem>CC(C)CCCCC=O</chem> | 156.26 | 11 | 1 | 0 | 4.08 | 9.50 | Activo |
| DPP4 | <chem>CC(C)CCCCC=O</chem> | 156.26 | 11 | 1 | 0 | 4.08 | 9.50 | Activo |
| DENV | <chem>CC(N)C(=O)C</chem> | 130.25 | 9 | 1 | 0 | -2.13 | 9.20 | Activo |
| DENV | <chem>CC(C)=C=C=C(C)C=C</chem> | 178.23 | 13 | 2 | 0 | 2.41 | 1.50 | Inactivo |
| Angiotensina II tipo I | <chem>CC(C)C(=O)O</chem> | 138.10 | 8 | 3 | 2 | 0.24 | 2.35 | Inactivo |

Mostrando registros del 1 al 5 de un total de 7 registros

Anterior 1 2 Siguiente

Fuente: Salas (2021).

Para los efectos de las capturas de pantalla del historial por diana terapéutica se coloco un ejemplo de una sola diana ya que las 4 (cuatro) vistas son similares entre si.

Gráfico 27. Captura de pantalla (Historial por diana terapéutica-administrador)

HOME PREDECIONES USUARIOS HISTORIAL CERRAR SESION

Mostrar 1 registros

Buscar:

| SMILES | Peso Molecular (g/mol) | Numero de átomos | H Donadores | H Aceptores | LogP | pIC50 | Actividad |
|--|------------------------|------------------|-------------|-------------|-------|-------|-----------|
| <chem>C1=CC(OC)C1=O1O1O</chem> | 175.40 | 8 | 2 | 1 | 1.34 | 9.50 | Activo |
| <chem>CC1C1P1=O1O1O</chem> | 138.10 | 8 | 3 | 2 | 0.24 | 2.35 | Inactivo |
| <chem>CC1OCCCCC1=O</chem> | 156.26 | 11 | 1 | 0 | 4.08 | 9.50 | Activo |
| <chem>CC(N)=CC(C)CC(Br)1</chem> | 130.25 | 9 | 1 | 0 | -2.13 | 9.20 | Activo |
| <chem>CCCCCCCCCCCCC(C)CC1=O1O1O</chem> | 284.44 | 20 | 3 | 1 | 4.50 | 3.50 | Inactivo |

Mostrando registros del 1 al 5 de un total de 6 registros

Anterior 1 2 Siguiente

Acetilcolinesterasa

Fuente: Salas (2021).

Gráfico 28. Captura de pantalla (Historial por diana terapéutica-usuario)

HOME PREDECIONES HISTORIAL CERRAR SESION

Mostrar 5 registros

Buscar:

| SMILES | Peso Molecular (g/mol) | Numero de átomos | H Donadores | H Aceptores | LogP | pIC50 | Actividad |
|-----------------------------------|------------------------|------------------|-------------|-------------|-------|-------|-----------|
| <chem>CC(N)=CC(C)CC(Br)1</chem> | 130.25 | 9 | 1 | 0 | -2.13 | 9.20 | Activo |
| <chem>CC1C1P1=O1O1O</chem> | 138.10 | 8 | 3 | 2 | 0.24 | 2.35 | Inactivo |
| <chem>CC1OCCCCC1=O</chem> | 156.26 | 11 | 1 | 0 | 4.08 | 9.50 | Activo |
| <chem>C1=CC(OC)C1=O1O1O</chem> | 175.40 | 8 | 2 | 1 | 1.34 | 9.50 | Activo |
| <chem>CCOC=OC=OC=C1OCC=O1O</chem> | 178.21 | 13 | 2 | 0 | 2.41 | 1.50 | Inactivo |

Mostrando registros del 1 al 5 de un total de 6 registros

Anterior 1 2 Siguiente

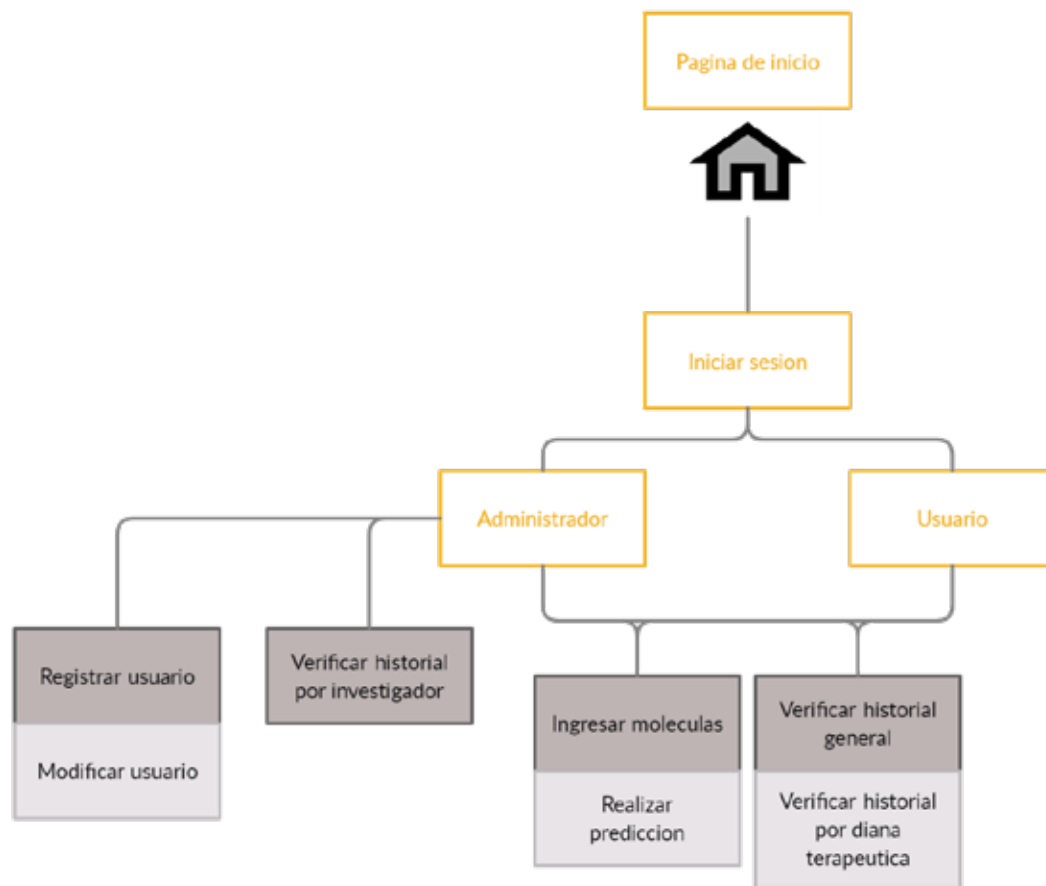
Acetilcolinesterasa

Fuente: Salas (2021).

4.3.8. Actividad VIII: Estructuración de las secciones de la plataforma.

En esta actividad se plasmó la forma esquemática el funcionamiento de cada módulo según al actor que realice la actividad y cuales actividades puede realizar en la aplicación. De igual manera, indica cuales son las direcciones que debe seguir para acceder a ellas

Gráfico 29. Carta estructurada (Mapa de navegación).



Fuente: Salas (2021).

4.4. Fase IV: Construcción de una aplicación web que prediga propiedades químicas y farmacéuticas para la determinación de posibles candidatos a fármaco utilizando redes neuronales

Para la construcción de la aplicación se dividió el proyecto en tres (3) fases, en primer lugar, la comprensión y desarrollo del servicio de procesamiento y gestión de las solicitudes del usuario o (Backend), por otra parte, el desarrollo de la interfaz del usuario (Frontend) cumpliendo debidamente con los requerimientos ya previamente determinados y por último el proceso de entrenamiento y prueba

de las redes neuronales, estos procesos se llevaron a cabo de manera simultánea siguiendo los esquemas realizados en la etapa anterior.

4.5. Fase V: Ejecución de un plan de pruebas de software para la verificación de el correcto funcionamiento de la aplicación web.

En el proceso de codificación y luego de este, se realizaron y aplicaron diversas pruebas en cada módulo, componente y servicio de la aplicación, con el objetivo final de comprobar que la plataforma tuviese una buena funcionalidad y validar así los resultados expresados en la misma. Se realizaron pruebas de caja blanca y caja negra.

Pruebas de caja negra:

En las pruebas de caja negra el Tester solo se centra en las entradas y salidas de la aplicación, sin preocuparse por el contenido interno. Lo que pase por dentro es indiferente, solo importa que, si se realiza cierta acción, la salida sea la indicada según los requerimientos.

Tabla 20. Inicio de sesión

| CASO DE PRUEBA | | |
|---------------------------|---|-------------------------|
| Número de prueba | Caso de Uso | Inicio de sesión |
| 1 | Estrategia | Prueba de caja negra |
| Descripción | El usuario desea ingresar en la plataforma | |
| Entradas | Correo y contraseña | |
| Resultado Esperado | El usuario ingresa exitosamente en la plataforma | |
| Resultado | Exitoso | |
| Observación | El usuario no tuvo dificultades e ingreso de manera exitosa al sistema. | |

Fuente: Salas (2021).

Tabla 21. Registro de usuario

| CASO DE PRUEBA | | |
|------------------------------|---|----------------------|
| Número de prueba 2 | Caso de Uso | Registro de usuario |
| | Estrategia | Prueba de caja negra |
| Descripción | El usuario desea registrar un usuario. | |
| Entradas | Nombre, correo, teléfono, escalafón, contraseña | |
| Resultado Esperado | El administrador registra exitosamente al usuario y este último tiene acceso a las funciones de la misma | |
| Resultado | Exitoso | |
| Observación | El administrador pudo registrar con éxito al usuario y posteriormente pudo acceder a las utilidades del sistema con facilidad | |

Fuente: Salas (2021).

Tabla 22. Dibujo de las moléculas a ingresar

| CASO DE PRUEBA | | |
|------------------------------|--|-------------------------|
| Número de prueba 3 | Caso de Uso | Dibujo de las moléculas |
| | Estrategia | Prueba de caja negra |
| Descripción | El usuario dibujara las moléculas en el espacio designado para ello. | |
| Entradas | Dibujo de la molécula | |
| Resultado Esperado | El usuario carga exitosamente la molécula y el sistema la transforma a estructura SMILES | |
| Resultado | Exitoso | |
| Observación | El usuario ingreso de manera sencilla la molécula que deseaba estudiar y el sistema la registro y transformo exitosamente. | |

Tabla 23. Ingreso de diana terapéutica para la predicción

| CASO DE PRUEBA | | |
|-------------------------------------|--|--|
| Número de prueba 4 | Caso de Uso | Ingreso de moléculas y diana terapéutica |
| | Estrategia | Prueba de caja negra |
| Descripción | El usuario tratara de ingresar la diana terapéutica para ala predicción | |
| Entradas | Selección de la diana terapéutica | |
| Resultado Esperado | El usuario carga exitosamente la diana terapéutica y realiza la predicción | |
| Resultado | Fallido | |
| Observación | El usuario tuvo dificultades al ingresar la diana terapéutica, expresando que la forma de selección era tediosa. | |
| Solución | Se edito el módulo cambiando el botón de la diana terapéutica por un botón desplegable dentro del módulo de la predicción. | |

Fuente: Salas (2021).

Tabla 24. Historial general y por diana terapéutica

| CASO DE PRUEBA | | |
|-------------------------------------|---|---|
| Número de prueba 5 | Caso de Uso | Historial general y por diana terapéutica |
| | Estrategia | Prueba de caja negra |
| Descripción | El usuario buscara el historial general y por diana terapéutica | |
| Entradas | Selección del tipo de historial que desea buscar | |

| | |
|---------------------------|--|
| Resultado Esperado | El usuario observa en la interfaz las predicciones realizadas con anterioridad según la opción ingresada |
| Resultado | Exitoso |
| Observación | El usuario selecciono el tipo de historial deseado y este fue mostrado por pantalla |

Fuente: Salas (2021).

Tabla 25. Historial según investigador

| CASO DE PRUEBA | | |
|------------------------------|---|------------------------------|
| Número de prueba 6 | Caso de Uso | Historial según investigador |
| | Estrategia | Prueba de caja negra |
| Descripción | El administrador buscara las predicciones realizadas según el investigador | |
| Entradas | Selección del investigador | |
| Resultado Esperado | El usuario consigue reflejar en la interfaz el historial de las predicciones por investigador | |
| Resultado | Fallido | |
| Observación | El usuario tuvo dificultades al escoger el investigador que al cual quería verificarle el historial | |
| Solución | Se edito el módulo cambiando el botón desplegable con los investigadores, colocando un input box como filtro. | |

Fuente: Salas (2021).

Pruebas de caja blanca:

Es el testing sobre el código fuente de la aplicación, y, en consecuencia, sobre los diferentes algoritmos y estructuras de datos utilizados.

Básicamente, el tester selecciona distintos valores de entrada para examinar cada uno de los posibles flujos de ejecución del programa y cerciorarse de que se devuelven los valores de salida adecuados.

Tabla 26. Privacidad de información

| CASO DE PRUEBA | | |
|------------------------------|---|-----------------------|
| Número de prueba 7 | Caso de Uso | Privacidad de datos |
| | Estrategia | Prueba de caja blanca |
| Descripción | El usuario ingresado al sistema trata de acceder al módulo de usuarios editando datos en cache o realizando peticiones al servidor con herramientas externas | |
| Entradas | Identificador del usuario, identificador de la página, entre otras cosas | |
| Resultado Esperado | El usuario no logra de ninguna forma acceder o ver la información del módulo de usuarios. | |
| Resultado | Exitoso | |
| Observación | Las rutas del servidor están protegidas por el uso de sesión y validaciones adicionales en base a los datos recibidos, por lo tanto, las consultas son rechazadas | |

Fuente: Salas (2021).

Tabla 27. Confirmación de la contraseña

| CASO DE PRUEBA | | |
|------------------------------|--|-----------------------|
| Número de prueba 8 | Caso de Uso | Privacidad de datos |
| | Estrategia | Prueba de caja blanca |
| Descripción | El usuario al cual el administrador está registrando, no confirma o confirma la contraseña erradamente | |

| | |
|---------------------------|---|
| Entradas | Nombre, correo, teléfono, escalafón, contraseña |
| Resultado Esperado | El usuario no se puede registrar hasta que las contraseñas no coincidan |
| Resultado | Exitoso |
| Observación | Los campos de contraseña están validados, para que el usuario solo pueda registrarse si las contraseñas coinciden |

Fuente: Salas (2021).

Tabla 28. Redireccionamiento por falta de datos de sesión

| CASO DE PRUEBA | | |
|------------------------------|---|---|
| Número de prueba 9 | Caso de Uso | Redireccionamiento por falta de datos de sesión |
| | Estrategia | Prueba de caja blanca |
| Descripción | El usuario trata de acceder a un módulo del sistema mediante la ruta sin haber iniciado sesión. | |
| Entradas | Ruta de la sección de organización. | |
| Resultado Esperado | El usuario no accede al módulo debido a que debe iniciar sesión primero y se le presentan mensajes de error para indicar esto | |
| Resultado | Exitoso | |
| Observación | El usuario no logra acceder al sistema hasta haber iniciado sesión | |

Fuente: Salas (2021).

Tabla 29. Predicción de candidatos a fármaco

| CASO DE PRUEBA | | |
|------------------------------|--------------------|------------------------------------|
| Número de prueba 5 | Caso de Uso | Predicción de candidatos a fármaco |
| | Estrategia | Prueba de caja blanca |

| | |
|---------------------------|---|
| Descripción | El usuario ingresa la molécula y escoge la diana terapéutica |
| Entradas | Molécula, diana terapéutica |
| Resultado Esperado | Se muestran los parámetros y se indica la actividad de la molécula |
| Resultado | Fallido |
| Observación | Algunos de los parámetros no daban los resultados esperados |
| Solución | Se realizó un cambio en las funciones utilizadas para el entrenamiento de la red neuronal y así poder obtener una mayor precisión en las predicciones |

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

Luego de examinar los resultados obtenidos en cada una de las fases ya descritas, junto con el desarrollo y despliegue del sistema, con sus respectivas pruebas, se llegó a una serie de desenlaces relacionados a los objetivos definidos en la presente investigación.

En la fase inicial o primera fase referente al diagnóstico de las plataformas similares al sistema desarrollado, aplicando técnicas de recolección de datos, se logró realizar una evaluación para la elaboración de sistemas de predicción para determinar posibles candidatos a fármaco, donde se logró examinar herramientas existentes para realizar ciertas funciones, además de conceptualizar los parámetros necesarios que determinan si un compuesto podría ser un buen candidato a fármaco.

En la segunda fase, en esta fase, se logró determinar haciendo uso del recurso de recolección de datos antes mencionado, los requerimientos que debía presentar la plataforma desarrollada, comenzando con los requerimientos funcionales del sistema, entre los más destacables, gestionar y administrar perfiles, ingreso de moléculas en formatos moleculares o SMILES, además de predicción de parámetros fisicoquímicos y farmacodinámicos de acuerdo a las diferentes dianas terapéuticas, aplicando algoritmos de aprendizaje automatizado, es decir, el uso de redes neuronales artificiales para el cálculo y estimación de los parámetros requeridos, registro de usuarios e historial de las predicciones realizadas entre otros.

Con respecto a los requerimientos no funcionales, se determinó como bases fundamentales la seguridad proporcionada en el mismo, que corresponde a la privacidad de los diversos usuarios y roles en el sistema presentado, además

también se crearon interfaces fáciles e intuitivas para que la herramienta fuese utilizada de la manera correcta.

En la tercera fase, se estableció el modelado para la aplicación de determinación de candidatos a fármaco mediante la metodología XP, en los cuales se elaboraron todos los diagramas y tablas requeridas por las bases de la ingeniería de software, en los que se pueden destacar, el diagrama de base de datos, la representación y especificación de los caso de uso, con su respectiva explicación por roles de usuario y la arquitectura del sistema, por otro lado, se plantearon y maquetaron los diseños que posteriormente fueron implementados en la plataforma, realizando un diseño base y estipulando la paleta de colores, para mantener la uniformidad a lo largo de toda la aplicación, también es importante destacar, que las tonalidades usadas, fueron designadas para dar una perspectiva de limpieza y estimular la creatividad.

Por su parte en la **cuarte fase,** se desarrolló la plataforma de predicción de candidatos a fármaco para las enfermedades Dengue, Alzheimer, Hipertensión y Diabetes, utilizando herramientas computacionales, comenzando con el frontend con HTML, CSS, JavaScript y el framework Bootstrap, para el sistema de dibujo de moléculas se utilizó un API llamado JSME, el cual permite el dibujo de estructuras moleculares para luego convertirlas en formato SMILES, y la base de datos se realizó mediante un entorno no relacional (NoSQL), como lo es firebase, también fue desarrollada la API que implementa las redes neuronales utilizando Tensorflow, el cual es una librería desarrollada por Google, la cual implementa los métodos necesarios para el entrenamiento de las redes neuronales, Keras, una librería que se ejecuta para Tensorflow para así agilizar el proceso de aprendizaje, para el manejo de datos se utilizó una librería de Python llamada Scikit-learn, la cual permitió realizar la separación de datos en aprendizaje y pruebas, además de poder seleccionar las características más relevantes de nuestra data, para el manejo de las estructuras químicas se utilizaron las librerías rdkit, openbabel y PaDEL, por

último fue utilizado el microframework Flask, como el marco de desarrollo web de Python para el manejo de las solicitudes y funciones de la API.

En la última fase, según lo estipulado en el trabajo de grado, se ejecutó el plan de pruebas para verificar el correcto funcionamiento de todos los módulos que integran la plataforma.

Con el desarrollo de esta aplicación se pueden ahorrar meses de investigación ya que la predicción del IC_{50} experimentalmente, puede llevar días, ya que se llevan a cabo por métodos húmedos lo cuales necesitan de mucha preparación, además de eso también se ahorra en dinero, ya que se reduce la necesidad de comprar materiales, equipos y reactivos, los cuales son costosos y finitos, además también se contribuye a la menor contaminación ya que muchos de estos reactivos son altamente contaminantes, por lo cual también se da una contribución no solo al área de química computacional y farmacológica si no también a la química verde.

5.2. Recomendaciones

En relación con la información que se obtuvo durante el desarrollo de las fases de proyecto, estructuradas mediante la metodología XP, para futuras aplicaciones del sistema propuesto, se recomienda aplicar las recomendaciones expuestas a continuación:

Teniendo en cuenta que los sistemas computacionales se encuentran en constante evolución y cambio, es importante tener en cuenta que los requisitos funcionales y no funcionales varían con el tiempo, lo que significa que durante el uso del sistema pueden realizarse actualizaciones y/o cambios en los diferentes módulos que lo componen.

La adición de nuevas dianas terapéuticas, puede ser muy beneficioso para la elaboración de sistemas con mayor cantidad de información que pueda ser enfocada a más líneas de investigación, teniendo un alcance mayor en las diferentes ramas de la investigación farmacológica.

Teniendo en cuenta la recomendación anterior, también pueden ser adicionados nuevos parámetros para el aprendizaje de las redes neuronales, tales como factores genéticos, los cuales pueden repercutir en mejores resultados para enfermedades con alta predisposición genética como lo es el cáncer, enfermedades autoinmunes, entre otros.

También se podrían aplicar otras herramientas del aprendizaje automatizado, como los algoritmos de bosques aleatorios (Random forest) o maquinas de vectores de soportes (Support vector machine) para comparar los rendimientos con respecto a las redes neuronales.

REFERENCIAS BIBLIOGRÁFICAS

Electrónicas:

- Bjerrum, E. J. (2017). **SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules.** Recuperado de: <http://arxiv.org/abs/1703.07076>
- Brown, F. K., Sherer, E. C., Johnson, S. A., Holloway, M. K., & Sherborne, B. S. (2017). **The evolution of drug design at Merck Research Laboratories.** *Journal of Computer-Aided Molecular Design*, 31(3), 255-266. Recuperado de: <https://doi.org/10.1007/s10822-016-9993-1>
- El Nacional. (2017). **Alzheimer.** Recuperado de: https://www.elnacional.com/sociedad/salud/pacientes-familiares-sufren-desmemoria-del-estado_210567/
- El Sevier. (2020). **Fases de desarrollo de un nuevo fármaco.** Recuperado de: <https://www.elsevier.com/es-es/connect/medicina/edu-fases-de-desarrollo-de-un-nuevo-farmaco>
- EVESCAM. (2017). **Estudio Venezolano de Salud Cardio-Metabólica (EVESCAM): Diseño e Implementación.** Recuperado de: http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0535-51332017000100006
- Hornos, L. (2017). **Introducción a SMILES.** Recuperado de:

<https://medium.com/@hexagono/introducción-a-smiles-dibujando-moléculas-en-el-bloc-de-notas-3c0df0d09c01>

- Indacea. (2018). **Diana terapéutica**. Recuperado de: <https://indacea.org/que-es-una-diana-terapeutica/>
- Kalliokoski, T., Kramer, C., Vulpetti, A., & Gedeck, P. (2013). **Comparability of Mixed IC 50 Data-A Statistical Analysis**. *PLoS ONE*, 8(4), 61007. Recuperado de: <https://doi.org/10.1371/journal.pone.0061007>
- Kuntz, I. D., Chen, K., Sharp, K. A., & Kollman, P. A. (1999). **The maximal affinity of ligands**. *Proceedings of the National Academy of Sciences of the United States of America*, 96(18), 9997-10002. Recuperado de: <https://doi.org/10.1073/pnas.96.18.9997>
- Merck Sharp and Dohme. (2016). **Proceso de investigación de un fármaco | MSD Salud**. Recuperado de: <https://www.msdsalud.es/informacion-practica/proceso-investigacion-farmaco.html>
- Mishra, H., Singh, N., Lahiri, T., & Misra, K. (2019). **A comparative study on the molecular descriptors for predicting drug-likeness of small molecules**. *Bioinformatics*, 3(9), 384-388. Recuperado de: <https://doi.org/10.6026/97320630003384>
- Moreno, E., Béjar, A., & Armengol, J. (2015). **Aprendizaje automático**. Recuperado de: www.edicionsupc.es
- OPS/OMS Venezuela. (2020). **Dengue**. Recuperado de: https://www.paho.org/venezuela/index.php?option=com_content&view=article&id=60:dengue&Itemid=212
- Organizacion Mundial de la Salud. (2016). **OMS | Preguntas y respuestas sobre la hipertensión. WHO**. Recuperado de: <http://www.who.int/features/qa/82/es/>
- Patankar, S. J., & Jurs, P. C. (2018). **Prediction of IC50 Values for ACAT Inhibitors from Molecular Structure**. *Journal of Chemical Information and Computer Sciences*, 40(3), 706-723. Recuperado de: <https://doi.org/10.1021/ci990125r>
- Qureshi, A., Tandon, H., & Kumar, M. (2015). **AVP-IC50Pred: Multiple machine learning techniques-based prediction of peptide antiviral activity in terms of half maximal inhibitory concentration (IC50)**. *Biopolymers*, 104(6), 753-763. Recuperado de: <https://doi.org/10.1002/bip.22703>
- Romano, F. Nissen, M. Paredes, H. Carlos, A. (2007). **ENFERMEDAD DE ALZHEIMER**. Recuperado de: <http://www.nlm.nih.gov/medlineplus/spanish/ency/article>
- Torrealba, C., y Rodríguez, Y. (2009, marzo). **Técnicas de Investigación**

Documental: LA RECOPIACIÓN DOCUMENTAL COMO TÉCNICA DE INVESTIGACIÓN.

Recuperado de:
<http://dani14238551.blogspot.com/2009/03/la-recopilacion-documental-como-tecnica.html>

Torres, E. M. (2008). **Dengue. Estudios Avancados**, 22(64), 33-52. Recuperado de:
<https://doi.org/10.1590/s0103-40142008000300004>

Universidad de la Rioja. (2020). **Análisis predictivo**. Recuperado de:
<https://www.unir.net/ingenieria/revista/noticias/analisis-predictivo/549205008626/>

Impresas:

American Diabetes Association. (2005). **Diagnosis and Classification of Diabetes Mellitus**. Volumen 18. Estados Unidos. Editorial: Asociación Americana de Diabetes

Arias, F. G. (2006). **El proyecto de investigación**. 6ta Edición. Caracas. Editorial: Episteme.

Figueroa, J., y Dávila, M. (2004). **Química Física del Ambiente y de los Procesos Medioambientales**. Primera edición. Barcelona. Editorial: Reverte

Henderson, D. (2007). **DAMA-DMBOK: Guía Del Conocimiento Para La Gestión De Datos**. Edición en Español. New Jersey. Editorial: DAMA International.

Hernández Sampieri, R., Fernández Collado, C., y del Pilar Baptista, M. (2010). **Metodología de la investigación**, 5ta Edición. Mexico DF. Editorial: Mc Graw Hill

Molina, J. (2007). **Implantación de Aplicaciones Informáticas de Gestión**. Primera edición. Madrid. Editorial Vision Net.

Osorio, F. L. (2008). **Base de datos relacionales**. Primera edición. Barcelona. Editorial: Instituto Tecnológico Metropolitano.

Perez A, O. A. (2011). **Enfoques metodológicos para el desarrollo de Software RUP – MSF – XP - SCRUM**. Volumen 10. Bogota. Editorial: Inventum.

Rojas, E. (2003). **El usuario de la Información**. Primera edición. Costa Rica. Editorial: Universidad Estatal a Distancia.

Mijares H y Garcia L. (2007). **Normas para la elaboración y presentación de los anteproyectos, proyectos y trabajos de grado**, Primera edición. Carabobo. Editorial: Universidad José Antonio Páez.

Villafranca, D. (2002). **Metodología de la Investigación**. San Antonio de los altos. Primera edición. Editorial Fundaca.

Voet, D., & Voet, J. G. (2006). **Bioquímica**. 3ra Edición. Buenos aires Editorial:
Medica Panamericana.