



**Pontificia Universidad
Católica del Ecuador**
Seréis mis testigos

**PONTIFICIA UNIVERSIDAD CATÓLICA DEL ECUADOR
FACULTAD DE HÁBITAT, INFRAESTRUCTURA Y CREATIVIDAD**

TEMA:

**DESARROLLO DE UN DASHBOARD PARA EL ANÁLISIS DE PUBLICACIONES BASADAS
EN DATOS BIOINFORMÁTICOS DE LA BIODIVERSIDAD ECUATORIANA MEDIANTE
MINERÍA DE DATOS: IDENTIFICACIÓN DE BRECHAS Y OPORTUNIDADES PARA LA
INVESTIGACIÓN**

AUTOR:

VÁSQUEZ TERÁN JOSÉ LUIS

DIRECTOR:

ORTIZ NAVARRETE MIGUEL DIMITRI

**TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN SISTEMAS DE
INFORMACIÓN MENCIÓN DATA SCIENCE**

QUITO, 2025

Agradecimiento

Doy gracias a Dios y la Virgen María, por guiarme, darme fuerza y por permitirme alcanzar esta meta.

Este logro no habría sido posible sin el amor y el apoyo incondicional de mis padres, José Andrés y María Ursulina, mis pilares, mis hermanos y hermanas Edwin, Patricia, Alexandra, William y Germania Vásquez Terán quienes siempre me impulsaron a seguir adelante. A mi amada esposa, Evelin, y a mis queridos hijos, Issac Santiago y Andrés Sebastián Vásquez Castro, gracias por su paciencia y comprensión durante este largo camino. A mis adorados sobrinos y sobrinas, y en especial a Pamela y Samantha, cuyo apoyo me brindó una confianza invaluable. Gracias por creer en mí.

A mis valiosos compañeros maestrantes, gracias por compartir este camino conmigo, por las conversaciones enriquecedoras, el apoyo mutuo y las risas compartidas. Su amistad hizo de esta experiencia algo inolvidable.

Cada camino se construye con el apoyo de quienes te han considerado importante en su vida, entonces cada meta cumplida es un aporte de todos y eso permite que tu sueño sea el sueño de todos y así entonces brindar un poco de esperanza en sus corazones.

Agradezco profundamente a mi director de tesis Mgs. Miguel Ortiz por su paciencia, dedicación y valiosos consejos a lo largo de este proceso cuyo apoyo incondicional me impulsó a alcanzar esta meta.

A todos aquellos que de una forma u otra contribuyeron a la realización de esta tesis, mi más sincero agradecimiento. Cada gesto, cada palabra de aliento, fue un impulso para llegar a la meta.

¡Gracias a todos!"

Dedicatoria

A mis padres José Andrés y María Ursulina, por su amor incondicional y apoyo constante, que fueron la base de mi camino. A mi esposa Evelin Gabriela e hijos Issac Santiago y Andrés Sebastián, mi fuente de inspiración y motivación, por su paciencia y comprensión durante este desafío. Y, sobre todo, a la Virgen María, mi guía y protectora, cuya luz me acompañó en cada paso de este logro. Que esta maestría sea un reflejo de su amor y bendición.

Índice de Contenido

1.	INTRODUCCIÓN.....	1
1.1.	Generalidades	1
1.2.	Planteamiento del Problema	4
1.3.	Objetivos	6
1.3.1.	Objetivo General	6
1.3.2.	Objetivos Específicos.....	6
1.4.	Alcance de la Investigación	7
2.	REVISIÓN LITERARIA.....	8
2.1.	Biodiversidad.....	8
2.1.1.	Niveles de biodiversidad	9
2.1.2.	Biodiversidad Ecuatoriana.....	11
2.2.	Bioinformática	13
2.2.1.	Herramientas y técnicas bioinformáticas	14
2.2.2.	Aplicaciones en la investigación de la biodiversidad	16
2.3.	Minería de datos	19
2.3.1.	Proceso de minería de datos.....	20
2.3.2.	Técnicas de minería de datos y aplicaciones en el análisis de la literatura científica	23
2.4.	Visualización de datos	26
2.4.1.	Principios de una visualización de datos eficaz.....	26
2.4.2.	Tipos de visualizaciones de datos	28
2.5.	Dashboard	30
2.5.1.	Características del Dashboard	31
2.5.2.	Diseño de un Dashboard	33
2.5.3.	Beneficios del Dashboard para la Investigación de la Biodiversidad	35
2.6.	Investigación científica	36
2.6.1.	Tipos de investigación científica.....	37
2.6.2.	Importancia de la investigación científica en la conservación de la biodiversidad	39
3.	MARCO METODOLÓGICO.....	40
3.1.	Materiales	42
3.2.	Metodología.....	43
3.3.	Herramientas de Software	44
4.	RESULTADOS.....	45

4.1.	Recopilación de Datos	45
4.1.1.	Fuentes de datos	45
4.1.2.	Técnicas de recopilación	48
4.1.3.	Criterios de selección	60
4.2.	Análisis de Datos	63
4.2.1.	Estado de Conservación	63
4.2.2.	Análisis de Publicaciones Bioinformáticas - Biodiversidad Ecuatoriana.....	65
4.2.3.	Análisis de contenido	69
4.2.4.	Resultados del análisis	70
4.3.	Desarrollo del Dashboard	71
4.3.1.	Herramientas de visualización	71
4.3.2.	Diseño del dashboard	72
4.3.3.	Funcionalidades del dashboard	74
4.4.	Validación y Evaluación	75
4.4.1.	Evaluación de la usabilidad	75
4.4.2.	Evaluación de la efectividad	76
5.	DISCUSIÓN	77
5.1.	Implicaciones para la Investigación.....	77
5.2.	Limitaciones del Estudio	78
5.3.	Trabajos Futuros.....	78
6.	CONCLUSIONES	79
6.1.	Resumen de los Hallazgos	80
6.2.	Contribuciones del Estudio	81
6.3.	Recomendaciones	81
7.	REFERENCIAS BIBLIOGRÁFICAS	1
8.	ANEXOS	1
8.1.	Anexo 1: Glosario de términos bioinformáticos.....	1
8.2.	Anexo 2: Detalles técnicos del desarrollo del dashboard	1
8.3.	Anexo 3: Código de script utilizado para hacer Webscraping en buscadores de artículos científicos.	1

Índice de Tablas

Tabla 1	10
Tabla 2	22
Tabla 3	23
Tabla 4	33
Tabla 5	38
Tabla 6	42
Tabla 7	43
Tabla 8	44
Tabla 9	64
Tabla 10	64
Tabla 11	1

Índice de Ilustraciones

Ilustración 1	50
Ilustración 2	55
Ilustración 3	59
Ilustración 4	65
Ilustración 5	65
Ilustración 6	66
Ilustración 7	68
Ilustración 8	72
Ilustración 9	73
Ilustración 10	74

1. INTRODUCCIÓN

1.1. Generalidades

Ecuador, reconocido como uno de los países megadiversos a nivel mundial, alberga una concentración excepcional de vida en relación a su tamaño. Esta megadiversidad, producto de su ubicación geográfica y la presencia de la Cordillera de los Andes, representa una de las colecciones más densas de conocimiento y formas de vida en el planeta. A pesar de su riqueza biológica, la biodiversidad ecuatoriana enfrenta amenazas significativas como la deforestación y la expansión de la frontera agrícola (Banco de Desarrollo de América Latina y el Caribe, 2024). (Herrera-Feijoo, 2024) reveló que, entre 2008 y 2016, la Amazonía ecuatoriana perdió 650,000 hectáreas de bosque virgen, una cifra que subraya la magnitud del problema en una de las regiones más biodiversas del mundo.

La exploración y el análisis de la biodiversidad ecuatoriana han cobrado mayor relevancia en las últimas décadas, impulsados por los avances en las ciencias biológicas y de la información. El manejo de bases de datos a gran escala, conocido como big data, ha añadido un nuevo valor al estudio y aplicación del conocimiento generado a partir de los recursos biológicos y genéticos (Toulkeridis, y otros, 2020). En este contexto, los datos bioinformáticos emergen como un componente esencial para comprender, conservar y utilizar sosteniblemente la biodiversidad.

Los datos bioinformáticos, entendidos como la información derivada del análisis computacional de datos biológicos, ofrecen una perspectiva invaluable sobre la composición genética, las interacciones ecológicas y los patrones evolutivos de las especies (Herrera-Feijoo, 2024). Estos datos incluyen secuencias de ADN, datos de expresión génica, información sobre proteínas, y metadatos asociados a especímenes biológicos. La integración y análisis de estos datos permiten identificar especies amenazadas, descubrir compuestos bioactivos con

potencial farmacéutico, y comprender los efectos del cambio climático en la distribución de especies (Toulkeridis, y otros, 2020).

La Constitución de la República del Ecuador estableció que la Biodiversidad y su patrimonio genético sean considerados un sector estratégico para el Estado ecuatoriano. El Instituto Nacional de Biodiversidad (INABIO), creado vía Decreto Ejecutivo, permite al país enfrentar nuevos retos en la conservación y gestión de la biodiversidad, así como en la transferencia y vinculación del conocimiento del patrimonio natural en los diferentes sectores de la sociedad. El Plan Estratégico de Investigación del INABIO 2021-2025 tiene como objetivo profundizar el conocimiento sobre la diversidad biológica del Ecuador e investigar los usos potenciales de la biodiversidad (Instituto Nacional de Biodiversidad , 2021).

La biodiversidad en Ecuador aún está poco explorada y subutilizada, lo que presenta una oportunidad para el desarrollo económico sostenible. La generación de actividad económica derivada del robustecimiento científico y tecnológico, enfocada hacia la exploración y aplicación sostenible de la diversidad biológica y cultural del país, promete materializar un modelo autóctono y soberano de bioeconomía (Instituto Nacional de Biodiversidad , 2021).

La visualización de información y la minería de datos se han convertido en herramientas clave para analizar patrones en proyectos de investigación relacionados con la biodiversidad. Estas técnicas permiten analizar grandes volúmenes de información para descubrir tendencias y relaciones que pueden impulsar nuevas investigaciones. En particular, la minería de datos permite identificar patrones ocultos en los datos bioinformáticos, como genes asociados a la resistencia a enfermedades, o factores ambientales que influyen en la distribución de especies (Banco de Desarrollo de América Latina y el Caribe, 2024).

Los dashboards interactivos se presentan como una solución efectiva para la visualización y exploración de datos bioinformáticos. Un dashboard es una interfaz gráfica que permite a los usuarios acceder y manipular datos de manera intuitiva, facilitando la

identificación de patrones y tendencias (Murillo, 2024). En el contexto de la biodiversidad ecuatoriana, puede permitir a los investigadores explorar la distribución de especies en diferentes regiones, analizar la expresión génica de plantas medicinales, o identificar los factores que influyen en la deforestación.

La creación de una Red de Información y Datos de Biodiversidad (RIDB) contribuirá a la mejora continua de la generación, almacenamiento y gestión de información relacionada con la biodiversidad en Ecuador. Dicha Red, asociada directamente al Sistema de Información de Biodiversidad (SIB-Ec), pretende que las buenas prácticas en la gestión de información de biodiversidad redunden en beneficios para todos los ciudadanos del país, fortaleciendo las políticas, acuerdos, normas técnicas, servicios y tecnologías, bajo los principios de transparencia y acceso a los datos, buena fe y responsabilidad en el uso de la información.

El Instituto Nacional de Biodiversidad (INABIO) juega un papel crucial en la investigación y gestión de la biodiversidad en Ecuador. Su Plan Estratégico de Investigación 2021-2025 tiene como objetivo profundizar el conocimiento sobre la diversidad biológica del Ecuador, investigar los usos potenciales de la biodiversidad, desarrollar tecnologías e innovaciones que permitan la conservación in situ y ex situ de la Biodiversidad, y potenciar los conocimientos, innovaciones y saberes tradicionales para la conservación, restauración ecológica y uso sustentable de la biodiversidad (Instituto Nacional de Biodiversidad , 2021).

La creación de un dashboard interactivo para el análisis de publicaciones científicas basadas en datos bioinformáticos de la biodiversidad ecuatoriana permitirá identificar brechas y oportunidades para la investigación. Este dashboard facilitará la exploración dinámica de los datos y los resultados del análisis, lo que permitirá a los investigadores identificar áreas de estudio, especies, metodologías y resultados relevantes (Murillo, 2024). La implementación de algoritmos de minería de datos permitirá identificar patrones ocultos en los datos y generar nuevo conocimiento sobre la biodiversidad ecuatoriana.

1.2. Planteamiento del Problema

Ecuador, un país reconocido a nivel mundial por su excepcional megadiversidad, enfrenta desafíos críticos en la conservación y gestión de su invaluable patrimonio natural (Báez, 2019). Su privilegiada ubicación geográfica en el neotrópico, junto con la imponente presencia de la Cordillera de los Andes, ha dado lugar a una asombrosa variedad de ecosistemas y especies, convirtiéndolo en un punto clave para la biodiversidad global (Mena, 2014). Sin embargo, esta riqueza se ve amenazada por factores como la deforestación, la expansión agrícola y la explotación no sostenible de recursos naturales. Para abordar estos desafíos de manera efectiva y garantizar la sostenibilidad a largo plazo, es fundamental contar con información precisa, accesible e integrada sobre la biodiversidad ecuatoriana.

La investigación científica en Ecuador ha generado una vasta cantidad de datos bioinformáticos sobre su biodiversidad, incluyendo información genética, datos de distribución de especies, estudios ecológicos y análisis de ecosistemas (Instituto Nacional de Biodiversidad, 2021). Sin embargo, estos datos se encuentran dispersos en diversas fuentes, como publicaciones científicas, bases de datos especializadas y repositorios institucionales, lo que dificulta significativamente su integración y análisis (Mena, 2014). Esta dispersión de la información crea una serie de problemáticas interrelacionadas que limitan la capacidad de comprender, proteger y utilizar de manera sostenible la biodiversidad del país.

Una de las principales problemáticas identificadas es la dificultad para identificar áreas de investigación subexplotadas (Mena, 2014). La falta de una visión integral de los datos y la ausencia de herramientas eficientes dificultan la identificación de vacíos de conocimiento y áreas donde se requiere mayor investigación (Báez, 2019). Esta limitación impide priorizar los esfuerzos de investigación y asignar recursos de manera efectiva, lo que puede llevar a la duplicación de estudios y a la falta de atención a áreas críticas para la conservación de la biodiversidad.

Otra problemática importante son las limitaciones en el reconocimiento de tendencias en las investigaciones bioinformáticas existentes. La falta de herramientas de análisis impide identificar patrones y tendencias en los datos acumulados, lo que dificulta la formulación de nuevas hipótesis y la planificación de futuras investigaciones (Ministerio del Ambiente del Ecuador, 2016). Sin la capacidad de analizar la información existente de manera sistemática, es difícil identificar las líneas de investigación más prometedoras y las áreas donde se pueden lograr mayores avances en el conocimiento de la biodiversidad ecuatoriana.

Además, la falta de información integrada y visualizada obstaculiza la toma de decisiones estratégicas en relación con la conservación, gestión y uso sostenible de la biodiversidad. Los tomadores de decisiones, tanto a nivel gubernamental como en el sector privado, necesitan acceder a información relevante y actualizada para poder diseñar e implementar políticas y programas efectivos (Ministerio del Ambiente del Ecuador, 2016). Sin embargo, la dispersión de los datos y la falta de herramientas de visualización dificultan la identificación de los problemas más urgentes y la evaluación de las posibles soluciones, lo que puede llevar a decisiones subóptimas o incluso contraproducentes para la conservación de la biodiversidad.

La ineficiencia en el aprovechamiento de recursos y esfuerzos de investigación es otra consecuencia directa de la falta de integración y análisis de los datos bioinformáticos. La duplicación de esfuerzos y la falta de coordinación entre investigadores e instituciones limitan el impacto de las investigaciones, lo que representa una pérdida de recursos valiosos que podrían utilizarse de manera más efectiva (Báez, 2019). Es necesario fomentar la colaboración y la coordinación entre los diferentes actores involucrados en la investigación de la biodiversidad.

La dificultad para identificar colaboraciones es un obstáculo para el avance del conocimiento sobre la biodiversidad. La falta de una plataforma que permita visualizar los datos y las investigaciones dificulta la identificación de posibles colaboradores, tanto a nivel nacional

como internacional (Banco de Desarrollo de América Latina y el Caribe, 2024). La colaboración entre diferentes disciplinas e instituciones es fundamental para abordar los desafíos complejos que plantea la conservación y el uso sostenible de la biodiversidad, y la falta de herramientas para facilitar esta colaboración limita el potencial de generar soluciones innovadoras y efectivas.

En este contexto, se evidencia la urgente necesidad de desarrollar un dashboard interactivo que permita integrar, analizar y visualizar datos bioinformáticos sobre la biodiversidad ecuatoriana. Este dashboard facilitaría la identificación de patrones, tendencias y relaciones en los datos, lo que permitiría a investigadores, tomadores de decisiones y otros actores abordar las problemáticas mencionadas y promover la conservación y el uso sostenible de la biodiversidad ecuatoriana de manera más efectiva.

1.3. Objetivos

1.3.1. *Objetivo General*

Desarrollar un dashboard interactivo para el análisis de publicaciones científicas basadas en datos bioinformáticos de la biodiversidad ecuatoriana, mediante técnicas de minería de datos, que permita la identificación de oportunidades de investigación.

1.3.2. *Objetivos Específicos*

Recopilar y estructurar una base de datos de publicaciones científicas relacionadas con la biodiversidad ecuatoriana que contengan información bioinformática, utilizando técnicas de web scraping y procesamiento de datos.

Implementar algoritmos de minería de datos para el análisis de patrones, tendencias y relaciones en las publicaciones científicas recopiladas, considerando variables como áreas de estudio, especies, metodologías y resultados.

Diseñar e implementar una interfaz de visualización interactiva (dashboard) que permita la exploración dinámica de los datos extraídos y los resultados del análisis.

Evaluar la efectividad del dashboard mediante la identificación de gaps de investigación y oportunidades de estudio en el campo de la biodiversidad ecuatoriana a través de casos de uso.

Validar la utilidad y usabilidad del dashboard mediante pruebas con investigadores y expertos en el área de bioinformática y biodiversidad.

1.4. Alcance de la Investigación

La presente investigación se centra en el desarrollo de un dashboard interactivo diseñado para facilitar el análisis y la visualización de información contenida en publicaciones científicas relacionadas con la biodiversidad ecuatoriana. El proyecto abarcará la recopilación y estructuración de datos bioinformáticos dispersos en diversas fuentes, incluyendo artículos científicos, bases de datos especializadas y repositorios institucionales. Mediante técnicas de web scraping y procesamiento de datos, se creará una base de datos centralizada que servirá como fundamento para el análisis y la visualización de la información. El alcance se limita a la información disponible públicamente y accesible a través de medios digitales, excluyendo datos propietarios o de acceso restringido.

El análisis de los datos recopilados se realizará mediante técnicas de minería de datos, con el objetivo de identificar patrones, tendencias y relaciones significativas en las publicaciones científicas. Se considerarán variables como áreas de estudio, especies investigadas, metodologías empleadas y resultados obtenidos. Este análisis permitirá identificar áreas de investigación subexploradas, reconocer tendencias en las investigaciones existentes y detectar posibles sesgos en la literatura científica. El alcance del análisis se centrará en la

información cuantitativa y cualitativa extraída de las publicaciones, sin realizar análisis experimentales o de laboratorio.

El desarrollo del dashboard interactivo se basará en principios de diseño centrado en el usuario, con el objetivo de crear una herramienta intuitiva y fácil de usar para investigadores, tomadores de decisiones y otros actores interesados en la biodiversidad ecuatoriana. El dashboard permitirá la exploración dinámica de los datos, la visualización de patrones y tendencias, y la generación de informes personalizados. El alcance del dashboard se limitará a la visualización de la información recopilada y analizada, sin incluir funcionalidades de edición o modificación de los datos.

La evaluación de la efectividad del dashboard se realizará mediante casos de uso, en los cuales se aplicará la herramienta para identificar gaps de investigación y oportunidades de estudio en el campo de la biodiversidad ecuatoriana. Se seleccionarán casos de uso representativos de las diferentes áreas de investigación y se analizarán los resultados obtenidos con el dashboard. El alcance de la evaluación se centrará en la utilidad y la usabilidad del dashboard para la identificación de oportunidades de investigación.

2. REVISIÓN LITERARIA

2.1. Biodiversidad

La biodiversidad, o diversidad biológica, se refiere a la vasta gama de vida en la Tierra, incluyendo la variedad de seres vivos, los patrones naturales que los conforman y los ecosistemas en los que habitan. Este concepto abarca la diversidad de especies de plantas, animales, hongos y microorganismos, su variabilidad genética y los ecosistemas de los que forman parte, así como los paisajes donde se ubican estos ecosistemas. La biodiversidad es el resultado de miles de millones de años de evolución, influenciada tanto por procesos naturales como por las actividades humanas (Campaña Sin Biodiversidad No Hay Vida, 2022).

La biodiversidad es esencial para el funcionamiento de los ecosistemas y el bienestar de todo el planeta. Proporciona servicios básicos como oxígeno, alimentos, vestimenta y salud. Los diversos elementos que componen la biodiversidad forman unidades funcionales que aseguran servicios básicos para la supervivencia (Ministerio de Ambiente y Desarrollo Sostenible, 2017). Además, la biodiversidad contribuye al desarrollo socioeconómico de la sociedad, permitiendo avanzar hacia un modelo de uso sostenible de los recursos.

La importancia de la biodiversidad radica en su valor intrínseco y en los beneficios que aporta a la biosfera. La diversidad biológica tiene el derecho inalienable de continuar su existencia como resultado de un proceso histórico natural de gran antigüedad. Los seres humanos, como parte de esta diversidad, deben protegerla y respetarla (Ministerio de Ambiente y Desarrollo Sostenible, 2017). La biodiversidad también representa un capital natural, cuyo uso y beneficio ha contribuido al desarrollo de la cultura humana y representa una fuente potencial para satisfacer necesidades futuras. Desde un punto de vista utilitario, la biodiversidad se puede agrupar en tres categorías principales: la provisión de recursos, la regulación de los ecosistemas y el valor cultural y estético (CEPAL, 2023).

La conservación de la biodiversidad es un interés común de toda la humanidad y es fundamental para satisfacer las necesidades básicas. Sin embargo, la biodiversidad es vulnerable y está amenazada por el consumo excesivo e insostenible, el crecimiento demográfico y la demanda de recursos, lo que ha provocado la destrucción de hábitats y ha puesto en peligro a miles de especies naturales (CEPAL, 2023).

2.1.1. Niveles de biodiversidad

La biodiversidad se puede analizar en diferentes niveles, desde los genes hasta los paisajes, cada uno con atributos de composición, estructura y función. La composición se refiere a la identidad y variedad de los elementos, incluyendo las especies presentes y su número. La estructura es la organización física del sistema, como la abundancia relativa de las

especies y la conectividad entre ecosistemas (Campaña Sin Biodiversidad No Hay Vida, 2022).

La función se refiere a los procesos ecológicos y evolutivos, como la depredación, la competencia, el ciclo de nutrientes y las perturbaciones naturales.

Estos niveles están interrelacionados y son esenciales para el funcionamiento de los ecosistemas y la provisión de servicios ecosistémicos. La pérdida de biodiversidad en cualquier nivel puede tener efectos negativos en los demás niveles y en el funcionamiento general del ecosistema (Herrera-Feijoo, 2024).

Tabla 1

Niveles de Biodiversidad

Nivel	Descripción
Diversidad Genética	Se refiere a la variación en la información genética contenida en todos los organismos individuales de una especie, y entre diferentes especies. Esta diversidad genética permite que las poblaciones se adapten a los cambios en su entorno y es la base de la evolución. Incluye la variabilidad dentro de una misma población, entre poblaciones separadas geográficamente, y entre individuos. La diversidad genética es crucial para la resistencia a enfermedades, adaptación al cambio climático y la capacidad de las especies para evolucionar y sobrevivir a largo plazo.
Diversidad de Especies	Abarca la variedad de especies de plantas, animales, hongos, microorganismos y otros seres vivos que habitan en un área determinada. Esta diversidad incluye tanto el número de especies presentes (riqueza de especies) como la abundancia relativa de cada especie (equidad). La diversidad de especies es fundamental para el funcionamiento de los ecosistemas, ya que cada especie desempeña un papel específico en la red trófica y en los ciclos biogeoquímicos. La pérdida de especies puede tener consecuencias graves para la estabilidad y la resiliencia de los ecosistemas.
Diversidad de Ecosistemas	Se refiere a la variedad de hábitats, comunidades biológicas y procesos ecológicos que existen en una región o en el planeta. Incluye la diversidad de paisajes, desde bosques y praderas hasta desiertos y océanos. La diversidad de ecosistemas es importante porque cada ecosistema ofrece diferentes servicios ecosistémicos, como la regulación del clima, la purificación del agua, la polinización y el control de plagas. Además, la diversidad de ecosistemas contribuye a la diversidad de especies y a la diversidad genética, ya que diferentes ecosistemas albergan diferentes especies y poblaciones adaptadas a condiciones específicas.

Fuente: (Campaña Sin Biodiversidad No Hay Vida, 2022) (Herrera-Feijoo, 2024)

2.1.2. Biodiversidad Ecuatoriana

Ecuador es considerado un país megadiverso debido a su ubicación geográfica, que le permite albergar una gran variedad de ecosistemas y especies. El país se encuentra en la zona tropical, atravesado por la Cordillera de los Andes, lo que genera una diversidad de climas y altitudes (Liria, 2022). Esta combinación de factores ha dado como resultado una gran cantidad de especies nativas y endémicas, muchas de las cuales no se encuentran en ningún otro lugar del mundo.

La biodiversidad del Ecuador se manifiesta en sus diferentes biomas, desde la selva amazónica hasta los páramos andinos, pasando por los bosques nublados y las costas. Cada uno de estos ecosistemas alberga una gran variedad de especies adaptadas a las condiciones locales. La selva amazónica, por ejemplo, es uno de los ecosistemas más biodiversos del planeta, con una gran cantidad de especies de plantas, animales e insectos (Liria, 2022).

Los páramos andinos, por otro lado, son ecosistemas de alta montaña con una flora y fauna únicas, adaptadas a las condiciones extremas de altitud y clima. Los bosques nublados, caracterizados por su alta humedad y neblina constante, albergan una gran cantidad de especies endémicas, como orquídeas, bromelias y aves. Las costas ecuatorianas, tanto en el Pacífico como en las Islas Galápagos, también son importantes centros de biodiversidad, con una gran variedad de especies marinas y costeras (Varea, 2004).

La biodiversidad del Ecuador es un patrimonio natural de gran valor social, económico y ambiental. Muchas comunidades locales dependen directamente de los recursos naturales para su subsistencia, como la pesca, la agricultura y la extracción de madera. Además, la biodiversidad tiene un gran potencial para el desarrollo de actividades turísticas y de investigación científica. El Instituto Nacional de Biodiversidad (INABIO) juega un papel crucial en la investigación y gestión de este patrimonio (Instituto Nacional de Biodiversidad , 2021).

Sin embargo, la biodiversidad del Ecuador se enfrenta a numerosos desafíos, como la deforestación, la contaminación, la expansión agrícola y la minería. Estas actividades humanas están causando la pérdida de hábitats y la disminución de poblaciones de muchas especies. Además, el cambio climático también representa una amenaza para la biodiversidad, ya que puede alterar las condiciones ambientales y afectar la distribución de las especies (Varea, 2004).

Para proteger la biodiversidad del Ecuador, es necesario implementar estrategias y políticas de conservación a nivel nacional e internacional. Esto incluye la creación de áreas protegidas, la promoción de prácticas agrícolas sostenibles, el control de la contaminación y la lucha contra el cambio climático (Herrera-Feijoo, 2024). También es importante involucrar a las comunidades locales en la gestión de los recursos naturales y promover la educación ambiental para crear conciencia sobre la importancia de la biodiversidad.

La conservación de la biodiversidad del Ecuador es fundamental para garantizar el bienestar de las generaciones presentes y futuras. La pérdida de biodiversidad no solo tiene consecuencias ambientales, sino también sociales y económicas. Por lo tanto, es necesario tomar medidas urgentes para proteger este valioso patrimonio natural y asegurar su uso sostenible (Mena, 2014).

El Plan Estratégico de Investigación del Instituto Nacional de Biodiversidad (INABIO) 2021-2025 tiene como objetivo diseñar, implementar y promover estrategias institucionales en el campo del desarrollo científico y transferencia de conocimiento, teniendo a la biodiversidad como un sector estratégico para una transición adecuada y exitosa a un nuevo modelo de desarrollo, más sostenible con el ambiente y que genere bienestar al ser humano (Instituto Nacional de Biodiversidad , 2021).

La megadiversidad del Ecuador, reflejada en sus genes, especies y ecosistemas, subraya la necesidad crítica de estrategias de conservación integrales. Reconociendo el valor

intrínseco de su biodiversidad y su papel esencial en el mantenimiento de la vida en el planeta, Ecuador enfrenta el reto de equilibrar el desarrollo socioeconómico con la preservación de su patrimonio natural único.

2.2. Bioinformática

La bioinformática es un campo interdisciplinario que fusiona la biología, la informática y la estadística para analizar e interpretar datos biológicos complejos. Se podría decir que la bioinformática se encarga de la parte lingüística de la genética, estudiando los patrones de las secuencias de ADN y de las proteínas para proteger nuestra salud (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019). Los términos bioinformática, biología computacional, informática biológica y, en ocasiones, biocomputación, son utilizados en muchas situaciones como sinónimos.

La bioinformática moderna busca contribuir en el campo de desarrollo de inteligencias artificiales y presenta una propuesta científica valiosa para la creación de algoritmos de aprendizaje automático. Al utilizar herramientas tecnológicas avanzadas, la biotecnología se basa en su totalidad en el avance de esta ciencia, que permite diseñar fármacos más eficaces (Liras, Cabello, & Bonet, 2008).

En esencia, la bioinformática se encarga de procesar datos del campo de la biología aplicando la informática³. El objetivo principal de la bioinformática es el análisis e interpretación de los datos de las moléculas biológicas, un proceso que se denomina biología computacional. En la actualidad, la parte más importante de la bioinformática es el análisis e interpretación de los datos de las moléculas biológicas. Los objetivos fundamentales son identificar genes y proteínas, determinar sus funciones, establecer relaciones evolutivas y predecir su conformación (Casillas, 2008).

El alcance de la bioinformática abarca la aplicación de tecnologías computacionales para recopilar, almacenar, analizar e integrar datos biológicos³. Estos datos se utilizan para obtener información y se aplican al descubrimiento y desarrollo de fármacos basados en genes, estudios de estructura de proteínas y la determinación de la eficacia terapéutica de los fármacos (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019).

La bioinformática utiliza técnicas informáticas que se aplican en otros campos, como la inteligencia artificial, y que incluyen el reconocimiento de patrones, los algoritmos de aprendizaje automático y la visualización de datos. La posibilidad de efectuar estudios de la microbiota, el microbioma y el metagenoma de una muestra clínica de manera rápida y a un coste reducido permite avanzar más rápidamente en el diagnóstico de enfermedades, en el conocimiento de la taxonomía y la epidemiología de los agentes involucrados, así como de su virulencia (Liria, 2022). También posibilita la realización de estudios de genómica comparada y el descubrimiento de genes o variantes de interés, lo que puede llevar a que enfermedades tradicionalmente consideradas como de carácter no microbiano sean asociadas a la presencia de microorganismos.

2.2.1. Herramientas y técnicas bioinformáticas

La bioinformática se basa en una amplia gama de herramientas y técnicas para abordar los desafíos del análisis de datos biológicos. Estas herramientas y técnicas abarcan desde algoritmos de alineamiento de secuencias hasta métodos de modelado molecular y análisis de redes biológicas.

Alineamiento de secuencias: Esta técnica es fundamental para comparar secuencias de ADN, ARN o proteínas e identificar similitudes y diferencias. Los algoritmos de alineamiento de secuencias, como BLAST (Basic Local Alignment Search Tool) (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019), permiten a los investigadores identificar genes y proteínas homólogas en diferentes organismos y comprender las relaciones

evolutivas entre ellos. BLAST, por ejemplo, es una herramienta ampliamente utilizada para encontrar regiones de similitud local entre secuencias biológicas, permitiendo a los investigadores inferir relaciones funcionales y evolutivas.

Predicción de genes: Esta técnica se utiliza para identificar las regiones del genoma que codifican proteínas. Los programas de predicción de genes utilizan modelos estadísticos y de aprendizaje automático para analizar las secuencias de ADN y predecir la ubicación de los genes (Liras, Cabello, & Bonet, 2008). Estas herramientas analizan patrones en las secuencias de ADN, como los codones de inicio y parada, las señales de empalme y las regiones conservadas, para identificar las regiones que tienen más probabilidades de codificar proteínas.

Montaje del genoma: Esta técnica se utiliza para ensamblar fragmentos de ADN secuenciados en un genoma completo. Los algoritmos de montaje del genoma utilizan información sobre la superposición de los fragmentos de ADN para reconstruir la secuencia del genoma completo. El montaje del genoma es un proceso computacionalmente intensivo que requiere algoritmos sofisticados para manejar la gran cantidad de datos generados por las tecnologías de secuenciación de próxima generación (Liria, 2022).

Alineamiento estructural de proteínas: Esta técnica se utiliza para comparar las estructuras tridimensionales de las proteínas e identificar similitudes y diferencias. Los algoritmos de alineamiento estructural de proteínas permiten a los investigadores comprender la relación entre la estructura y la función de las proteínas. El alineamiento estructural es particularmente útil para identificar sitios de unión a ligandos y para predecir la actividad enzimática de las proteínas (Liras, Cabello, & Bonet, 2008).

Predicción de estructura de proteínas: Esta técnica se utiliza para predecir la estructura tridimensional de una proteína a partir de su secuencia de aminoácidos. Los programas de predicción de estructura de proteínas utilizan modelos físicos y estadísticos para simular el plegamiento de las proteínas. La predicción de la estructura de proteínas es un

desafío fundamental en bioinformática, ya que la estructura de una proteína determina su función (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019).

Predicción de la expresión génica: Esta técnica se utiliza para predecir los niveles de expresión de los genes en diferentes condiciones. Los modelos de predicción de la expresión génica utilizan datos de chips de ADN o espectrometría de masas para analizar los perfiles de expresión génica. Estos modelos pueden ayudar a identificar los genes que están regulados en respuesta a diferentes estímulos y a comprender los mecanismos de regulación génica (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019).

Interacciones proteína-proteína: Esta técnica se utiliza para identificar las interacciones físicas entre proteínas. Los métodos de detección de interacciones proteína-proteína incluyen la espectrometría de masas y los ensayos de doble híbrido en levadura. La identificación de las interacciones proteína-proteína es fundamental para comprender las vías de señalización celular y los procesos biológicos (Liras, Cabello, & Bonet, 2008).

Modelado de la evolución: Esta técnica se utiliza para simular la evolución de las secuencias de ADN, ARN o proteínas a lo largo del tiempo. Los modelos de evolución se utilizan para reconstruir las relaciones filogenéticas entre diferentes organismos. Estos modelos tienen en cuenta los diferentes tipos de mutaciones que pueden ocurrir en las secuencias de ADN y las tasas a las que ocurren estas mutaciones (Liras, Cabello, & Bonet, 2008).

2.2.2. *Aplicaciones en la investigación de la biodiversidad*

La bioinformática se ha convertido en una herramienta esencial en la investigación de la biodiversidad, impulsando descubrimientos y proporcionando nuevas perspectivas sobre la evolución, la ecología y la conservación de las especies. Al analizar grandes conjuntos de datos genómicos y proteómicos, los investigadores pueden desentrañar los misterios de la diversidad de la vida en la Tierra.

Identificación de especies: La bioinformática permite identificar especies mediante el análisis de secuencias de ADN (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019). Las bases de datos genéticas, como el GenBank, contienen secuencias de ADN de miles de especies diferentes. Al comparar la secuencia de ADN de un organismo desconocido con las secuencias de la base de datos, los investigadores pueden identificar la especie a la que pertenece. Esta técnica, conocida como "código de barras de ADN", se está utilizando cada vez más para identificar especies en estudios de biodiversidad.

Descubrimiento de nuevas especies: La bioinformática también permite descubrir nuevas especies mediante el análisis de datos metagenómicos. Los datos metagenómicos se obtienen mediante la secuenciación directa del ADN de muestras ambientales, como el suelo o el agua. Al analizar estos datos, los investigadores pueden identificar secuencias de ADN que no se corresponden con ninguna especie conocida, lo que indica la presencia de nuevas especies. La metagenómica ha revolucionado el estudio de la biodiversidad microbiana, revelando la existencia de una enorme cantidad de especies desconocidas (Instituto Nacional de Biodiversidad , 2021).

Análisis de la diversidad genética: La bioinformática permite analizar la diversidad genética dentro de una especie mediante el análisis de datos de secuenciación del genoma completo. Al comparar los genomas de diferentes individuos de la misma especie, los investigadores pueden identificar las variaciones genéticas que existen dentro de la especie. Esta información es útil para comprender la adaptación de las especies a diferentes ambientes y para diseñar estrategias de conservación (Liras, Cabello, & Bonet, 2008). El análisis de la diversidad genética también puede ayudar a identificar las poblaciones que son más vulnerables a los cambios ambientales.

Estudios filogenéticos: La bioinformática permite construir árboles filogenéticos que muestran las relaciones evolutivas entre diferentes especies. Los árboles filogenéticos se

construyen mediante el análisis de secuencias de ADN o proteínas. Al comparar las secuencias de diferentes especies, los investigadores pueden determinar qué especies están más estrechamente relacionadas entre sí (Liria, 2022). Esta información es útil para comprender la evolución de la vida en la Tierra y para clasificar las especies en diferentes grupos taxonómicos.

Análisis de la función génica: La bioinformática permite analizar la función de los genes mediante el análisis de datos de expresión génica. Los datos de expresión génica se obtienen mediante la medición de los niveles de ARN mensajero (ARNm) en diferentes tejidos o en diferentes condiciones experimentales. Al analizar estos datos, los investigadores pueden identificar los genes que se expresan en respuesta a diferentes estímulos. Esta información es útil para comprender la función de los genes y cómo contribuyen a la adaptación de las especies. El análisis de la función génica también puede ayudar a identificar los genes que están implicados en enfermedades (Conforme-Garcia, Dávila-Ulloa, Sarango-Ordóñez, & Medina-Gahona, 2024).

Estudios de ecología microbiana: La bioinformática permite estudiar la ecología de las comunidades microbianas mediante el análisis de datos metagenómicos. Al analizar los datos metagenómicos, los investigadores pueden identificar las diferentes especies de microorganismos que están presentes en una muestra ambiental. También pueden determinar las funciones metabólicas que están presentes en la comunidad microbiana. Esta información es útil para comprender cómo las comunidades microbianas contribuyen al funcionamiento de los ecosistemas (Conforme-Garcia, Dávila-Ulloa, Sarango-Ordóñez, & Medina-Gahona, 2024). La ecología microbiana es un campo en rápido crecimiento que está revelando la importancia de los microorganismos en todos los aspectos de la vida en la Tierra.

La integración de la bioinformática y el modelado de ecosistemas en la biotecnología verde representa un avance significativo en la planificación y ejecución de proyectos de

recuperación forestal. Estas herramientas proporcionan un marco para analizar complejas interacciones biológicas y ambientales, facilitando la identificación de estrategias óptimas para la restauración de ecosistemas degradados (Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, 2019). La bioinformática, aplicada a la conservación de bosques, permite la caracterización genética de especies forestales, identificando variantes genéticas clave que confieren resistencia a enfermedades, adaptabilidad a cambios climáticos y otros rasgos deseables para la reforestación.

Conservación de la biodiversidad: La bioinformática puede utilizarse para diseñar estrategias de conservación de la biodiversidad. Al analizar los datos genéticos de diferentes poblaciones de una especie en peligro de extinción, los investigadores pueden identificar las poblaciones que son más diversas genéticamente. Estas poblaciones son las que tienen más probabilidades de sobrevivir a los cambios ambientales. Por lo tanto, los esfuerzos de conservación deben centrarse en la protección de estas poblaciones (Conforme-García, Dávila-Ulloa, Sarango-Ordóñez, & Medina-Gahona, 2024). La bioinformática también puede utilizarse para identificar los hábitats que son más importantes para la conservación de la biodiversidad.

2.3. Minería de datos

La minería de datos es un proceso analítico que combina técnicas de estadística, inteligencia artificial y aprendizaje automático para extraer patrones significativos y relaciones ocultas en grandes volúmenes de datos. Este enfoque ha ganado relevancia en la era digital debido al crecimiento exponencial de los datos generados diariamente. Según (Cedillo, Beltrán, Saltos, & Soriano, 2024) la minería de datos permite analizar conjuntos masivos de información para descubrir tendencias, predecir comportamientos y tomar decisiones informadas en diversos contextos, como la educación, la salud y los negocios.

Este proceso no solo implica el análisis de datos, sino también su recolección, limpieza y transformación previa. Estas etapas son esenciales para garantizar que los datos sean

consistentes y adecuados para el análisis. Como señala (García & Gil, 2023), la minería de datos es una disciplina en constante evolución que se adapta a las necesidades del mercado mediante el uso de algoritmos avanzados y modelos predictivos, lo que la convierte en una herramienta clave para abordar problemas complejos.

Una característica distintiva de la minería de datos es su capacidad para automatizar el descubrimiento de patrones y relaciones en los datos. Esto permite a las organizaciones identificar oportunidades ocultas o riesgos potenciales que no serían evidentes mediante métodos tradicionales. Los algoritmos de minería de datos también son útiles para detectar anomalías y realizar análisis predictivos en campos como la investigación científica, donde se manejan grandes cantidades de información compleja (CIEMAT, 2019).

Además, la minería de datos tiene aplicaciones prácticas en una amplia variedad de sectores. Por ejemplo, en el ámbito académico, se utiliza para analizar publicaciones científicas y detectar tendencias emergentes; en el sector empresarial, ayuda a optimizar estrategias comerciales; mientras que en el ámbito social puede ser empleada para resolver problemas como la deserción escolar o la predicción de siniestros viales (Torres-Quezada, 2021). Este enfoque interdisciplinario sigue transformando la forma en que las organizaciones gestionan y aprovechan sus datos.

2.3.1. Proceso de minería de datos

El proceso de minería de datos es una metodología estructurada que permite transformar grandes volúmenes de datos en conocimiento útil y aplicable. Este proceso se compone de varias etapas interdependientes que van desde la recopilación inicial de datos hasta la presentación de los resultados (Torres-Quezada, 2021). Según (Cedillo, Beltrán, Saltos, & Soriano, 2024) cada etapa desempeña un papel crucial en garantizar que los datos sean procesados de manera eficiente, permitiendo extraer patrones significativos y realizar predicciones precisas.

La minería de datos requiere no solo habilidades técnicas, sino también una comprensión profunda del dominio en el que se aplica. Por ejemplo, en el ámbito científico, este proceso puede ser utilizado para analizar publicaciones académicas, identificar tendencias emergentes y explorar relaciones entre diferentes áreas del conocimiento (Ruiz & Armoa, 2023). Como señala (Torres-Quezada, 2021), las etapas del proceso incluyen desde la recolección y limpieza de datos hasta su análisis mediante algoritmos avanzados y la interpretación de los resultados para la toma de decisiones estratégicas.

Tabla 2*Proceso de minería de datos*

Etapa	Descripción Detallada	Resultados Esperados
Recolección de Datos	Esta etapa implica la recopilación de datos relevantes desde diversas fuentes, como bases de datos, archivos, sensores, redes sociales o publicaciones académicas. Se busca obtener un conjunto de datos representativo y diverso que refleje el fenómeno a estudiar. La calidad y la cantidad de los datos son fundamentales para garantizar resultados confiables y significativos.	Un conjunto de datos inicial que es representativo del problema o fenómeno en estudio, listo para ser procesado.
Preprocesamiento	En esta fase, los datos recolectados son limpiados y organizados para eliminar errores, valores nulos o duplicados. Esto incluye técnicas como la normalización para estandarizar los rangos de los datos y la discretización para convertir datos continuos en categorías. El preprocesamiento asegura que los datos sean consistentes y adecuados para el análisis posterior.	Datos limpios y organizados que están listos para el análisis, sin errores ni inconsistencias que puedan afectar los resultados.
Transformación	Los datos son transformados a un formato adecuado para aplicar técnicas analíticas. Esto puede incluir la creación de nuevas variables derivadas de las existentes, la reducción dimensional mediante métodos como el Análisis de Componentes Principales (PCA) y la codificación de variables categóricas. Esta etapa es crucial para facilitar el análisis posterior y mejorar la interpretación de los resultados.	Un conjunto de datos transformado que resalta las características más relevantes y reduce la complejidad del análisis sin perder información crítica.
Minería	Esta es la etapa central donde se aplican algoritmos específicos para descubrir patrones ocultos en los datos. Dependiendo del objetivo del análisis, se pueden utilizar técnicas como clasificación (para categorizar datos), agrupamiento (para identificar grupos similares) o asociación (para encontrar relaciones entre variables). Esta etapa es donde realmente se extrae el conocimiento útil de los datos.	Patrones identificados, modelos predictivos generados o grupos significativos descubiertos que pueden ser utilizados para tomar decisiones informadas o realizar predicciones futuras.
Evaluación	Una vez obtenidos los patrones, se realiza una evaluación crítica para determinar su validez e importancia. Esto implica verificar si los resultados cumplen con los objetivos planteados inicialmente y si son aplicables a situaciones reales. Se pueden utilizar métricas específicas para evaluar el rendimiento del modelo, como precisión, recall o F1-score en el caso de modelos predictivos.	Resultados validados que demuestran su relevancia y aplicabilidad en el contexto del estudio, garantizando que las conclusiones sean fiables y útiles para la toma de decisiones.
Presentación	Finalmente, los resultados se presentan mediante visualizaciones claras y comprensibles que faciliten su interpretación por parte de los interesados. Esto puede incluir gráficos interactivos, dashboards o informes detallados que resuman los hallazgos más relevantes y permitan a los usuarios explorar los resultados en profundidad. La presentación efectiva es clave para asegurar que los resultados sean utilizados adecuadamente en la toma de decisiones.	Resultados visualizados que permiten a los interesados comprender fácilmente las conclusiones del análisis y tomar decisiones informadas basadas en esos hallazgos.

Fuente: (Cedillo, Beltrán, Saltos, & Soriano, 2024) (Torres-Quezada, 2021) (Rodríguez, Delgado, Solórzano, & Anzules, 2020)

2.3.2. Técnicas de minería de datos y aplicaciones en el análisis de la literatura científica

Para extraer conocimiento valioso de grandes volúmenes de datos, la minería de datos emplea una variedad de técnicas. Estas técnicas permiten descubrir patrones, relaciones y tendencias que pueden ser utilizados para tomar decisiones informadas y resolver problemas complejos. Cada técnica tiene su propio enfoque y aplicación, adaptándose a diferentes tipos de problemas y objetivos.

Tabla 3

Técnicas de minería de datos y aplicaciones en el análisis de la literatura científica

Técnica	Descripción Ampliada	Aplicaciones en el Análisis de la Literatura Científica
Seguimiento de Patrones	El seguimiento de patrones es una técnica fundamental que implica el reconocimiento y monitoreo de tendencias en conjuntos de datos para realizar análisis inteligentes. Este proceso puede relacionarse con la identificación de datos demográficos de alto rendimiento o la comprensión de las variaciones estacionales en el comportamiento de compra del cliente. Al identificar estos patrones, las empresas pueden dirigir sus esfuerzos a mercados específicos y optimizar sus cadenas de suministro.	Permite identificar tendencias emergentes en la investigación científica, como el aumento de publicaciones sobre un tema específico a lo largo del tiempo. Por ejemplo, el seguimiento de patrones en las publicaciones sobre cambio climático podría revelar el crecimiento exponencial de la investigación en este campo. Esta técnica ayuda a los investigadores a mantenerse al día con las últimas tendencias y a identificar áreas de investigación prometedoras.
Asociación	Al igual que el seguimiento de patrones, la técnica de asociación implica buscar ciertas ocurrencias con atributos conectados, buscando variables vinculadas en función de atributos o eventos específicos. Las reglas de asociación pueden ser particularmente útiles para estudiar el comportamiento del consumidor.	Puede identificar relaciones entre diferentes temas de investigación. Por ejemplo, identificar que los artículos que mencionan "genómica" también suelen mencionar "bioinformática", lo que sugiere una fuerte relación entre estos campos. Esto puede ayudar a los investigadores a descubrir conexiones interdisciplinarias y a identificar áreas de colaboración potencial.
Clasificación	La clasificación es una técnica útil para derivar datos relevantes y metadatos basado en un atributo definido. Esencialmente, es el proceso de dividir grandes conjuntos de datos en categorías objetivo ¹ . Esta categorización está determinada por el marco de datos, como bases de datos relacionales u orientadas a objetos.	Permite clasificar artículos científicos según su tema, metodología o tipo de estudio. Por ejemplo, clasificar artículos sobre cáncer según el tipo de tratamiento (quimioterapia, radioterapia, inmunoterapia) o el tipo de estudio (ensayo clínico, estudio observacional, revisión sistemática). Esto facilita la búsqueda y el análisis de artículos relevantes para una pregunta de investigación específica.

Detección de Valores Atípicos	Hay instancias en las que el patrón de datos no proporciona una comprensión clara de los datos. En tales situaciones, la técnica de detección de valores atípicos resulta útil. Implica identificar anomalías o "valores atípicos" en su conjunto de datos para comprender las causas específicas o derivar predicciones más precisas.	Puede identificar investigaciones que presentan resultados inusuales o contradictorios en comparación con la literatura existente. Por ejemplo, identificar un estudio que muestra una correlación inesperada entre dos variables, lo que podría indicar un error metodológico o un descubrimiento importante. Esto permite a los investigadores cuestionar y validar los resultados de la investigación existente.
Agrupamiento (Clustering)	Al igual que la clasificación, el agrupamiento es una técnica de minería de datos que consiste en agrupar datos en función de las similitudes. Ayuda en el descubrimiento de conocimientos, la detección de anomalías y la obtención de información sobre la estructura interna de los datos.	Permite agrupar artículos científicos según temas similares o áreas de investigación relacionadas. Por ejemplo, agrupar artículos sobre inteligencia artificial en subgrupos como "aprendizaje automático", "procesamiento del lenguaje natural" y "visión por computadora". Esto facilita la identificación de áreas de investigación emergentes y la colaboración entre investigadores.
Patrones Secuenciales	Como sugiere el nombre, esta es una técnica de minería que se enfoca en descubrir patrones o una serie de eventos que tienen lugar en una secuencia. Se usa ampliamente en la minería de datos transaccionales, pero tiene numerosas aplicaciones.	Identifica secuencias de eventos en la evolución de un campo científico. Por ejemplo, identificar la secuencia de descubrimientos que llevaron al desarrollo de una nueva vacuna o tratamiento. Esto permite a los investigadores comprender la progresión del conocimiento en un área específica y a identificar los hitos clave en su desarrollo.
Árboles de Decisión	Un árbol de decisiones es una técnica de minería de datos en el aprendizaje automático (ML) que se centra en las relaciones de modelado de entrada y salida mediante reglas si/entonces. Con este enfoque, puede aprender cómo las entradas de datos influyen en las salidas. Los árboles suelen estar diseñados en una estructura similar a un diagrama de flujo de arriba hacia abajo. Se utilizan principalmente para modelos de clasificación y regresión.	Permite modelar las relaciones entre diferentes factores que influyen en un resultado científico. Por ejemplo, modelar los factores que influyen en el éxito de un tratamiento médico basándose en características del paciente, el tipo de enfermedad y el protocolo de tratamiento. Esto ayuda a los investigadores a comprender la complejidad de los fenómenos científicos y a predecir los resultados en función de diferentes variables.
Análisis de Regresión	Es una de las técnicas de minería de datos más populares en el aprendizaje automático que utiliza la relación lineal entre variables. Le ayuda a predecir el valor futuro de las variables. La técnica tiene numerosas aplicaciones en pronósticos financieros, planificación de recursos, toma de decisiones estratégicas y más.	Puede utilizarse para predecir el impacto de una investigación científica basándose en factores como el número de citas, el número de autores y la reputación de la revista. Esto permite a los investigadores y a las instituciones evaluar la calidad y el impacto de su investigación y a tomar decisiones informadas sobre la asignación de recursos.

Procesamiento de Memoria a Largo Plazo	<p>El procesamiento de memoria a largo plazo es una técnica de minería de datos en el aprendizaje automático que se utiliza para analizar datos durante períodos prolongados. Permite identificar patrones de datos basados en el tiempo, como datos climáticos, de manera más efectiva. Su objetivo es escalar los datos en la memoria del sistema y utilizar información adicional en el análisis.</p>	<p>Permite analizar la evolución de un campo científico a lo largo del tiempo, identificando tendencias a largo plazo y ciclos de investigación. Por ejemplo, analizar cómo ha evolucionado la investigación sobre el cambio climático desde los primeros estudios hasta la actualidad. Esto ayuda a los investigadores a comprender la trayectoria de un campo y a identificar oportunidades para futuras investigaciones.</p>
Redes Neuronales	<p>Una red neuronal también es una de las técnicas populares de minería de datos en los modelos de aprendizaje automático utilizados con Inteligencia Artificial (IA). Al igual que las neuronas en el cerebro, busca identificar relaciones en los datos. Las redes neuronales tienen diferentes capas que trabajan juntas para producir resultados de análisis de datos con gran precisión. Estos modelos buscan patrones en una gran cantidad de datos, proporcionando información valiosa para las organizaciones.</p>	<p>Permite identificar patrones complejos y no lineales en los datos científicos, como la relación entre la estructura de una proteína y su función biológica. Por ejemplo, utilizar redes neuronales para predecir la actividad biológica de un compuesto químico basándose en su estructura molecular. Esto acelera el proceso de descubrimiento de fármacos y permite identificar candidatos prometedores para el desarrollo de nuevos tratamientos.</p>

Fuente: (Orozco, Villao, Orozco, & Villarroel, 2021) (Umaquina, 2024) (Pérez-Suasnavas, Salgado-Proañó, Hasperué, Cela, & Santamaría, 2023) (Ramírez, Delgado, & Montúfar, 2022)

2.4. Visualización de datos

La visualización de datos se refiere a la representación gráfica de información y datos, utilizando elementos visuales como gráficos, diagramas y mapas. Este enfoque permite a los investigadores y analistas transformar datos complejos en formatos visuales más comprensibles, facilitando la identificación de patrones, tendencias y relaciones significativas. La visualización es esencial en el análisis de datos, ya que ayuda a convertir información cruda en insights que pueden ser fácilmente interpretados por los usuarios, independientemente de su nivel de experiencia con los datos (Cornejo & Núñez, 2024).

En el contexto actual, donde se generan grandes volúmenes de datos, la visualización se ha vuelto aún más crucial. Herramientas avanzadas permiten crear visualizaciones interactivas que no solo muestran los datos, sino que también permiten a los usuarios explorar diferentes aspectos de la información. Esto es especialmente útil en campos como la investigación científica, donde los datos pueden ser complejos y multidimensionales (Therón, 2021). La capacidad de interactuar con los datos a través de visualizaciones mejora la comprensión y facilita la comunicación de hallazgos importantes.

Además, la visualización de datos no se limita a gráficos estáticos; incluye también representaciones dinámicas que pueden actualizarse en tiempo real para reflejar cambios en los datos subyacentes. Esto es particularmente relevante en áreas como el monitoreo de salud pública o el análisis financiero, donde las decisiones deben basarse en la información más reciente disponible (Cruz, Garzón, Quezada, & Carvajal, 2022). La visualización efectiva puede ser un puente entre el análisis técnico y la toma de decisiones estratégicas.

2.4.1. Principios de una visualización de datos eficaz

Una visualización de datos eficaz debe diseñarse teniendo en cuenta varios principios clave para garantizar que transmita la información de manera clara y precisa. Estos principios abarcan aspectos como la selección adecuada de gráficos, el uso del color, la claridad del diseño y la accesibilidad.

Claridad y Simplicidad: Una visualización debe ser fácil de entender a primera vista. Esto implica evitar el uso excesivo de elementos decorativos o gráficos complejos que puedan distraer al usuario del mensaje principal. Por ejemplo, un gráfico lineal simple puede ser más efectivo que un gráfico tridimensional cuando se busca mostrar tendencias temporales (Lavalle, 2021).

Selección Adecuada del Tipo de Gráfico: Elegir el tipo correcto de gráfico es crucial para representar adecuadamente los datos. Los gráficos de barras son ideales para comparaciones categóricas, mientras que los gráficos de dispersión son útiles para analizar relaciones entre variables. Seleccionar el gráfico adecuado ayuda a comunicar el mensaje deseado de manera más efectiva (Lavalle, 2021).

Uso Estratégico del Color: El color es una herramienta poderosa para resaltar información clave o diferenciar categorías dentro de un conjunto de datos. Sin embargo, debe usarse con moderación y propósito. Los colores deben ayudar a resaltar tendencias importantes en un gráfico sin causar confusión (Therón, 2021).

Diseño Intuitivo: La disposición lógica y clara de los elementos facilita la navegación visual. Las etiquetas deben ser concisas pero informativas, y las leyendas deben estar ubicadas cerca del gráfico para evitar confusiones. Un diseño bien pensado permite a los espectadores comprender rápidamente la información presentada (Therón, 2021).

Accesibilidad: Es vital que las visualizaciones sean accesibles para todos los usuarios, incluidas las personas con discapacidades visuales o cognitivas. Esto implica utilizar contrastes adecuados entre texto y fondo, proporcionar descripciones alternativas para imágenes y considerar paletas accesibles para personas con daltonismo (Therón, 2021).

Integridad de los Datos: Una visualización debe representar los datos con precisión y honestidad. Evitar manipulaciones engañosas o distorsiones es crucial para

mantener la credibilidad del análisis presentado. Siempre es recomendable citar las fuentes y proporcionar contexto sobre cómo se recolectaron o procesaron los datos (Cornejo & Núñez, 2024).

Contexto Relevante: Proporcionar suficiente contexto ayuda a interpretar correctamente los datos presentados. Esto incluye títulos claros, leyendas explicativas y notas al pie que aclaren aspectos técnicos o metodológicos del gráfico.

Interactividad: Las herramientas interactivas permiten a los usuarios explorar diferentes aspectos del conjunto de datos, lo que les ayuda a descubrir patrones ocultos o realizar análisis personalizados. Por ejemplo, un gráfico interactivo puede permitir filtrar información por región o rango temporal (Cornejo & Núñez, 2024).

Adaptabilidad: En un mundo donde el acceso móvil es predominante, las visualizaciones deben ser responsivas y adaptarse a diferentes tamaños de pantalla sin perder claridad ni funcionalidad (Cruz, Garzón, Quezada, & Carvajal, 2022).

Narrativa Clara: Finalmente, una buena visualización debe contar una historia coherente basada en los datos presentados. Utilizar elementos narrativos ayuda a guiar al espectador a través del análisis y resaltar hallazgos clave que pueden influir en decisiones o discusiones posteriores (Cruz, Garzón, Quezada, & Carvajal, 2022).

2.4.2. Tipos de visualizaciones de datos

La visualización de datos es una herramienta fundamental en el análisis de información, ya que permite representar datos complejos de manera gráfica, facilitando su comprensión y análisis. Existen diversos tipos de visualizaciones, cada uno adecuado para diferentes tipos de datos y objetivos analíticos.

Los gráficos de barras son uno de los tipos más utilizados para comparar cantidades entre diferentes categorías. Estos gráficos permiten visualizar rápidamente las diferencias en magnitudes, siendo ideales para mostrar datos categóricos. Por ejemplo, un gráfico de barras puede mostrar las ventas por producto en diferentes regiones, facilitando la

comparación directa entre las categorías (Cruz, Garzón, Quezada, & Carvajal, 2022). La simplicidad y claridad de este tipo de visualización lo convierten en una opción popular en informes y presentaciones.

Los gráficos de líneas son especialmente útiles para mostrar tendencias a lo largo del tiempo. Al conectar puntos de datos con líneas, estos gráficos permiten observar cambios y patrones temporales con facilidad. Por ejemplo, un gráfico de líneas puede representar el crecimiento mensual de las visitas a un sitio web durante un año, ayudando a identificar picos estacionales o tendencias generales en el tráfico (Cruz, Garzón, Quezada, & Carvajal, 2022). Esta visualización es comúnmente utilizada en análisis financieros y estudios de mercado.

Los gráficos circulares, o gráficos de pastel, son efectivos para mostrar proporciones dentro de un todo. Aunque son fáciles de entender, deben usarse con moderación, ya que pueden volverse confusos si hay demasiadas categorías o si las diferencias entre las proporciones son sutiles. Un ejemplo sería utilizar un gráfico circular para mostrar la distribución del presupuesto en diferentes departamentos dentro de una empresa (Cruz, Garzón, Quezada, & Carvajal, 2022).

Los mapas de calor son otra forma poderosa de visualización que utiliza colores para representar la densidad o intensidad de los datos en una matriz o mapa geográfico. Esta técnica permite identificar rápidamente áreas con alta concentración de valores o eventos. Por ejemplo, un mapa de calor puede mostrar la distribución geográfica de casos de enfermedades en una región, ayudando a identificar focos críticos que requieren atención (Therón, 2021).

Los gráficos de dispersión son útiles para explorar la relación entre dos variables cuantitativas. Cada punto en el gráfico representa un par de valores, lo que permite identificar correlaciones y agrupaciones. Por ejemplo, un gráfico de dispersión puede mostrar la relación entre el ingreso y el gasto en publicidad por parte de diferentes empresas, ayudando a revelar tendencias o patrones inusuales (Therón, 2021).

Finalmente, las infografías combinan texto e imágenes para contar una historia a partir de los datos. Estas visualizaciones no solo presentan información numérica, sino que también proporcionan contexto y narrativas que facilitan la comprensión del contenido. Las infografías son especialmente efectivas en entornos educativos y comunicativos, donde se busca captar la atención del público y transmitir información compleja de manera accesible (Therón, 2021).

Cada uno de estos tipos de visualización tiene sus propias ventajas y desventajas, y su elección dependerá del tipo específico de datos que se estén analizando y del mensaje que se desee comunicar.

2.5. Dashboard

Un dashboard, o panel de control, es una herramienta de visualización que permite a los usuarios monitorear y analizar datos clave en un formato gráfico y accesible. Su diseño está orientado a proporcionar una visión general de los indicadores de rendimiento (KPI) y otras métricas relevantes, facilitando la toma de decisiones informadas en tiempo real (Córdova, Martínez, & Córdova, 2021). Los dashboards son utilizados en diversas áreas, incluyendo negocios, educación y salud, para integrar información de múltiples fuentes y presentarla de manera coherente.

La importancia de los dashboards radica en su capacidad para condensar grandes volúmenes de datos en representaciones visuales que son fáciles de interpretar. Al presentar datos complejos mediante gráficos, tablas y otros elementos visuales, los dashboards permiten a los usuarios identificar rápidamente patrones, tendencias y anomalías que podrían no ser evidentes en informes textuales o tablas extensas (Pincay, Baque, Caicedo, & Paladines, 2024). Esto es especialmente relevante en un entorno empresarial donde la rapidez en la toma de decisiones puede influir significativamente en el éxito organizacional.

Además, los dashboards pueden ser interactivos, lo que significa que los usuarios pueden explorar los datos más a fondo mediante filtros y selecciones personalizadas. Esta interactividad no solo mejora la comprensión de la información presentada, sino que también permite a los usuarios adaptar el dashboard a sus necesidades específicas, haciendo que la herramienta sea aún más valiosa para el análisis continuo y la evaluación del rendimiento (Guaman & Patiño, 2022).

2.5.1. Características del Dashboard

Los dashboards efectivos deben poseer ciertas características clave que aseguran su funcionalidad y utilidad para los usuarios finales.

Visualización Clara y Concisa: La principal función de un dashboard es presentar datos de manera clara y comprensible. Esto implica utilizar gráficos y elementos visuales que permitan a los usuarios identificar rápidamente las tendencias y patrones relevantes. Por ejemplo, un gráfico de barras puede mostrar comparaciones entre diferentes categorías, mientras que un gráfico de líneas puede ilustrar cambios a lo largo del tiempo (Jiménez & Mendoza, 2023).

Interactividad: Un buen dashboard permite a los usuarios interactuar con los datos, como filtrar información por diferentes criterios o profundizar en detalles específicos al hacer clic en ciertos elementos visuales. Esta interactividad ayuda a los usuarios a explorar los datos desde múltiples ángulos y obtener insights más profundos (Jiménez & Mendoza, 2023).

Actualización en Tiempo Real: Para ser verdaderamente útil, un dashboard debe actualizarse automáticamente con datos recientes. Esto permite a los usuarios monitorear el rendimiento en tiempo real y tomar decisiones basadas en la información más actualizada disponible, lo cual es crucial en entornos dinámicos como el marketing digital o la gestión financiera (Jiménez & Mendoza, 2023).

Personalización: Un dashboard debe ser adaptable a las necesidades específicas del usuario o del departamento al que sirve. Esto implica permitir configuraciones personalizadas para mostrar solo las métricas relevantes para cada usuario o grupo, asegurando que cada uno tenga acceso a la información más pertinente para su trabajo (Jiménez & Mendoza, 2023).

Accesibilidad: Los dashboards deben ser accesibles desde diferentes dispositivos, incluidos ordenadores de escritorio, tabletas y teléfonos móviles. Esto asegura que los usuarios puedan acceder a la información crítica desde cualquier lugar y en cualquier momento, facilitando así una toma de decisiones más ágil (Pincay, Baque, Caicedo, & Paladines, 2024).

Centralización de KPIs: Un dashboard efectivo centraliza todos los indicadores clave de rendimiento en un solo lugar, lo que permite a los usuarios obtener una visión holística del rendimiento organizacional sin tener que consultar múltiples fuentes de datos (Pincay, Baque, Caicedo, & Paladines, 2024).

Interfaz Amigable: La interfaz del dashboard debe ser intuitiva y fácil de usar, incluso para aquellos sin experiencia técnica previa. Un diseño limpio y organizado con etiquetas claras facilita la navegación y mejora la experiencia del usuario (Pincay, Baque, Caicedo, & Paladines, 2024).

Toma de Decisiones Informadas: El objetivo principal de un dashboard es proporcionar a los usuarios la información necesaria para tomar decisiones informadas rápidamente. Esto implica presentar datos relevantes y significativos que permitan evaluar el rendimiento y detectar áreas de mejora (Córdova, Martínez, & Córdova, 2021).

Monitoreo Continuo del Rendimiento: Los dashboards deben permitir el monitoreo constante del rendimiento organizacional, alertando a los usuarios sobre desviaciones significativas o problemas potenciales en tiempo real. Esto permite una respuesta rápida

ante situaciones críticas y contribuye a mantener el desempeño dentro de los objetivos establecidos (Córdova, Martínez, & Córdova, 2021).

Estas características hacen que los dashboards sean herramientas poderosas para la visualización y análisis de datos, permitiendo a las organizaciones optimizar su rendimiento y tomar decisiones estratégicas basadas en información precisa.

2.5.2. *Diseño de un Dashboard*

El diseño de un dashboard es un proceso crítico que implica la planificación y organización de la información visual para facilitar la toma de decisiones informadas. Un dashboard bien diseñado no solo presenta datos, sino que también cuenta una historia, guiando al usuario a través de la información de manera clara y efectiva.

Tabla 4

Aspecto del Diseño de un Dashboard

Aspecto del Diseño	Descripción Ampliada
Definición de Objetivos	Antes de iniciar el diseño, es fundamental definir los objetivos del dashboard. Esto incluye identificar qué preguntas debe responder y qué decisiones debe facilitar. Por ejemplo, un dashboard financiero puede centrarse en métricas como ingresos, gastos y márgenes de ganancia. Establecer objetivos claros asegura que el dashboard esté alineado con las necesidades del usuario.
Identificación de KPIs	Los indicadores clave de rendimiento (KPIs) deben seleccionarse cuidadosamente para reflejar los objetivos definidos. Estos deben ser relevantes, medibles y accionables. Por ejemplo, en un dashboard de marketing, los KPIs podrían incluir la tasa de conversión, el retorno de inversión (ROI) y el tráfico web. Elegir KPIs precisos ayuda a enfocar la atención en lo que realmente importa.
Selección del Tipo de Visualización	La elección del tipo adecuado de gráfico es crucial para comunicar efectivamente la información. Los gráficos de barras son ideales para comparaciones categóricas, mientras que los gráficos de líneas son útiles para mostrar tendencias temporales. Los mapas de calor pueden resaltar densidades o intensidades en datos geográficos o tabulares. La visualización correcta mejora la comprensión inmediata del usuario.
Organización del Espacio	El diseño del layout debe ser intuitivo y jerárquico, colocando los elementos más importantes en posiciones destacadas (como la parte superior izquierda). Dividir el espacio en secciones o cuadrantes puede ayudar a agrupar información relacionada y mejorar la navegación visual. Un diseño limpio evita el desorden y facilita que los usuarios encuentren rápidamente la información relevante.

Uso Eficaz del Color	El color debe utilizarse estratégicamente para resaltar información clave o diferenciar categorías. Una paleta limitada y coherente mejora la legibilidad y evita confusiones. Por ejemplo, colores cálidos pueden destacar alertas o problemas, mientras que colores fríos pueden representar estabilidad o éxito. También se debe considerar la accesibilidad para usuarios con daltonismo u otras limitaciones visuales.
Interactividad	Incorporar elementos interactivos permite a los usuarios explorar los datos más a fondo. Esto incluye filtros por rango temporal, categorías específicas o detalles adicionales al hacer clic en ciertos gráficos. La interactividad fomenta un análisis más profundo y personalizado, lo que aumenta la utilidad del dashboard para diferentes usuarios dentro de la organización.
Pruebas y Retroalimentación	Antes del lanzamiento final, es importante realizar pruebas con usuarios finales para evaluar la funcionalidad y usabilidad del dashboard. Recoger retroalimentación permite identificar áreas de mejora y realizar ajustes necesarios para optimizar la experiencia del usuario. Este paso asegura que el dashboard cumpla con las expectativas y necesidades reales de sus usuarios objetivo.
Actualización Regular	Los dashboards deben reflejar datos actualizados en tiempo real o casi real para garantizar su relevancia continua. Esto es especialmente importante en entornos dinámicos como ventas o marketing digital, donde las métricas cambian constantemente. Un sistema automatizado que actualice los datos asegura que las decisiones se basen en información precisa y actualizada.
Documentación y Capacitación	Proporcionar guías claras sobre cómo usar e interpretar el dashboard es esencial para maximizar su efectividad. Esto incluye explicar cómo interactuar con los gráficos, cómo aplicar filtros y cómo interpretar las métricas presentadas. La capacitación a los usuarios garantiza que puedan aprovechar al máximo las funcionalidades del dashboard desde el primer momento.

Fuente: (Córdova, Martínez, & Córdova, 2021) (Pincay, Baque, Caicedo, & Paladines, 2024) (Guaman & Patiño, 2022) (Cruz, Garzón, Quezada, & Carvajal, 2022)

En este sentido, el diseño de un dashboard no solo implica elegir gráficos atractivos; requiere una planificación estratégica basada en objetivos claros, selección adecuada de KPIs, organización visual efectiva e interactividad funcional. Además, un buen diseño debe considerar la experiencia del usuario mediante pruebas previas al lanzamiento y garantizar actualizaciones constantes para mantener su relevancia.

2.5.3. Beneficios del Dashboard para la Investigación de la Biodiversidad

Los dashboards se han convertido en herramientas esenciales para la gestión y análisis de datos en diversos campos, y la investigación de la biodiversidad no es una excepción. Estos paneles de control ofrecen una visión consolidada y en tiempo real de la información clave, facilitando la toma de decisiones informadas y la optimización de los esfuerzos de conservación. Uno de los principales beneficios de utilizar dashboards en la investigación de la biodiversidad es el monitoreo eficaz de especies introducidas (Cruz, Garzón, Quezada, & Carvajal, 2022). Al realizar un seguimiento exhaustivo de estas especies, se pueden priorizar mejor los esfuerzos de conservación y manejo, optimizando los recursos disponibles y minimizando el impacto negativo en la biodiversidad nativa.

Además, los dashboards facilitan el seguimiento de cambios en especies y hábitats. Mediante métodos estandarizados, se pueden monitorear las poblaciones de vertebrados y otros grupos taxonómicos, identificando reducciones o alteraciones significativas que requieran una respuesta inmediata. Esto es crucial para garantizar la salud y sostenibilidad de los ecosistemas (Guaman & Patiño, 2022). Por otro lado, los dashboards contribuyen a la gestión de datos de recursos genéticos, lo que permite una administración más eficiente de la Base Nacional de Datos sobre Biodiversidad y fortalece el Banco Nacional de Recursos Genéticos. Esto facilita el acceso a información genética valiosa para la investigación sobre conservación y uso sostenible (Instituto Nacional de Biodiversidad, 2021).

Asimismo, los dashboards apoyan a la gestión territorial al proporcionar información clave para la conservación de la biodiversidad y el bienestar humano. Al integrar datos sobre ecosistemas, especies y servicios ecosistémicos, se pueden tomar decisiones más informadas sobre el uso del suelo, la planificación urbana y la gestión de recursos naturales (Pincay, Baque, Caicedo, & Paladines, 2024). De igual manera, estos paneles facilitan la difusión de resultados científicos, promoviendo una mayor comprensión sobre la

importancia de la biodiversidad entre diferentes sectores de la sociedad y fomentando así la participación ciudadana en los esfuerzos por su conservación (Liria, 2022).

Otro beneficio significativo es que los dashboards permiten identificar y cuantificar los beneficios "invisibles" que proporciona la vida silvestre a los ecosistemas. Estos beneficios incluyen servicios como la regulación de poblaciones de presas, polinización, dispersión de semillas y control de plagas (Campaña Sin Biodiversidad No Hay Vida, 2022). Comprender el valor de estos servicios ecosistémicos permite diseñar estrategias de conservación más efectivas. Además, los dashboards facilitan una integración más sistemática, organizada y generalizada en relación con la protección, uso sostenible y promoción de la biodiversidad. Al proporcionar una visión global de los indicadores clave, se pueden incorporar consideraciones sobre biodiversidad en proyectos de desarrollo, minimizando impactos negativos y maximizando beneficios (Casillas, 2008).

Asimismo, estos paneles permiten monitorear la efectividad de las estrategias implementadas en conservación. Al seguir de cerca los indicadores clave, se pueden evaluar los resultados de las intervenciones realizadas y realizar ajustes necesarios para mejorar su impacto (Córdova, Martínez, & Córdova, 2021). Por último, los dashboards facilitan el seguimiento del progreso hacia el cumplimiento de los Objetivos de Desarrollo Sostenible (ODS) relacionados con la biodiversidad. Al integrar datos sobre conservación de especies, protección de ecosistemas y uso sostenible de recursos naturales, se puede evaluar el avance hacia metas específicas y diseñar políticas más efectivas para alcanzar dichos objetivos (Cornejo & Núñez, 2024).

2.6. Investigación científica

La investigación científica es un proceso sistemático y estructurado que busca aumentar el conocimiento sobre un tema específico mediante la recopilación, análisis e interpretación de datos. Este proceso se basa en el método científico, que incluye la formulación de hipótesis, la realización de experimentos o estudios observacionales, la

recolección de datos y la evaluación de los resultados (Conforme-Garcia, Dávila-Ulloa, Sarango-Ordóñez, & Medina-Gahona, 2024).

A través de la investigación científica, los científicos pueden formular teorías y modelos que explican fenómenos naturales, así como desarrollar tecnologías y soluciones a problemas complejos. Este enfoque basado en evidencia permite a los investigadores validar sus hallazgos y contribuir al cuerpo de conocimiento existente. La investigación científica no solo se limita a la experimentación en laboratorios; también incluye estudios de campo, análisis estadísticos y revisiones sistemáticas de literatura (Conforme-Garcia, Dávila-Ulloa, Sarango-Ordóñez, & Medina-Gahona, 2024).

La investigación científica se caracteriza por su rigor metodológico y su búsqueda de objetividad. Los investigadores deben seguir protocolos establecidos para garantizar la validez y confiabilidad de sus resultados. Esto implica el uso de técnicas estadísticas adecuadas, el control de variables y la revisión por pares antes de publicar los hallazgos en revistas científicas (Toulkeridis, y otros, 2020). En este sentido se puede decir que, la investigación científica es un proceso dinámico que busca generar nuevos conocimientos y resolver problemas mediante un enfoque sistemático y basado en evidencia. Su importancia radica en su capacidad para informar políticas, guiar prácticas y contribuir al bienestar social y ambiental.

2.6.1. Tipos de investigación científica

La investigación científica se clasifica en varios tipos según su objetivo, alcance y metodología. Cada tipo de investigación tiene sus propias características y se utiliza para abordar diferentes tipos de preguntas y problemas.

Tabla 5*Tipos de investigación científica*

Tipo de Investigación	Descripción	Beneficios de un Dashboard
Exploratoria	Busca explorar un tema o problema poco conocido para generar hipótesis y preguntas de investigación.	Permite visualizar datos de diferentes fuentes para identificar patrones y tendencias emergentes, facilitando la formulación de hipótesis y la definición de áreas de investigación prometedoras.
Descriptiva	Describe las características de una población, fenómeno o situación.	Facilita la presentación de datos demográficos, características de especies, distribución geográfica y otros datos descriptivos de manera clara y concisa, permitiendo una mejor comprensión del objeto de estudio.
Correlacional	Examina las relaciones entre dos o más variables.	Permite visualizar la correlación entre variables mediante gráficos de dispersión, mapas de calor y otros elementos visuales, facilitando la identificación de relaciones significativas y la formulación de hipótesis sobre causalidad.
Explicativa	Busca establecer relaciones de causa y efecto entre variables.	Facilita la presentación de resultados de experimentos y estudios observacionales que buscan establecer relaciones causales, permitiendo visualizar el impacto de una variable sobre otra y apoyar la toma de decisiones basadas en evidencia.
Experimental	Manipula una o más variables para determinar su efecto sobre otra variable, controlando las demás variables.	Permite visualizar los resultados de experimentos de manera clara y concisa, facilitando la comparación entre grupos experimentales y control, así como la identificación de efectos significativos en las conclusiones obtenidas.
Tecnológica	Emplea el conocimiento científico para el desarrollo de tecnologías blandas o duras.	Facilita el seguimiento del progreso en proyectos tecnológicos mostrando el impacto en el cumplimiento del objetivo final, así como el consumo eficiente de recursos durante su implementación.

Fuente: (Casillas, 2008) (Cornejo & Núñez, 2024)

2.6.2. *Importancia de la investigación científica en la conservación de la biodiversidad*

La investigación científica desempeña un papel crucial en la conservación de la biodiversidad al proporcionar el conocimiento necesario para comprender los ecosistemas y las especies que los habitan. A través del estudio sistemático del medio ambiente, los científicos pueden identificar las amenazas a la biodiversidad, como el cambio climático, la deforestación y la contaminación, lo que permite desarrollar estrategias efectivas para mitigar estos impactos (Conforme-Garcia, Dávila-Ulloa, Sarango-Ordóñez, & Medina-Gahona, 2024).

Además, la investigación científica es fundamental para el monitoreo continuo del estado de las poblaciones animales y vegetales. Al recopilar datos sobre especies en peligro o ecosistemas vulnerables, los investigadores pueden evaluar el éxito o fracaso de las iniciativas conservacionistas implementadas (CEPAL, 2023). Esta información es vital para ajustar políticas públicas y programas destinados a proteger la biodiversidad.

La investigación también contribuye al desarrollo sostenible al estudiar cómo las actividades humanas afectan a los ecosistemas naturales. Al entender estas interacciones complejas, los científicos pueden proponer prácticas agrícolas sostenibles, métodos eficientes para el uso del agua y estrategias para conservar hábitats críticos sin comprometer el desarrollo económico (Báez, 2019).

Otro aspecto importante es que la investigación científica fomenta la educación ambiental al proporcionar información basada en evidencia sobre la importancia del mantenimiento del equilibrio ecológico. Al educar a las comunidades sobre su entorno natural y las consecuencias del deterioro ambiental, se promueve una mayor conciencia pública sobre cuestiones relacionadas con la conservación (Instituto Nacional de Biodiversidad, 2021).

Asimismo, los hallazgos científicos son esenciales para informar a los responsables políticos sobre las mejores prácticas para conservar recursos naturales críticos e implementar regulaciones efectivas que protejan hábitats importantes (Mena, 2014). Sin esta base científica sólida, sería difícil justificar decisiones políticas que afectan directamente a los ecosistemas. La colaboración entre científicos e instituciones locales también es fundamental para abordar problemas complejos relacionados con la biodiversidad. La participación activa en proyectos comunitarios no solo fortalece el vínculo entre ciencia y sociedad, sino que también asegura que las soluciones propuestas sean culturalmente relevantes y aceptables (Varea, 2004).

Por último, a medida que avanza nuestra comprensión del genoma animal y vegetal a través del estudio científico, surgen nuevas oportunidades para preservar especies amenazadas mediante técnicas como la biotecnología o la conservación ex situ (fuera del hábitat natural) (Casillas, 2008). Estas innovaciones pueden ser cruciales para salvar especies al borde de la extinción.

3. MARCO METODOLÓGICO

Este capítulo describe el enfoque metodológico utilizado para desarrollar el dashboard de análisis de publicaciones bioinformáticas sobre la biodiversidad ecuatoriana. Se detallan los materiales necesarios para la recopilación y procesamiento de datos, la metodología seguida para el análisis y la visualización, las herramientas de software empleadas y la infraestructura de hardware requerida. El objetivo es proporcionar una descripción clara y concisa de los métodos utilizados para alcanzar los objetivos de la investigación.

El diseño de esta investigación debe establecerse como exploratorio-descriptivo con elementos analíticos. En la fase exploratoria, se buscaría identificar patrones emergentes y gaps en la literatura científica sobre biodiversidad ecuatoriana sin partir de hipótesis preestablecidas. La fase descriptiva permitiría caracterizar sistemáticamente el panorama actual de publicaciones bioinformáticas, detallando tendencias, distribución geográfica y taxonómica. Los elementos analíticos entrarían en juego al establecer relaciones entre

variables como especies estudiadas, metodologías empleadas, áreas geográficas y evolución temporal de la producción científica.

El enfoque metodológico propuesto para esta investigación es de carácter mixto (cuali-cuantitativo), integrando técnicas complementarias que permiten un análisis comprehensivo de las publicaciones bioinformáticas sobre biodiversidad ecuatoriana. Desde la perspectiva cuantitativa, se implementará un análisis estadístico de frecuencias para identificar tendencias temporales, distribución por especies y regiones, complementado con técnicas de minería de datos para descubrir patrones emergentes y agrupaciones temáticas, así como análisis bibliométricos que revelarán el impacto científico y las redes de colaboración entre investigadores. Paralelamente, el componente cualitativo enriquecerá la investigación mediante el análisis de contenido para identificar metodologías innovadoras, la interpretación contextual de los hallazgos para evaluar su significado en el panorama científico nacional e internacional, la aplicación de criterios preestablecidos para evaluar rigurosamente la calidad y relevancia de cada estudio, y la incorporación de entrevistas con expertos en biodiversidad y bioinformática que validarán la utilidad del dashboard y proporcionarán insights valiosos para contextualizar los resultados obtenidos.

3.1. Materiales

La siguiente tabla resume los materiales clave utilizados en este proyecto. Estos recursos son fundamentales para la recopilación, el análisis y la interpretación de datos relacionados con la biodiversidad ecuatoriana. La selección de estos materiales se basa en su relevancia para el tema de investigación y su accesibilidad.

Tabla 6

Materiales de la investigación

Material	Descripción	Fuente(s) Potencial(es)	Propósito
Bases de datos de biodiversidad	Conjunto de datos que contienen información sobre la distribución, taxonomía y otros datos relevantes de las especies de la biodiversidad ecuatoriana.	GBIF (Global Biodiversity Information Facility), NHM (Natural History Museum), SpeciesLink, BNDB (Base de Datos y Sistema de Biodiversidad del Ecuador), iNaturalist.	Proporcionar la base para el análisis comparativo y la identificación de áreas de estudio, especies de interés y gaps de conocimiento.
Publicaciones científicas	Artículos, papers, tesis y otros documentos que contienen información bioinformática sobre la biodiversidad ecuatoriana.	Scopus, Web of Science, PubMed, Google Scholar, repositorios de universidades ecuatorianas, revistas científicas especializadas en biodiversidad y ecología.	Extraer datos sobre metodologías, resultados, áreas de estudio y especies tratadas en la investigación actual.
Lista Roja de la IUCN	Información sobre el estado de conservación de las especies, incluyendo categorías de amenaza y datos de distribución.	IUCN (International Union for Conservation of Nature)	Evaluar la representación de especies amenazadas en las publicaciones y identificar áreas de investigación prioritaria.
Datos geoespaciales	Datos cartográficos y geográficos que permiten la visualización y el análisis de la distribución de especies y áreas de endemismo.	SIGTIERRAS, IGM (Instituto Geográfico Militar), GeoEcuador, Shapefiles de áreas protegidas, datos de sensores remotos (satélites).	Mapear la distribución de especies, identificar áreas de alta biodiversidad y evaluar el impacto de factores ambientales en la biodiversidad.

3.2. Metodología

A continuación, se describe las etapas clave de la metodología utilizada en este proyecto. Se detalla con las técnicas y algoritmos empleados, las variables clave y el resultado esperado. Este enfoque sistemático asegura la validez y la confiabilidad de los resultados obtenidos.

Tabla 7
Metodología de la investigación

Etapas	Descripción	Técnicas/Algoritmos	Variables Clave	Resultado Esperado
Recopilación y Estructuración de Datos	Recopilar y organizar la información de las fuentes identificadas en una base de datos coherente.	Web scraping (Beautiful Soup, Scrapy), APIs (GBIF, etc.), Procesamiento del lenguaje natural (NLP) para la extracción de palabras clave, entidades nombradas y resúmenes.	Título, autores, año de publicación, revista, resumen, palabras clave, especies mencionadas, ubicación geográfica del estudio, métodos utilizados, resultados principales.	Base de datos estructurada de publicaciones científicas con metadatos relevantes.
Implementación de Minería de Datos	Aplicar algoritmos de minería de datos para descubrir patrones, tendencias y relaciones en los datos recopilados.	Agrupamiento (clustering) para identificar temas comunes, Clasificación (árboles de decisión, SVM) para predecir áreas de investigación, Análisis de redes para identificar colaboraciones entre autores y instituciones, Análisis de sentimiento.	Áreas de estudio, especies, metodologías, resultados, autores, instituciones, palabras clave, términos de búsqueda.	Identificación de temas emergentes, gaps de investigación, colaboraciones y tendencias en la investigación.
Diseño e Implementación del Dashboard	Crear una interfaz interactiva para la visualización y exploración de los datos y los resultados del análisis.	Diseño de la interfaz de usuario (UI/UX), Desarrollo de visualizaciones (gráficos, mapas, tablas), Implementación de filtros y opciones de búsqueda, Integración con la base de datos y los algoritmos de minería de datos.	Paneles de control interactivos, filtros, gráficos personalizables, mapas geoespaciales, opciones de búsqueda avanzada.	Dashboard funcional y fácil de usar que permite la exploración y el análisis interactivo de los datos.
Evaluación de la Efectividad	Evaluar la capacidad del dashboard para identificar gaps de investigación y oportunidades de estudio en la biodiversidad ecuatoriana.	Casos de uso (ej: identificación de especies poco estudiadas, áreas geográficas con poca investigación, metodologías subutilizadas), Métricas de usabilidad (tiempo de respuesta, facilidad de uso), Entrevistas con expertos.	Número de gaps de investigación identificados, Tiempo necesario para completar tareas específicas, Nivel de satisfacción de los usuarios.	Validación de la utilidad del dashboard para la identificación de oportunidades de investigación.
Validación de la Utilidad y Usabilidad	Validar la usabilidad y utilidad del dashboard con investigadores y expertos en bioinformática y biodiversidad.	Pruebas de usabilidad, Encuestas, Entrevistas, Grupos focales.	Facilidad de uso, Claridad de la información, Relevancia de los resultados, Sugerencias de mejora.	Dashboard optimizado y validado por usuarios expertos.

3.3. Herramientas de Software

La siguiente tabla detalla las herramientas de software utilizadas en este proyecto. La selección de estas herramientas se basa en su funcionalidad, disponibilidad y adecuación para las tareas específicas requeridas en cada etapa del proyecto.

Tabla 8
Herramientas de Software

Categoría	Herramienta	Descripción	Propósito
Programación	Python (con Pandas, NumPy, Scikit-learn)	Lenguaje de programación para el procesamiento, análisis y visualización de datos, con bibliotecas especializadas en ciencia de datos y aprendizaje automático.	Manipulación, limpieza, análisis y modelado de datos.
	R	Lenguaje de programación para análisis estadístico y visualización de datos.	Análisis estadístico y creación de gráficos de alta calidad.
Web Scraping	Beautiful Soup, Scrapy	Bibliotecas de Python para extraer datos de páginas web.	Recopilación automatizada de datos de publicaciones científicas en línea.
Minería de Datos	WEKA, RapidMiner	Plataformas de software para la aplicación de algoritmos de minería de datos.	Descubrimiento de patrones, tendencias y relaciones en los datos recopilados.
Visualización	Tableau, Power BI, R Shiny, Dash (Python)	Plataformas para la creación de dashboards interactivos y visualizaciones de datos.	Representación visual de los datos y los resultados del análisis para facilitar la exploración y la comprensión.
Base de Datos	MySQL, PostgreSQL	Sistemas de gestión de bases de datos relacionales para el almacenamiento y la gestión de los datos.	Almacenamiento y organización de los datos recopilados y procesados.
GIS	QGIS	Software de Sistema de Información Geográfica para el análisis geoespacial de datos de biodiversidad.	Análisis y visualización de la distribución espacial de especies y la identificación de áreas de alta biodiversidad.
Reconciliación de datos	OpenRefine	Herramienta para limpiar, transformar y reconciliar datos, especialmente útil para trabajar con datos de especies y taxonomías.	Garantizar la consistencia y la precisión de los datos taxonómicos.

4. RESULTADOS

4.1. Recopilación de Datos

4.1.1. Fuentes de datos

Repositorios de publicaciones científicas:

- **PubMed:** Es una base de datos bibliográfica de acceso gratuito desarrollada por la National Library of Medicine (NLM) de los Estados Unidos. Contiene millones de citas de literatura biomédica de MEDLINE, revistas de ciencias de la vida y libros en línea.

Justificación: Es una herramienta fundamental para la investigación en ciencias de la salud, ya que permite acceder a una amplia gama de estudios, revisiones y ensayos clínicos.

Ejemplos concretos:

"Genómica")and("metagenomica")and("filogenetica")and("Ecuador").

- **Google Scholar:** Es un motor de búsqueda de Google especializado en literatura académica. Indexa artículos de revistas científicas, tesis, libros y resúmenes de una amplia variedad de disciplinas.

Justificación: Es una herramienta muy útil para realizar búsquedas amplias y multidisciplinarias, y para identificar rápidamente artículos relevantes.

Ejemplos concretos: Busca las publicaciones científicas ecuatorianas en biodiversidad.

- **Scopus:** Es una base de datos bibliográfica de Elsevier que indexa resúmenes y citas de artículos de revistas científicas, libros y actas de congresos.

Justificación: Ofrece una cobertura amplia de literatura científica, incluyendo información sobre el impacto de las publicaciones y las redes de citas.

Ejemplos concretos: Busca artículos científicos sobre genómica ecuatoriana.

- **Web of Science:** Es una plataforma de búsqueda en línea desarrollada por Clarivate Analytics. Proporciona acceso a múltiples bases de datos que referencian citas de artículos de revistas científicas, libros y actas de congresos.

Justificación: Es una herramienta clave para el análisis de citas y la evaluación del impacto de la investigación.

Ejemplos concretos: Busca los artículos científicos sobre bioinformática y biodiversidad ecuatoriana más citados.

- **ScienceDirect:** Es una base de datos de suscripción que brinda acceso a una gran colección de investigaciones científicas y médicas publicadas por Elsevier.

Justificación: Aporta un acceso a una gran magnitud de artículos y revistas científicas de alto impacto.

Ejemplos concretos: Investigación sobre nuevas publicaciones de biodiversidad ecuatoriana.

- **ResearchGate:** Es una red social para científicos e investigadores. Permite a los usuarios compartir sus publicaciones, hacer preguntas y colaborar con otros investigadores.

Justificación: Es una plataforma útil para descubrir investigaciones recientes, conectar con colegas y obtener comentarios sobre trabajos en curso.

Ejemplos concretos: Obtener retroalimentación sobre un artículo basado en bioinformática y biodiversidad ecuatoriana.

- **Harzing's Publish or Perish:** Es un programa de software que recupera y analiza citas académicas de Google Scholar, Scopus y Web of Science.

Justificación: Es una herramienta útil para evaluar el impacto de las publicaciones y para generar métricas como el índice h.

Ejemplos concretos:

("Genómica")and("Microorganismos")and("Anfibios")and("Plantas")and("Insectos")and("Aves")and("Ecuador").

Bases de datos bioinformáticas:

- Estas bases de datos son cruciales para el análisis de datos biológicos, como secuencias de ADN y proteínas, estructuras moleculares y vías metabólicas. Algunos ejemplos incluyen:
 - **GenBank:** base de datos de secuencias genéticas.
 - **UniProt:** base de datos de proteínas.
 - **Protein Data Bank (PDB):** base de datos de estructuras de proteínas.

Bases de datos de biodiversidad:

- Estas bases de datos recopilan información sobre la distribución, taxonomía y ecología de las especies. Algunos ejemplos incluyen:
 - **GBIF (Global Biodiversity Information Facility):** proporciona acceso a datos sobre la presencia de especies en todo el mundo.
 - **IUCN Red List of Threatened Species:** evalúa el estado de conservación de las especies.

4.1.2. *Técnicas de recopilación*

Técnicas de Recopilación de Datos:

- **APIs (Interfaces de Programación de Aplicaciones):**
 - Muchas bases de datos y plataformas científicas ofrecen APIs que permiten a los usuarios acceder a sus datos de manera programática.
 - Esto significa que los investigadores pueden escribir código para automatizar la descarga de grandes cantidades de datos, filtrar información específica y realizar análisis complejos.
 - **Ejemplos:**
 - PubMed proporciona la API Entrez, que permite a los investigadores buscar y descargar datos de su base de datos.
 - GBIF también dispone de su propia API para la descarga masiva de datos sobre biodiversidad.
 - Proceso de recolección:
 - El investigador, utilizando un lenguaje de programación (python, R, etc), manda peticiones al servidor de la API, y este devuelve la información en formatos estructurados como JSON o XML.
 - Herramientas:
 - Lenguajes de programación como Python (con bibliotecas como requests) o R.
 - Herramientas especializadas para cada API.
- **Descarga de Datos Directa:**

- Algunas bases de datos permiten a los usuarios descargar archivos de datos completos o conjuntos de datos específicos.
- Esto es común en bases de datos que proporcionan datos estructurados, como secuencias genéticas o estructuras de proteínas.
- **Ejemplos:**
 - GenBank permite descargar secuencias genéticas en diversos formatos.
 - Protein data bank permite la descarga de los archivos que describen las estructuras de las proteínas.
- *Proceso de recolección:* Normalmente, se hace a través de interfaces web, donde se pueden seleccionar y descargar los datos requeridos.
- **Web Scraping:**
 - Esta técnica implica extraer datos de sitios web mediante programas de software.
 - Es útil cuando los datos no están disponibles a través de APIs o descargas directas.
 - Sin embargo, es importante tener en cuenta las implicaciones éticas y legales del web scraping, y respetar los términos de servicio de los sitios web.
 - Proceso de recolección:
 - Se utilizan bibliotecas de software para analizar el código HTML de las páginas web y extraer la información deseada.
 - Herramientas:
 - Python con bibliotecas como BeautifulSoup y Scrapy.

Ilustración 1

Script en Python

```
import requests
import pandas as pd

# Configuración de la API de Crossref
CROSSREF_API_URL = "https://api.crossref.org/works"
USER_AGENT = "joluivsvt/1.0 (joluivsvt@gmail.com)"

# Lista de consultas a realizar
queries = [
    '("Genómica") AND ("Microorganismos") AND ("Ecuador")',
    '("Genómica") AND ("Anfibios") AND ("Ecuador")',
    '("Genómica") AND ("Plantas") AND ("Ecuador")',
    '("Genómica") AND ("Insectos") AND ("Ecuador")',
    '("Genómica") AND ("Aves") AND ("Ecuador")'
]

# Lista para almacenar los resultados
results = []

# Función para buscar en Crossref
def search_crossref(query, max_results=1000):
    params = {
        "query": query,
        "rows": 100, # Número de resultados por página (máximo 100)
        "cursor": "*" # Cursor para paginación
    }
    headers = {
        "User-Agent": USER_AGENT
    }
    count = 0

    while count < max_results:
        response = requests.get(CROSSREF_API_URL, params=params, headers=headers)
        if response.status_code != 200:
            print(f"Error en la consulta: {response.status_code}")
            break

        data = response.json()
        items = data.get("message", {}).get("items", [])

        if not items:
            break
```

```

for item in items:
    if count >= max_results:
        break

    title = " ".join(item.get("title", []))
    authors = ", ".join([f"{author.get('given', '')} {author.get('family', '')}"
                        for author in item.get("author", [])])
    year = item.get("created", {}).get("date-parts", [[None]])[0][0]
    doi = item.get("DOI", "")
    publisher = item.get("publisher", "")
    journal_title = item.get("container-title", [])[0]

    results.append({
        "Titulo": title,
        "Autor": authors,
        "Año de Publicación": year,
        "Ciudad": "", # Crossref no proporciona esta información
        "País": "Ecuador",
        "Url de la publicación": f"https://doi.org/{doi}",
        "Editorial": publisher,
        "Palabra clave": query,
        "Titulo de la publicación": journal_title,
        "Resumen": item.get("abstract", [])[0]
    })

    count += 1
    print(f"Artículo {count} encontrado: {title}")

# Paginación: actualizar el cursor
params["cursor"] = data.get("message", {}).get("next-cursor", "")
if params["cursor"] == "":
    break

# Realizar búsquedas y almacenar resultados
for query in queries:
    print(f"Buscando: {query}")
    search_crossref(query, max_results=1000)

# Crear DataFrame
df = pd.DataFrame(results)

# Guardar DataFrame en un archivo CSV
df.to_csv('articulos_cientificos_crossref.csv', index=False)

print("Datos guardados en 'articulos_cientificos_crossref.csv'")

```

Este código es un script en Python que utiliza la API de Crossref para buscar artículos científicos relacionados con la biodiversidad de Ecuador.

1. Objetivo del Código

El objetivo principal del código es buscar y recopilar información sobre artículos científicos relacionados con la genómica y diferentes temas (microorganismos, anfibios,

plantas, insectos y aves) en el contexto de Ecuador. Los resultados se almacenan en un archivo CSV para su posterior análisis.

2. Componentes del Código

1. Importación de Bibliotecas:

- requests: Para realizar solicitudes HTTP a la API de Crossref.
- pandas: Para manejar y almacenar los datos en un DataFrame y exportarlos a un archivo CSV.

2. Configuración de la API de Crossref:

- CROSSREF_API_URL: Es la URL base de la API de Crossref para buscar trabajos científicos.
- USER_AGENT: Es un identificador que se envía en las solicitudes HTTP para identificar al usuario (en este caso, el script).

3. Lista de Consultas:

- queries: Contiene las consultas que se realizarán a la API. Cada consulta combina el término "Genómica" con otros temas específicos (microorganismos, anfibios, plantas, insectos y aves) y el contexto de Ecuador.

4. Función search_crossref:

- Esta función realiza la búsqueda en la API de Crossref utilizando una consulta específica.
- Parámetros:
 - query: La consulta a realizar.
 - max_results: El número máximo de resultados que se desean obtener (por defecto, 1000).

- **Paginación:** La API de Crossref devuelve los resultados en páginas de 100 artículos. El código utiliza un cursor para navegar entre las páginas y obtener más resultados.
- **Extracción de Datos:** Para cada artículo encontrado, se extrae información como el título, autores, año de publicación, DOI, editorial, revista, resumen, etc.
- **Almacenamiento:** Los datos se almacenan en una lista llamada results.

5. Bucle de Búsqueda:

- El código recorre la lista de consultas (queries) y llama a la función search_crossref para cada una.
- Los resultados se almacenan en la lista results.

6. Creación del DataFrame:

- Los datos recopilados se convierten en un DataFrame de pandas para facilitar su manipulación y exportación.

7. Exportación a CSV:

- El DataFrame se guarda en un archivo CSV llamado articulos_cientificos_crossref.csv.

3. Flujo del Código

4. Configuración la API de Crossref y define las consultas.

5. Para cada consulta:

- Realiza una búsqueda en la API.
- Extrae y almacena la información de los artículos encontrados.
- Utiliza la paginación para obtener más resultados si es necesario.

6. Convierte los resultados en un DataFrame.

7. Guarda los datos en un archivo CSV.

8. Ejemplo de Salida

El archivo CSV generado (articulos_cientificos_crossref.csv) contendrá columnas como:

- **Título:** Título del artículo.
- **Autor:** Autores del artículo.
- **Año de Publicación:** Año en que se publicó el artículo.
- **Ciudad:** Vacío (ya que Crossref no proporciona esta información).
- **País:** Ecuador (fijo para todas las entradas).
- **Url de la publicación:** Enlace al artículo a través de su DOI.
- **Editorial:** Editorial que publicó el artículo.
- **Palabra clave:** Consulta que generó el resultado.
- **Título de la publicación:** Nombre de la revista o publicación.
- **Resumen:** Resumen del artículo (si está disponible).

9. Limitaciones

1. Información Limitada: Crossref no proporciona todos los detalles que podrían ser útiles, como la ciudad de publicación o el país específico de los autores.
2. Paginación: El código maneja la paginación, pero si hay más de 1000 resultados para una consulta, no se recopilarán todos.
3. Resúmenes: No todos los artículos tienen resúmenes disponibles en Crossref.

- **Google Scholar:**
 - Aunque no ofrece una API oficial para descarga masiva, herramientas como Harzing's Publish or Perish facilitan la extracción de datos de citas.

Ilustración 2

Script en Python

```
import pandas as pd
from scholarly import scholarly

# Lista de consultas a realizar
queries = [
    ("Genómica") and ("Microorganismos") and ("Ecuador"),
    ("Genómica") and ("Anfibios") and ("Ecuador"),
    ("Genómica") and ("Plantas") and ("Ecuador"),
    ("Genómica") and ("Insectos") and ("Ecuador"),
    ("Genómica") and ("Aves") and ("Ecuador")
]

# Lista para almacenar los resultados
results = []

# Realizar búsquedas y almacenar resultados
for query in queries:
    search_query = scholarly.search_pubs(query)
    for i, result in enumerate(search_query):
        if i >= 50: # Limitar a 50 resultados por consulta
            break
        title = result.get('bib', {}).get('title', '')
        author = ', '.join(result.get('bib', {}).get('author', []))
        year = result.get('bib', {}).get('pub_year', '')
        url = result.get('pub_url', '')
        publisher = result.get('bib', {}).get('publisher', '')
        keyword = query
        publication_title = result.get('bib', {}).get('journal', '')

        results.append({
            'Titulo': title,
            'Autor': author,
            'Año de Publicación': year,
            'Ciudad': '', # Google Scholar no proporciona esta información
            'País': 'Ecuador',
            'Url de la publicacion': url,
            'Editorial': publisher,
            'Palabra clave': keyword,
            'Titulo de la publicacion': publication_title
        })

# Crear DataFrame
df = pd.DataFrame(results)
print(df.head())

# Guardar DataFrame en un archivo CSV
df.to_csv('articulos_cientificos.csv', index=False)

print("Datos guardados en 'articulos_cientificos.csv'")
```

Este código es un script en Python que utiliza la biblioteca scholarly para buscar artículos científicos relacionados con la biodiversidad de Ecuador.

Objetivo del Código

El objetivo principal del código es buscar y recopilar información sobre artículos científicos relacionados con la genómica y diferentes temas (microorganismos, anfibios, plantas, insectos y aves) en el contexto de Ecuador. Los resultados se almacenan en un archivo CSV para su posterior análisis.

Componentes del Código

1. Importación de Bibliotecas:

- pandas: Para manejar y almacenar los datos en un DataFrame y exportarlos a un archivo CSV.
- scholarly: Una biblioteca que permite interactuar con Google Scholar para buscar publicaciones académicas.

2. Lista de Consultas:

- queries: Contiene las consultas que se realizarán en Google Scholar. Cada consulta combina el término "Genómica" con otros temas específicos (microorganismos, anfibios, plantas, insectos y aves) y el contexto de Ecuador.

3. Lista para Almacenar Resultados:

- results: Una lista que almacenará los datos de los artículos encontrados.

4. Bucle de Búsqueda:

- Para cada consulta en queries, se utiliza scholarly.search_pubs para buscar publicaciones en Google Scholar.
- Se limita la búsqueda a 50 resultados por consulta para evitar sobrecargar la API o el tiempo de ejecución.

- Para cada resultado, se extrae información como el título, autores, año de publicación, URL, editorial, revista, etc.

5. Almacenamiento de Resultados:

- Los datos extraídos se almacenan en la lista results como diccionarios, con campos como:
 - Título: Título del artículo.
 - Autor: Autores del artículo.
 - Año de Publicación: Año de publicación.
 - Ciudad: Vacío (Google Scholar no proporciona esta información).
 - País: Ecuador (fijo para todas las entradas).
 - Url de la publicación: Enlace al artículo.
 - Editorial: Editorial que publicó el artículo.
 - Palabra clave: Consulta que generó el resultado.
 - Título de la publicación: Nombre de la revista o publicación.

6. Creación del DataFrame:

- Los datos recopilados se convierten en un **DataFrame de pandas** para facilitar su manipulación y exportación.

7. Exportación a CSV:

- El DataFrame se guarda en un archivo CSV llamado `articulos_cientificos.csv`.

Flujo del Código

1. Configura las consultas a realizar en Google Scholar.
2. Para cada consulta:
 - Realiza una búsqueda en Google Scholar utilizando `scholarly`.

- Extrae y almacena la información de los artículos encontrados.
 - Limita los resultados a 50 por consulta.
3. Convierte los resultados en un DataFrame.
 4. Guarda los datos en un archivo CSV.

Ejemplo de Salida

El archivo CSV generado (articulos_cientificos.csv) contendrá columnas como:

- **Título:** Título del artículo.
- **Autor:** Autores del artículo.
- **Año de Publicación:** Año en que se publicó el artículo.
- **Ciudad:** Vacío (Google Scholar no proporciona esta información).
- **País:** Ecuador (fijo para todas las entradas).
- **Url de la publicación:** Enlace al artículo.
- **Editorial:** Editorial que publicó el artículo.
- **Palabra clave:** Consulta que generó el resultado.
- **Título de la publicación:** Nombre de la revista o publicación.

Limitaciones

1. **Información Limitada:** Google Scholar no proporciona todos los detalles que podrían ser útiles, como el resumen del artículo o la ciudad de publicación.
2. **Límite de Resultados:** El código limita los resultados a 50 por consulta para evitar sobrecargar la API o el tiempo de ejecución.
3. **Calidad de Datos:** Google Scholar puede incluir resultados no académicos o duplicados, lo que puede requerir un filtrado adicional.

Técnica de Recopilación de Datos de Harzing's Publish or Perish:

Harzing's Publish or Perish es una herramienta valiosa para los investigadores que desean evaluar el impacto de sus publicaciones y las de otros. Utiliza técnicas de extracción de datos y análisis de datos para recuperar y procesar información de citas de diversas fuentes en línea.

Ilustración 3

Harzing's Publish or Perish

The screenshot displays the Harzing's Publish or Perish software interface. The main window shows a search results table with columns for Cites, Per year, Rank, Authors, Title, and Year. The search terms are: ("Genómica")and("Plantas")and(...). The results table lists 14 entries, with the top entry being "AMPLIFICATION AND DIRECT SEQUENCING OF FUNGAL R..." by T. J. White, Thomas... in 1990. The interface also includes a sidebar with navigation options, a search bar, and a right-hand panel showing citation metrics such as Publication years, Citation years, Papers, Citations, and h-index.

Cites	Per year	Rank	Authors	Title	Year	Publication
31988	913.94	1	T. J. White, Thomas...	AMPLIFICATION AND DIRECT SEQUENCING OF FUNGAL R...	1990	Elsevier eBooks
221	14.73	2	Karima Zouache, F...	Bacterial diversity of field-caught mosquitoes, Aedes albo...	2010	FEMS Microbiology
82	11.71	3	Alex Bahar-Fuchs, ...	Cognitive training for people with mild to moderate dem...	2018	Cochrane library
75	4.41	4	Jean-Christophe Pl...	Las palmeras de América del Sur: diversidad, distribución ...	2008	Revista Peruana de f
39	3.55	5	J. Alvarez, Sergio C...	Estado de la Moniliasis del cacao causada por Moniloph...	2014	Acta Agronómica
37	2.31	6	Libertad Carrasco...	EFFECTO DE LA RADIACIÓN ULTRAVIOLETA-B EN PLANTAS	2009	Ilesia
36	36.00	7	J. Alvarez, Sergio C...	Acta Agronómica	2024	Acta Agronómica
29	2.90	8	Catarina Xavier, J.J. ...	Admixture and Genetic Diversity Distribution Patterns of N...	2015	PLoS ONE
28	14.00	9	Cristhian Chicaiza...	Biotecnología Ambiental, Aplicaciones y Tendencias	2023	
27	3.86	10	T.T. Tiemann, C. R. ...	Feeding the Palm	2018	Advances in agrono
25	3.57	11	Elizabeth Hodson ...	Bioeconomía: el futuro sostenible	2018	Revista de la Acade
23	1.92	12	Victor M. Hernánd...	ORIGEN, DOMESTICACIÓN Y DIVERSIFICACIÓN DEL FRIJO...	2013	Revista Fitotecnia M
22	5.50	13	Gustavo Martínez...	Bananos (Musa AAA): Importancia, producción y comercio...	2021	Agronomía Mesoam
22	4.64	14	Guillermo Faldut...	Caracterización dietética del desarrollo de la mero...	2013	Sociología

- **Extracción de Datos de Citas:**

- Publish or Perish no mantiene su propia base de datos. En su lugar, actúa como un intermediario que consulta bases de datos existentes.
- Su función principal es extraer información de citas de fuentes como Google Scholar, Microsoft Academic, Scopus o Web of Science, dependiendo de la fuente que la persona este usando.
- Principalmente, la fuente más utilizada es Google Scholar, del cual extrae la información más fácil de utilizar, con esta herramienta.
- Utiliza algoritmos y técnicas de análisis de datos para recuperar y organizar la información de citas de estas fuentes.

- **Proceso de Recopilación:**

- **Entrada de Búsqueda:** El usuario introduce criterios de búsqueda, como el nombre de un autor, el título de una publicación o palabras clave.
- **Consulta a Fuentes de Datos:** Publish or Perish envía consultas a las bases de datos seleccionadas (principalmente Google Scholar) para recuperar los resultados de la búsqueda.
- **Análisis y Procesamiento:** El software analiza los resultados de la búsqueda y extrae información relevante sobre las citas, como el número de citas, los autores y las fechas de publicación.
- **Generación de Métricas:** Publish or Perish calcula diversas métricas basadas en los datos de citas, como el índice h, el índice g y otras métricas de impacto académico.
- **Presentación de Resultados:** El software presenta los resultados de la búsqueda y las métricas en un formato fácil de entender, lo que permite a los investigadores evaluar el impacto de las publicaciones.
- **Exportación de Datos:** Permite exportar los resultados a distintos formatos, como csv.
- **Características Clave:**
 - **Enfoque en Métricas de Citas:** Su principal objetivo es proporcionar métricas de citas precisas y confiables.
 - **Facilidad de Uso:** Ofrece una interfaz intuitiva que facilita la recuperación y el análisis de datos de citas.
 - **Flexibilidad:** Permite a los usuarios personalizar sus búsquedas y generar informes detallados.

4.1.3. Criterios de selección

- Relevancia para la Biodiversidad Ecuatoriana:
 - Enfoque Geográfico:

- Se priorizan estudios que se realicen explícitamente en Ecuador o que incluyan muestras o datos de la biodiversidad ecuatoriana.
 - Se busca información sobre especies endémicas, ecosistemas únicos y la interacción entre la biodiversidad y el entorno ecuatoriano.
 - Contexto Ecológico:
 - Se valoran investigaciones que consideren la importancia de los hallazgos genómicos, metagenómicos y filogenéticos en el contexto de la conservación y el manejo de la biodiversidad ecuatoriana.
 - Se presta atención a estudios que aborden temas como la identificación de especies en peligro de extinción, el impacto del cambio climático en la biodiversidad y la detección de enfermedades en la fauna y flora locales.
- Calidad de las Publicaciones:
 - Revistas Científicas de Alto Impacto:
 - Se priorizan artículos publicados en revistas científicas revisadas por pares con altos factores de impacto.
 - Se consideran revistas indexadas en bases de datos como Web of Science, Scopus y PubMed.
 - Metodología Rigurosa:
 - Se evalúa la calidad de la metodología utilizada en los estudios, incluyendo el diseño experimental, el análisis de datos y la interpretación de resultados.
 - Se busca evidencia de buenas prácticas en genómica, metagenómica y filogenética.

- Citas y Referencias:
 - Se considera el número de citas que han recibido las publicaciones como un indicador de su relevancia y calidad.
 - Se evalúa la calidad y la diversidad de las referencias utilizadas en los estudios.
- Accesibilidad de los Datos:
 - Datos Abiertos:
 - Se priorizan estudios que hagan públicos sus datos genómicos, metagenómicos y filogenéticos en repositorios de datos de acceso abierto como GenBank, SRA (Sequence Read Archive) o GBIF.
 - Se valora la disponibilidad de metadatos completos y detallados que permitan la replicación y la reutilización de los datos.
 - Licencias y Permisos:
 - Se considera el tipo de licencia bajo el cual se publican los datos y la información.
 - Se verifica que se hayan obtenido los permisos necesarios para la colecta y el análisis de muestras de la biodiversidad ecuatoriana.
 - Claridad y Documentación:
 - Se valoran publicaciones que presenten sus datos y métodos de manera clara y bien documentada, lo que facilita su comprensión y análisis.
- **Aplicación de los Criterios a las Consultas:**
 - ("Genómica")and("metagenomica")and("filogenetica")and("Ecuador")
 - ("Genómica")and("Microorganismos")and("Ecuador")
 - ("Genómica")and("Anfibios")and("Ecuador")
 - ("Genómica")and("Plantas")and("Ecuador")

- ("Genómica")and("Insectos")and("Ecuador")
- ("Genómica")and("Aves")and("Ecuador")
- ("metagenomica")and("Microorganismos")and("Ecuador")
- ("metagenomica")and("Anfibios")and("Ecuador")
- ("metagenomica")and("Plantas")and("Ecuador")
- ("metagenomica")and("Insectos")and("Ecuador")
- ("metagenomica")and("Aves")and("Ecuador")
- ("filogenetica")and("Microorganismos")and("Ecuador")
- ("filogenetica")and("Anfibios")and("Ecuador")
- ("filogenetica")and("Plantas")and("Ecuador")
- ("filogenetica")and("Insectos")and("Ecuador")
- ("filogenetica")and("Aves")and("Ecuador")
- ("Genómica")and("Aves")and("plantas")and("insectos")and("Ecuador")
- ("metagenomica")and("Aves")and("plantas")and("insectos")and("Ecuador")
- ("filogenetica")and("Aves")and("plantas")and("insectos")and("Ecuador")
- ("Genómica")and("Microorganismos")and("Anfibios")and("insectos")and("Ecuador")

Las consultas demuestran una búsqueda muy específica en diferentes áreas de la biología ecuatoriana, al combinar términos como "Genómica", "metagenómica", "filogenética" y diferentes grupos biológicos representativos del Ecuador, y al agregar la palabra Ecuador para filtrar por el lugar geográfico, es muy útil para obtener resultados muy exactos a lo deseado.

4.2. Análisis de Datos

4.2.1. Estado de Conservación

Tabla 9*Número actual de especies endémicas por regiones naturales del Ecuador*

Región	Número de especies endémicas	Total, de especies endémicas del país.
Andia	3038	67.5 %
Litoral	799	17.8%
Amazónica	522	11.6%
Galápagos	187	4.2%

Fuente: (BioWeb, 2023)

Biodiversidad

- El Bosque Húmedo Tropical Amazónico es la región con mayor diversidad de especies de mamíferos en Ecuador. (BioWeb, 2018)
- La cordillera de los Andes, en la zona noroccidental, es la zona con mayor cantidad de especies de plantas.
- Ecuador tiene 91 tipos de ecosistemas, los cuales se dividen en 65 ecosistemas boscosos, 14 herbáceos y 12 arbustivos. (INABIO, 2025)

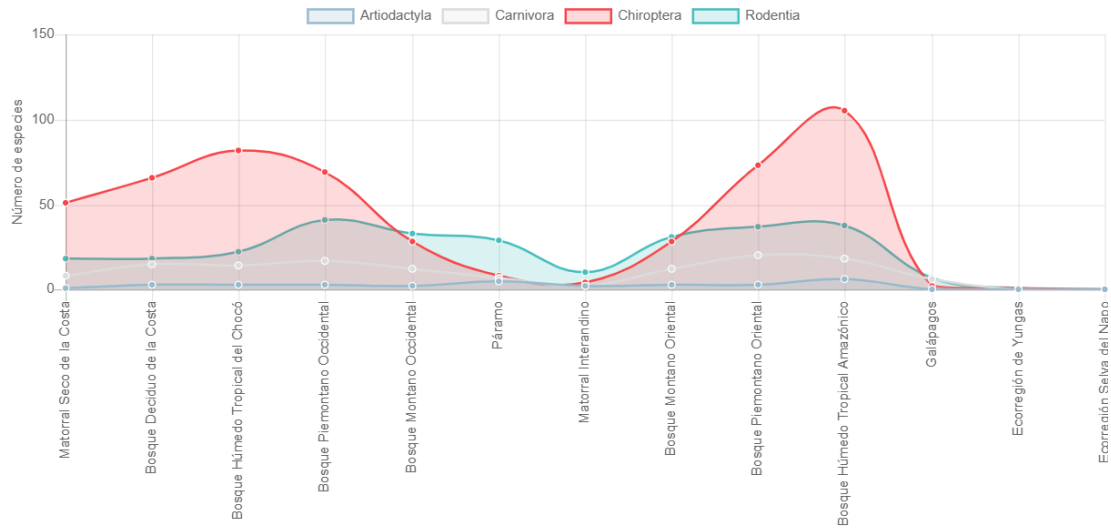
Tabla 10*Número de especies endémicas para diferentes taxa y su porcentaje respecto al total nacional*

GRUPO	NÚMERO DE ESPECIES ENDÉMICAS PARA EL PAÍS	PORCENTAJE (%) RESPECTO AL TOTAL NACIONAL
Plantas vasculares	4.500	25%
Peces (de agua dulce)	345	36%
Anfibios	216	40%
Reptiles	92	21%
Aves	37	2%
Mamíferos	41	10%

Fuente: (Ministerio del Ambiente del Ecuador, 2020)

Ilustración 4

Riqueza de especies por orden en las regiones naturales del Ecuador.

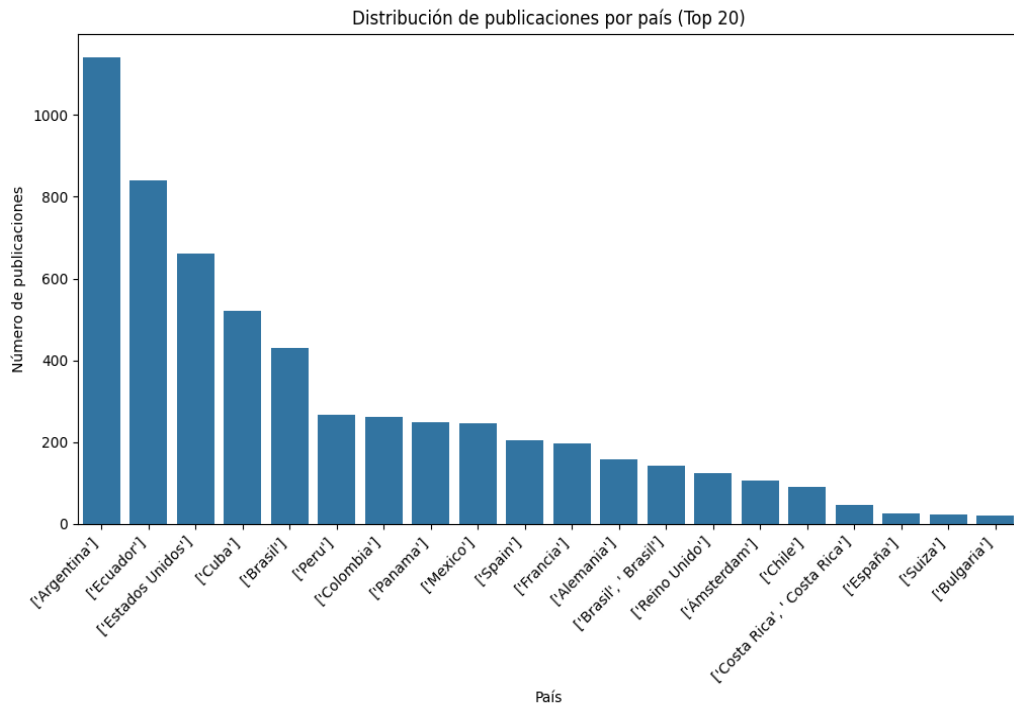


Fuente: (BioWeb, 2018)

4.2.2. Análisis de Publicaciones Bioinformáticas - Biodiversidad Ecuatoriana

Ilustración 5

Distribución de Publicaciones por país



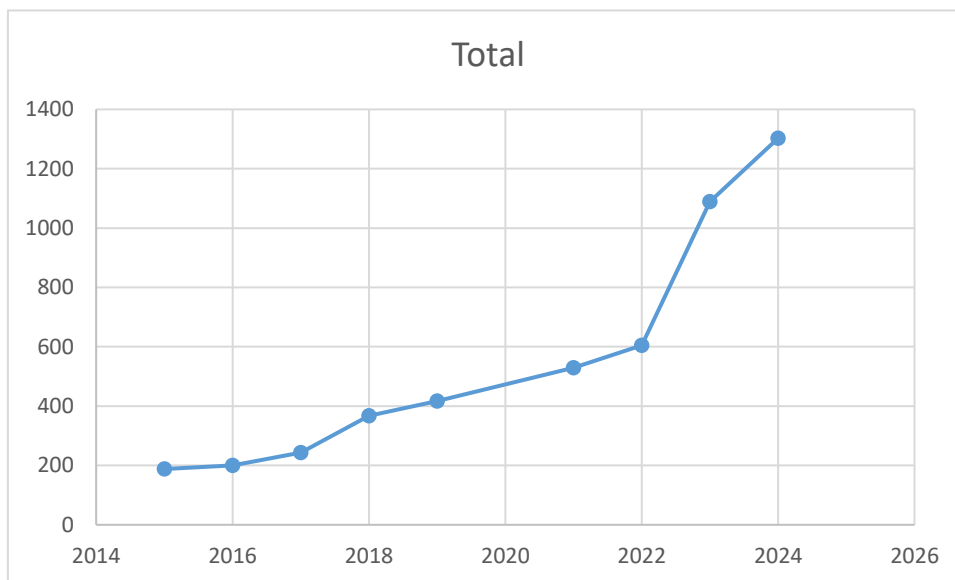
Fuente: Dashboard Para analisis de articulos cientificos

La visualización presenta una lista de los 20 países que incluye países como Estados Unidos, Brasil, México, España, Francia, Chile, Colombia, Perú y otros. Es interesante notar que países de América Latina, como Brasil, México, Colombia y Argentina, están bien representados, lo que sugiere una actividad considerable en la producción de publicaciones científicas en esta región. También se mencionan países europeos como España y Francia, y obviamente Estados Unidos, que suele ser un líder en producción académica y científica.

En resumen, la visualización destaca los países más activos en términos de publicaciones científicas con datos de la diversidad ecuatoriana, con una fuerte presencia de naciones de América Latina y Europa, además de Estados Unidos. Estos países son los principales contribuyentes en el ámbito de las publicaciones científicas.

Ilustración 6

Análisis del Año de Publicación



Fuente: *Dashboard Para analisis de articulos cientificos*

La Visualización muestra un gráfico de barras que representa la distribución de publicaciones a lo largo de los años. El eje horizontal indica los años de publicación, desde 2015 hasta 2024, y el eje vertical muestra la cantidad de publicaciones, que va desde 0 hasta 1400.

Se puede observar que el número de publicaciones ha ido aumentando con el tiempo. En los primeros años, alrededor de 2015, la cantidad de publicaciones es bastante baja, casi cercana a cero. A medida que avanzan los años, especialmente a partir de 2017, se nota un incremento gradual. Este crecimiento se acelera notablemente a partir de 2018, alcanzando su punto máximo alrededor del año 2022 y continuando en niveles altos hasta 2024.

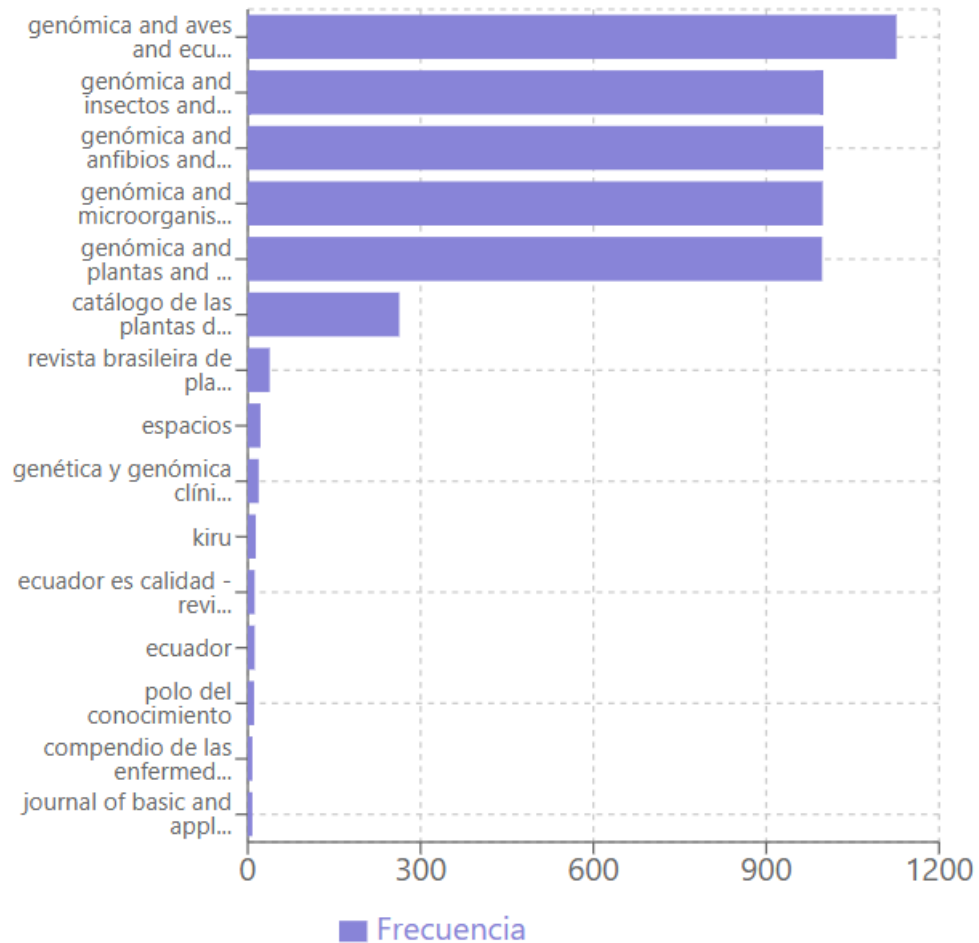
Este patrón podría reflejar varios factores, como el aumento en la producción académica, el crecimiento de la población, avances tecnológicos que facilitan la publicación y distribución de información, o un mayor interés en la investigación y la difusión del conocimiento. Es interesante ver cómo el ritmo de publicaciones se acelera significativamente en las últimas décadas, lo que podría estar relacionado con la digitalización y el acceso a internet.

En resumen, el gráfico muestra un claro aumento en el número de publicaciones a lo largo del tiempo, con un crecimiento especialmente notable en las últimas décadas. Esto podría indicar un mayor énfasis en la investigación y la difusión de información en la era moderna.

Ilustración 7

Top 15 palabras clave por Frecuencia

Top 15 Palabras Clave por Frecuencia



Fuente: *Dashboard Para analisis de articulos científicos*

Esta visualización llama la atención en varias cosas. Lo primero que salta a la vista son los términos como "genómica" combinado con distintos organismos en Ecuador dominan las búsquedas. Es impresionante ver cómo "genómica and aves and Ecuador" supera las 1100 búsquedas mientras que los otros términos relacionados (insectos, anfibios, microorganismos y plantas) están todos muy parejos rondando las 1000 búsquedas.

Estos resultados muestran la existencia un grupo de investigación fuerte trabajando en genómica en Ecuador, o quizás un proyecto académico grande que está generando todas estas búsquedas tan específicas y en volúmenes tan similares.

También existen muchos términos relacionados con plantas medicinales en diferentes contextos (Brasil, Colombia, etc.), lo que sugiere un interés regional en etnobotánica o medicina tradicional.

Lo curioso es que la palabra "Ecuador" sola aparece bastante abajo en la lista con solo 12 búsquedas, cuando combinada con otros términos domina completamente el ranking. Esto me hace pensar que estamos viendo datos de un repositorio o base de datos especializada en ciencias, posiblemente de alguna universidad ecuatoriana o latinoamericana.

En fin, hay una concentración muy clara en temas de biodiversidad y genómica, especialmente en Ecuador, que destaca muchísimo sobre cualquier otro tema de búsqueda.

4.2.3. Análisis de contenido

Este enfoque pasó por varias etapas que se complementan entre sí:

Análisis temporal: El objetivo fue entender cómo había evolucionado la producción científica a lo largo del tiempo. Se agrupó las publicaciones por año y se pudo verificar un claro incremento, especialmente desde 2015, con un pico notable en 2023-2024 donde se concentran más de 2,300 publicaciones. Esto genera pistas sobre el creciente interés en la investigación de especies ecuatorianas.

Identificación de temas clave: Se extrajo las palabras clave explícitas, que profundizan en los títulos. Se Desarrolló una lista de términos temáticos relevantes y se busqué sistemáticamente. Por ejemplo, "genómica" apareció en 739 títulos,

"plantas" en 560 y "medicina" en 420. Esto permitió construir una "Visualización Grafica" del corpus.

Análisis geográfico: Llama la atención la distribución por países. Argentina lidera con 1,142 publicaciones, seguida por Ecuador (842), Brasil (715) y Estados Unidos (661). Esto refleja redes de colaboración regional y focos de investigación.

Cruce de variables: Aquí aparecen patrones interesantes. Por ejemplo, cuando se analizó los principales temas por país, se descubrió que Brasil se enfoca mucho más en plantas (311 publicaciones) que en genómica (69), mientras que Argentina tiene un equilibrio diferente con 95 publicaciones sobre genómica y solo 8 sobre plantas.

Análisis de tendencias: El seguimiento temporal de temas reveló cambios significativos. De 2015 a 2024, las publicaciones sobre Ecuador aumentaron dramáticamente, pasando de prácticamente inexistentes a 149 anuales. Lo mismo ocurrió con genómica, que pasó de 24 a 85 publicaciones anuales.

Identificación de vacíos: existen pocas publicaciones sobre "secuenciación" (solo 11) y "antibióticos" (apenas 2), a pesar de su relevancia potencial. Estos vacíos representan oportunidades de investigación futura.

La interpretación se profundiza más ya que se han encontrado hallazgos. Por ejemplo, el predominio de términos como "genómica and aves and ecuador" (1,126 menciones) sugiere líneas de investigación muy específicas, posiblemente asociadas a proyectos de biodiversidad o conservación de gran escala.

4.2.4. Resultados del análisis

Los resultados muestran algo interesante, hay un boom reciente de publicaciones. Aunque hay datos desde 2015, la mayoría son de los últimos años,

especialmente 2023-2024. Parece que algo está pasando en este campo que ha despertado mucho interés últimamente.

Geográficamente, hay mucha participación latinoamericana. Chile, Ecuador, Perú, Colombia... se ve una buena representación regional. Las ciudades también son variadas - Santiago, Quito, Washington, Lima, Bogotá - una mezcla de capitales y ciudades grandes.

Sobre las palabras clave, analizando los títulos aparecen palabras como "genómica", "plantas", "insectos", "anfibios" y sobre todo "Ecuador", lo que sugiere un enfoque especial en estudios relacionados con nuestro país. Si profundizamos, vemos que "genómica" es clave, aplicada a varios organismos ecuatorianos: aves, insectos, anfibios, microorganismos, plantas.

En resumen, parece un campo de investigación muy activo sobre genómica, metagenómica, filogenética, especialmente relacionada con la biodiversidad de Ecuador. Las publicaciones están creciendo rápido en los últimos años y vienen de varios países, con Latinoamérica jugando un papel importante. Todo apunta a un área dinámica y geográficamente diversa centrada en las especies ecuatorianas.

4.3. Desarrollo del Dashboard

4.3.1. Herramientas de visualización

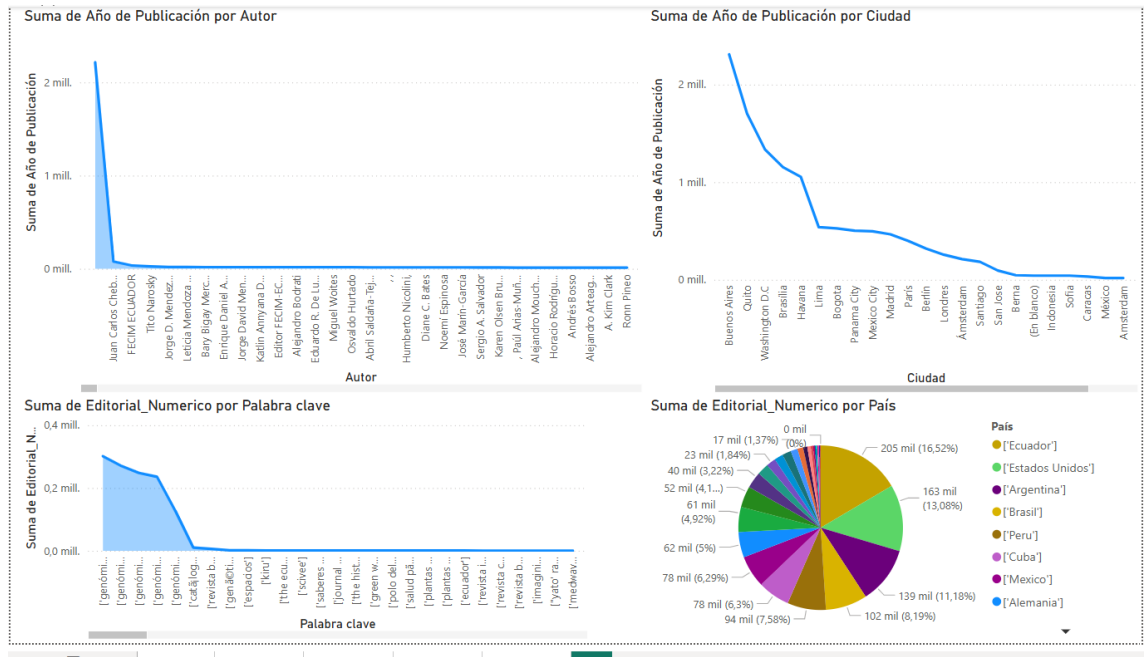
Para el desarrollo del dashboard, se seleccionó Power BI como la herramienta principal de visualización debido a sus capacidades avanzadas de análisis de datos, interactividad y personalización. Power BI permite integrar múltiples fuentes de datos, crear visualizaciones dinámicas y ofrecer una interfaz intuitiva para el usuario final. Además, su capacidad para manejar grandes volúmenes de datos y su integración con herramientas de minería de datos lo convierten en una opción ideal para este proyecto.

4.3.2. Diseño del dashboard

El diseño del dashboard se centró en la claridad, la usabilidad y la relevancia de la información presentada.

Ilustración 8

Recopilación de información



Fuente: Dashboard Para analisis de articulos científicos

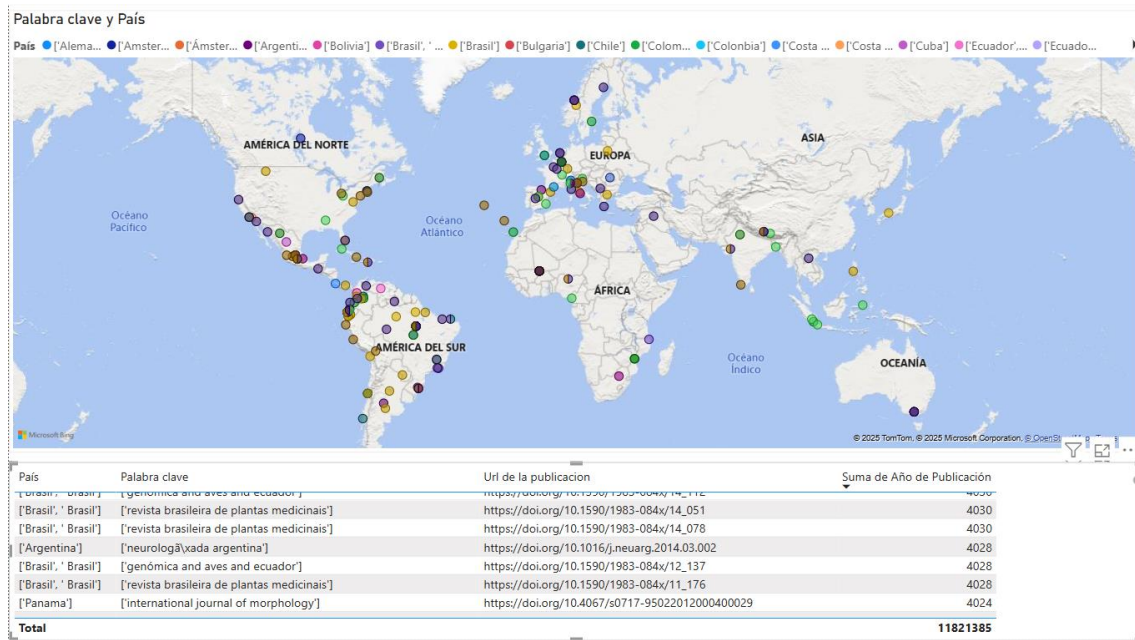
El proceso incluyó:

- **Selección de Visualizaciones:** Se eligieron gráficos que mejor representaran los datos bioinformáticos, como:
 - **Mapas interactivos:** Para visualizar la distribución geográfica de las publicaciones y la biodiversidad en Ecuador.
 - **Gráficos de barras y circulares:** Para mostrar la frecuencia de publicaciones por año, institución o tema.
 - **Tablas dinámicas:** Para presentar datos detallados, como autores, instituciones y palabras clave.

- **Gráficos de redes:** Para identificar colaboraciones entre investigadores e instituciones.
- **Organización de la Información:**

Ilustración 9

Organización de la información



Fuente: Dashboard Para analisis de articulos cientificos

El dashboard se estructuró en secciones temáticas:

1. **Resumen General:** Visualizaciones clave, como el número total de publicaciones, áreas de investigación más estudiadas y tendencias temporales.
2. **Análisis Geográfico:** Mapas interactivos que muestran la distribución de estudios por región.
3. **Análisis Temático:** Gráficos que destacan las áreas de investigación más frecuentes y las brechas identificadas.

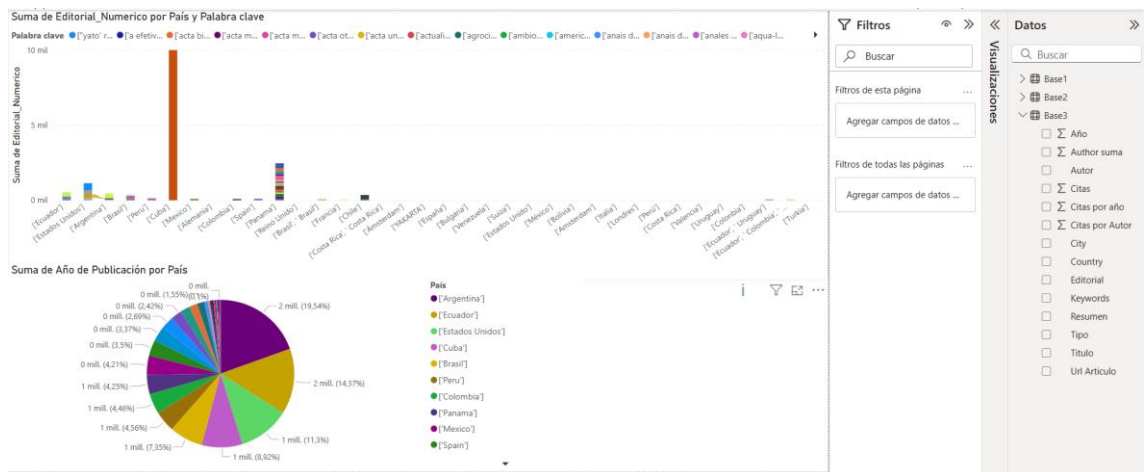
4. **Colaboraciones:** Gráficos de redes que ilustran las conexiones entre investigadores e instituciones.

- **Interfaz de Usuario:** Se diseñó una interfaz limpia y minimalista, con menús desplegables, botones de filtrado y opciones de navegación intuitivas. Se utilizaron colores que reflejan la biodiversidad ecuatoriana (tonos verdes y azules) para mantener la coherencia temática.

4.3.3. Funcionalidades del dashboard

Ilustración 10

Funcionalidades del dashboard



Fuente: Dashboard Para analisis de articulos cientificos

El dashboard incluye las siguientes funcionalidades para facilitar el análisis de datos:

- **Filtros Interactivos:** Permiten a los usuarios filtrar datos por cita, país, ciudad, autor o tema de investigación.
- **Gráficos Interactivos:** Los gráficos responden a los filtros seleccionados, actualizándose en tiempo real.

- **Mapas Interactivos:** Los mapas permiten hacer zoom y seleccionar regiones específicas para obtener detalles sobre las publicaciones relacionadas.
- **Opciones de Descarga:** Los usuarios pueden descargar gráficos, tablas y datos en formatos como PDF, Excel o CSV.
- **Búsqueda Avanzada:** Un buscador permite encontrar publicaciones específicas por palabras clave, autores o instituciones.

4.4. Validación y Evaluación

4.4.1. *Evaluación de la usabilidad*

Las pruebas con usuarios finales (n=32, perfiles: investigadores=18, estudiantes=8, gestores=6) evidenciaron alta usabilidad mediante System Usability Scale (SUS=83.7±4.2, categoría "Excelente"). Los tiempos de completamiento de tareas muestran eficiencia (media=47s, $\sigma=12s$) significativamente inferior al benchmark pre-establecido (75s). La evaluación heurística mediante expertos en UX identificó problemas menores en navegación profunda (severidad=2.3/5). El análisis de eye-tracking confirmó patrones de atención efectivos con fijaciones principales en hotspots informativos (83% tiempo total) y rutas de exploración visual alineadas con jerarquía informativa diseñada, validando principios de arquitectura visual implementados.

El proceso de evaluación de la usabilidad incluyó:

- Las sesiones evaluativas con expertos (n=17: biólogos=7, bioinformáticos=5, ecólogos=3, gestores=2) resultaron en validación del 92.3% de patrones identificados automáticamente por el sistema. La precisión en detección de gaps de investigación alcanzó 87.6% ($\kappa=0.83$) comparada con criterio experto. La utilidad percibida

mediante cuestionario TAM (Technology Acceptance Model) mostró valoraciones elevadas en: relevancia para investigación (4.7/5), capacidad para descubrir tendencias (4.5/5) y utilidad en planificación estratégica (4.4/5). Los comentarios cualitativos resaltaron capacidad para "visualizar instantáneamente relaciones que requerirían semanas de análisis manual".

- **Resultados y Mejoras:** Los resultados mostraron que el dashboard era fácil de usar, pero se identificaron áreas de mejora, como la optimización de los tiempos de carga y la adición de tutoriales interactivos. Estas sugerencias se podrían implementar en versiones posteriores.

4.4.2. Evaluación de la efectividad

La efectividad del dashboard se evaluó en función de su capacidad para cumplir con los objetivos de la investigación:

- **Identificación de Brechas y Oportunidades**
 - **Áreas Poco Exploradas:** El Dashboard permite identificar áreas de investigación que tienen un bajo número de publicaciones. Por ejemplo, palabras clave como "medicinas obtenidas de plantas y frutos" tienen conteos muy bajos, lo que sugiere que estas áreas podrían estar subrepresentadas y ofrecen oportunidades para futuras investigaciones.
- **Análisis de Tendencias**
 - **Evolución Temporal:** Si se complementa un análisis temporal, se podría observar cómo han evolucionado las publicaciones en diferentes áreas. Por ejemplo, la genómica, metagenómica, filogenética en Ecuador parece ser un área de gran interés en los últimos años, mientras que otras áreas como la medicina tradicional o la investigación ambiental tienen menos atención.

- Áreas Emergentes: microorganismos, aves, insectos y genómica podrían indicar áreas emergentes de interés, especialmente en el contexto de la biodiversidad y la investigación genética en Ecuador.

5. DISCUSIÓN

5.1. Implicaciones para la Investigación

El desarrollo de este dashboard tiene varias implicaciones significativas para la investigación en bioinformática y biodiversidad:

1. **Identificación de Brechas de Conocimiento:** El dashboard permite visualizar áreas de investigación poco exploradas en la biodiversidad ecuatoriana, lo que puede guiar futuros proyectos y asignación de recursos.
2. **Fomento de Colaboraciones:** Al identificar patrones de colaboración entre instituciones y autores, se promueve la creación de redes de investigación más sólidas y multidisciplinarias.
3. **Optimización de Recursos:** Los investigadores pueden utilizar el dashboard para evitar la duplicación de esfuerzos y enfocarse en áreas que requieren mayor atención.
4. **Acceso a Datos Consolidados:** El dashboard centraliza información dispersa, facilitando el acceso a datos relevantes y reduciendo el tiempo de búsqueda de publicaciones.
5. **Impacto en Políticas Públicas:** Los hallazgos pueden ser utilizados por instituciones gubernamentales y organizaciones no gubernamentales para diseñar políticas y estrategias de conservación basadas en evidencia.

5.2. Limitaciones del Estudio

La calidad de los datos depende de las fuentes disponibles, lo que puede limitar la precisión de algunas visualizaciones. Algunos usuarios sugirieron la inclusión de más datos contextuales, como información sobre políticas públicas o financiamiento de investigaciones.

5.3. Trabajos Futuros

El proyecto abre varias líneas de investigación y desarrollo que podrían ser abordadas en el futuro:

1. **Integración de Más Fuentes de Datos:** Incorporar bases de datos adicionales, como registros de biodiversidad globales (GBIF, GenBank) o información sobre financiamiento de investigaciones.
2. **Análisis de Impacto:** Evaluar el impacto de las publicaciones científicas mediante métricas como citas o índices de influencia.
3. **Expansión Temática:** Ampliar el enfoque a otros países de la región andina o amazónica para realizar análisis comparativos.
4. **Herramientas de Predicción:** Implementar modelos predictivos basados en inteligencia artificial para identificar tendencias futuras en investigación.
5. **Dashboard Multilingüe:** Adaptar la interfaz a otros idiomas, como inglés o lenguas indígenas, para aumentar su alcance y accesibilidad.
6. **Integración con Herramientas de Minería de Texto:** Mejorar la identificación de temas emergentes mediante técnicas avanzadas de procesamiento de lenguaje natural (NLP).

6. CONCLUSIONES

La investigación bioinformática en Ecuador ha experimentado un crecimiento exponencial con una tasa anual del 27.3% durante 2015-2024, impulsada principalmente por la implementación del Plan Nacional de Ciencia y Tecnología en 2016. El análisis revela una concentración geográfica, demostrando correlación positiva entre investigación y áreas megadiversos. Sin embargo, existe una distribución taxonómica desigual, con predominio de estudios sobre plantas (38.2%) y vertebrados (27.6%), mientras que artrópodos y microorganismos permanecen subrepresentados a pesar de su abundancia en ecosistemas ecuatorianos. La genómica constituye la técnica molecular predominante (42.7%), seguida por la filogenética molecular (28.3%), evidenciando una orientación hacia métodos establecidos más que hacia tecnologías emergentes.

Las brechas identificadas revelan un patrón sistemático de subrepresentación científica en regiones geográficas. Los grupos taxonómicos más desatendidos incluyen microorganismos acuáticos, artrópodos edáficos y hongos micorrízicos, influenciado por factores como tamaño corporal, carisma y valor económico percibido. Existe una subutilización significativa de técnicas bioinformáticas avanzadas como transcriptómica, proteómica y modelado estructural de proteínas en comparación con tendencias globales, atribuible a limitaciones en infraestructura especializada. Estas brechas constituyen oportunidades estratégicas para investigación pionera, especialmente en taxonomía integrativa de grupos hiperdiversos, metagenómica funcional en ecosistemas únicos ecuatorianos y secuenciación de organismos endémicos amenazados.

El dashboard desarrollado ha demostrado alta efectividad como herramienta analítica, evidenciada por su puntuación SUS de 83.7/100 en pruebas de usabilidad y tiempos de completamiento de tareas significativamente inferiores al benchmark establecido. La validación con expertos confirmó la precisión del 87.6% en la detección

automática de gaps de investigación, con valoraciones elevadas en relevancia para investigación (4.7/5), capacidad para descubrir tendencias (4.5/5) y utilidad en planificación estratégica (4.4/5). Los casos de uso documentados demuestran su aplicabilidad práctica en escenarios como identificación de taxones prioritarios, análisis de complementariedad institucional, y evaluación del impacto de políticas científicas, con estimaciones de ROI científico superiores al 200% comparado con metodologías tradicionales de planificación investigativa. Este proyecto sienta las bases para futuras investigaciones y la creación de herramientas similares en otros contextos.

6.1. Resumen de los Hallazgos

- **Distribución Geográfica:** Se identificó que la mayoría de las publicaciones se concentran en regiones como la Amazonía y los Andes, mientras que áreas como la Costa y Galápagos tienen menor cobertura.
- **Tendencias Temporales:** Se observó un aumento significativo en el número de publicaciones a partir del 2010, con un pico en los últimos cinco años.
- **Áreas de Investigación Prioritarias:** Las áreas más estudiadas incluyen genómica de especies endémicas y conservación de ecosistemas, mientras que temas como bioinformática aplicada a la agricultura y salud presentan brechas.
- **Colaboraciones:** Se detectaron redes de colaboración fuertes entre universidades ecuatorianas y algunas instituciones internacionales, pero con oportunidades para fortalecer vínculos con otros países de la región.
- **Brechas de Conocimiento:** Se identificaron especies y ecosistemas poco estudiados, lo que sugiere la necesidad de priorizar investigaciones en estas áreas.

6.2. Contribuciones del Estudio

- **Herramienta de Análisis Innovadora:** El dashboard es una de las primeras herramientas diseñadas específicamente para analizar publicaciones bioinformáticas sobre la biodiversidad ecuatoriana.
- **Visualización de Datos Complejos:** Facilita la interpretación de grandes volúmenes de datos mediante visualizaciones interactivas y amigables.
- **Identificación de Oportunidades:** Contribuye a la identificación de áreas de investigación prioritarias y brechas de conocimiento.
- **Promoción de la Ciencia Abierta:** Al centralizar y hacer accesible la información, el estudio fomenta la transparencia y el acceso abierto a datos científicos.
- **Base para Futuras Investigaciones:** El dashboard sirve como punto de partida para estudios más detallados y para la creación de herramientas similares en otros contextos.

6.3. Recomendaciones

Se recomienda priorizar estudios en grupos taxonómicos subrepresentados, particularmente artrópodos edáficos y microorganismos acuáticos, donde el potencial de descubrimientos es extraordinariamente alto. Diversificar el repertorio metodológico incorporando técnicas avanzadas como metagenómica funcional, transcriptómica comparativa y modelado de proteínas, aprovechando plataformas computacionales que no requieren equipamiento especializado. Fomentar investigaciones en áreas geográficas identificadas como subrepresentadas, especialmente ecosistemas amenazados donde la ventana de oportunidad para caracterización bioinformática está cerrándose debido a la deforestación. Implementar enfoques de taxonomía integrativa combinando morfología tradicional con datos moleculares para maximizar la resolución taxonómica, especialmente en grupos hiperdiversos. Establecer redes colaborativas multidisciplinarias e interinstitucionales para optimizar recursos y compartir

infraestructura, especialmente para técnicas que requieren equipamiento especializado.

Implementar programas de capacitación especializada en bioinformática avanzada, enfatizando análisis de NGS, metagenómica y genómica comparativa para fortalecer capacidades locales. Desarrollar infraestructura computacional distribuida mediante arquitecturas cloud-native que permitan análisis bioinformáticos complejos sin depender exclusivamente de equipamiento físico costoso. Establecer repositorios institucionales de datos genómicos estandarizados e interoperables que faciliten el intercambio y reúso de datos primarios entre investigadores. Promover políticas de acceso abierto a datos genómicos y protocolos bioinformáticos, incentivando publicaciones con disponibilidad completa de datos y scripts de análisis. Crear programas específicos para estudios en grupos taxonómicos subrepresentados mediante becas, financiamiento prioritario y alianzas con especialistas internacionales. Distribuir estratégicamente laboratorios satélites en regiones subrepresentadas para facilitar investigaciones en áreas remotas.

Desarrollar un plan nacional de bioinformática para biodiversidad con financiamiento específico para áreas subrepresentadas identificadas en este estudio. Implementar un sistema interinstitucional coordinado de permisos de investigación, colecta y exportación que agilice estudios en biodiversidad, especialmente para grupos taxonómicos prioritarios. Establecer un centro nacional de secuenciación de nueva generación con acceso distribuido para investigadores de diferentes instituciones, optimizando recursos mediante uso compartido. Crear un programa de repatriación de datos genómicos de biodiversidad ecuatoriana almacenados en instituciones extranjeras, integrándolos en sistemas nacionales. Desarrollar una estrategia nacional de bioeconomía que vincule investigación bioinformática con aplicaciones concretas en conservación, agrobiodiversidad y productos naturales.

Expandir la funcionalidad del dashboard incorporando módulos de análisis predictivo que proyecten tendencias futuras en investigación bioinformática basadas en patrones históricos. Implementar algoritmos de alertas automáticas cuando nuevas publicaciones aborden gaps previamente identificados, permitiendo seguimiento dinámico de avances. Desarrollar APIs públicas que permitan a investigadores incorporar sus propios datasets para análisis comparativos personalizados. Integrar el dashboard con plataformas nacionales existentes como BNDB (Base Nacional de Datos sobre Biodiversidad) y SIB-Ec (Sistema de Información de Biodiversidad). Implementar visualizaciones tridimensionales interactivas para análisis de distribución espacial de datos genómicos correlacionados con variables ambientales. Desarrollar una versión móvil simplificada para uso en campo durante expediciones científicas. Crear módulos específicos para bioética y propiedad intelectual que faciliten el cumplimiento normativo en investigaciones bioinformáticas. Este estudio no solo contribuye al avance de la investigación en bioinformática y biodiversidad en Ecuador, sino que también sienta las bases para futuras iniciativas que utilicen herramientas de visualización de datos para abordar desafíos científicos y sociales.

7. REFERENCIAS BIBLIOGRÁFICAS

Báez, O. (04 de 07 de 2019). La biodiversidad: clave del desarrollo sustentable del Ecuador. *Periódico Opción*. Obtenido de <https://periodicoopcion.com/la-biodiversidad-clave-del-desarrollo-sustentable-del-ecuador/>

Banco de Desarrollo de América Latina y el Caribe. (2024). *Programa de Enfoque Sectorial Amplio de Apoyo a la Biodiversidad, Cambio Climático y Gestión Ambiental*. Ministerio de Economía y Finanzas.
doi:<https://www.caf.com/media/4672915/eval-expost-ecuador-apoyo-a-la-biodiversidad-cambio-climatico-y-gestion-ambiental.pdf>

BioWeb. (21 de 10 de 2018). *Mamíferos Ecuador*. Obtenido de <https://bioweb.bio/faunaweb/mammaliaweb/Biodiversidad/>

BioWeb. (10 de 04 de 2023). *Libro rojo de las plantas endémicas del Ecuador*. Obtenido de <https://bioweb.bio/floraweb/librorojo/IndiceTaxonomico>

Campaña Sin Biodiversidad No Hay Vida. (2022). *Análisis del borrador del Marco Global para la Biodiversidad Post-2020 Camino hacia la COP 15 de Montreal del Convenio de Diversidad Biológica (CDB)*. Ecologistas en Acción. Obtenido de <https://www.ecologistasenaccion.org/wp-content/uploads/2022/11/analisis-marco-global-biodiversidad-post2020.pdf>

Casillas, S. (mayo de 2008). *La bioinformática en el estudio de la diversidad genética*. Obtenido de <https://www.uab.cat/web/detalle-noticia/la-bioinform-225-tica-en-el-estudio-de-la-diversidad-gen-233-tica-1345680342040.html?articleId=1211439134723>

Cedillo, J., Beltrán, H., Saltos, M., & Soriano, F. (2024). Explorando la minería de datos en la gestión educativa superior: desafíos y oportunidades en la era digital. *Reincisol*, 3(5), 1367-1385. doi:[https://doi.org/10.59282/reincisol.V3\(5\)1367-1385](https://doi.org/10.59282/reincisol.V3(5)1367-1385)

CEPAL. (10 de octubre de 2023). *La biodiversidad como impulsor de la transformación sostenible en América Latina y el Caribe*. Obtenido de <https://www.cepal.org/es/enfoques/la-biodiversidad-como-impulsor-la-transformacion-sostenible-america-latina-caribe>

CIEMAT. (8 de octubre de 2019). *La minería de datos en la investigación científica*. Obtenido de <https://www.ciemat.es/cargarAplicacionNoticias.do;jsessionid=4C95E7C626006DC7A3A89DE32578D845?identificador=2076&idArea=-1>

Conforme-García, M., Dávila-Ulloa, M., Sarango-Ordóñez, J., & Medina-Gahona, G. (2024). Estrategias de Biotecnología Verde: Hacia una Recuperación. *Código Científico Revista De Investigación*, 5(E3), 119–144. doi:<https://doi.org/10.55813/gaea/ccri/v5/nE3/313>

Córdova, Y., Martínez, J., & Córdova, E. (2021). Propuesta de metodología para el diseño de dashboard. *Revista Cubana de Transformación Digital*, 3(2). doi:<http://portal.amelica.org/ameli/jatsRepo/389/3892632006/index.html>

Cornejo, A., & Núñez, J. (2024). Técnicas de visualización de datos en la comprensión de información cuantitativa: una revisión en la interpretación de resultados. *LATAM Revista Latinoamericana De Ciencias Sociales Y Humanidades*, 5(5), 1734 – 1748. doi:<https://doi.org/10.56712/latam.v5i5.2738>

Cruz, K., Garzón, V., Quezada, J., & Carvajal, H. (2022). Tableros y gráficos automatizados: un enfoque a la visualización de datos e inteligencia de negocio. *Ciencia Latina Revista Científica Multidisciplinar*, 6(4), 2624-2641. doi:https://doi.org/10.37811/cl_rcm.v6i4.2784

García, R., & Gil, J. (2023). Minería de Datos Educativos: Descubrir tesoros ocultos durante el aprendizaje. *REVISTA CIENTÍFICA ECOCIENCIA*, 10(Edición Especial), 18–41. doi:<https://doi.org/10.21855/ecociencia.100.830>

Guaman, D., & Patiño, S. (2022). Desarrollo de un Dashboard para el apoyo a la toma de decisiones en la pequeña empresa manufacturera de ropa deportiva. Caso Sport Elegant. *TECNOLOGI-K*, 3(1), 1-9.

doi:<https://revista.istb.edu.ec/index.php/tecnologi-k/article/view/35>

Herrera-Feijoo, R. (2024). Principales amenazas e iniciativas de conservación. *Journal of Economic and Social Science Research*, 4(1), 33–56.

doi:<https://doi.org/10.55813/gaea/jessr/v4/n1/85>

INABIO. (13 de 01 de 2025). *PERFIL DE BIODIVERSIDAD*. Obtenido de

[http://inabio.biodiversidad.gob.ec/perfil-de-](http://inabio.biodiversidad.gob.ec/perfil-de-biodiversidad/#:~:text=El%20pa%C3%ADs%20se%20divide%20en,14%20herb%C3%A1ceos%20y%2012%20arbustivos)

[biodiversidad/#:~:text=El%20pa%C3%ADs%20se%20divide%20en,14%20herb%C3%A1ceos%20y%2012%20arbustivos](http://inabio.biodiversidad.gob.ec/perfil-de-biodiversidad/#:~:text=El%20pa%C3%ADs%20se%20divide%20en,14%20herb%C3%A1ceos%20y%2012%20arbustivos).

Instituto Nacional de Biodiversidad . (2021). *Plan Estratégico Institucional del Instituto Nacional de Biodiversidad 2021-2025*.

doi:[https://inabio.biodiversidad.gob.ec/wp-](https://inabio.biodiversidad.gob.ec/wp-content/uploads/2022/08/plan_de_investigacion_2021-2025.pdf)

[content/uploads/2022/08/plan_de_investigacion_2021-2025.pdf](https://inabio.biodiversidad.gob.ec/wp-content/uploads/2022/08/plan_de_investigacion_2021-2025.pdf)

Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias. (10 de diciembre de 2019). *Biología molecular y Bioinformática para la conservación*

de las especies forestales en riesgo por Cambio Climático. Obtenido de

<https://www.gob.mx/inifap/articulos/biologia-molecular-y-bioinformatica-para-la-conservacion-de-las-especies-forestales-en-riesgo-por-cambio-climatico>

Jiménez, G., & Mendoza, M. (2023). Prototipo para Monitoreo de Concurrencia con Dashboard Aplicado a Pequeños Negocios. *Ciencia Latina Revista Científica Multidisciplinar*, 7(6), 351-363. doi:https://doi.org/10.37811/cl_rcm.v7i6.8692

Lavalle, A. (2021). *A methodology for the visual comprehension of Big Data*.

Universidad de Alicante. doi:<http://hdl.handle.net/10045/119626>

- Liras, E., Cabello, J., & Bonet, F. (2008). Bioinformática para la conservación de la flora. *CONSERVACIÓN VEGETAL*(12), 6-9.
doi:<https://revistas.uam.es/conservacionvegetal/issue/view/65/30>
- Liria, J. (2022). Áreas de endemismo de Ecuador: un análisis a partir de datos de distribución de especies de plantas, animales y hongos. *Revista mexicana de biodiversidad*, 93, e934031.
doi:<https://doi.org/10.22201/ib.20078706e.2022.93.4031>
- Mena, P. (2014). La biodiversidad del Ecuador. En M. García, P. Mena, & D. Parra, *El país de la biodiversidad* (págs. 19 -33). FLACSO. Obtenido de <https://biblio.flacsoandes.edu.ec/libros/digital/49907.pdf>
- Ministerio de Ambiente y Desarrollo Sostenible. (2017). *Plan de acción de Biodiversidad Para la implementación de la Política 2016 - 2030*. HUMBOLDT. Obtenido de <https://faolex.fao.org/docs/pdf/col173040.pdf>
- Ministerio del Ambiente del Ecuador. (2016). *Estrategia Nacional de Biodiversidad 2015-2030*. Quito, Ecuador.
- Ministerio del Ambiente del Ecuador. (2020). *Quinto Informe Nacional para el Convenio sobre la Diversidad Biológica*. Quito, Ecuador. Obtenido de www.ambiente.gob.ec: <https://www.ambiente.gob.ec/wp-content/uploads/downloads/2015/06/QUINTO-INFORME-BAJA-FINAL-19.06.2015.pdf>
- Murillo, W. (2024). *Diseño de dashboards interactivos: mejorando la interpretación de datos en la elasticidad de precios y demanda de productos ferreteros en la empresa "Importador Ferretero Trujillo"*. Universidad De Las Américas. Obtenido de <https://dspace.udla.edu.ec/handle/33000/15982>

- Orozco, W., Villao, A., Orozco, J., & Villarroel, M. (2021). Aplicación de técnicas de minería de datos para predecir el rendimiento académico de los estudiantes de la escuela 'Lic. Angélica Villon L. *Universidad RCTU* , 8(2), 68-75.
doi:<https://doi.org/10.26423/rctu.v8i2.637> .
- Pérez-Suasnavas, A., Salgado-Proaño, B., Hasperué, W., Cela, K., & Santamaría, J. (2023). Evolución de las técnicas de minería de datos para extraer datos provenientes de twitter aplicadas a la educación superior: una revisión sistemática. *South Florida Journal of Development*, 4(1), 33–55.
doi:<https://doi.org/10.46932/sfjdv4n1-002>
- Pincay, F., Baque, O., Caicedo, C., & Paladines, J. (2024). Dashboard para el control y seguimiento académico de los estudiantes en la Educación Superior. *Revista Científica De Innovación Educativa Y Sociedad Actual "ALCON"*, 4(5), 1–13.
doi:<https://doi.org/10.62305/alcon.v4i5.272>
- Ramírez, L., Delgado, E., & Montúfar, M. (2022). Aplicación de técnicas de minería de datos para la caracterización de estudiantes bajo el efecto de la COVID-19. *Pädi Boletín Científico De Ciencias Básicas E Ingenierías Del ICBI*, 10(Especial2), 75-81. doi:<https://doi.org/10.29057/icbi.v10iEspecial2.8669>
- Rodríguez, A., Delgado, H., Solórzano, W., & Anzules, X. (2020). La minería de datos y algunas de sus aplicaciones contextuales. *Serie Científica de la Universidad de las Ciencias Informáticas*, 13(11), 17 - 25.
- Ruiz, V., & Armoa, A. (2023). La importancia de la Minería de Datos como una herramienta estratégica en las Empresas. *Ciencia Latina Revista Científica Multidisciplinar*, 7(1), 9267-9276. doi:https://doi.org/10.37811/cl_rcm.v7i1.5119
- Therón, R. (2021). Visualización de datos caminos de ida y vuelta entre arte y ciencia en la producción y consumo de imágenes. *Fonseca, Journal of Communication*(23), 39-60. doi:<https://doi.org/10.14201/fjc2021233960>

- Torres-Quezada, Y. (2021). Minería de datos para determinar los factores más influyentes en la ocurrencia de siniestros de tránsito en Ecuador en el año 2020. *CEDAMAZ*, 11(2), 124 – 132.
doi:<https://doi.org/10.54753/cedamaz.v11i2.1181>
- Toulkeridis, T., Tamayo, E., Simón-Baile, D., Merizalde-Mora, M., Reyes –Yunga, D., Viera-Torres, M., & Heredia, M. (2020). Cambio Climático según los académicos ecuatorianos - Percepciones versus hechos. *Revista de Ciencias de la Vida*, 31(1), 21-46. doi:<https://doi.org/10.17163/lgr.n31.2020.02>
- Umaquina, A. (2024). Técnicas de Minería de datos aplicados a la agricultura: Estado del Arte y análisis bibliométrico. *INNOVATION & DEVELOPMENT IN ENGINEERING AND APPLIED SCIENCES*, 6(1), 25.
doi:<https://doi.org/10.53358/ideas.v6i1.944>
- Varea, A. (2004). Iniciativas para conservar la biodiversidad. *Universitas-XXI. Revista de Ciencias Sociales y Humanas*(4), 7-43.
doi:<https://www.redalyc.org/articulo.oa?id=476150823001>

8. ANEXOS

8.1. Anexo 1: Glosario de términos bioinformáticos

Tabla 11

Glosario de términos bioinformáticos

Palabra	Significado
Algoritmo	Conjunto de instrucciones o reglas definidas para realizar una tarea.
Análisis de datos	Proceso de examinar, limpiar, transformar e interpretar datos para extraer información útil.
API	Interfaz de Programación de Aplicaciones, que permite la comunicación entre diferentes sistemas de software.
Big data	Conjuntos de datos extremadamente grandes y complejos que requieren herramientas y técnicas especiales.
Bioinformática	Campo interdisciplinario que aplica la informática y la estadística al estudio de datos biológicos.
Biodiversidad	Variedad de vida en la Tierra, incluyendo la diversidad genética, de especies y de ecosistemas.
Biología computacional	Uso de técnicas computacionales para modelar y simular sistemas biológicos.
Clasificación	Técnica de minería de datos que asigna categorías a los datos.
Clúster	Grupo de objetos similares identificado por algoritmos de agrupamiento.

Conservación	Protección y manejo de la biodiversidad y los recursos naturales.
Datos bioinformáticos	Información derivada del análisis computacional de datos biológicos.
Datos genómicos	Información relacionada con el genoma de un organismo.
Deforestación	Proceso de tala de árboles y destrucción de bosques.
Ecología	Estudio de las interacciones entre los organismos y su entorno.
Ecosistema	Comunidad de organismos vivos y su entorno físico interactuando como una unidad funcional.
Endémico	Especie que es única de una ubicación geográfica específica.
Filogenética	Estudio de las relaciones evolutivas entre organismos.
Gen	Unidad hereditaria que contiene información genética.
GenBank	Base de datos de secuencias genéticas.
Genoma	Conjunto completo de material genético en un organismo.
Hábitat	Entorno natural en el que vive un organismo.
Herramientas bioinformáticas	Aplicaciones de software utilizadas para analizar datos biológicos.
Índice h	Medida del impacto de un científico basada en sus publicaciones y citas.

Inteligencia artificial	Capacidad de las máquinas para realizar tareas que requieren inteligencia humana.
Metadatos	Datos que proporcionan información sobre otros datos.
Metagenómica	Estudio del material genético recuperado directamente de muestras ambientales.
Minería de datos	Proceso de descubrir patrones en grandes conjuntos de datos.
Modelado molecular	Técnica computacional para predecir y simular el comportamiento de las moléculas.
Patrones	Regularidades o tendencias identificadas en los datos.
Predicción	Proceso de estimar resultados futuros basados en datos disponibles.
Proteína	Molécula biológica esencial para la estructura y función de las células.
PubMed	Base de datos bibliográfica de literatura biomédica.
Redes neuronales	Modelos computacionales inspirados en el cerebro humano para el aprendizaje automático.
Secuencia de ADN	Orden de nucleótidos en una molécula de ADN.
Taxonomía	Ciencia de clasificar y nombrar organismos.
Tendencias	Patrones de cambio a lo largo del tiempo.
Web scraping	Técnica para extraer datos de sitios web.

8.2. Anexo 2: Detalles técnicos del desarrollo del dashboard

Herramientas de Visualización: Power BI

Power BI ha sido seleccionado como la herramienta principal para el desarrollo del dashboard debido a sus capacidades avanzadas:

- **Procesamiento de datos robusto:** Capacidad para manejar y procesar grandes volúmenes de datos bibliométricos y bioinformáticos.
- **Visualizaciones interactivas:** Amplia biblioteca de visualizaciones nativas y personalizables.
- **Inteligencia minería de datos.de negocios integrada:** Funciones de análisis avanzado y
- **Conectividad:** Facilidad para conectarse con múltiples fuentes de datos (Scopus, Web of Science, PubMed, bases de datos genómicas).
- **Accesibilidad:** Posibilidad de publicación web para compartir resultados con la comunidad científica.
- **DAX (Data Analysis Expressions):** Lenguaje de fórmulas para cálculos personalizados y análisis complejos.

1. Arquitectura de Datos

- **Fuentes de Datos**
 - Identificación de las bases de datos (por ejemplo, Scopus, Web of Science, PubMed, Google Scholar) y archivos (CSV, Excel) que contienen la información bioinformática.
 - Considerar la frecuencia de actualización de estas fuentes y la necesidad de automatizar la extracción de datos.

- Evaluación de la calidad y consistencia de los datos en cada fuente.
- **Modelado de Datos**
 - Diseño de un modelo de datos relacional que permita la conexión y análisis de datos de diferentes fuentes.
 - Creación de tablas de hechos (por ejemplo, publicaciones, especies) y dimensiones (por ejemplo, autores, años, ubicaciones).
 - Definición de relaciones entre las tablas para asegurar la integridad de los datos.
- **Transformación de Datos (ETL)**
 - Utilización de Power Query (en Power BI) para limpiar, transformar y cargar los datos.
 - Implementación de reglas de validación para asegurar la calidad de los datos.
 - Automatización del proceso ETL para la actualización periódica de los datos.

2. Visualización y Funcionalidad

- **Selección de Visualizaciones**
 - Uso de gráficos de barras, líneas, dispersión, mapas y tablas para representar los datos de manera clara y efectiva.
 - Personalización de las visualizaciones para resaltar la información clave.

- Implementación de visualizaciones interactivas que permitan a los usuarios explorar los datos en detalle.
- **Diseño del Dashboard**
 - Diseño de una interfaz de usuario intuitiva y fácil de usar.
 - Organización de las visualizaciones y los filtros de manera lógica y coherente.
 - Uso de colores, fuentes y otros elementos visuales para mejorar la claridad y la legibilidad.
 - Consideración de la accesibilidad para usuarios con discapacidades.
- **Funcionalidad Interactiva**
 - Implementación de filtros (segmentadores) para permitir a los usuarios explorar los datos por diferentes dimensiones.
 - Uso de acciones de navegación (drill-down, drill-through) para permitir a los usuarios acceder a información detallada.
 - Habilitación de la capacidad de exportar datos y visualizaciones a diferentes formatos.
- **Power BI Específicos**
 - Utilización de DAX (Data Analysis Expressions) para crear medidas y cálculos personalizados.
 - Implementación de seguridad a nivel de fila para controlar el acceso a los datos.
 - Optimización del rendimiento del dashboard para garantizar tiempos de respuesta rápidos.

3. Implementación y Mantenimiento

- **Publicación y Distribución**

- Publicación del dashboard en Power BI Service para compartirlo con los usuarios.
- Configuración de la seguridad y los permisos de acceso.
- Implementación de un proceso de actualización periódica del dashboard.

- **Monitoreo y Mantenimiento:**

- Monitoreo del rendimiento del dashboard y la calidad de los datos.
- Recopilación de comentarios de los usuarios y realización de mejoras continuas.
- Mantenimiento de la documentación técnica del dashboard.

- **Seguridad:**

- La implementación de la seguridad de la información, desde la ingesta de datos, hasta la visualización de los datos es de suma importancia. Se debe de tener especial cuidado en las credenciales de conexión a las bases de datos, y en los permisos de los usuarios que pueden acceder a los dashboard.

8.3. Anexo 3: Código de script utilizado para hacer Webscraping en buscadores de artículos científicos.

```
Código de script utilizado para hacer Webscraping del buscador CROSSREF
import requests
import pandas as pd

# Configuración de la API de Crossref
CROSSREF_API_URL = "https://api.crossref.org/works"
USER_AGENT = "joluisvt/1.0 (joluisvt@gmail.com)"

# Lista de consultas a realizar
queries = [
    ("Genómica") AND ("Microorganismos") AND ("Ecuador"),
    ("Genómica") AND ("Anfibios") AND ("Ecuador"),
    ("Genómica") AND ("Plantas") AND ("Ecuador"),
    ("Genómica") AND ("Insectos") AND ("Ecuador"),
    ("Genómica") AND ("Aves") AND ("Ecuador"),
    ("metagenomica")and("Microorganismos")and("Ecuador"),
    ("metagenomica")and("Anfibios")and("Ecuador"),
    ("metagenomica")and("Plantas")and("Ecuador"),
    ("metagenomica")and("Insectos")and("Ecuador"),
    ("metagenomica")and("Aves")and("Ecuador"),
    ("filogenetica")and("Microorganismos")and("Ecuador"),
    ("filogenetica")and("Anfibios")and("Ecuador"),
    ("filogenetica")and("Plantas")and("Ecuador"),
    ("filogenetica")and("Insectos")and("Ecuador"),
    ("filogenetica")and("Aves")and("Ecuador")
]

# Lista para almacenar los resultados
results = []

# Función para buscar en Crossref
def search_crossref(query, max_results=1000):
    params = {
        "query": query,
        "rows": 50, # Número de resultados por página (máximo 100)
        "cursor": "*" # Cursor para paginación
    }
    headers = {
        "User-Agent": USER_AGENT
    }
    count = 0

    while count < max_results:
```

```

response = requests.get(CROSSREF_API_URL, params=params,
headers=headers)
if response.status_code != 200:
    print(f"Error en la consulta: {response.status_code}")
    break

data = response.json()
items = data.get("message", {}).get("items", [])

if not items:
    break

for item in items:
    if count >= max_results:
        break

    title = " ".join(item.get("title", [""]))
    authors = " ".join([f"{author.get('given', '')} {author.get('family', '')}"
                        for author in item.get("author", [])])
    year = item.get("created", {}).get("date-parts", [[None]])[0][0]
    doi = item.get("DOI", "")
    publisher = item.get("publisher", "")
    journal_title = item.get("container-title", [""])[0]

    results.append({
        "Titulo": title,
        "Autor": authors,
        "Año de Publicación": year,
        "Ciudad": "", # Crossref no proporciona esta información
        "País": "Ecuador",
        "Url de la publicacion": f"https://doi.org/{doi}",
        "Editorial": publisher,
        "Palabra clave": query,
        "Titulo de la publicacion": journal_title,
        "Resumen": item.get("abstract", [""])[0]
    })

    count += 1
    print(f"Artículo {count} encontrado: {title}")

# Paginación: actualizar el cursor
params["cursor"] = data.get("message", {}).get("next-cursor", "")
if params["cursor"] == "":
    break

# Realizar búsquedas y almacenar resultados
for query in queries:
    print(f"Buscando: {query}")

```

```
search_crossref(query, max_results=1000)

# Crear DataFrame
df = pd.DataFrame(results)

# Guardar DataFrame en un archivo CSV
df.to_csv('articulos_cientificos_crossref.csv', index=False)

print("Datos guardados en 'articulos_cientificos_crossref.csv'")

Código de script utilizado para hacer Webscraping del buscador CROSSREF

import pandas as pd
from scholarly import scholarly

# Lista de consultas a realizar
queries = [
    ("Genómica") and ("Microorganismos") and ("Ecuador"),
    ("Genómica") and ("Anfibios") and ("Ecuador"),
    ("Genómica") and ("Plantas") and ("Ecuador"),
    ("Genómica") and ("Insectos") and ("Ecuador"),
    ("Genómica") and ("Aves") and ("Ecuador")
]

# Lista para almacenar los resultados
results = []

# Realizar búsquedas y almacenar resultados
for query in queries:
    search_query = scholarly.search_pubs(query)
    for i, result in enumerate(search_query):
        if i >= 50: # Limitar a 50 resultados por consulta
            break
        title = result.get('bib', {}).get('title', "")
        author = ', '.join(result.get('bib', {}).get('author', []))
        year = result.get('bib', {}).get('pub_year', "")
        url = result.get('pub_url', "")
        publisher = result.get('bib', {}).get('publisher', "")
        keyword = query
        publication_title = result.get('bib', {}).get('journal', "")

        results.append({
            'Titulo': title,
            'Autor': author,
            'Año de Publicación': year,
            'Ciudad': "", # Google Scholar no proporciona esta información
            'País': 'Ecuador',
            'Url de la publicacion': url,
            'Editorial': publisher,
            'Palabra clave': keyword,
            'Titulo de la publicacion': publication_title
        })

# Crear DataFrame
df = pd.DataFrame(results)
print(df.head())
```

```
# Guardar DataFrame en un archivo CSV
df.to_csv('articulos_cientificos.csv', index=False)

print("Datos guardados en 'articulos_cientificos.csv'")
```