Imagen que contiene señal, reloj, firmar

Descripción generada automáticamente

**INSTITUTO POLITÉCNICO NACIONAL**

**(IPN)**

**ESCUELA SUPERIOR DE CÓMPUTO**

**(ESCOM)**

*Trabajo Terminal*

**“Diseño optimizado de la rizobacterias para la mejora en el crecimiento y salud de una planta utilizando redes neuronales y algoritmos genéticos”**

2025-A-150

*Presentan*

**Emmanuel Alvarez Cañedo**

**Ángel Gabriel Guadiana Martínez**

**José Carlos Legorreta Flores**

*Director*

***Dr. Daniel Molina P*é*rez***

***Icono

Descripción generada automáticamente***

|  |  |  |  |
| --- | --- | --- | --- |
| Imagen que contiene señal, reloj, firmar  Descripción generada automáticamente | **INSTITUTO POLITÉCNICO NACIONAL**  **ESCUELA SUPERIOR DE CÓMPUTO**  **SUBDIRECCIÓN ACADÉMICA** | | ***Icono  Descripción generada automáticamente*** |
| No. de TT: 2025-A-150 | | **27 NOVIEMBRE 2024** | |
| Documento técnico | | | |

**“Diseño optimizado de la rizobacterias para la mejora en el crecimiento y salud de una planta utilizando redes neuronales y algoritmos genéticos”**

2025-A-150

*Presentan*

**Emmanuel Alvarez Cañedo**

**Ángel Gabriel Guadiana Martínez**

**José Carlos Legorreta Flores**

*Directores*

***Dr. Molina Pérez Daniel***

**RESUMEN**

El propósito de este Trabajo Terminal (TT) es el diseño de rizobacterias optimizadas para el crecimiento y la salud de las plantas, aplicando técnicas como redes neuronales con algoritmos genéticos para la optimización; la investigación busca contribuir a la agricultura sostenible con una propuesta de un método hibrido para la mejora de producción agrícola, así en este trabajo se va a mencionar términos de PGRB. (Rizobacterias Promotoras del Crecimiento en Plantas) y el IPR (índice de Promoción del Crecimiento).

**Palabras clave –** algoritmos genéticos, bioinformática, redes neuronales, bacterias.

# Agradecimientos

# Índice

(Generarlo automáticamente en Word. Hasta un tercer nivel. )

# Índice de tablas

(Generarlo automáticamente en Word)

# Índice de figuras

(Generarlo automáticamente en Word)

# Glosario de términos

|  |  |
| --- | --- |
| **Término** | Definición |
| **ADN**  **(ácido desoxirribonucleico)** | Es un segmento biológico que contiene detalles genómicos |
| **Bicatenaria** | Que está formado por dos cadenas que entrelazadas y forma la estructura del ADN. |
| **ARN (Ácido Ribonucleico)** | Es una molécula monocatenaria se encarga de la síntesis de proteína y de las funciones regulatorias. |
| **Monocatenaria** | Es solo se forma por una cadena, suelen ser cadenas de nucleótidos con su respectiva base nitrogenada. |
| **Algoritmos Genéticos** | Método que ayuda optimizar soluciones en base con procesos estocásticos; se basa en la selección natural y genética utiliza operadores como selección, cruzamiento, y mutación |
| **Redes Neuronales Artificiales:** | Son un modelo computacional que imita el funcionamiento y la estructura de las redes neuronales biológicas. |
| **Rizobacterias** | Son bacteria benéfica que habitan en la zona de las raíces para ayudar en el crecimiento de la planta mediante proceso bioquímicos |
| **Rizosfera** | Es el área que rodea a las raíces de una planta en el suelo y se caracteriza por ser un lugar de interacción química y biológica importante. |
| **Fijación Biológica de Nitrógeno** | Microorganismo ayudan a la realización de la fijación de nitrógeno, convirtiendo el nitrógeno atmosférico en formas utilizables por las plantas, |
| **Solubilización de Fosfatos y Potasio** | La capacidad de ciertos microorganismos de liberar fósforo y potasio del suelo a partir de compuestos insolubles, haciéndolos accesibles para la asimilación por las plantas. |
| **Fitohormonas** | Pequeñas sustancias producidas por microorganismos, como auxinas, giberelinas y citoquininas, que ayudan a determinar el crecimiento de las plantas. |
| **Biopelículas** | Grupos de multicelular crean comunidades más grandes, viviendo. En la rizosfera, las biopelículas protegen la salud de las raíces |  |
| **DNA2Vec** | Técnica computacional que representa fragmentos de ADN como vectores en un espacio dimensional. |
| **Pla-BA-db** | Es una base de datos para cual implica la gestión de información en ficheros relativos a bacterias asociadas a plantas. |
| **Índice de promoción del crecimiento** | Métrica cuantitativa del efecto positivo de una rizobacterias en la planta basada en factores de crecimiento como la biomasa y la longitud. |
|  |  |
| **Secuenciación del ADN** | Método de laboratorio que determina el orden exacto de los nucleótidos en una molécula de ADN. |
| **Batch** | Una unidad de datos que se asigna durante la formación de modelos de aprendizaje profundo. Son lotes porciones minúsculas para simplificar el proceso de cálculo |
| **Embebido** | Representaciones vectoriales de datos sofisticadas tales como términos, imágenes o secuencias genéticas en un espacio matemático. |
| **Cianuro de Hidrógeno** | Compuesto bioquímico y volátil que muchas bacterias producen un desecho |

# Abreviaturas

| Abreviatura | Significado |
| --- | --- |

|  |  |
| --- | --- |
| **PGPR** | Plant Growth-Promoting Rhizobacteria (Rizobacterias promotoras del crecimiento de plantas) |

|  |  |  |  |
| --- | --- | --- | --- |
| **PGPT** | |  | | --- | | Plant Growth-Promoting Traits (Rasgos promotores del crecimiento vegetal) |  |  | | --- | |  | |

|  |  |
| --- | --- |
| IPR | Índice Promotoras de crecimiento |

|  |  |  |  |
| --- | --- | --- | --- |
| **ADN** | |  | | --- | | Deoxyribonucleic Acid (Ácido Desoxirribonucleico) |  |  | | --- | |  | |

|  |  |
| --- | --- |
| **ARN** | Ácido Desoxirribonucleico (Ácido Ribonucleico) |

|  |  |
| --- | --- |
| **AG** | Algoritmo Genético |

|  |  |
| --- | --- |
| **CNN** | Convolutional Neural Network (Red neuronal convolucional) |

|  |  |
| --- | --- |
| **RNN** | Recurrent Neural Network (Red neuronal Recurrente) |

|  |  |
| --- | --- |
| **IA** | Inteligencia Artificial |
| IAA | acético ácido indol acético |
| HCN | Cianuro de Hidrógeno |

# Capítulo 1 Introducción

El uso de rizobacterias en la agricultura tiene múltiples ventajas cuando se contrasta con los métodos tradicionales, como el hecho de que son naturales y no requieren productos químicos que puedan resultar nocivos de forma general, además de que pueden llegar a proporcionar un correcto estado general de la salud del suelo. Es así como se sigue en los pasos de la agricultura regenerativa/de conservación y ecológica (sostenible), las cuales intentan restablecer el equilibrio del ecosistema agrícola.

La agricultura sostenible ha surgido como una de las respuestas frente a las necesarias soluciones de la seguridad alimentaria y de la crisis del medioambiente en el siglo xxi, el cual está marcado por el crecimiento extremo de la población, la presión excesiva de los recursos naturales y la preocupación por los efectos negativos de la agricultura convencional. La exigencia implica hallar prácticas agrícolas que incrementen la producción, al mismo tiempo que ayudan a recuperar la salud del suelo, del agua, de la biodiversidad.

* 1. Antecedentes

En la agricultura sostenible se usan métodos que ayuda al crecimiento de una planta con la ayuda de rizobacterias que son beneficiosas en ella, ya que pertenecen a un grupo de bacterias que tiene efectos positivos en las plantas porque suelen habitar en la rizosfera, es un ecosistema clave que se encuentra entre la raíz y el suelo en donde ocurre la mayoría proceso bioquímicos que benefician a la planta como puede ser el intercambio de nutrientes, fijación de nitrógeno que se trata de convertir el nitrógeno en nutrientes que puede digerir la planta así como la producción de fitohormonas que ayuda a un mejor desarrollo de la planta, son algunos beneficios muy importantes que por ello toma interés este trabajo terminal.

El estudio en la universidad de durango se comprobó el impacto que tiene la rizobacteria en el crecimiento del chile jalapeño mediante cultivos entres los mejores resultados se mostró mejoras tanto en las hojas como en el chile este hallazgo sugiere que las rizobacteria puede ser una alternativa más viable y sostenible sin perjudicar tanto al suelo y a nosotros mismos, por los fertilizantes artificiales que puede recibir una planta [1].

En este trabajo se opta por usar métodos menos agresivos como el uso de rizobacterias en las cuales son bacterias beneficiosas para las plantas que habitan en la rizosfera, Para estudiar y evaluar la eficacia de estas rizobacterias, se usará bioinformática para el estudio genómico del microorganismo a través de datos computacionales, Mediante algoritmos se podrá procesar ese volumen de datos para optimizar evaluar y seleccionar, después se usarla modelos de Aprendizaje profundo para una clasificación más precisa [2].

* 1. **Planteamiento del problema**

El aumento de la población mundial y la consiguiente demanda de alimentos se ha convertido a la agricultura sostenible en una prioridad global debido a la necesidad de satisfacer la demanda creciente de alimentos mientras se minimiza el impacto ambiental. En este contexto, las rizobacterias, microorganismos que colonizan la rizosfera y establecen relaciones simbióticas con las plantas, se utilizan regularmente para el mejoramiento del crecimiento de las plantas. Esta idea ocurrió porque existen bacterias programadas contra el cáncer que se usan para un ámbito más médico, nuestro trabajo es para un ámbito más biológico pero los principios son los mismos [3].

La mejora del crecimiento de las plantas por parte de los PGPR (Rizobacterias promotoras del crecimiento de plantas) se cuantifica como un aumento en plántulas, vigor, biomas y proliferación del sistema de raíces y rendimiento en diversas especies de plantas. Desde su reconocimiento como un subconjunto importante de microorganismos colonizadores de raíces, en las últimas tres décadas, se han llevado a cabo varios estudios, a un ritmo exponencial.

En este aspecto la salud en las plantas también es una necesidad como lo resalta, La salud de las plantas depende, además de la nutrición, de una correcta interacción con las bacterias benéficas, esa interacción puede ocurrir a nivel de rizosfera (región de suelo y microorganismos que rodean las raíces de las plantas) [4]. Con esto, se comprende la importancia de las bacterias en la región del suelo cercanas a las raíces de las plantas tanto para el crecimiento como para la salud, región en la cual las rizobacterias habitan. Lo anterior, sugiere un sistema en donde el desempeño de la planta se puede ver afectado por los organismos cercanos a ella, en donde sugiere que, en condiciones cambiantes y estresantes, las plantas se asocian estrechamente con estos microbios de diversas formas, como la absorción de nutrientes, el crecimiento y desarrollo de las plantas, y la aptitud para la salud de las plantas. En lugar de funcionar y evolucionar de forma independiente, las plantas y sus microorganismos asociados coexisten como parte de un sistema biológico.

* 1. **Propuesta de solución**

Se va a recurrir una Base de datos de bacterias asociadas a las plantas (PLaBA-db) que nos ayudar a visualizar los efectos positivos de la bacteria a la planta, se necesita hacer una limpieza a la base de datos para solo trabajar con datos que en realidad sean relevantes, mediante técnicas de bioinformática se hará un estudio genómico del microrganismo a través de datos computacionales.

Existen archivos que contienes la secuencia genómica representadas con letras de la bacteria que queremos analizar y evaluar se tendrá que extraer los datos genómicos y poder hacer una clasificación de PGPR (Rizobacterias promotoras del crecimiento de plantas), posteriormente se hace un embebido de los datos genómico para reducir su alta dimensionalidad y pasarlo a una red neuronal con una capa convolucional (CNN) para identificar patrones de la secuencia genómica, también se adaptarla una red neuronal recurrente (RNN) que se caracteriza por tener memoria que nos ayudara a la clasificación del genoma y nos brindara un vector de características que es la relación entre beneficio que puede tener la rizobacteria generada en una planta.

Se usará un algoritmo genético que se inspira en el proceso de la selección natural se hará un proceso en donde se va a inicia una población aleatoria de genomas por consiguiente se aplicaran operadores de optimización, evaluación y selección esto nos dará una secuencia genómica que representa una rizobacteria con ciertos rasgos promotoras del crecimiento vegetal que más se acerque a las características que deseamos obtener al final para la evaluación del genoma resultante del algoritmo genético pasara por la misma red neuronal para una clasificación más precisa de la secuencia genómica resultante, así poder maximizar los efectos de una rizobacteria a una planta.

* 1. **Objetivo general**

El objetivo de este trabajo terminal es diseñar rizobacterias para mejorar el crecimiento y la salud de las plantas, teniendo como objetivo el mejoramiento de los parámetros de biocontrol, biofertilización y biorremediación que brinda una rizobacteria a la planta. se aplicarán técnicas como de redes neuronales y algoritmos genéticos, que consistirá en identificar combinaciones específicas de características genéticas.

* 1. **Objetivos específicos**
* Desarrollar un análisis del estado del arte enfocado en las rizobacterias, complementado con información relevante sobre las técnicas basadas en algoritmos genéticos y redes neuronales.
* Tomar requisitos para el desarrollo tomando en cuenta restricciones de diseño de la rizobacteria y el entorno de la planta como entradas para los algoritmos planteados.
* Hacer uso y adaptar algoritmos con nuestras necesidades de este trabajo que ayudara en el diseño de las rizobacterias, utilizando técnicas de redes neuronales y algoritmos genéticos.
* Implementar de forma práctica como conteo de PGPT (Rasgos promotores del crecimiento vegetal) o simular los efectos de una rizobacteria diseñada por los algoritmos para comprobar su efectividad.
  1. **Justificación**

A forma que pasa el tiempo se observó una gran producción de alimentos por la razón de que existe una expansión demográfica muy grande por ende hay una producción en masa en alimentos que conlleva a que se acelere la forma de producción para que sea lo más ágil posible

Por ello se contempló este trabajo porque se necesita una solución innovadora y sostenible que pueda abordar la problemática en particular, para esto se propuso estas tecnologías una red neuronal que usara como clasificadora tanto un algoritmo genético que se inspira en el proceso de la selección natural para la generación de secuencias genómicas.

Con este enfoque se pretende traer un beneficio a la naturaleza para reducir la huella de carbono por el mal uso de los fertilizantes químicos y en relación también nos traerá grandes beneficios a nosotros los humanos porque estaríamos consumiendo alimentos que no estarán contaminados.

Por esta manera se podría lograr aumenta la producción de alimentos de una manera sostenible sustituyendo los fertilizantes químicos y optar por otra alternativa que hasta podrían lograr mejores resultados de esta forma poder contribuir a la agricultura sostenible [5].

Haciendo referencia a un posible mejoramiento de lo que ya existe al respecto, dicha investigación propone una intervención innovadora que, a la vez, aporta en prácticas de fertilización y de protección de cultivos que sin duda existen pero que, quizás, puedan ser superadas. Al diseñar rizobacterias aptas para favorecer el crecimiento y la salud de plantas, se puede también llegar a la posibilidad de reducir su uso y de mejorar la salud del suelo al mismo tiempo.

* 1. **Organización del documento**

{Una muy breve descripción de cada capítulo, un párrafo por cada capítulo, utilizar diferentes formas de redacción por cada capítulo. Se recomienda usar la taxonomía incremental de Marzano y Kendall}

En el **capítulo 1**, se describe la problemática junto con la …

En el **capítulo 2**, se revisan los trabajos similares al proyecto …

En el **capítulo 3**, se describen los conceptos básicos que se …

En el **capítulo 4**, se analizan …

} [Un párrafo por cada capítulo]

# Capítulo 2. Estado del arte

Con la ayuda de antecedentes se validó, que fuera una buena opción centrarnos en las rizobacteria en particular por que se tuvo en cuenta que entre varias bacterias que tenían características similares la rizobacteria fue la que obtuvo mejores resultados.

2.1**Articulo sobre la** evaluación de las bacterias rizosféricas de vida libre por sus múltiples actividades promotoras del crecimiento de las plantas [6]

Se aislaron un total de 72 bacterias de diferente cepa pero que contribuyan a un cierto beneficio en la planta pertenecientes cepas como Azotobacter, Pseudomonas fluorescentes, Mesorhizobium y Bacillus se colocaron en diferentes raíces de plantas y suelos rizosféricos en las cercanías. Estos aislamientos se examinaron en vitro para determinar sus rasgos promotores del crecimiento de las plantas, como la producción de ácido indol acético ácido indol acético (IAA), amoniaco (NH3), cianuro de hidrógeno (HCN), sideróforo, solubilización de fosfato y actividad antifúngica. Más del 80% de los aislados de Azotobacter, Pseudomonas fluorescentes y Mesorhizobium produjeron IAA, mientras que solo el 20% de los aislados de Bacillus fue productor de IAA. La solubilización de fosfato se detectó comúnmente en los aislados de Bacillus (80%), seguido de Azotobacter (74,47%), Pseudomonas (55,56%) y Mesorhizobium (16,67%) como se observa en la figura 2.1.1 las actividades directas que tuvo en los aislamientos de prueba.

Imagen que contiene electrónica, computadora

Descripción generada automáticamente

Figura 2.1.1 Actividades directas de los aislamientos de prueba.[6]

En el aspecto de que también cumple la característica de control de plagas por ejemplo una plaga muy común en las plantas es de tipo fúngica esto quiere decir que son hongos que habitan en la rizosfera y le quita ciertos nutrientes que deberían ser de la planta, se observó que las rizobacterias tiene un gran control sobre ellas se le suele definir como actividades indirectas de la rizobacteria.

La producción de sideróforos y la actividad antifúngica de estos aislados, excepto Mesorhizobium, se exhibieron en un 12,77 % de los aislados. La producción de HCN fue un rasgo más común en Pseudomonas (88,89 %) y Bacillus (50 %). Sobre la base de múltiples actividades promotoras del crecimiento de las plantas, once aislados bacterianos (siete Azotobacter, tres Pseudomonas y un Bacillus) como se observa en la figura 2.2.1.

Diagrama

Descripción generada automáticamente

Figura 2.1.2 Actividades indirectas de los aislamientos de prueba. [6]

Como se puede observar nos dio un indicativo de que las rizobacterias son una buena opción para ser estudiada y analizarla para ser estudiada de una manera de una forma más detallada ya puede ser gran aporte en la agricultura sostenible gracias a estas bacterias es posible una simbiosis y contribuir en la mejora de absorción de nutrientes esenciales, como nitrógeno y fósforo, lo que no solo aumenta el rendimiento de los cultivos, sino que también reduce la dependencia de fertilizantes químicos.

## 2.2 Artículo de la actividad benéfica de los ácidos orgánicos producidos por rizobacterias que solubilizan fosfato [7]

Se tuvo que estudiar que hace que la rizobacteria sea tan especial, se trata de su producción de ácidos orgánicos de bajo peso molecular por las rizobacterias es uno de los mecanismos más conocidos de solubilización de fosfatos en el suelo, un proceso que hace que el fósforo esté disponible para la nutrición de las plantas. La capacidad de solubilizar los fosfatos de los ácidos orgánicos.

Durante la última década, se ha registrado considerablemente el conocimiento relacionado con los microorganismos solubilizadores de fosfato, se relación con las bacterias promotoras del crecimiento vegetal o PGPR (rizobacterias promotoras del crecimiento vegetal). Se trata de bacterias que colonizan de manera libre en el suelo, que son capaces de adaptarse y persistir en la rizosfera de la planta y de inducir su crecimiento o desarrollo. Las rizobacterias pueden ser de tipo benéfico o antagónico respecto a la planta. Los PGPR son capaces de solubilizar fósforo inorgánico a partir de diferentes compuestos los ácidos que producen se muestran en la figura 2.2.2.

Diagrama

Descripción generada automáticamente con confianza media

 Figura 2.2.2. Ácidos orgánicos alifáticos producidos por bacterias solubilizadoras de fosfato [7].

Se dotó de un gen implicado en la biosíntesis de un cofactor fundamental para el correcto funcionamiento de ciertas bacterias que oxidan alcoholes, aldehídos y azúcares. Este gen se expresa en una bacteria, la cual tiene la capacidad de sintetizar una enzima específica, pero no puede producir el cofactor que debe unirse a la misma para dar lugar.

Se analizo el funcionamiento del gen de la bacteria que produce dicho factor en dicha cual favorece la disolución de minerales. Por otro lado, la capacidad para solubilizar compuestos sirve para que se sigan manteniendo las bacterias funcionales en condiciones de temperaturas extremas. También se ha puesto de manifiesto que las bacterias inmovilizadas favorecen la producción de ácidos orgánicos, lo que mejora la disolución de fosfatos.

## 2.3 Articulo de selección y caracterización de rizobacterias promotoras del crecimiento en plántulas de chile poblano [8]

Para este anteceden se tuvo en cuenta en ya enfocarnos en la rizobacteria y las diferentes cepas que existen para ello se estudiaron de forma aislada cada cepa, se estudiaron los resultados en una gran que tuvo sobre la planta. En el antecedente se recalca aislaron de las cepas de rizobacteria. El objetivo del presente estudio, con la finalidad de evaluar el potencial que tienen estas cepas de rizobacterias para promover el crecimiento de plántulas de chile poblano.

Las cepas se caracterizaron por su capacidad de solubilizar fosfatos, fijar nitrógeno y producir auxinas, pues fueron las 10 cepas que obtuvieron mejores resultados en el crecimiento de las plantas y que finalmente fueron identificadas a nivel molecular. Se aislaron un total de 215 cepas de rizobacterias, de las cuales 74 resultaron ser solubilizadoras de fosfato, siendo así la totalidad de las cepas que se identificaron por las 10 que mejor respondieron a los ensayos de crecimiento de plántulas de chile poblano como se visualiza en la figura 2.3.1.

Imagen que contiene Diagrama

Descripción generada automáticamente

Figura 2.3.1 Solubilización de fosfatos por las cepas de bacterias solubilizadoras de fosfatos aislados en suelo rizosférico de chile poblano.

La utilización de diferentes cepas de rizobacterias promotoras del crecimiento, en general tiene cualidades para solubilizar fosfatos y segregar auxinas, representa una opción concreta para incrementar la calidad de las plántulas de chile poblano. Se comprobó el efecto beneficioso, ya que, implementados de forma individual lograron una mayor altura, producción de biomasa y más hojas. El uso de estas bacterias supone una alternativa sostenible para disminuir la dependencia insumos sintéticos de glucosa, y por lo tanto reducir los efectos negativos que son generados por estos insumos sobre el medio ambiente y la salud de los humanos.

## 2.4 Aplicación en Identificar bacterias PGPR mediante bioinformática [9]

La identificación de las bacterias promotoras del crecimiento vegetal conocidas como PGPR, se llevó a cabo a partir de herramientas de bioinformática. Para ello, se aislaron cuatro cepas bacterianas a partir de las muestras ambientales que mostraron propiedades características de las PGPR; estas propiedades son la fijación de nitrógeno, la solubilización de fósforo, la producción de fitohormonas como las auxinas y la generación de sideróforos.

La extracción del ADN genómico se llevó a cabo utilizando protocolos estandarizados, garantizando la pureza y calidad del ADN genómico previamente aislado. Con el ADN genómico aislado se amplificó el gen marcador microbiano 16S rRNA mediante la reacción en cadena de la polimerasa (PCR) utilizando cebadores universales diseñados para flanquear las regiones hipervariables del gen, el cual se considera ampliamente usado en estudios de identificación bacteriana y estudios filogenéticos debido a su alta conservación entre especies y alta variabilidad en regiones específicas.

Una vez realizadas las amplificaciones, las muestras se enviaron a su secuenciación. Dependiendo de los recursos, se utilizó tecnología de secuenciación. Las secuencias obtenidas fueron analizadas con análisis bioinformático. Esto incluye un primer análisis de la calidad de las secuencias (análisis inicial para eliminar secuencias erróneas), una alineación inicial de las secuencias y una posterior comparación un repositorio de secuencias de cepas, se realizó un análisis filogenético que determinó la identificación de las especies bacterianas como se observa en la Figura 2.4.1.

Interfaz de usuario gráfica, Texto, Aplicación

Descripción generada automáticamente

Figura 2.4.1 Resultado del análisis de cepas en una base de datos bacteriana. [9]

El análisis determinó que las cuatro cepas corresponden a géneros bacterianos que son conocidos por su actividad PGPR, como Bacillus, Pseudomonas y Azospirillum. Esta información resulta esencial para el desarrollo de aplicaciones biotecnológicas para propósitos centrados en la mejora del crecimiento y rendimiento de cultivos agrícolas.

## 2.5 Aplicación en clasificación taxonómica de bacterias usando machine learning [10]

La tarea de realizar pronósticos a partir de secuencias de ADN supone un desafío notable por diversas razones, tales como la reducida longitud de las lecturas, la carencia de la continuidad, la fragmentación de los datos, y la inmensa diversidad de tipos de bacterias que existen en nuestro entorno. Con el objetivo de dar respuesta a estos contratiempos, han aparecido modelos de aprendizaje automático (machine learning) que permiten la extracción de las características relevantes en este tipo de datos.

Este trabajo pretende ser una aproximación al estudio y aplicación de modelos de aprendizaje profundo con vistas a proporcionar una estrategia correcta y adecuada para la clasificación taxonómica de las bacterias desde las secuencias de ADN, sirviendo de nuestra propuesta una manera adecuada y correcta para abordar la clasificación taxonómica de las bacterias utilizando secuencias de ADN a partir de las aportaciones que se han ido realizando en la ciencia para el procesamiento de datos biológicos a partir de las técnicas más avanzadas en machine learning.

La representación vectorial construida a partir de k-mers ha sido efectiva para la reducción de dimensionalidad en proceso de usar k-mers en el análisis de secuencias de ADN implica dividir las secuencias de nucleótidos (A, T, C, G) en fragmentos de longitud fija, denominados k-mers, para representar las secuencias de manera que los modelos de aprendizaje automático puedan extraer características significativas como se observa en la figura 2.5.1.

Texto

Descripción generada automáticamente

Figura 2.5.1 representación vectorial construida a partir de k-mers [10].

Se menciono que el procesamiento de los datos, el entrenamiento y la evaluación de los modelos de aprendizaje profundo se realizarán mediante el entorno el uso de herramienta como Python, el cual incorpora bibliotecas en particular, TensorFlow y Keras conocidas por las posibilidades que éstas ofrecen para la construcción, entrenamiento y despliegue de modelos de aprendizaje profundo escalables y eficaces.

Se entrenó una red neuronal multicapa (MLP) por cada combinación de k-mer y taxón, se evaluó el modelo final con un conjunto de validación, utilizando como métricas para evaluar el rendimiento del modelo las siguientes:

* Exactitud: Es el porcentaje de predicciones correctas que realiza el modelo sobre todas las clases. Se habla de exactitud cuando se considera el rendimiento general del modelo.
* Precisión: Mide la confianza que se tiene en que el modelo afirma que una instancia pertenece a una clase. Se pregunta cuántas de aquellas instancias que el modelo afirma pertenecen a una clase existen realmente.
* Recall: Es la capacidad que tiene el modelo para detectar todas las instancias de una clase. Se pregunta cuántas de todas las instancias que pertenecen a una clase se etiquetaron correctamente. Si necesitas poner más información sobre la forma en la que se calcularon estas métricas

Estas métricas se aplicaron para calificar la red neuronal multicapa, que posteriormente permite profundizar en el análisis del rendimiento clasificador como se muestra en la figura 2.5.2.

Gráfico

Descripción generada automáticamente

Figura 2.5.2 Métricas se aplicaron para calificar la red neuronal multicapa. [10]

## 2.6 Aplicación de algoritmos genéticos aplicados a búsqueda de motifs o pgpr en secuencias de ADN [11]

En esta aplicación del algoritmo genético se presentó la comparación de un par de enfoques para la solución del problema de localizar PGPR en secuencias de ADN mediante algoritmos genéticos. El primero de los enfoques que se presenta corresponde a una búsqueda por posición, en donde se evalúan los patrones obtenidos a partir de posiciones comenzadas y se determina si corresponde al PGPR a localizar en las secuencias. El segundo enfoque corresponde a la búsqueda por patrón, en donde se evalúan todas las posibles cadenas a ser generadas mediante longitud fija de nucleótidos, que serán comparadas de inicio a final en las secuencias de ADN para encontrar el PGPR.

Los algoritmos genéticos (AG) reducen todas las posibilidades y seleccionan las más relevantes para encontrar la solución. Los resultados nos muestran qué método es más rápido para encontrar los PGPR, es decir, cuál de ellos determina mejor estos patrones con los datos simulados. Las simulaciones de datos se realizarán utilizando secuencias de ADN aleatorias, en las que se insertan PGPR en posiciones aleatorias de forma manual en cada secuencia.

El enfoque propuesto en este trabajo pone de manifiesto la eficiente respuesta que pueden tener los algoritmos genéticos para una tarea compleja como la búsqueda de patrones (PGPR), su efectividad se descompone en situaciones complejas o donde el número de secuencias en un conjunto de datos es elevado. Los algoritmos genéticos, a pesar de ser más complejos en sus descripciones algorítmicas, muestran una adaptabilidad y calidad creciente en la medida que aumentan los datos. Por lo tanto, este trabajo muestra cómo los algoritmos genéticos responden de forma más eficiente a una tarea compleja, viene a ser una ventaja en aquellas tareas de optimización y búsqueda en grandes volúmenes de datos biológicos en la figura 2.6.1 muestra una gráfica de la efectividad.

Gráfico, Gráfico en cascada

Descripción generada automáticamente

Figura 2.6.2 Grafica de efectividad de algoritmos genéticas [11].

En el desarrollo se llevaron a cabo 30 ejecuciones para cada grupo de secuencias de ADN, con lo que se pudieron recuperar, en la mayoría de los casos, el PGPR. En general nos proporcionó buenos resultados con respecto a las indicaciones en la literatura sobre el método de muestreo de Gibbs, que (desde un punto de vista personal) es más rápido que los algoritmos genéticos, pero no necesariamente mejor. Los algoritmos genéticos han de generar poblaciones, las cuales son evaluadas con respecto a su capacidad, después han de mutar y cruzar los individuos de nuevas generaciones, reevaluarlos, y así en cada generación

Con respecto a la precisión, los resultados obtenidos de los algoritmos genéticos son marginalmente más complejos por su interactividad, tienen mayor capacidad para encontrar patrones de ADN significativos, los algoritmos genéticos, a partir de ir evaluando y corrigiendo continuamente las poblaciones de soluciones, producen mayor eficacia en la identificación de los patrones esperados (PGPR). Esto se aprecia además para volúmenes grandes de datos, en el cual el nivel de precisión registrado por los algoritmos genéticos es un factor determinante que conlleva a buenos resultados y resultados útiles como se muestra en la figura 2.6.2 la precisión que tuvo el algoritmo genético.

Gráfico, Gráfico de barras

Descripción generada automáticamente

Figura 2.6.2 Grafica de precisión de los algoritmos genéticos [11].

los algoritmos genéticos pueden parecer más complejos debido a su naturaleza iterativa y su necesidad de generar y evaluar múltiples generaciones, los resultados obtenidos muestran que ofrecen una mayor precisión en la identificación de patrones en las secuencias de ADN. Si bien el método de muestreo de Gibbs es más rápido, especialmente cuando se trabaja con volúmenes pequeños de datos, los algoritmos genéticos tienen una ventaja significativa en escenarios donde la precisión y la capacidad para manejar grandes cantidades de información son cruciales. Esto refuerza mi creencia de que, a pesar de la mayor complejidad, los algoritmos genéticos son una herramienta valiosa para resolver problemas de búsqueda de patrones en datos biológicos.

## 2.7 Trabajo Terminal Z

El Trabajo Terminal de Diseño Optimizado de Rizobacterias para el Aumento en el Crecimiento y Salud de una Planta Utilizando Redes Neuronales y Algoritmos Genéticos realiza una puesta en práctica de tecnologías avanzadas como las redes neuronales artificiales y los algoritmos genéticos para el diseño de rizobacterias aplicadas en el ámbito agrícola. Las rizobacterias ahí descritas, organismos que habitan en la rizosfera de la planta, se encuentran involucradas en el proceso de crecimiento vegetal mediante la fijación de nitrógeno, la producción de fitohormonas y el biocontrol de patógenos. Este proyecto destaca también por ser una propuesta interdisciplinar, involucrando, por tanto, la biología, la bioinformática y las técnicas computacionales para resolver un problema muy relevante en materia de sostenibilidad agrícola.

El objetivo general de este trabajo ha sido el de la mejora del diseño de rizobacterias mediante la optimización de parámetros como la biocontrol, la biofertilización y la biorremediación mediante el uso de redes neuronales para predecir las características funcionales y de algoritmos genéticos para optimizar estas funciones siguiendo las restricciones ambientales y biológicas; se han utilizado herramientas como PyTorch para implementar la arquitectura de las redes neuronales, Biopython para manejar secuencias genómicas y datos biológicos, así como PLaBA-db que proporciona información sobre las cepas de bacterias asociadas a las plantas.

Empleo de Redes Neuronales y Algoritmos Genético, para detectar patrones en datos genómicos de gran tamaño, se entrenan las redes neuronales, al tiempo que los algoritmos genéticos buscan combinaciones óptimas de características bacterianas.

Dataset Complejo y Estructurado la PLaBA-db contiene más de 5,500 registros sobre cepas bacterianas, que están acompañados de información de tipo genómico y funcional. Aplicaciones Simuladas: El diseño se ha probado en una forma completamente virtual, que simula los efectos de las rizobacterias mejoradas en los cultivos bajo diferentes condiciones.

Sostenibilidad Agrícola permite reducir la necesidad de utilizar agroquímicos nocivos, fomentando la implementación de prácticas agrícolas de tipo ecológico, optimización de recursos sirve para identificar rizobacterias que presenten unas características concretas y adaptadas a diferentes condiciones ambientales y cultivos, adaptabilidad esta metodología puede ser aplicada a múltiples cultivos y diversas regiones con unos mínimos cambios, complejidad técnica se presenta un alto requerimiento de conocimientos avanzados en biología, programación y aprendizaje automático, limitaciones de la simulación aunque los resultados virtuales obtenidos son amplios y positivos, aún deben ser validados ampliamente en condiciones de campo reales.

La redes neuronales y algoritmos genéticos, no solo se presentan innovadoras, sino también versátiles, abriendo puertas hacia aplicaciones futuras en distintas disciplinas. Sin embargo, todavía es necesario el realizar pruebas en campo para validar los hallazgos y asegurar que las posibles soluciones planteadas sean aplicables en este sentido, así como escalables.

La interdisciplinariedad y el enfoque en la sostenibilidad confirman que se trata de un trabajo de aporte importante y actual en agricultura. En futuras versiones, sería interesante poder añadir estrategias de modelado híbrido, que mejoraran la exactitud de la simulación, así como la interacción entre rizobacterias y otros microorganismos presentes en el suelo. Este acercamiento hace patente como la tecnología en la agricultura puede ser capaz de transformar un modelo más eficiente y sostenible.

## 2.8 Tabla comparativa

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Título del trabajo | Uso de PGPR | Aplicación de bioinformática | Impacto en sustentabilidad ambiental | Enfoque en desarrollo agrícola | Innovación tecnológica |
| Efecto de rizobacterias promotoras del crecimiento vegetal en el desarrollo y rendimiento del chile jalapeño | ✔ | ✘ | ✘ | ✔ | ✘ |
| The role of bioinformatics in agricultural research | ✘ | ✔ | ✘ | ✘ | ✔ |
| Ya se puede programar una bacteria como una computadora | ✘ | ✔ | ✘ | ✘ | ✔ |
| Sustentabilidad ambiental, del concepto a la práctica | ✘ | ✘ | ✔ | ✘ | ✘ |
| Diseño ambiental y producción de conocimiento interdisciplinario | ✘ | ✘ | ✔ | ✘ | ✘ |
| Producción de HCN por Pseudomonas | ✔ | ✘ | ✘ | ✘ | ✔ |
| Estrategias para el desarrollo de fertilizantes biológicos en la agricultura | ✔ | ✘ | ✔ | ✔ | ✘ |
| Evaluación de la calidad de plántulas de chile poblano inoculadas con rizobacterias promotoras de crecimiento | ✔ | ✘ | ✘ | ✔ | ✘ |
| Identificación de las bacterias PGPR mediante bioinformática | ✔ | ✔ | ✘ | ✘ | ✔ |
| Algoritmos genéticos aplicados a búsqueda de motifs en secuencias de ADN | ✘ | ✔ | ✘ | ✘ | ✔ |
| ADN y ARN: concepto, diferencias y funciones | ✘ | ✔ | ✘ | ✘ | ✘ |

**Tabla 1**. Tabla comparativa entre los principales artículos

La tabla comparativa que se ha presentado pone de manifiesto las principales características de los trabajos analizados en relación con la propuesta y, aunque se pueden observar que ciertos trabajos dan cabida a aspectos clave de la propuesta PGPR, bioinformática, sustentabilidad la mayoría de los trabajos no dan una solución que abarca todas las diferentes áreas que tienen relación de investigación.

# Capítulo 3. Marco Teórico

## 3.1. Ácidos nucleicos

El ácido desoxirribonucleico (ADN) y el ácido ribonucleico (ARN) son biomoléculas esenciales para el almacenamiento y transmisión de la información genética. Aunque ambos son ácidos nucleicos y comparten algunas similitudes estructurales, como estar compuestos por nucleótidos, sus funciones y estructuras específicas difieren considerablemente [12].

### **3.1.1.** ADN

El ácido desoxirribonucleico es un material que contiene la información genética y hereditaria de los seres vivos, y codifica para la producción de proteínas [12][13].

## **3.1.1.1** **Estructura y función**

Según [11] los ácidos nucleicos son una molécula bicatenaria, lo que significa que está formado por dos cadenas complementarias que se enrollan entre sí. Cada nucleótido está compuesto por un grupo fosfato, un azúcar desoxirribosa y una de las cuatro bases nitrogenadas: adenina (A), guanina (G), citosina (C) o timina (T)[14]. La secuencia de estas bases codifica la información genética utilizada para la síntesis de proteínas y el desarrollo de los organismos. El ADN se encuentra principalmente en el núcleo de las células eucariotas, aunque también se puede localizar en las mitocondrias y cloroplastos.

El NIH [4] dice que su función principal es la de almacenar la información genética hereditaria y dirigir el desarrollo y funcionamiento de los seres vivos a través de la transcripción y posterior traducción de las proteínas. Durante la replicación celular, el ADN se copia para asegurar que cada nueva célula tenga una copia exacta de la información genética [15].

## 3.1.2. ARN

El ácido ribonucleico (ARN) es un ácido presente en todas las células vivas que tiene similitudes estructurales con el ADN. Sin embargo, a diferencia del ADN, es más frecuente que el ARN esté formado por una única cadena [16]

## 3.1.2.1. Estructura y función

El ARN es una molécula generalmente monocatenaria (de una sola cadena) y está compuesta por ribosa en lugar de desoxirribosa como azúcar, y uracilo (U) en lugar de timina como base nitrogenada. Existen varios tipos de ARN, cada uno con una función específica dentro de la célula: ARN mensajero (ARNm), ARN de transferencia (ARNt) y ARN ribosomal (ARNr), entre otros. El ARNm es responsable de llevar la información genética desde el ADN en el núcleo hasta los ribosomas, donde se lleva a cabo la síntesis de proteínas. El ARNt participa en la traducción al unir los aminoácidos durante la formación de proteínas, mientras que el ARNr es un componente estructural de los ribosomas [11].

A diferencia del ADN, el ARN no almacena información genética a largo plazo, sino que actúa como intermediario en la expresión génica. Además, algunas formas de ARN pueden tener funciones catalíticas y reguladoras en la célula [17].

## 3.1.3. Diferencias entre ADN y ARN

Las principales diferencias entre el ADN y el ARN se encuentran el tipo de azúcar (desoxirribosa en el ADN y ribosa en el ARN), las bases nitrogenadas (el ARN tiene uracilo en lugar de timina), y su estructura general (el ADN es bicatenario mientras que el ARN suele ser monocatenario) [7]. El ADN se encuentra principalmente en el núcleo celular, mientras que el ARN puede estar tanto en el núcleo como en el citoplasma [3]. Las funciones también varían significativamente: el ADN almacena la información genética a largo plazo, mientras que el ARN está involucrado en la síntesis de proteínas y en diversas funciones regulatorias y catalíticas [11].

## 3.2. **Secuenciación del ADN**

Es un proceso fundamental en la biología molecular y la genómica que consiste en determinar el orden exacto de los nucleótidos (adenina [A], guanina [G], citosina [C] y timina [T]) en una molécula de ADN. Este procedimiento permite a los científicos comprender mejor el código genético de los organismos y ha sido clave en el avance de numerosas áreas como la medicina, la biotecnología y la evolución [18, 19, 20].

## 3.2.1. Técnicas de secuenciación

A lo largo del tiempo, se han desarrollado diferentes métodos para la secuenciación del ADN. Dos de los enfoques más conocidos son el método de Sanger y las tecnologías de secuenciación de nueva generación (NGS) [18].

Método de Sanger: Introducido en la década de 1970 por Frederick Sanger, este método es un enfoque de secuenciación basado en la terminación de la cadena. Se utiliza un conjunto de nucleótidos modificados (didioxinucleótidos) que provocan la terminación de la síntesis de ADN en puntos específicos. Luego, se separan los fragmentos resultantes mediante electroforesis para determinar la secuencia. A pesar de ser un método más antiguo, sigue siendo útil para secuencias cortas y con alta precisión [18, 19].

Secuenciación de Nueva Generación (NGS): Estas técnicas surgieron para superar las limitaciones del método de Sanger, ofreciendo una capacidad de secuenciación mucho mayor, velocidad y reducción de costos. La NGS incluye una variedad de tecnologías, como la secuenciación por síntesis, la secuenciación por ligación y la secuenciación de nano poros. Estas plataformas permiten la secuenciación masiva y paralela de millones de fragmentos de ADN, lo que es esencial para proyectos como la secuenciación del genoma humano y estudios genómicos a gran escala [18,20].

## 3.2.2 Aplicaciones de la Secuenciación del ADN

La secuenciación del ADN ha revolucionado la biología moderna, brindando una comprensión detallada del genoma de diversos organismos. Algunas de las aplicaciones más importantes incluyen [18, 19, 20]:

Medicina de precisión: La secuenciación del ADN permite identificar mutaciones genéticas que pueden estar asociadas con enfermedades hereditarias o susceptibilidad a ciertos trastornos. Esta información es fundamental para el desarrollo de tratamientos personalizados en enfermedades como el cáncer y trastornos genéticos [18].

Diagnóstico de enfermedades: A través de la secuenciación, es posible diagnosticar infecciones virales o bacterianas con mayor rapidez y precisión, así como identificar variantes genéticas responsables de enfermedades [18].

Estudios evolutivos: La secuenciación del ADN ha permitido a los científicos reconstruir árboles filogenéticos y comprender la historia evolutiva de las especies al comparar secuencias genómicas de diferentes organismos [19].

Investigación en biodiversidad: Se utiliza en estudios de ecología y biodiversidad para identificar especies y analizar su diversidad genética, especialmente en proyectos como la secuenciación de microbiomas o el análisis de ADN ambiental (eDNA) [20] como se observa en la Figura 3.2.2.1.

Diagrama, Forma, Polígono

Descripción generada automáticamente

Figura 3.1.2.1 [Secuenciación de genoma completo](http://cnx.org/contents/GFy_h8cu@10.7:5l844Z38@7/Whole-Genome-Sequencing)

## 3.3. Rizobacterias

Son bacterias que habitan en la rizosfera es un área de suelo que se encuentra unida a la raíz esa zona se caracteriza por tener una gran interacción única y muy dinámica en los procesos bioquímicos que pasa entre las raíces de la plantas y microorganismos del suelo tal como se muestra en la figura 3.2.2.1[21].

Diagrama

Descripción generada automáticamente

Ilustración Figura 3.2.2.1 Ciclo de la rizobacteria [21]

### 3.3.1. PGPR

En el proceso de establecimiento de las relaciones con rizobacterias las plantas invierten hasta el 20% de las fuentes de carbono obtenidas durante la fotosíntesis, esto a cambio de la mejora de la arquitectura de raíz tal y como se muestra en la figura 1, la absorción de nutrientes y estimulación del sistema inmune de la planta llevados a cabo por las PGPR que se conoce como Plant Growth-Promoting Rhizobacteria o, en español, rizobacterias promotoras del crecimiento vegetal.

### 3.3.1.1. Mecanismos de acción

Los mecanismos de acción de las bacterias promotoras del crecimiento vegetal (PGPR) en las plantas son diversos y se pueden clasificar en dos grandes grupos: los mecanismos extracelulares (ePGPR) y los intracelulares (iPGPR) [21].

Los mecanismos extracelulares ocurren fuera de las raíces, principalmente en la rizosfera o en los espacios intercelulares del córtex de la raíz. Por otro lado, los mecanismos intracelulares se llevan a cabo dentro de las células de la planta. Dependiendo de cómo influyen en el crecimiento vegetal, estos mecanismos se dividen en dos tipos principales: directos e indirectos [21].

Los mecanismos directos son aquellos que ocurren dentro de la planta y afectan directamente su metabolismo. Esto se logra principalmente a través de la modificación de la expresión genética de la planta, lo que puede resultar en un aumento en la absorción de nutrientes, la producción de fitohormonas, o la mejora de la capacidad de la planta para tolerar el estrés [22].

*Los mecanismos indirectos, en cambio, actúan fuera de la planta y no modifican directamente su metabolismo, pero contribuyen a su crecimiento al protegerla de patógenos o mejorar las condiciones del suelo. Estos mecanismos incluyen la producción de antibióticos, sideróforos (compuestos que secuestran hierro y limitan su disponibilidad para patógenos) y la competencia por nutrientes esenciales que de otro modo serían utilizados por organismos nocivos [21].*

### 3.3.1.2. Mecanismos de mejoramiento

Son un grupo especializado de rizobacterias que tienen la capacidad de mejorar el crecimiento y salud de las plantas mediante diversos mecanismos, las PGPR tienen 3 grupos de acción como se muestra en la ilustración 2 que dotan a la planta de diferentes mejoras entre los cuales se destacan:

### 3.3.1.2.1. Fijación biológica de nitrógeno (FBN)

La fijación biológica de nitrógeno es super importante porque el nitrógeno (N) es un nutriente fundamental para varios procesos vitales en las plantas, como la creación de aminoácidos, proteínas y ácidos nucleicos. Aunque hay mucho nitrógeno en la atmósfera en forma de N2, las plantas no lo pueden absorber directamente [13]. Para que esto sea posible, ciertos microorganismos, como las bacterias de la familia Rizobiaceae, transforman el nitrógeno en amonio (NH4+), que es la forma que las plantas sí pueden utilizar [14]. Todo este proceso depende de una enzima llamada nitrogenasa, que es muy sensible al oxígeno. Por eso, estos microorganismos han desarrollado formas de proteger su actividad, como compartimentar las células [23].

Este proceso es importante para mejorar la fertilidad del suelo sin tener que usar tanto fertilizante artificial.

### 3.3.1.2.2. Solubilización de fosfatos

El fósforo (P) es el segundo nutriente más importante después del nitrógeno, pero no siempre está disponible en el suelo porque suele estar en formas difíciles de disolver, ya sea en compuestos orgánicos o inorgánicos [24]. Las bacterias solubilizadoras de fosfato (PSB) ayudan a liberar este fósforo mediante la secreción de ácidos orgánicos, como el ácido glucónico o el 2-cetoglucónico, que "quitan" el fósforo de los minerales en el suelo [23]. También usan enzimas como fosfatasas y fitasas para liberar el fósforo de compuestos orgánicos. Este proceso mejora el uso que las plantas hacen de los fertilizantes fosfatados, lo que permite que se necesiten menos fertilizantes químicos [24].

### 3.3.1.2.3. Solubilización de potasio

El potasio (K) es importante para procesos como la fotosíntesis, la regulación de las estomas, el transporte de azúcares y la síntesis de proteínas. Aunque hay mucho potasio en el suelo, generalmente está en formas poco solubles. Algunas bacterias pueden solubilizar estos minerales y liberar el potasio para que las plantas lo puedan usar. Esto lo logran mediante la producción de ácidos orgánicos como el cítrico, oxálico y málico, que descomponen los minerales que contienen potasio. Entre las bacterias que tienen esta capacidad se encuentran las de los géneros Bacillus, Burkholderia y Paenibacillus. Este proceso es especialmente útil en cultivos como el maíz, trigo y tabaco [24].

### 3.3.1.2.4. Producción de fitohormonas

Las fitohormonas son sustancias que regulan el crecimiento y afectan muchos aspectos del desarrollo de las plantas. Algunas rizobacterias que promueven el crecimiento vegetal (PGPR) pueden producir fitohormonas como auxinas, citocininas, giberelinas, ácido abscísico y etileno. Estas hormonas ayudan a que las plantas absorban mejor los nutrientes y sean más resistentes al estrés [24].

### 3.3.1.2.4.1 Tipos de fitohormonas relevantes

**Auxinas:** Estimulan el crecimiento de las raíces, aumenta el tamaño, peso y la ramificación de las raíces, lo que permite a las plantas explorar mejor el suelo en busca de nutrientes [25].

**Citocininas:** Promueven la división celular y retrasan el envejecimiento de las hojas. También estimulan la formación de raíces adventicias y la expansión de las hojas [25]

**Giberelinas:** Ayudan en la germinación, el crecimiento del tallo, la floración y la formación de frutos. Son importantes para la tolerancia al estrés y el crecimiento de las plantas en condiciones difíciles [25].

**Ácido abscísico (ABA):** Sirve para aumentar la respuesta de la planta frente al estrés, como la sequía y la salinidad. Induce el cierre de las estomas para evitar la pérdida de agua y activa genes relacionados con la tolerancia al estrés [25].

### 3.3.1.2.5. Producción de exopolisacáridos y biopelículas

Ayudan a formar biopelículas en la superficie de las raíces. Estas biopelículas mejoran la adherencia de los microorganismos y crean una barrera física que protege a las plantas de situaciones difíciles, como la sequía o la alta salinidad. Además, los EPS también pueden activar las defensas de las plantas y mejorar la estructura del suelo, facilitando la retención de agua y nutrientes. Bacterias como Rhizobium y Bacillus son conocidas por su capacidad para producir EPS. [25].

### 3.3.2 Interacciones planta-microorganismos

Las interacciones entre las plantas y los microorganismos se desarrollan en tres zonas principales de la planta: la filosfera, la endosfera y la rizosfera. La filosfera abarca las partes aéreas como los tallos, hojas, flores y frutos, mientras que la endosfera se refiere al sistema interno de transporte de la planta. En cambio, la rizosfera es la región del suelo que rodea las raíces y es donde se concentra la mayor actividad microbiana como se muestra en la figura 3.3.2.1[26].

Esta zona, aunque no tiene límites precisos, suele extenderse unos pocos milímetros desde la superficie de las raíces. La geometría y extensión de la rizosfera pueden variar según factores ambientales y la cantidad de exudados que las raíces de las plantas secretan. Estos exudados incluyen compuestos como ácidos orgánicos, azúcares y aminoácidos, que influyen fuertemente en la actividad biológica y química de la rizosfera. [27].

La rizosfera es una de las áreas más ricas en microorganismos del suelo, conteniendo poblaciones densas de bacterias, hongos, levaduras y protozoos. De hecho, la cantidad de bacterias en esta región puede alcanzar cifras extremadamente altas, con hasta mil millones de células bacterianas por gramo de suelo [27].

**figura 3.3.2.1 Mecánicas de las PGPR [21]**

### 3.3.3 PGPR como alternativa

En la investigación Efecto de rizobacterias promotoras del crecimiento vegetal en el desarrollo y rendimiento del chile jalapeño [16], se explora el impacto de varias rizobacterias, el estudio se realizó en un invernadero en Gómez Palacio, Durango, México, y evaluó el efecto de estas bacterias en diferentes etapas de crecimiento mediante pruebas in vitro y en semilleros. Entre los resultados destacados, la rizobacteria S. plymuthica mostró mejoras en altura y número de hojas en plántulas, mientras que la rizobacteria S. marcescens favoreció el crecimiento en altura, número de hojas, área foliar y biomasa seca en condiciones de maceta. Además, la rizobacteria S. liquefaciens incrementó la cantidad y el tamaño de frutos.

Estos hallazgos se muestran de una forma más clara e la ilustración 3 y sugieren que el uso de rizo bacterias como biofertilizantes puede ser una alternativa sostenible para mejorar la producción de chile jalapeño se muestra en la graficas de la figura 3.3.3.1 [28].

Imagen que contiene Diagrama

Descripción generada automáticamente

figura 3.3.3 Grafica comparativa de efectos [28]

El estudio Inoculación de rizobacterias promotoras del crecimiento vegetal en tomate (Solanum lycopersicum L.) cv. afrodita en invernadero [19], Muestra que con el uso de rizobacterias los diámetros ecuatorial y polar de los frutos de los diferentes tratamientos, así como el contenido de sólidos solubles, presentaron diferencias altamente significativas, del mismo modo, se encontraron significancias estadísticas en espesor de pericarpio y firmeza de fruto por efecto de la interacción sustratos. En el contenido de fenoles totales y capacidad antioxidante se registró diferencia significativa en la interacción sustratos. En el peso de fruto se encontró diferencia significativa.

Bacterias rizosféricas con beneficios potenciales en la agricultura, concluye que Las rizobacterias representan una alternativa biotecnológica en la agricultura principalmente por la gran cantidad de mecanismos moleculares que permiten mejorar el rendimiento y la salud de las plantas. Los inoculantes a base de rizobacterias son una alternativa biotecnológica en la agricultura sustentable, que permite incrementar el rendimiento, disminuyendo los costos de producción en la práctica agrícola [30]

También se evaluó el efecto de la inoculación de tres cepas de rizobacterias promotoras de crecimiento vegetal (RPCV) Arthrobacter pokkaliiJLB4, Pseudomonas tolaasiiP61 y Pseudomonas tolaasiiA46 en plantas de pakchoi en condiciones controladas en  un  macrotúnel, [31]

Los resultados obtenidos mostraron que la cepa Pseudomonas tolaasiiP61 fue la que tuvo el mayor efecto en el peso seco, así como en el contenido comparados con el control, con incrementos de 11.5%, 10.8% y 12.3%, respectivamente. La cepa Arthrobacter pokkalii JLB4 provocó mayor crecimiento de la raíz, el volumen radical fue superior al testigo. Sin embargo, no tuvo efecto en el peso seco de parte aérea. Los resultados sugieren que el uso combinado de fertilizante y RPCV conduce a mayor acumulación de biomasa y de nutrientes como el P, K y S que la sola aplicación de fertilizante. [31]

## 3.4 Algoritmos Genéticos

Son técnicas de optimización y búsqueda inspiradas en los procesos de evolución natural. Estos algoritmos son parte del campo más amplio de la inteligencia artificial y la computación evolutiva, y son especialmente útiles para resolver problemas complejos donde las soluciones pueden no ser fácilmente identificables a través de métodos tradicionales [32].

### 3.4.1 Fundamentos de los Algoritmos Genéticos

Los AG se basan en la idea de la "supervivencia del más apto", un concepto que se refiere a cómo los individuos que poseen características más adecuadas para su entorno tienden a sobrevivir y reproducirse. En el contexto de los AG, los individuos se representan como "cromosomas" y su conjunto de soluciones posibles se llama "población"[33].

### 3.4.2 Partes de un algoritmo genético

El algoritmo genético consta de 8 partes que siguen una secuencia cíclica tal y como se muestra en la figura 3.4.2.1

Representación: En un AG, cada solución potencial se representa como un cromosoma. Esto puede ser en forma de cadenas binarias, números reales o estructuras más complejas, dependiendo del problema a resolver [34].

Población: Un AG comienza con una población inicial de cromosomas, que pueden ser generados aleatoriamente o a partir de soluciones conocidas [34].

Evaluación: Cada cromosoma se evalúa utilizando una función de aptitud que mide qué tan buena es la solución representada por ese cromosoma en comparación con el problema que se está resolviendo [34].

Selección: Los cromosomas se seleccionan para la reproducción basándose en su aptitud. Métodos comunes incluyen la selección por torneo, donde se eligen aleatoriamente dos cromosomas y se selecciona el de mayor aptitud, o la selección por ruleta, donde la probabilidad de selección está proporcional a la aptitud [34].

Cruzamiento (Crossover): Los cromosomas seleccionados se combinan para crear nuevos "descendientes". Este proceso se asemeja a la reproducción sexual, donde se intercambian partes de los cromosomas de los padres para generar nuevas soluciones [34].

Mutación: Para mantener la diversidad genética en la población, se aplican cambios aleatorios a algunos cromosomas. Esto puede incluir invertir un bit en una cadena binaria o cambiar un valor en un cromosoma numérico [34].

Reemplazo: Los nuevos descendientes se incorporan a la población, reemplazando a algunos de los cromosomas menos aptos, y el ciclo se repite [34].

**Convergencia y Terminado**

El proceso de selección, cruzamiento y mutación se repite durante un número determinado de generaciones o hasta que se alcanza un criterio de convergencia, como un valor de aptitud deseado. A medida que avanza el algoritmo, la población tiende a mejorar, encontrando soluciones más cercanas al óptimo [33, 34].

Diagrama

Descripción generada automáticamente

Figura 3.4.2.1 Proceso de un algoritmo genético [24].

### 3.4.3 Variantes de Algoritmos genéticos

Los algoritmos genéticos (AG) han evolucionado para adaptarse a distintos tipos de problemas y optimizaciones, dando lugar a varias variantes que modifican o mejoran el enfoque clásico. Estas variantes están diseñadas para mejorar la eficiencia, precisión o capacidad de exploración de los AG tradicionales, enfrentando diferentes desafíos en problemas complejos [34].

### 3.4.3.1. Algoritmos Genéticos Paralelos

Los AG paralelos están diseñados para aprovechar las arquitecturas de computación distribuida o de múltiples núcleos. En lugar de trabajar con una única población, estos AG dividen la población en subpoblaciones, conocidas como "islas", que evolucionan independientemente durante un número determinado de generaciones. De manera periódica, las islas intercambian individuos, lo que ayuda a mantener la diversidad genética y evitar la convergencia prematura [34].

Este enfoque mejora la exploración del espacio de búsqueda, ya que cada subpoblación puede explorar diferentes áreas del espacio de soluciones. También reduce el tiempo de procesamiento al distribuir la carga computacional [34].

### 3.4.3.2. Algoritmos Genéticos Adaptativos

Los AG adaptativos ajustan dinámicamente sus parámetros de operación, como las tasas de mutación y cruzamiento, en función del estado actual de la búsqueda. Por ejemplo, si se detecta que la población ha convergido prematuramente (es decir, que todos los individuos son similares entre sí), la tasa de mutación puede aumentar para introducir más diversidad [34].

Este enfoque es útil para evitar la "estancación" en óptimos locales y permite que el algoritmo sea más flexible y eficiente en problemas con paisajes de búsqueda complejos [34].

### 3.4.3.3. Algoritmos Genéticos Basados en Elitismo

En los AG con elitismo, se garantiza que los mejores individuos de la población actual se mantengan en la siguiente generación, independientemente del proceso de selección y reproducción. Este enfoque asegura que la solución óptima encontrada no se pierda debido a mutaciones o cruzamientos desfavorables [34].

El elitismo es útil para acelerar la convergencia hacia el óptimo, pero puede reducir la diversidad genética si se utiliza en exceso, lo que puede llevar a la convergencia prematura [34].

### 3.4.3.4 Algoritmos Genéticos Multiobjetivo

Los problemas de optimización multiobjetivo requieren encontrar soluciones que optimicen múltiples objetivos simultáneamente, que a menudo están en conflicto entre sí. Los AG multiobjetivo, como el NSGA-II (Non-dominated Sorting Genetic Algorithm), utilizan estrategias para identificar un conjunto de soluciones no dominadas, conocidas como el frente de Pareto. Estas soluciones representan un compromiso entre los diferentes objetivos, permitiendo que el usuario elija según las prioridades del problema [34].

Este tipo de AG es particularmente útil en ingeniería y diseño, donde se deben considerar múltiples criterios (por ejemplo, costo, rendimiento y durabilidad) [34].

### 3.4.3.5. Algoritmos Genéticos Híbridos

Los AG híbridos combinan el enfoque de los algoritmos genéticos con otras técnicas de optimización, como algoritmos locales, búsquedas o redes neuronales. Estos algoritmos buscan aprovechar las fortalezas de diferentes métodos para mejorar la eficiencia y efectividad de la búsqueda. Por ejemplo, un AG híbrido puede utilizar una búsqueda local para refinar las soluciones encontradas por el algoritmo genético, mejorando la velocidad de convergencia hacia el óptimo global [34].

### 3.4.4 Estrategias para los operadores genéticos

### 3.4.4.1 Métodos de Representación

La representación define cómo las soluciones (o individuos) se codifican en el AG. Algunas formas comunes de representar los cromosomas son:

Representación Binaria: Cada cromosoma es una cadena de bits (0s y 1s). Es el tipo de representación más común en los AG clásicos [35].

Representación de Números Reales: Cada gen del cromosoma es un número real, útil para problemas de optimización continua [35].

Representación Permutacional: Se utiliza en problemas como el de la ruta del viajero, donde el cromosoma representa una permutación de elementos [35].

Representación Basada en Árboles: Usada en programación genética, donde cada cromosoma se representa como un árbol (por ejemplo, para representar funciones matemáticas) [35].

### 3.4.4.2 Métodos de Selección

La selección determina qué individuos de la población actual se usarán para producir la próxima generación. Algunos métodos comunes incluyen:

Selección por Ruleta (Roulette Wheel): La probabilidad de que un individuo sea seleccionado es proporcional a su aptitud. Se asigna una "rueda" donde cada sector es proporcional a la aptitud del individuo [25].

Los individuos con mayor aptitud tienen más probabilidades de ser seleccionados [35].

Selección por Torneo: Se eligen al azar varios individuos de la población, y el que tiene la mejor aptitud gana el "torneo" y es seleccionado. Este método puede ser ajustado para favorecer la selección de individuos más aptos [35].

Selección Elitista: Garantiza que los mejores individuos de la generación actual se mantendrán en la siguiente. Ayuda a conservar las soluciones óptimas encontradas [35].

Selección de Ranking: Los individuos se ordenan por aptitud y se asignan probabilidades de selección basadas en su rango, no en su aptitud absoluta. Esto evita la dominancia de los individuos con aptitudes extremadamente altas [35].

Métodos de Reproducción (Cruzamiento)

El cruzamiento o crossover combina el material genético de dos padres para producir descendientes. Algunos de los métodos más comunes son:

Cruzamiento de Punto Único (Single-point Crossover): Se selecciona un punto de corte al azar en el cromosoma de los padres, y los segmentos de los padres se intercambian en ese punto para producir dos descendientes tal y como lo muestra la figura 3.4.4.2 [35].

Imagen que contiene Diagrama

Descripción generada automáticamente

Figura 3.4.4.2 Cruzamiento de un punto [35]

Cruzamiento de Dos Puntos (Two-point Crossover): Dos puntos de corte se seleccionan en los cromosomas de los padres, y el segmento intermedio se intercambia. La Figura 3.4.4.3 muestra como este método introduce mayor variación que el cruzamiento de un solo punto [33].

*Tabla

Descripción generada automáticamente*

Figura 3.4.4.3 Cruzamiento en dos puntos [35]

*Cruzamiento Uniforme: Cada gen del descendiente se selecciona aleatoriamente de uno de los padres. Esto permite una mayor mezcla de genes en cada generación tal y como lo ejemplifica la* Figura 3.4.4.*4 [25].*

*Imagen que contiene Diagrama

Descripción generada automáticamente*

Figura 3.4.4.4 Cruzamiento uniforme [35]

Cruzamiento en Permutación: Utilizado en problemas donde la representación es una permutación, como el problema del vendedor viajero, para generar descendientes válidos (sin duplicados en las permutaciones) [35].

### 3.4.4.3 Métodos de Mutación

La mutación introduce variación al modificar genes en los cromosomas. Los métodos de mutación más comunes incluyen:

Mutación de Bit (Bit-flip Mutation) mostrada en la Figura 3.4.4.1 muestra en representaciones binarias, se selecciona un gen o más al azar y se invierte (de 0 a 1 o de 1 a 0) [35].

Imagen de la pantalla de un celular con texto e imagen

Descripción generada automáticamente con confianza baja

Figura 3.4.4.1 Mutación bit-flip [35]

Mutación Gaussiana: En representaciones con números reales, los genes se modifican agregando un valor aleatorio basado en una distribución normal (gaussiana) [25].

Mutación por Intercambio (Swap Mutation): En representaciones de permutación, dos genes se seleccionan y se intercambian [35].

Mutación por Inversión: Un segmento del cromosoma se selecciona al azar y se invierte [25].

## 3.5 Redes Neuronales

Una red neuronal es un modelo computacional que imita la organización de las neuronas en el cerebro, con varias capas de nodos interconectados. Este tipo de red tiene la capacidad de aprender a partir de datos, permitiendo su entrenamiento para identificar patrones, clasificar información y predecir eventos futuros. [36]

### 3.5.1 Neuronas artificiales

Son unidades computacionales simples que tienen señales de entrada (x1, x2, xn) multiplicadas por un peso y sumadas para ser evaluadas por una función de activación [37].

Las redes neuronales se denominan redes neuronales artificiales (ANN) o redes neuronales simuladas (SNN). Son un subconjunto del aprendizaje automático [38].

Consta de una capa de entrada, una o varias capas ocultas y una capa de salida. Las capas están interconectadas mediante nodos, o neuronas; cada capa utiliza la salida de la capa anterior como entrada [35].

Imagen que contiene Diagrama

Descripción generada automáticamente

Figura 3.5.1.1 Neurona artificial [37]

### 3.5.2 Funciones de activación

Según [39] es una transformación no lineal del valor de salida de una neurona antes de que el valor se pase como entrada a los cálculos de la siguiente capa de la red neuronal, para asegurar que los resultados sean lo lineales, algunas de las funciones más relevantes son:

***Relu:*** Lleva a cabo una operación de umbral no lineal en la que cualquier valor de entrada inferior a cero como se muestra en la figura 3.5.1.2 se establece en cero [40].

Esta operación es equivalente a:

Imagen que contiene Texto

Descripción generada automáticamente

Figura 3.5.1.2 **Relu** [37]

***Softmax:*** Generalización de la regresión logística que permite aplicarla en escenarios multiclase. A la hora de predecir la clase de una muestra, el modelo calcula un score para cada clase y hace pasar el vector de scores por la función softmax como se muestra en la figura 3.5.1.3 (también llamada exponencial normalizada) [31].

Esta función viene definida por la siguiente fórmula:

Texto

Descripción generada automáticamente

Figura 3.5.1.3 **Softmax** [37]

### 3.5.3 Arquitectura de una Red Neuronal

Se denomina arquitectura a la topología, estructura o patrón de conexionado de una red neuronal. [42]

Se distinguen distintos tipos de capas: la capa de entrada o sensorial está compuesta por neuronas que reciben datos o señales del entorno; la capa de salida es aquella cuyas neuronas proporcionan la respuesta de la red neuronal; las capas ocultas no tienen conexión directa con el entorno, proporcionando a la red neuronal grados de libertad adicionales como se muestra en la figura 3.5.3.1 [42,43].

Diagrama, Esquemático

Descripción generada automáticamente

figura 3.5.3.1 Arquitectura de una red neuronal [43]

***Capa de Entrada:*** Recibe los datos brutos.

***Capas Ocultas***: Son una o más capas intermedias que procesan la información. Cada neurona en estas capas realiza operaciones matemáticas sobre las entradas que recibe y produce una salida, la cual es transmitida a la siguiente capa.

***Capa de Salida:*** Genera el resultado final del modelo, que puede ser una predicción en función de los datos de entrada.

### 3.5.4 Algoritmo de aprendizaje

El aprendizaje o proceso de entrenamiento de una red neuronal se entiende como la etapa mediante la cual la red neuronal adquiere la capacidad de desempeñar las funciones específicas que requiere el proble el algoritmo de aprendizaje ajusta adaptativamente la diferencia entre el output deseado y el que la red neuronal genera. Así, la red aprende a realizar automáticamente una tarea prescrita [43].

Mediante el aprendizaje, los pesos de las conexiones a la red se modifican, y se dice que la red ha aprendido cuando los valores de los pesos permanecen estables o tienen un error mínimo. Los criterios para cambiar el valor de los pesos de conexión son determinados por la regla de aprendizaje. [42,43]

El aprendizaje de una RNA se basa en ajustar los pesos de las conexiones entre las neuronas mediante un proceso iterativo conocido como entrenamiento. Este proceso implica:

Forward Propagation: Los datos fluyen desde la capa de entrada hasta la de salida a través de las capas ocultas, generando una predicción [43].

Cálculo de Error (Función de Costo): La predicción generada por la red se compara con el valor real (ground truth) usando una función de costo, que mide el error de la predicción que se pretende abordar [42, 43].

Para la regresión se debe usar **Error Cuadrático Medio (Mean Squared Error, MSE), ya que es el indicador que muestra la desviación estándar de los valores residuales (errores de predicción). Los valores residuales son una medida de la distancia de los puntos de datos de la línea de regresión; RMSE es una medida de cuál es el nivel de dispersión de estos valores residuales, le indica el nivel de concentración de los datos en la línea de mejor ajuste, como se indica en la fórmula siguiente se muestra en la figura 3.5.4.1** **[44].**

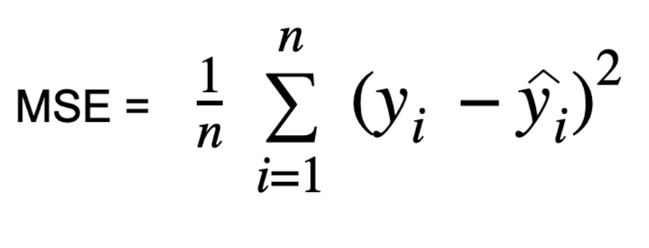


figura 3.5.4.1 usar **Error Cuadrático Medio** [44]

**n**: Número total de observaciones.

**yᵢ**: Valor real (observado) de la i-ésima observación.

**ŷᵢ**: Valor predicho para la i-ésima observación.

**Σ**: Sumatoria desde i = 1 hasta n.

**Entropía cruzada, es** Utilizada para tareas de clasificación, [35] dice que es una métrica que se puede utilizar para reflejar la precisión de los pronósticos probabilísticos. La cross-entropy tiene fuertes vínculos con la estimación de máxima verosimilitud. La cross-entropy es de suma importancia para los sistemas de pronóstico modernos, ya que es instrumental para lograr pronósticos superiores, incluso para métricas alternativas, su fórmula es se muestra en la figura 3.5.4.2:

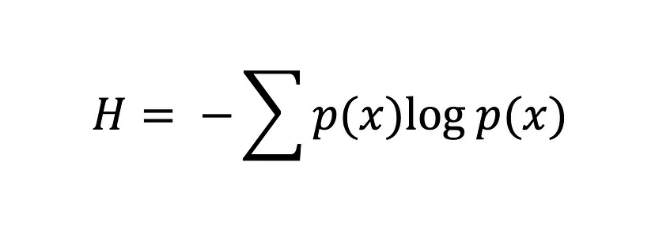


figura 3.5.4.1 Entropía [44]

Donde:

**H (p, q)**: Entropía cruzada entre la distribución verdadera (**p**) y la distribución predicha (**q**).

**pᵢ**: Probabilidad verdadera de la clase **i** (valor real, generalmente 1 para la clase correcta y 0 para las demás en clasificación one-hot).

**qᵢ**: Probabilidad predicha para la clase **i**.

**Σ**: Sumatoria sobre todas las clases posibles.

**Backpropagation (Retropropagación del Error):** El error se propaga de vuelta a través de la red para ajustar los pesos de las conexiones utilizando técnicas de optimización como el gradiente descendente. [32]

### 3.5.5 Tipos de redes neuronales 3.5.5.1 Redes Neuronales Convolucionales (CNN)

Las Redes Neuronales Convolucionales (Convolutional Neural Networks, CNN) son un tipo de arquitectura diseñada específicamente para procesar datos estructurados espacialmente, como imágenes, series temporales o secuencias de datos. Se destacan por su capacidad de extraer patrones locales mediante operaciones de convolución [46, 47].

Principales Componentes de una CNN

Capas Convolucionales: Estas capas aplican filtros o kernels que extraen características locales del input. Cada filtro recorre la entrada y produce un mapa de características que resalta patrones específicos, como bordes, texturas o motivos en datos genómicos [46].

Capas de Pooling: Estas capas reducen la dimensionalidad de los mapas de características, permitiendo que la red enfoque su atención en las características más relevantes y reduzca el sobreajuste. El max pooling selecciona el valor más alto en una región, mientras que el average pooling toma el promedio [46].

Capas Densas: Una vez que las capas convolucionales han extraído las características, las capas densas procesan esa información para realizar tareas como clasificación o regresión [36].

## 3.5.5.2 Redes Neuronales Recurrentes

Según [48] Redes Neuronales Recurrentes (Recurrent Neural Networks, RNN) son arquitecturas diseñadas para modelar secuencias de datos al capturar dependencias temporales o contextuales. Esto se logra mediante conexiones recurrentes que permiten que la salida de una capa al tiempo dependa tanto de la entrada actual como de su estado anterior.

Principales Componentes de una RNN

Estado Oculto (Hidden State): Una RNN mantiene un vector de estado oculto​, que almacena información relevante de los pasos anteriores y se actualiza en cada paso de tiempo [49].

Celdas de Memoria (LSTM/GRU): Para superar el problema del gradiente desvanecido en las RNN tradicionales, se introdujeron celdas como LSTM (Long Short-Term Memory) y GRU (Gated Recurrent Unit), que regulan qué información conservar o olvidar mediante puertas (gates)[48].

## 3.6 Bach

El procesamiento por lotes es una técnica que utiliza lotes para procesar grandes volúmenes de datos. En lugar de introducir todo el conjunto de entrenamiento en un modelo a la vez, se dividen los datos en lotes y se realiza una secuencia de trabajos unificados de entrenamiento consecutivo del modelo en un lote tras otro [50].

El tamaño del lote es uno de los hiperparámetros más importantes en el entrenamiento de aprendizaje profundo y representa la cantidad de muestras utilizadas en un paso hacia adelante y hacia atrás a través de la red y tiene un impacto directo en la precisión y la eficiencia computacional del proceso de entrenamiento. El tamaño del lote se puede entender como un equilibrio entre precisión y velocidad [51].

Las iteraciones son la cantidad de lotes necesarios para completar una época que se utiliza para medir el progreso del proceso de entrenamiento. El recuento de iteraciones es igual a la cantidad de lotes en una época y se calcula dividiendo la cantidad total de muestras en el conjunto de datos de entrenamiento por el tamaño del lote [50].

### 3.7. Embebido

Transforma la información en representaciones numéricas que captan las similitudes y relaciones entre conceptos, agilizando el trabajo de los sistemas. [52]

Es un medio de representar objetos como texto, imágenes y audio como puntos en un espacio vectorial continuo donde las ubicaciones de esos puntos en el espacio son semánticamente significativas para los algoritmos de machine learning. [54]

Un de estos embeddings es “dna2vec” (Representaciones vectoriales consistentes de k-meros de longitud variable) aplicada en [53] como un método para entrenar representaciones distribuidas de k-meros de longitud variable. El método se basa en el modelo de incrustación de palabras word2vec, que se entrena en una red neuronal de dos capas poco profunda.

Las representaciones más comunes de una secuencia larga de ADN es dividirla en componentes k-meros más cortos. Desafortunadamente, la codificación vectorial sencilla de k-meros como un vector one-hot es vulnerable a la maldición de la dimensionalidad. [52]

# Capítulo 4. Análisis

En este capítulo se incluyen todas las fases de análisis relacionadas con el trabajo terminal

## 4.1 Requerimientos

El presente trabajo terminal determina requerimientos funcionales y no funcionales que desglosaremos en esta sección con el índice 4.1.1 y 4.1.2 respectivamente.

### 4.1.1 Requerimientos funcionales

En la Tabla 2, se desglosan los requerimientos funcionales del trabajo terminal.

|  |  |  |  |
| --- | --- | --- | --- |
| RF# | Descripción | Número y nombre de caso de uso | En caso de tener regla de negocio |
| RF1 | La red neuronal solicita los datos | CU1 solicita conjunto de datos |  |
| RF2 | La red obtiene el genoma | CU2 buscar genoma |  |
| RF3 | La red obtiene la lista de PGPTs | CU3 buscar PGPT list |  |
| RF4 | La red obtiene el conteo de PGPTs | CU4 buscar PGPT count |  |
| RF5 | La red obtiene el ambiente | CU5 buscar environment |  |
| RF6 | Entrenar la red neuronal | CU6 Clasificación PGPT |  |
| RF7 | El algoritmo genético solicita el vector de características deseadas | CU7 Solicitar vector de características deseadas |  |
| RF8 | El algoritmo genético generara secuencias aleatorias | CU8 Algoritmo |  |
| RF9 | El algoritmo genético evaluara los efectos de las secuencias generadas aleatoriamente a través de la red neuronal entrenada | CU9? |  |
| RF10 | El algoritmo genético seleccionara los genomas que tengan características similares a el vector de características deseadas | CU10? |  |
| RF11 | El algoritmo genético combinara características de los genomas seleccionados, así como aplicar la mutación | CU11? |  |
| RF12 | El algoritmo genético repetirá su proceso hasta cumplir con un determinado número de repeticiones | CU12 | RN1 |

**Tabla 2**. Requerimientos funcionales, elaboración propia.

### 4.1.2 Requerimientos no funcionales

En la Tabla 3, se desglosan los requerimientos no funcionales del trabajo terminal.

|  |  |  |  |
| --- | --- | --- | --- |
| RNF# | Descripción | Propiedad | Aplicación |
| RNF1 | Tiempo máximo de ejecución del algoritmo genético | Rendimiento | Tiempo de terminación máximo de 2 minutos |
| RNF2 | El tiempo máximo de entrenamiento de la red neuronal | Rendimiento | Tiempo máximo de entrenamiento de 12 horas |

**Tabla 3**. Requerimientos no funcionales, elaboración propia.

### 4.1.3 Reglas de negocio

En la Tabla 2, se desglosan las reglas de negocio propuestas para el trabajo terminal.

|  |  |  |  |
| --- | --- | --- | --- |
| **RN#** | **Descripción** | **Propiedad** | **Aplicación** |
| RN1 | El parámetro que defina las generaciones del algoritmo genético no tiene que tener un valor mayor a 200 | Rendimiento | Generaciones máximas del algoritmo de 200 |

**Tabla 4**. Reglas de negocio, elaboración propia.

## 4.2 Análisis de herramientas a usar

En este apartado se presentará una comparativa de diferentes herramientas a usar en este Trabajo Terminal.

### 4.2.1 Análisis de sistemas operativos

A continuación, se desglosa una comparativa de los sistemas operativos en la Tabla 5.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Sistema operativo | Rendimiento | Compatibilidad con herramientas de bioinformática | Facilidad de aprendizaje | Familiaridad con el equipo |
| Linux | Alto | Media | Baja | Media |
| Windows | Medio | Alta | Alta | Alta |
| Mac | Medio | Media | Alta | Baja |

**Tabla 5**. Análisis de sistemas operativos, elaboración propia.

### 4.2.2 Análisis de lenguajes de programación

En la Tabla 6, se muestra el análisis de lenguajes de programación.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lenguaje | Rendimiento | Bibliotecas par a redes neuronales | Bibliotecas para algoritmos genéticos | Comunidad/ Soporte |
| Python | X | √ | √ | √ |
| C | √ | √ | X | √ |
| Ruby | √ | X | X | X |
| Java | X | √ | X | √ |
| JavaScript | X | √ | X | √ |

**Tabla 6**. Análisis de lenguajes de programación, elaboración propia.

## 4.3 Análisis de riesgos

En la Tabla 7, se muestra el análisis de riesgos detectados para el desarrollo del trabajo terminal.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Identificador del riesgo** | **Descripción** | **Categoría** | **Probabilidad** | **Gestión del riesgo** |
| R01 | Perdida de avances en el código | Técnico | Moderada | Realizar un respaldo en la nube todos los viernes |
| R02 | Retraso en el entrenamiento de la red neuronal | Técnico | Moderada |  |
| R03 | Indisposición para trabajar por parte de un compañero de equipo | Personal | Baja | Realizar una reunión en línea para repartir asignaciones |
| R04 | Insatisfacción por parte de un sinodal referente al documento técnico | Organizacional | Alta | Solicitar revisiones parciales y específicas, así como plantear una métrica para su satisfacción |
| R05 | Problemas en la obtención de datos | Técnico | Baja | Solicitar la ayuda de un experto en bioinformática |
| R06 | Falta de recursos computacionales para entrenar el modelo | Técnico | Media | Solicitar acceso a la Escuela Superior de Computo a servidores para el entrenamiento |
| R07 | Datos insuficientes para entrenar el modelo | Técnico | Media | Ampliar el paradigma del proyecto e incluir otros tipos de bacterias asociadas a plantas |

**Tabla 7**. Análisis de riesgos, elaboración propia.

## 4.4 Análisis de factibilidad

Se evaluaron las factibilidades técnica, operativa, económica y legal del proyecto para determinar su viabilidad y garantizar su implementación exitosa.

### 4.4.1 Factibilidad técnica

El presente trabajo terminal cuenta con los siguientes equipos mostrados en la Tabla 8.

|  |  |
| --- | --- |
| **Nombre del equipo** | **características** |
| MSI | Nombre del dispositivo MSI  Procesador 12th Gen Intel(R) Core(TM) i5-12450H 2.00 GHz  RAM instalada 16.0 GB (15.7 GB usable)  Identificador de dispositivo 3A88E38A-29F2-4492-8435-D0E5B796EF86  Id. del producto 00342-21075-52069-AAOEM  Tipo de sistema Sistema operativo de 64 bits, procesador basado en x64  Lápiz y entrada táctil La entrada táctil o manuscrita no está disponible para esta pantalla |
| DESKTOP-MQ4R9C4 | Procesador Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz 2.59 GHz  RAM instalada 16.0 GB  Id. del dispositivo 9B1BE3C9-AC06-4FEF-B3A7-39B84EC5AA05  Id. del producto 00327-30810-23040-AAOEM  Tipo de sistema Sistema operativo de 64 bits, procesador x64  Lápiz y entrada táctil La entrada táctil o manuscrita no está disponible para esta pantalla |
| ASUS TUF GAMING | **LAPTOP ASUS TUF GAMING FX507VU-LP180 GeForce RTX™ 4050 I7-13620H/ 16GB/ SSD1TB/ V6GB/ FREE/ 15.6**  **Procesador**: Core i7-13620H **Memoria**: 16 GB DDR5-4800 **Almacenamiento**: SSD 1 TB **Pantalla**: 15,6 pulgadas FHD **Tarjeta de vídeo**: GeForce RTX™ 4050 6GB |

**Tabla 8**. Análisis de lenguajes de factibilidad técnica, elaboración propia.

### 4.4.2 Factibilidad operativa

A continuación, se publicó una encuesta en un grupo de biotecnología en una red social para recabar datos sobre la opinión de este grupo en particular sobre la importancia de esta investigación, es importante notar que en la primera parte del instrumento se explicara el proyecto de forma resumida. Para este propósito, se diseñaron las siguientes preguntas:

1. **¿Considera relevante esta investigación para el avance de la biotecnología?**
   * *Explicación:* Es fundamental conocer si el grupo objetivo percibe la relevancia de la investigación, ya que esto valida el impacto potencial del trabajo terminal.
2. **¿Cree que los resultados de esta investigación podrían ser aplicados en la práctica?**
   * *Explicación:* Identificar el potencial de aplicabilidad ayuda a justificar la utilidad del proyecto y refuerza su factibilidad.
3. **¿Qué tan familiarizado está con las técnicas utilizadas en el proyecto, como redes neuronales y algoritmos genéticos?**
   * *Explicación:* Esta pregunta evalúa el nivel de conocimiento del grupo objetivo, lo que puede indicar si es necesario incluir más antecedentes en el documento técnico.
4. **¿Considera que el enfoque del proyecto es innovador en comparación con otros trabajos similares?**
   * *Explicación:* Determina si el grupo percibe innovación en el proyecto, lo cual puede influir en su interés y apoyo.
5. **¿Qué tan importante considera que es el diseño de microorganismos para la mejora del crecimiento vegetal?**
   * *Explicación:* Evalúa la percepción del impacto del tema específico dentro del ámbito de la biotecnología.
6. **¿Qué sugerencias tendría para mejorar el alcance del proyecto?**
   * *Explicación:* Permite obtener retroalimentación específica y útil para ajustar los objetivos y enfoque del proyecto.
7. **¿Está de acuerdo con la metodología técnica propuesta para este trabajo terminal?**
   * *Explicación:* Evalúa la aceptación de las técnicas y pasos descritos, lo que puede señalar posibles áreas de mejora.
8. **¿Recomendaría explorar colaboraciones con otros campos relacionados para enriquecer el proyecto?**
   * *Explicación:* Identifica oportunidades interdisciplinarias que podrían incrementar el valor del trabajo.

Estas preguntas fueron seleccionadas y diseñadas para recabar información específica que permita fortalecer los fundamentos del trabajo terminal y ajustarlo según la percepción y necesidades del grupo objetivo.

Los resultados fueron obtenidos y proporcionaron la siguiente grafica mostrada en la FIGURA X.

### 4.4.3 Factibilidad económica

En esta sección, se realizó un análisis de factibilidad económica dividida en diferentes aspectos, las cantidades se encontrarán en la moneda local (pesos mexicanos).

**4.4.3.1 Recursos humanos**

Se contemplaron los siguientes roles para el trabajo terminal mostrados en la Tabla 9, los datos fueron recabados de [54] y [55].

|  |  |  |  |
| --- | --- | --- | --- |
| **Núm.** | **Cargo** | **Costo individual** | **Costo Total** |
| 1 | Bioinformático | $8,000.00 | $8,000.00 |
| 2 | Desarrollador de IA | $27,103.00 | $54,206.00 |
|  | Total |  | $62,206.00 |

**Tabla 9**. Análisis de recursos humanos, elaboración propia.

**4.4.3.2 Recursos tecnológicos**

Se consideraron los siguientes recursos tecnológicos mostrados en la Tabla 10.

|  |  |  |  |
| --- | --- | --- | --- |
| **Cantidad** | **Descripción** | **Costo/hora** | **Total** |
| 1 | Computadora MSI: Procesador 12th Gen Intel Core i5-12450H, RAM 16 GB, sistema operativo x64. | $20.00 | $2,800.00 (140 h) |
| 1 | Computadora DESKTOP-MQ4R9C4: Procesador Intel Core i7-9750H, RAM 16 GB, sistema operativo x64. | $20.00 | $2,800.00 (140 h) |
| 1 | Laptop ASUS TUF GAMING: Core i7-13620H, 16 GB DDR5, SSD 1 TB, GeForce RTX 4050, pantalla FHD. | $25.00 | $3,500.00 (140 h) |
| 1 | Licencia de Microsoft Office |  | $800.00 |
|  | Total |  | $9,900.00 |

**Tabla 10**. Análisis de recursos tecnológicos, elaboración propia.

**4.4.3.3 Recursos materiales**

Se consideraron los siguientes recursos materiales mostrados en la Tabla 11.

|  |  |  |  |
| --- | --- | --- | --- |
| **Cantidad** | **Descripción** | **Costro unitario** | **Total** |
| 2 | Impresión de documento técnico (uno por semestre) | $300.00 | $600.00 |
| 440 | Transporte público para 2 personas (11 meses, 20 días hábiles/mes, $10 por día por persona) | $10.00 | $4,400.00 |
|  | Total |  | $5,00.00 |

**Tabla 11**. Análisis de recursos materiales, elaboración propia.

**4.4.3.4 Flujo de pago**

Se consideraron los siguientes recursos tecnológicos mostrados en la TABLA.

|  |  |
| --- | --- |
| **Recursos** | **Costo** |
| Recursos humanos | $62,206.00 |
| Recursos tecnológicos | $9,900.00 |
| Recursos materiales | $5,00.00 |
| Total | $77,106.00 |

**Tabla 12**. Flujo de pago, elaboración propia.

**4.4.4 Factibilidad legal [56]**

El proyecto fue evaluado conforme a las normativas vigentes en México y se verificó su cumplimiento con la "Ley de Bioseguridad de Organismos Genéticamente Modificados" (LBOGM), la cual regula el uso, manejo y diseño de organismos genéticamente modificados (OGMs).

**Aspectos legales considerados:**

1. **Uso de OGMs:** De acuerdo con el Artículo 1 de la LBOGM, el proyecto se clasifica dentro de actividades de uso confinado, garantizando la evaluación de riesgos y medidas de bioseguridad.
2. **Protección del medio ambiente:** El diseño de rizobacterias está enfocado en mejorar cultivos sin causar daños al medio ambiente, cumpliendo los principios del Artículo 9.
3. **Propiedad intelectual:** Se respetarán las normativas sobre propiedad intelectual aplicables a los procesos biotecnológicos.

**Cláusula de responsabilidad:** El equipo desarrollador del presente trabajo terminal no se hará responsable por el uso indebido del algoritmo fuera de los propósitos éticos y legales especificados en este proyecto. Cualquier modificación que resulte en consecuencias negativas será responsabilidad de quien la implemente.

**4.5 Análisis de sostenibilidad**

El trabajo terminal considera tanto los aspectos ambientales como los económicos y sociales, asegurando que el impacto general sea positivo y contribuya al desarrollo sostenible.

**Evaluación del impacto ambiental:**

1. **Reducción en el uso de fertilizantes químicos:** El diseño de rizobacterias optimizadas busca disminuir la dependencia de fertilizantes convencionales, reduciendo la contaminación del suelo y las emisiones de gases de efecto invernadero.
2. **Uso responsable de recursos:** Se planea emplear energía eficiente y minimizar el desperdicio de materiales durante el desarrollo del proyecto.
3. **Manejo adecuado de desechos:** Se implementarán protocolos para la disposición segura de cualquier material residual generado durante las pruebas experimentales.

**Estrategias para minimizar impactos:**

* Uso de infraestructura computacional eficiente que optimice el consumo energético.
* Divulgación de los descubrimientos realizados en el trabajo terminal.

# Capítulo 5. Diseño

## 5.1 Diagrama de arquitectura del software, sistema, aplicación

### 5.1.1 Diagrama de contexto

El diagrama de contexto muestra el sistema a desarrollar y las interacciones principales con actores externos y otros sistemas como se muestra en la figura 5.1.1.1. En este caso, el sistema está compuesto por dos módulos principales:

* **Red Neuronal**: Interactúa con el conjunto de datos, obtiene información genómica y entrena un modelo para clasificar PGPTs.
* **Algoritmo Genético**: Utiliza el modelo entrenado para generar y evaluar genomas con base en un vector de características deseadas.

**Actores:**

1. **Usuario Científico**: Proporciona el conjunto de datos y define los parámetros del entrenamiento y vector de características deseadas.
2. **Base de Datos Genómica**: Fuente de datos para los procesos de búsqueda y entrenamiento.

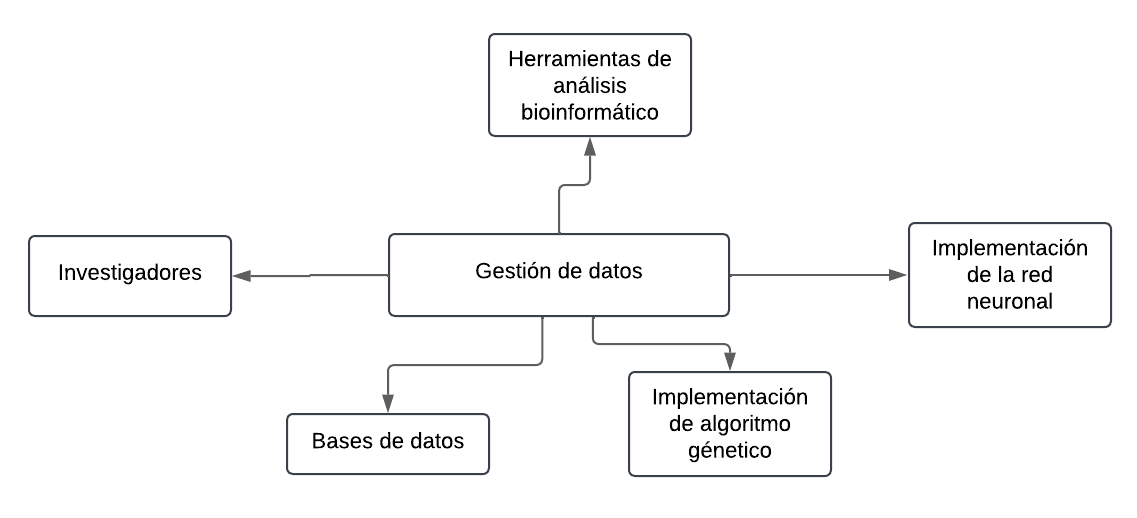


figura 5.1.1.1 Diagrama Contexto. Elaboracion Propia

### 5.1.2 Diagrama de contenedores

El sistema se divide en los siguientes contenedores como se muestra en la figura 5.1.1.2:

* **Backend**: Realiza la gestión de datos, entrenamiento de la red neuronal y ejecución del algoritmo genético.
* **Almacenamiento**: Bases de datos que contienen los genomas, las listas de PGPTs y los resultados del entrenamiento.

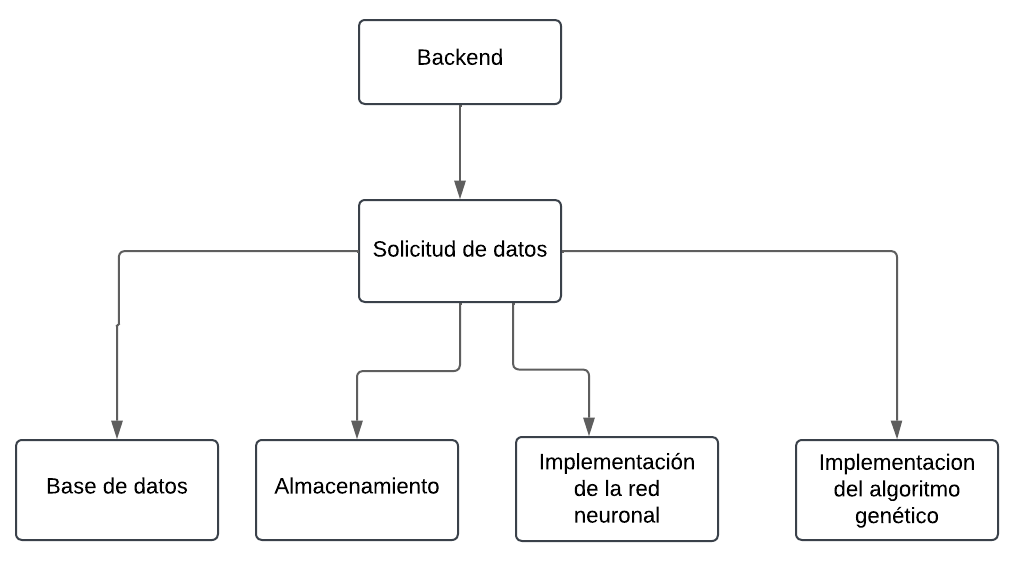


figura 5.1.2.2 Diagrama Contenedores.. Elaboracion Propia

### 5.1.3 Diagrama de componentes

Dentro del contenedor de solicitación de datos como se muestra en la figura 5.1.3.4, se identifican los siguientes componentes:

1. **Módulo de Red Neuronal**: Responsable de entrenar el modelo y realizar clasificaciones.
2. **Módulo de Algoritmo Genético**: Genera secuencias aleatorias, evalúa y selecciona genomas.
3. **Almacenar Datos genómicos**: permitir la consulta de datos y resultados.
4. **Gestor de Datos**: Administracion bases de datos.

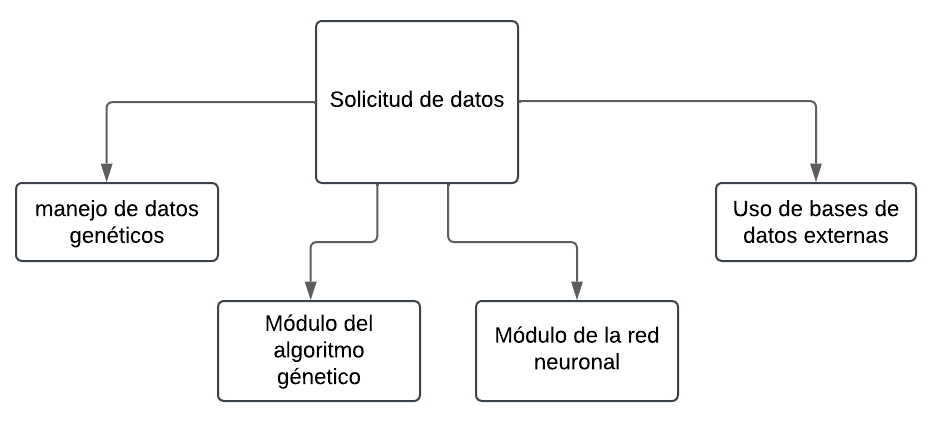
****

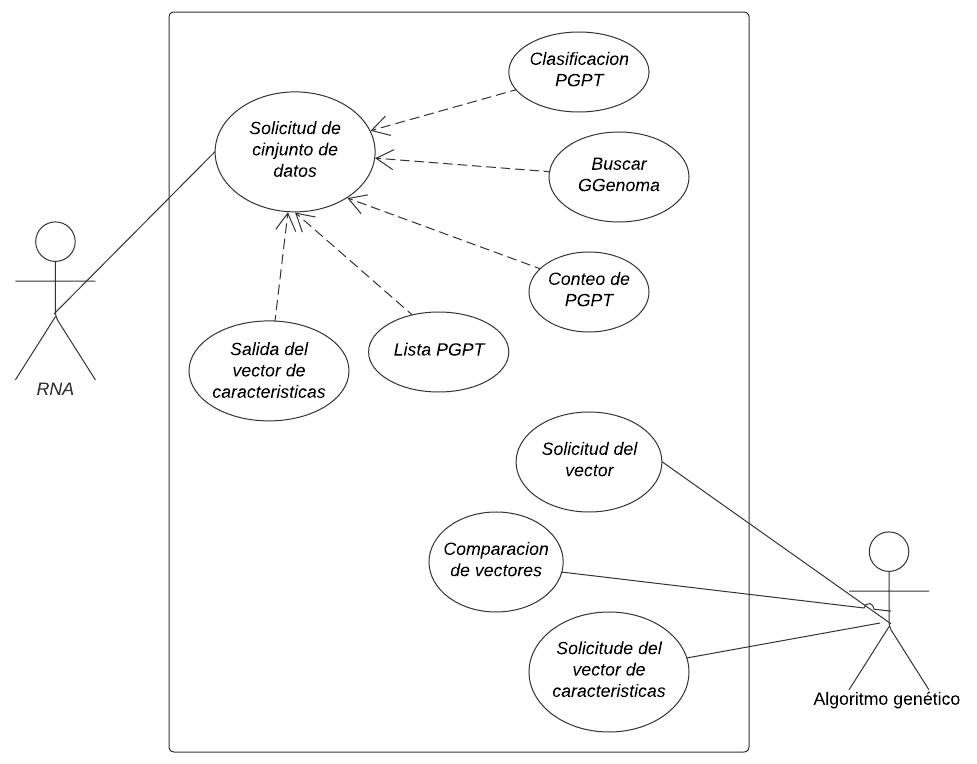
figura 5.1.3.3 Diagrama Contenedores.. Elaboracion Propia

## 5.2 Casos de uso

### 5.2.1 Diagrama de casos de uso

Se utiliza notación UML para identificar los casos de uso principales y sus relaciones con los actores como se muestra en la figura 5.2.1.4. Los casos de uso clave son:

1. Solicitar conjunto de datos.
2. Buscar genoma.
3. Buscar lista de PGPTs.
4. Buscar conteo de PGPTs.
5. Buscar información ambiental.
6. Clasificación de PGPTs (entrenar red neuronal).
7. Generar secuencias genéticas aleatorias.
8. Evaluar secuencias generadas.
9. Seleccionar genomas similares al vector deseado.
10. Aplicar mutación y combinación de genomas.



## 5.2.2 Especificación de Casos de Uso

**Caso de Uso No. 1: Solicitar Conjunto de Datos**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Solicitar Conjunto de Datos** | **Identificador: CU1** |
| **Actores** | Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF1 - La red neuronal solicita los datos |  |
| **Precondición** | La lista de datos debe estar disponible en el sistema. |  |
| **Postcondición** | Se obtiene un conjunto de datos listo para el análisis. |  |
| **Descripción** | La red neuronal solicita el conjunto de datos que será utilizado como entrada para los procesos futuros. |  |
| **Resumen** | El sistema entrega un conjunto de datos solicitado por la red neuronal. |  |

Tabla 13. Flujo del caso de uso "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Red neuronal | Solicita un conjunto de datos. |
| 2 | Sistema | Valida la disponibilidad de los datos. |
| 3 | Sistema | Entrega el conjunto solicitado. |
| 4 | Red neuronal | Confirma la recepción de los datos. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si la lista de datos no está disponible, el sistema informa del fallo a la red neuronal. |
| 3 | Si los datos no pueden ser entregados, el sistema detiene el proceso y solicita entradas. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso No. 2: Buscar Genoma**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Buscar Genoma** | **Identificador: CU2** |
| **Actores** | Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF2 - La red obtiene el genoma |  |
| **Precondición** | El sistema debe contar con una base de datos de genomas. |  |
| **Postcondición** | Se recupera el genoma solicitado para el análisis. |  |
| **Descripción** | La red neuronal solicita la información de un genoma específico que será analizado en los procesos posteriores. |  |
| **Resumen** | El sistema localiza y entrega el genoma requerido. |  |

Tabla 15. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Red neuronal | Solicita información de un genoma. |
| 2 | Sistema | Valida el acceso a los datos. |
| 3 | Sistema | Extrae el genoma solicitado desde la base de datos. |
| 4 | Sistema | Entrega el genoma solicitado. |
| 5 | Red neuronal | Confirma la recepción del genoma. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si no se tiene acceso a los datos del genoma, el sistema informa del fallo a la red neuronal. |
| 3 | Si el genoma no está disponible, el sistema informa la ausencia de información. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso No. 3: Buscar Lista de PGPTs**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Buscar Lista de PGPTs** | **Identificador: CU3** |
| **Actores** | Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF3 - La red obtiene la lista de PGPTs |  |
| **Precondición** | El sistema debe contar con una base de datos de PGPTs. |  |
| **Postcondición** | Se obtiene una lista de PGPTs lista para el análisis. |  |
| **Descripción** | La red neuronal solicita una lista completa de PGPTs disponibles para realizar análisis y planificación posteriores. |  |
| **Resumen** | El sistema localiza y entrega la lista completa de PGPTs solicitada por la red neuronal. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Red neuronal | Solicita la lista de PGPTs. |
| 3 | Sistema | Extrae la lista de PGPTs desde la base de datos. |
| 4 | Sistema | Entrega la lista solicitada a la red neuronal. |
| 5 | Red neuronal | Confirma la recepción de la lista. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si no se tiene acceso a la lista de PGPTs, el sistema informa la falla a la red neuronal. |
| 3 | Si los datos solicitados no están disponibles, el sistema devuelve un mensaje indicando error. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Buscar Conteo de PGPTs** | **Identificador: CU4** |
| **Actores** | Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF4 - La red obtiene el conteo de PGPTs |  |
| **Precondición** | La lista de PGPTs debe estar disponible en el sistema. |  |
| **Postcondición** | Se obtiene el conteo total de PGPTs para el análisis. |  |
| **Descripción** | La red neuronal solicita el conteo de PGPTs disponibles, lo cual es esencial para planificar procesos posteriores. |  |
| **Resumen** | El sistema cuenta y devuelve el número total de PGPTs existentes. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Red neuronal | Solicita el conteo total de PGPTs. |
| 2 | Sistema | Recibe la solicitud y verifica los datos disponibles. |
| 3 | Sistema | Realiza el conteo de PGPTs almacenados. |
| 4 | Sistema | Envía el número total a la red neuronal. |
| 5 | Red neuronal | Recibe y confirma el conteo. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si la lista de PGPTs no está disponible, el sistema informa del error a la red neuronal. |
| 4 | Si el conteo no puede ser calculado por un fallo, el sistema detiene el proceso e informa del problema. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 5: Buscar Ambiente**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Buscar Ambiente** | **Identificador: CU5** |
| **Actores** | Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF5 - La red obtiene el ambiente |  |
| **Precondición** | El sistema debe tener acceso a una base de datos que contenga información sobre ambientes. |  |
| **Postcondición** | Se obtiene información del ambiente seleccionado y se almacena temporalmente en la memoria de la red neuronal. |  |
| **Descripción** | La red neuronal solicita información del ambiente específico que será utilizado en el análisis de datos. |  |
| **Resumen** | El sistema recupera información del ambiente requerido y lo entrega a la red neuronal. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Red neuronal | Solicita información sobre el ambiente requerido. |
| 2 | Sistema | Recibe la solicitud y valida el acceso a los datos ambientales. |
| 3 | Sistema | Extrae información del ambiente desde la base de datos. |
| 4 | Sistema | Envía los datos del ambiente a la red neuronal. |
| 5 | Red neuronal | Confirma la recepción y prepara la información para el análisis. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si no se tiene acceso a los datos del ambiente, el sistema informa del fallo a la red neuronal. |
| 4 | Si los datos del ambiente solicitado no están disponibles, el sistema devuelve un mensaje indicando la ausencia de información. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 6: Clasificación PGPT**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Clasificación PGPT** | **Identificador: CU6** |
| **Actores** | Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF6 - Entrenar la red neuronal |  |
| **Precondición** | La red neuronal debe haber recibido los datos necesarios para el entrenamiento. |  |
| **Postcondición** | Se completan las iteraciones del entrenamiento y la red queda preparada para clasificar los PGPTs. |  |
| **Descripción** | La red neuronal utiliza los datos obtenidos para realizar un entrenamiento supervisado o no supervisado con el fin de clasificar los PGPTs. |  |
| **Resumen** | La red neuronal se entrena utilizando los datos genómicos y ambientales, con el objetivo de clasificar los PGPTs de manera precisa. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Red neuronal | Carga los datos obtenidos para su procesamiento. |
| 3 | Sistema | Ejecuta las iteraciones necesarias para completar el entrenamiento. |
| 4 | Red neuronal | Genera un modelo final basado en los datos procesados. |
| 5 | Red neuronal | Confirma la finalización del entrenamiento y guarda el modelo resultante. |

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 3 | Si ocurre un fallo durante las iteraciones, la red pausa el entrenamiento y solicita nuevos parámetros. |
| 4 | Si el modelo final no es válido, se reinicia el entrenamiento desde el inicio. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 7: Solicitar Vector de Características Deseadas**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Solicitar Vector de Características Deseadas** | **Identificador: CU7** |
| **Actores** | Algoritmo genético |  |
| **Tipo** | Primario |  |
| **Referencias** | RF7 - El algoritmo genético solicita el vector de características deseadas |  |
| **Precondición** | El sistema debe contar con un modelo entrenado en la red neuronal. |  |
| **Postcondición** | Se obtiene el vector de características deseadas que define las condiciones para generar genomas específicos. |  |
| **Descripción** | El algoritmo genético solicita un vector de características deseadas, el cual sirve como base para la generación de nuevas secuencias. |  |
| **Resumen** | El sistema entrega el vector de características deseadas requerido por el algoritmo genético para iniciar su proceso. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Algoritmo genético | Solicita el vector de características deseadas al sistema. |
| 2 | Sistema | Verifica si el modelo entrenado de la red neuronal está disponible. |
| 3 | Sistema | Genera el vector de características deseadas a partir del modelo. |
| 4 | Sistema | Envía el vector generado al algoritmo genético. |
| 5 | Algoritmo genético | Recibe y valida el vector para iniciar su proceso. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si no existe un modelo entrenado, el sistema informa al algoritmo genético que el vector no puede ser generado. |
| 4 | Si el vector no puede ser calculado, el sistema solicita ajustes en el modelo de entrenamiento. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 8: Generar Secuencias Aleatorias**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Generar Secuencias Aleatorias** | **Identificador: CU8** |
| **Actores** | Algoritmo genético |  |
| **Tipo** | Primario |  |
| **Referencias** | RF8 - El algoritmo genético generará secuencias aleatorias |  |
| **Precondición** | El algoritmo genético debe haber recibido el vector de características deseadas. |  |
| **Postcondición** | Se generan secuencias aleatorias que serán evaluadas en pasos posteriores. |  |
| **Descripción** | El algoritmo genético genera secuencias genéticas iniciales de manera aleatoria, utilizando el vector de características deseadas como base. |  |
| **Resumen** | El sistema genera un conjunto inicial de secuencias aleatorias para su evaluación y procesamiento. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Algoritmo genético | Solicita la generación de secuencias aleatorias. |
| 2 | Sistema | Valida que el vector de características deseadas esté disponible. |
| 3 | Sistema | Genera un conjunto de secuencias iniciales de manera aleatoria. |
| 4 | Sistema | Envía las secuencias generadas al algoritmo genético. |
| 5 | Algoritmo genético | Confirma la recepción de las secuencias para proceder con su evaluación. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si el vector no está disponible, el sistema solicita al usuario definir uno nuevo. |
| 4 | Si las secuencias no pueden generarse por un error técnico, el sistema reinicia el proceso. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 9: Evaluar Efectos de Secuencias Generadas**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Evaluar Efectos de Secuencias Generadas** | **Identificador: CU9** |
| **Actores** | Algoritmo genético, Red neuronal |  |
| **Tipo** | Primario |  |
| **Referencias** | RF9 - El algoritmo genético evaluará los efectos de las secuencias generadas a través de la red neuronal entrenada |  |
| **Precondición** | Las secuencias generadas deben estar disponibles y la red neuronal debe estar entrenada. |  |
| **Postcondición** | Se evalúan los efectos de las secuencias generadas en función del modelo entrenado. |  |
| **Descripción** | El algoritmo genético utiliza la red neuronal entrenada para evaluar las secuencias generadas y determinar su efectividad en relación con el vector de características deseadas. |  |
| **Resumen** | Las secuencias generadas son evaluadas por la red neuronal, obteniendo resultados que serán utilizados para la selección de las mejores secuencias. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Algoritmo genético | Envía las secuencias generadas a la red neuronal para su evaluación. |
| 2 | Red neuronal | Recibe las secuencias y las procesa según su modelo entrenado. |
| 3 | Red neuronal | Calcula los efectos de cada secuencia y genera un reporte de evaluación. |
| 4 | Red neuronal | Envía los resultados al algoritmo genético. |
| 5 | Algoritmo genético | Recibe y analiza los resultados para seleccionar las secuencias más adecuadas. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si la red neuronal no puede procesar las secuencias, el sistema detiene el proceso e informa del fallo. |
| 4 | Si los resultados no son válidos, el sistema solicita nuevas iteraciones. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 10: Seleccionar Genomas con Características Similares**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Seleccionar Genomas con Características Similares** | **Identificador: CU10** |
| **Actores** | Algoritmo genético |  |
| **Tipo** | Primario |  |
| **Referencias** | RF10 - El algoritmo genético seleccionará los genomas que tengan características similares al vector de características deseadas |  |
| **Precondición** | Los resultados de la evaluación de secuencias deben estar disponibles. |  |
| **Postcondición** | Se seleccionan los genomas más cercanos al vector de características deseadas. |  |
| **Descripción** | El algoritmo genético analiza los resultados de la evaluación y selecciona los genomas que mejor cumplan con los criterios establecidos. |  |
| **Resumen** | El sistema filtra las secuencias generadas para seleccionar aquellas que cumplan con las características deseadas. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Algoritmo genético | Analiza los resultados obtenidos de la evaluación. |
| 2 | Algoritmo genético | Filtra las secuencias que cumplen con el vector de características deseadas. |
| 3 | Sistema | Almacena las secuencias seleccionadas para su procesamiento posterior. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si ninguna secuencia cumple con los criterios, el sistema genera nuevas secuencias para reiniciar el proceso. |

**Caso de Uso Nro. 11: Combinación y Mutación de Genomas Seleccionados**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Combinación y Mutación de Genomas Seleccionados** | **Identificador: CU11** |
| **Actores** | Algoritmo genético |  |
| **Tipo** | Primario |  |
| **Referencias** | RF11 - El algoritmo genético combinará características de los genomas seleccionados y aplicará mutación |  |
| **Precondición** | Los genomas seleccionados deben estar disponibles. |  |
| **Postcondición** | Se generan nuevas secuencias combinadas y mutadas listas para evaluación. |  |
| **Descripción** | El algoritmo genético toma los genomas seleccionados y aplica operadores genéticos como cruce y mutación para generar nuevas secuencias. |  |
| **Resumen** | Se crean nuevas secuencias genéticas a partir de las características de los genomas seleccionados. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Algoritmo genético | Realiza cruces entre los genomas seleccionados. |
| 2 | Algoritmo genético | Aplica mutaciones aleatorias a las secuencias generadas. |
| 3 | Sistema | Almacena las nuevas secuencias para su evaluación posterior. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si el cruce o la mutación fallan, el sistema utiliza genomas alternativos. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Caso de Uso Nro. 12: Repetir Proceso Hasta Cumplir Repeticiones**

|  |  |  |
| --- | --- | --- |
| **Caso de Uso** | **Repetir Proceso Hasta Cumplir Repeticiones** | **Identificador: CU12** |
| **Actores** | Algoritmo genético, Sistema |  |
| **Tipo** | Primario |  |
| **Referencias** | RF12 - El algoritmo genético repetirá su proceso hasta cumplir con un determinado número de repeticiones |  |
| **Precondición** | El sistema debe estar configurado con un número de repeticiones establecido. |  |
| **Postcondición** | Se completan todas las iteraciones, y el sistema genera resultados finales. |  |
| **Descripción** | El sistema ejecuta el proceso completo de generación, evaluación, selección, combinación y mutación de manera iterativa hasta alcanzar el número de repeticiones definido. |  |
| **Resumen** | El algoritmo genético repite su flujo de trabajo completo hasta cumplir con el número requerido de iteraciones. |  |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

**Curso Normal**

|  |  |  |
| --- | --- | --- |
| **Nro.** | **Ejecutor** | **Paso o Actividad** |
| 1 | Sistema | Inicializa el proceso iterativo. |
| 2 | Algoritmo genético | Ejecuta las fases de generación, evaluación, selección, combinación y mutación. |
| 3 | Sistema | Verifica si se ha alcanzado el número de repeticiones. |
| 4 | Sistema | Si no se alcanzó, reinicia el proceso; si se alcanzó, finaliza y guarda los resultados. |

**Cursos Alternos**

|  |  |
| --- | --- |
| **Nro.** | **Descripción de acciones alternas** |
| 2 | Si una iteración falla, el sistema informa el error, pero continúa con las siguientes iteraciones. |

Tabla 14. Flujo del curso normal "Buscar Genoma", elaboración propia.

## 5.3 Diagramas de secuencia

Se representará la interacción entre componentes durante el proceso de clasificación de PGPTs y generación de secuencias genéticas, destacando la solicitud de datos, el módulo de red neuronal, y el módulo de algoritmo genético.

[Diagrama

Descripción generada automáticamente](https://mlai.lucid.app/plugin/images/aiplugin_3f5e8114-8c7d-471e-a991-6cc622bc5004)

figura 5.2.1.4 Diagrama de secuencia. elaboración Propia

## 5.4 Diagramas de código

### 5.4.1 Diagramas de clases

Se modelará la estructura de clases del sistema, incluyendo:

* **Clase RedNeuronal**: Contiene atributos y métodos relacionados con el entrenamiento y predicción.
* **Clase AlgoritmoGenetico**: Define los procesos de generación, evaluación y mutación.

#### 5.4.2 Modelo de datos

Se incluirá un diagrama entidad-relación que muestra las tablas y relaciones principales:

* Tabla **Genomas**: Almacena información genómica.
* Tabla **Resultados**: Guarda los resultados de clasificación y evaluación.

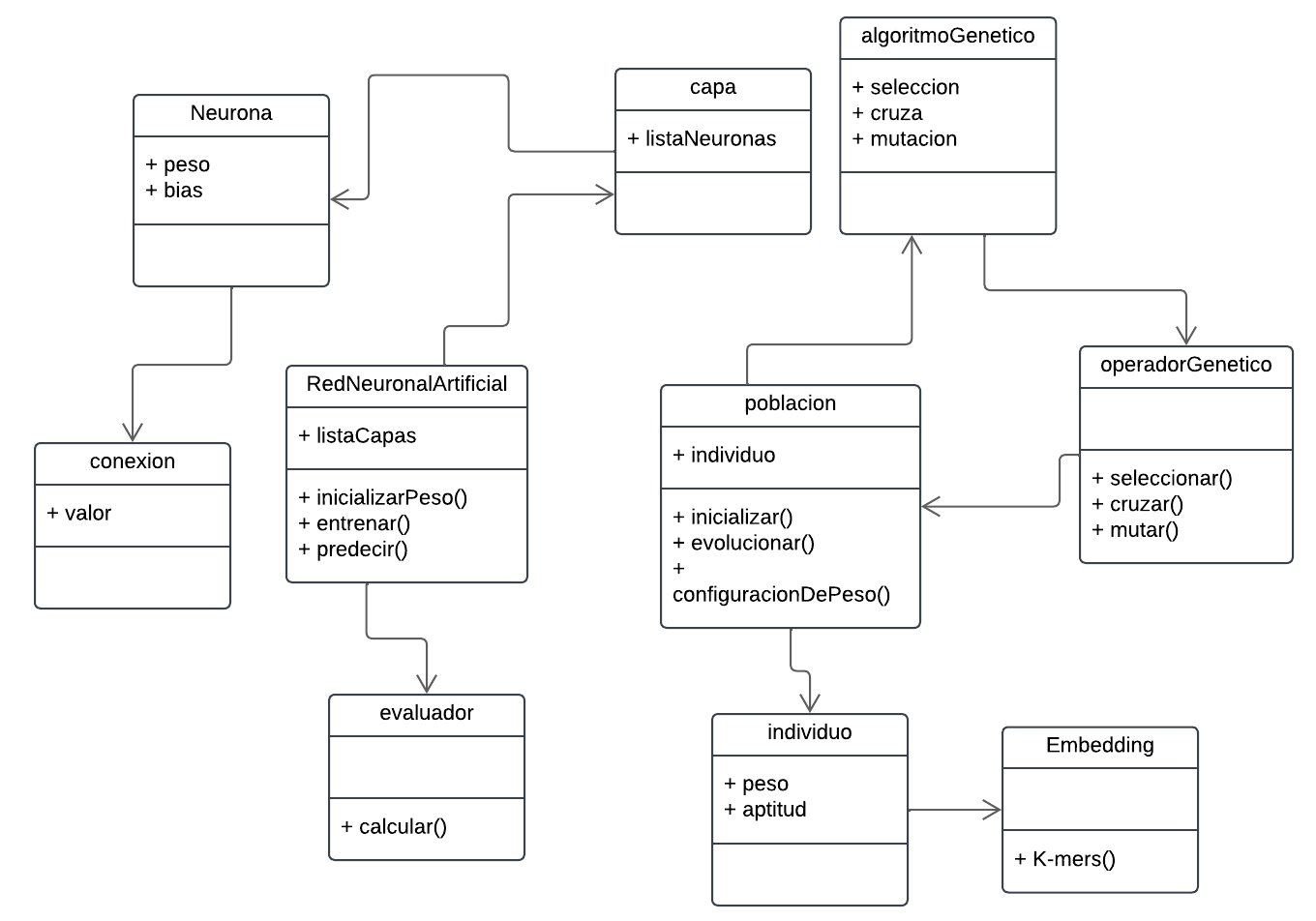


figura 5.2.1.4 Diagrama de secuencia. elaboración Propia

# Capítulo 9. Referencias

[1] M. . Camacho-Rodríguez, J. J. . Almaraz-Suárez, C. . Vázquez-Vázquez, A. Angulo-Castro, M. E. Ríos-Vega, y A. González-Mancilla, «Efecto de rizobacterias promotoras del crecimiento vegetal en el desarrollo y rendimiento del chile jalapeño», Remexca, vol. 13, n.º 28, pp. 185–196, sep. 2022.

[2] J. R. S. M. García, A. R. Gómez, and A. C. D. González, "The role of bioinformatics in agricultural research," *BMC Bioinformatics*, vol. 19, no. 1, p. 123, Oct. 2018. doi: 10.1186/s12859-018-2182-6.

[3] E. Pérez, "Ya se puede programar una bacteria como una computadora," Infobae, August 24, 2017. [Online]. Available: <https://www.infobae.com/america/vice/2017/08/24/ya-se-puede-programar-una-bacteria-como-unacomputadora/>

[4] H. Nava-Meléndez, M. Villicaña-Ortiz, y G. M. Carmona-Lara, «Sustentabilidad ambiental, del concepto a la práctica: Una oportunidad para la implementación de la evaluación ambiental estratégica en México», *Revista Mexicana de Ciencias Ambientales*, vol. 11, no. 3, pp. 333-345, 2020. Disponible: [https://www.scielo.org.mx](https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0187-57792020000300333).

[5] S. Funtowicz y J. Ravetz, «Diseño ambiental y producción de conocimiento interdisciplinario», *División de Ciencias y Artes para el Diseño, UAM-Xochimilco*, 2020. Disponible: <https://www.scielo.org.mx>.

[6] F. Blaine, Metting Jr. «Producción de HCN por Pseudomonas», *Physiology and Molecular Biology of Plants*, vol. 12, núm. 134, pp. 4–18, 2006. Disponible: <https://www.sciencedirect.com/science/article/pii/S0944501306000437>.

[7] A. Martínez-Rodríguez, R. L. López-Hernández y M. E. López-Meyer, "Estrategias para el desarrollo de fertilizantes biológicos en la agricultura", *Revista Mexicana de Ciencias Agrícolas*, vol. 1, no. 1, pp. 1-9, 2010. Disponible en: <https://www.scielo.org.mx/scielo.php?pid=S0187-57792010000100007&script=sci_arttext>.

[8] J. A. Gómez-Velasco, A. M. Rodríguez-Barrientos, E. A. Martínez-Álvarez, M. A. López-Valdez y G. O. Martínez-Álvarez, "Evaluación de la calidad de plántulas de chile poblano inoculadas con rizobacterias promotoras de crecimiento", *Revista Mexicana de Ciencias Agrícolas*, vol. 8, no. 3, pp. 463-474, 2017. Disponible en: <https://www.scielo.org.mx/scielo.php?script=sci_arttext&pid=S0188-49992017000300463>.

[9] Compostando Ciencia, "Identificación de las bacterias PGPR mediante bioinformática", *Compostando Ciencia*, 11 de abril de 2016. Disponible en: <https://www.compostandociencia.com/2016/04/identificacion-las-bacterias-pgpr-mediante-bioinformatica/>.

[10] M. A. Castañeda Reyes, "Algoritmos genéticos aplicados a búsqueda de motifs en secuencias de ADN," Tesis de Maestría en Ciencias de la Computación, Instituto Tecnológico y de Estudios Superiores de Monterrey, Atizapán de Zaragoza, Edo. de Méx., noviembre 2016.

[11] CTSA. "ADN y ARN: concepto, diferencias y funciones," *Universidad Internacional de Valencia*, [Online]. Available:  <https://www.universidadviu.com/int/actualidad/nuestros-expertos/adn-y-arn-concepto-diferencias-y-funciones>

[12] “¿Qué es el ADN?: MedlinePlus Genetics.” https://medlineplus.gov/spanish/genetica/entender/basica/adn/

[13] “Repaso de la estructura y replicación del ADN (artículo) | Khan Academy,” *Khan Academy*. https://es.khanacademy.org/science/ap-biology/gene-expression-and-regulation/replication/a/hs-dna-structure-and-replication-review

[14] “Ácido desoxirribonucleico (ADN),” *Genome.gov*. <https://www.genome.gov/es/genetics-glossary/%C3%81cido-desoxirribonucleico-ADN>

[15] F. L. && J. Matías, “ADN: estructura molecular.” <https://atlasgeneticsoncology.org/teaching/209035/adn-estructura-molecular>

[16] “Ácido ribonucleico (ARN),” *Genome.gov*. <https://www.genome.gov/es/genetics-glossary/%C3%81cido-ribonucleico-ARN>

[17] Universidad Internacional de Valencia. "Diferencias entre el ADN y el ARN," *CTS Alimentaria*, [Online]. Available: <https://ctsalimentaria.com/diferencias-entre-el-adn-y-el-arn>

[18] "Secuenciación del ADN," *National Human Genome Research Institute*, <https://www.genome.gov/es/about-genomics/fact-sheets/Secuenciacion-del-ADN>).

[19] "Secuenciación," *Wikipedia*, <https://es.wikipedia.org/wiki/Secuenciaci%C3%B3n>

[20] "Secuenciación del ADN," *National Human Genome Research Institute*, <https://www.genome.gov/es/genetics-glossary/Secuenciacion-del-ADN>

[21] "Algoritmos genéticos," *Conogasi*, <https://conogasi.org/articulos/algoritmos-geneticos/> (accedido: 11-oct-2024).

[22] "Algoritmo genético," *Wikipedia*, <https://es.wikipedia.org/wiki/Algoritmo_gen%C3%A9tico> (accedido: 11-oct-2024).

[23] B. Martín del Brío y C. Serrano Cinca, «Fundamentos de redes neuronales artificiales: hardware y software», Scire, vol. 1, n.º 1, pp. 103–125, jun. 1995.

[24] P. R. F. Ocaris and F. C. Horacio, “LAS REDES NEURONALES y LA EVALUACIÓN DEL RIESGO DE CRÉDITO.” <http://www.scielo.org.co/scielo.php?script=sci_arttext&pid=S1692-33242007000100007>

[25] N. Raynaud, J. P. Deneault, S. M. Lengeler, and G. R. Mercey, "Selective biocontrol of plant pathogens by rhizobacteria," PLoS One, vol. 11, no. 4, p. e0152478, 2016. [Online]. Available: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0152478>.

[26] D. Benjumea Muñoz, "Aislamiento y caracterización de rizobacterias promotoras del crecimiento vegetal en el cultivo de banano," Universidad de Sevilla, 2017. [Online]. Available: <https://idus.us.es/bitstream/handle/11441/65140/BENJUMEA%20MU%C3%91OZ%2C%20DANIEL.pdf>.

[27] “Interacciones planta-microorganismos,” *INTECH*. <https://intech.conicet.gov.ar/investigacion/interacciones-planta-microorganismos/>

[28] M. . Camacho-Rodríguez, J. J. . Almaraz-Suárez, C. . Vázquez-Vázquez, A. . Angulo-Castro, M. E. . Ríos-Vega, y A. . González-Mancilla, «Efecto de rizobacterias promotoras del crecimiento vegetal en el desarrollo y rendimiento del chile jalapeño», *Remexca*, vol. 13, n.º 28, pp. 185–196, sep. 2022.

[29] L. M. Laura, M. P. R. A, H. I. Montserrat, A. M. S. M, and P. A. J. R, “Caracterización de rizobacterias aisladas de tomate y su efecto en el crecimiento de tomate y pimiento.” https://www.scielo.org.mx/scielo.php?script=sci\_arttext&pid=S0187-73802013000100007

[30] A. Rodríguez-Sahagún, A. Velasco-Jiménez, O. Castellanos-Hernández, G. Acevedo-Hernández y R. Clarenc Aarland, “Bacterias rizosféricas con beneficios potenciales en la agricultura”, *terra* , vol. 38, núm. 2, págs. 333–345, abril de 2020.

[31] S. E. . Cazares-Esquivel, M. P. . Carballo-Sánchez, y J. J. . Almaraz-Suarez, «Efecto de rizobacterias promotoras del crecimiento vegetal en parámetros agronómicos y de macronutrientes en pak choi», *Remexca*, vol. 14, n.º 4, pp. 633–639, jun. 2023.

[32] "Algoritmos genéticos," *Conogasi*, <https://conogasi.org/articulos/algoritmos-geneticos/> (accedido: 11-oct-2024).

[33] “Algoritmos genéticos,” *Conogasi*, Sep. 21, 2018. <https://conogasi.org/articulos/algoritmos-geneticos/>

[34] "Algoritmo genético," *Wikipedia*, <https://es.wikipedia.org/wiki/Algoritmo_gen%C3%A9tico> (accedido: 11-oct-2024).

[35] F. Sancho, "Algoritmos Genéticos," *Universidad de Sevilla*, <https://www.cs.us.es/~fsancho/Blog/posts/Algoritmos_Geneticos.md.html> (accedido: 11-oct-2024).

[36] “¿Qué es una red neuronal?,” MATLAB & Simulink. <https://la.mathworks.com/discovery/neural-network.html>

[37] J. Brownlee, “Crash Course On Multi-Layer Perceptron Neural Networks”, Machine Learning Mastery, 2016. [Online]. Available: https://machinelearningmastery.com/neural-networks-crash-course/

[38] “Multi-Layer Neural Network”, Unsupervised Feature Learning and Deep Learning Tutorial. [Online]. Available: http://ufldl.stanford.edu/tutorial/supervised/MultiLayerNeuralNetworks/

[39] “Redes neuronales: Funciones de activación,” *Google for Developers*. https://developers.google.com/machine-learning/crash-course/neural-networks/activation-functions?hl=es-419

[40] “relu - Aplicar la activación de unidad lineal rectificada - MATLAB - MathWorks América Latina.” https://la.mathworks.com/help/deeplearning/ref/dlarray.relu\_es.html

[41] “Regresión Softmax | Interactive Chaos.” <https://interactivechaos.com/es/manual/tutorial-de-machine-learning/regresion-softmax>

[42] “Everything you need to know about batches in Machine Learning | CloudFactory Computer Vision Wiki,” *CloudFactory Computer Vision Wiki*. https://wiki.cloudfactory.com/docs/mp-wiki/training-parameters/everything-you-need-to-know-about-batches-in-machine-learning

[43] “SPSS Modeler Subscription,” Aug. 17, 2021. https://www.ibm.com/docs/es/spss-modeler/saas?topic=networks-neural-model

[44] “Oracle® Fusion Cloud EPM Trabajo con Planning.” https://docs.oracle.com/cloud/help/es/pbcs\_common/PFUSU/insights\_metrics\_RMSE.htm#PFUSU-GUID-FD9381A1-81E1-4F6D-8EC4-82A6CE2A6E74

[45] L. Team, “Definición de Cross-Entropy.” https://www.lokad.com/es/definicion-de-entropia-cruzada/

[46] “¿Qué son las redes neuronales convolucionales? | 3 cosas que debe saber,” *MATLAB & Simulink*. https://la.mathworks.com/discovery/convolutional-neural-network.html

[47] Equipo editorial de IONOS, “Convolutional Neural Network: ¿qué es una red neuronal convolucional (CNN)?,” *IONOS Digital Guide*, Sep. 03, 2024. https://www.ionos.mx/digitalguide/paginas-web/desarrollo-web/convolutional-neural-network/

[48] “¿Qué es una RNN?: Explicación sobre redes neuronales recurrentes: AWS,” *Amazon Web Services, Inc.* https://aws.amazon.com/es/what-is/recurrent-neural-network/

[49] “Procesamiento de Lenguaje Natural con DeepLearning.” <https://www.tamps.cinvestav.mx/~fpech/doc/DLScr/PLN.html>

[50] “Épocas, tamaño de lote, iteraciones: su importancia.” <https://www.sabrepc.com/blog/Deep-Learning-and-AI/Epochs-Batch-Size-Iterations?srsltid=AfmBOoq6jXjPk_tD2LHuQSWkOeyVYwhGmrn68pXYu-2CPjcEgnZ0y9qy>

[51] S. Navarro, “¿Qué es el batch size en Deep Learning? [2024],” *KeepCoding Bootcamps*, May 22, 2024. <https://keepcoding.io/blog/batch-size-deep-learning/>

[52] E, “Los embeddings: una tecnología presente en nuestro día a día,” *BBVA AI Factory*, Oct. 24, 2024. https://www.bbvaaifactory.com/es/behind-daily-life-embeddings-in-action/

[53] P. Ng, “dna2vec: Consistent vector representations of variable-length k-mers,” *arXiv.org*, Jan. 23, 2017. <https://arxiv.org/abs/1701.06279>

[54] Glassdoor - Bioinformatics Salary Glassdoor, "Bioinformatics sueldo," [Enlace]. Disponible en: https://www.glassdoor.com.mx/Sueldos/bioinformatics-sueldo-SRCH\_KO0,14.htm. [Accedido: 7-ene-2025].

[55] Glassdoor - Desarrollador de Inteligencia Artificial Salary Glassdoor, "Desarrollador de inteligencia artificial sueldo," [Enlace]. Disponible en: https://www.glassdoor.com.mx/Sueldos/desarrollador-de-inteligencia-artificial-sueldo-SRCH\_KO0,40.htm. [Accedido: 7-ene-2025].

[56] Diputados - Ley General de Bienes Nacionales Cámara de Diputados, "Ley General de Bienes Nacionales," [Enlace]. Disponible en: <https://www.diputados.gob.mx/LeyesBiblio/pdf/LBOGM.pdf>. [Accedido: 7-ene-2025].