

Nuevo enfoque de optimización metaheurística: aplicación en granjas eólicas

Dylan Samuel Cantillo Arrieta
dept. Ingeniería de Sistemas
Universidad del Norte
Barranquilla, Colombia
dilanc@uninorte.edu.co

Breyunner Hurtado Acuña
dept. Ingeniería de Sistemas
Universidad del Norte
Barranquilla, Colombia
breyunnerh@uninorte.edu.co

Tutor

Elias David Niño Ruiz
dept. Ingeniería de Sistemas
Universidad del Norte
Barranquilla, Colombia
enino@uninorte.edu.co

Asesor

Alfonso Manuel Mancilla Herrera
dept. Ingeniería de Sistemas
Universidad del Norte
Barranquilla, Colombia
amancill@uninorte.edu.co

November 24, 2023

Abstract

El presente proyecto de grado aborda la necesidad de mejorar los métodos de optimización con un método novedoso que combina técnicas de optimización metaheurística, como algoritmos genéticos, restricción al cono y generación de direcciones aleatorias. Se propone una aplicación en el contexto de las energías renovables, en particular en el diseño y operación de granjas eólicas. El objetivo principal es superar las limitaciones de los métodos tradicionales en términos de miopía y velocidad de convergencia de los algorítmicos metaheurísticos. Se destaca la complejidad de los sistemas reales y la interacción de factores geográficos y meteorológicos en la optimización de granjas eólicas. La investigación busca validar la eficacia y eficiencia del nuevo enfoque mediante la implementación y validación de un prototipo en Python, utilizando datos históricos de viento en Colombia. Se espera que este enfoque revolucione la forma en que se abordan los problemas de optimización en campos diversos.

This graduate project addresses the need to enhance optimization methods using a novel approach that combines metaheuristic optimization techniques, such as genetic algorithms, cone constraint handling, and random direction generation. The proposed application is within the realm of renewable energies, specifically in the design and operation of wind farms. The primary objective is to overcome the limitations of traditional methods in terms of convergence speed and myopia. The complexity of real systems and the interaction of geographical and meteorological factors in wind farm optimization are emphasized. The research aims to validate the effectiveness and efficiency of the new approach through the implementation and validation of a prototype in Python, utilizing historical wind data from Colombia. It is anticipated that this approach will revolutionize how optimization problems are addressed across various fields.

Keywords: *Optimización; Metaheurística; Granjas eólicas; Predicción; Viento; Algoritmos genéticos*

List of Figures

1	Árbol de formulación del problema (optimización)	7
2	Árbol de formulación del problema (energía eólica)	8
3	Metodología matemática	10
4	Requerimientos funcionales	12
5	Arquitectura lógica de la solución	18
6	Arquitectura física de la solución	19
7	Pipeline de ingeniería de características	24
8	Comparación Dixon-Price con x_0 conocido	27
9	Comparación Dixon-Price con x_0 aleatorio	28
10	Comparación Rosenbrock con x_0 conocido	28
11	Comparación Rosenbrock con x_0 aleatorio	29
12	Valores Verdaderos vs Valores Predichos	31
13	Dashboard información general Colombia	32
14	Dashboard por regiones	33
15	Dashboard por departamentos	34
16	Dashboard por municipios	35
17	Dashboard por región y mes	36
18	Dashboard por departamento y mes	37

List of Tables

1	Criterios de búsqueda	13
2	Librerías y bibliotecas empleadas	20
3	Funciones de prueba	22
4	Métricas Dixon-Price con x_0 conocido	27
5	Métricas Dixon-Price con x_0 aleatorio	27
6	Métricas Rosenbock con x_0 conocido	28
7	Métricas Rosenbock con x_0 aleatorio	29

Introducción

La energía eólica se ha establecido firmemente como una de las fuentes de energía renovable más prometedoras en la actualidad, con un crecimiento expansivo en todo el mundo (Asghar et al., 2023). A pesar de sus ventajas, la eficiencia de las granjas eólicas sigue siendo un desafío crucial debido a la necesidad de optimizar los parámetros de los aerogeneradores (turbinas), una tarea sumamente compleja debido a la multitud de variables involucradas, debido a la infraestructura y a las condiciones meteorológicas.

Sin embargo, la optimización de sistemas complejos no se limita exclusivamente al ámbito de la energía eólica. Como señala Dietz et al. (2020), la optimización de sistemas complejos es un área fundamental de la investigación de operaciones en la actualidad, ya que permite encontrar soluciones óptimas o casi óptimas a problemas con muchas variables, restricciones e interdependencias. No obstante, los métodos de optimización clásicos, como el método de Newton-Rapson, Descenso del Gradiente, Región de Confianza, entre otros, presentan limitaciones significativas cuando se aplican a problemas no lineales, no convexos y multidimensionales. Debido a quedar atrapados en óptimos locales en lugar de alcanzar soluciones óptimas globales (Jin, 2021).

Es en este contexto que las metaheurísticas han surgido como alternativas prometedoras para superar estas limitaciones, como lo señala el enfoque de Ezugwu (2021). Algunos ejemplos de metaheurísticas incluyen algoritmos genéticos, optimización por enjambre de partículas y búsqueda tabú. Aunque estos enfoques han demostrado ser eficaces en una amplia gama de problemas, en particular, los algoritmos genéticos aún tienen un margen de mejora en términos de velocidad de convergencia y la capacidad de evitar la convergencia prematura o la miopía.

El presente proyecto de grado se centra en presentar un enfoque novedoso de optimización metaheurística diseñado específicamente para mitigar la miopía y mejorar la convergencia en comparación con las técnicas existentes. Esta propuesta combina algoritmos genéticos, exploración aleatoria del espacio de búsqueda y restricción al cono de confianza, entre otras prácticas matemáticas. Para validar su eficacia, se implementa y se pone a prueba esta metodología en el contexto de la optimización de granjas eólicas y la predicción de vientos en Colombia. Este problema, que además de su significativa complejidad, presenta una gran importancia práctica debido a los beneficios ambientales que ofrece. Además, como sugiere Wiser et al. (2021) se estima una reducción de los costos de la energía eólica entre el 37% y 49% para el 2050, es decir, una realidad en la que debemos trabajar.

Justificación

En un mundo cada vez más complejo y dinámico, la optimización de sistemas y procesos se ha convertido en una necesidad crucial para diversas disciplinas, abarcando campos desde la ingeniería hasta las energías renovables. Uno de los más beneficiados se los ha llevado la investigación científica. Además, la constante evolución y crecimiento de la tecnología ha planteado desafíos cada vez más multidimensionales. En este contexto, la optimización desempeña un papel fundamental al permitir encontrar soluciones óptimas o subóptimas en situaciones donde las variables, restricciones y objetivos pueden ser numerosos, interdependientes y sujetos a cambios constantes.

En este plano real de recursos limitados es imperativo encontrar una solución óptima basada en información imperfecta o incompleta. Los métodos de optimización tradicionales (orgánicos) han sido valiosos en muchos escenarios, pero a menudo demuestran una insuficiencia en términos de precisión por la complejidad en problemas multidimensionales o no lineales, Santillán, López, Ortiz, and Ruiz (2008) establece que aproximadamente un 90% de los problemas en este plano real se comportan como sistemas no lineales y los métodos con menos complejidad no logran superar las soluciones subóptimas o si acaso llegan a una.

La aparición de metaheurísticas para resolver tales problemas de optimización es uno de los logros más notables de las últimas dos décadas en la investigación operativa (Gendreau & Potvin, 2005). Aunque se ha demostrado ampliamente en la literatura, Gutjahr (2009), que los algoritmos genéticos son eficaces en la solución de problemas de optimización complejos, estos algoritmos aún tienen un margen para mejorar en términos de velocidad de convergencia y capacidad para evitar caer en la miopía.

De acuerdo con Herrera-Quintero, Jalil-Naser, Banse, and Samper-Zapater (2015) para esta sociedad, enfocado principalmente en Colombia, se requiere aplicar soluciones TIC más eficientes y efectivas a problemas complejos que afectan a la vida cotidiana, tales como la logística, el transporte, la plantación urbana, la toma de decisiones, entre otros. De manera mas específica, estos avances en las técnicas de optimización podrían posibilitar notables beneficios socioeconómicos, como las energías renovables, planteando una trascendencia para la sociedad en mejora de la economía regional y cubrir las demandas de electricidad a comunidades remotas. Además, como establece Wiser et al. (2021) en los avances de la energía eólica, la necesidad de reducción de costos generales de la energía por medio de la predicción del viento para mayor rentabilidad y eficiencia a las granjas eólicas se presenta como un gran beneficio social a largo plazo. Uno de los beneficios mas importantes, por lo que todos los países a nivel mundial están luchando, es la reducción del uso de combustibles fósiles.

A pesar de la existencia de varios softwares para estimar el potencial eólico de una

región o país, de acuerdo con el estudio Gualtieri (2020) hay una falta significativa de herramientas que integren técnicas de optimización metaheurística avanzadas en la gestión de granjas eólicas. Asimismo, no todas las técnicas de optimización son aplicables o eficientes en todos los contextos geográficos y climatológicos.

Estos alcances podrían aplicarse en una amplia gama de campos y problemas prácticos en un futuro por la búsqueda de generalidad de las soluciones en los problemas de optimización. Aunque el estudio se centra en granjas eólicas de Colombia, la investigación lograría contribuir al desarrollo de nuevas técnicas de optimización abriendo las puertas a enfoques más robustos y confiables, y su efectividad en la mitigación de la miopía y mejora de la velocidad de convergencia. Por consiguiente, gracias a la revisión sistemática de literatura, el nuevo enfoque de optimización permitiría la prueba de nuevas técnicas y enfoques en el campo de la innovación informática, posibilitando el desarrollo de algoritmos más sofisticados y eficientes. Esto puede abrir el camino para nuevas aplicaciones y avances en el campo de la inteligencia artificial y la optimización computacional.

La conveniencia de este trabajo radica en su potencial para revolucionar la forma en que se abordan los problemas de optimización en diversas disciplinas sumergiéndose en el ahorro de tiempo y recursos de la búsqueda de soluciones óptimas para contribuir al avance científico y tecnológico. Se espera que este trabajo arroje luz sobre la eficacia de la combinación de diferentes técnicas de metaheurística para abordar problemas complejos de optimización enfocado en granjas eólicas y la predicción del viento. Además, se espera obtener conocimientos sobre cómo estas técnicas pueden superar las limitaciones de los métodos tradicionales y qué desafíos específicos pueden abordar de manera más efectiva.

Descripción del problema

La industria colombiana se encuentra en una etapa crucial en la que la implementación de técnicas de optimización desempeña un papel fundamental para potenciar su eficiencia y competitividad. A partir de investigaciones recientes, es evidente que la optimización se ha vuelto una necesidad en diversos sectores, ya sea para la reducción de costos o para contribuir al desarrollo sostenible del país.

Estas investigaciones sugieren, colectivamente, la necesidad y beneficios de la implementación de técnicas de optimización en la industria colombiana. Como mencionan Higuera Alzate (2018), SÁCHICA (2020) y Cerveleón and Ledesma (2022), en industrias del cemento, consumo energético y sector carbón, respectivamente, se incentiva la eficiencia y optimización de procesos como herramienta principal para el progreso del avance tecnológico y económico, así como en la reducción de costos y sostenibilidad de las empresas. Por otro lado, los procesos operacionales deficientes, la mala administración de recursos y una deficiente programación puede llevar a un sector, como la aviación comercial en

Colombia (Carreño Ávila, 2019), a una urgente necesidad de implementación de técnicas de optimización.

Esta problemática se fundamenta en la absoluta necesidad de incorporar en la industria colombiana técnicas de optimización avanzadas. Con ello entra a discusión, como se plantea en el árbol de formulación del problema (Fig. 1), las limitaciones intrínsecas de los enfoques tradicionales de optimización, tal como han señalado investigaciones previas de Santillán et al. (2008), González, Osborne, and Lawrence (2016) y Jiang, Chai, González, and Garnett (2019). Asimismo, se requiere un esfuerzo constante para mejorar la velocidad de convergencia de los métodos metaheurísticos existentes, como lo destaca Gutjahr (2009) y Allawi, Ibraheem, and Humaidi (2019).

Además, otro tipo de objeto de estudio son la gran complejidad de los problemas del mundo real (en su gran mayoría problemas no lineales), los bastos espacios de búsqueda y la multivariabilidad, presentados en las diferentes estudios Singh, Brownlee, and Cairns (2022), Bettinger, Brun, and Boyer (2022) y Gonzales, Ortega, and Blanco (2022).

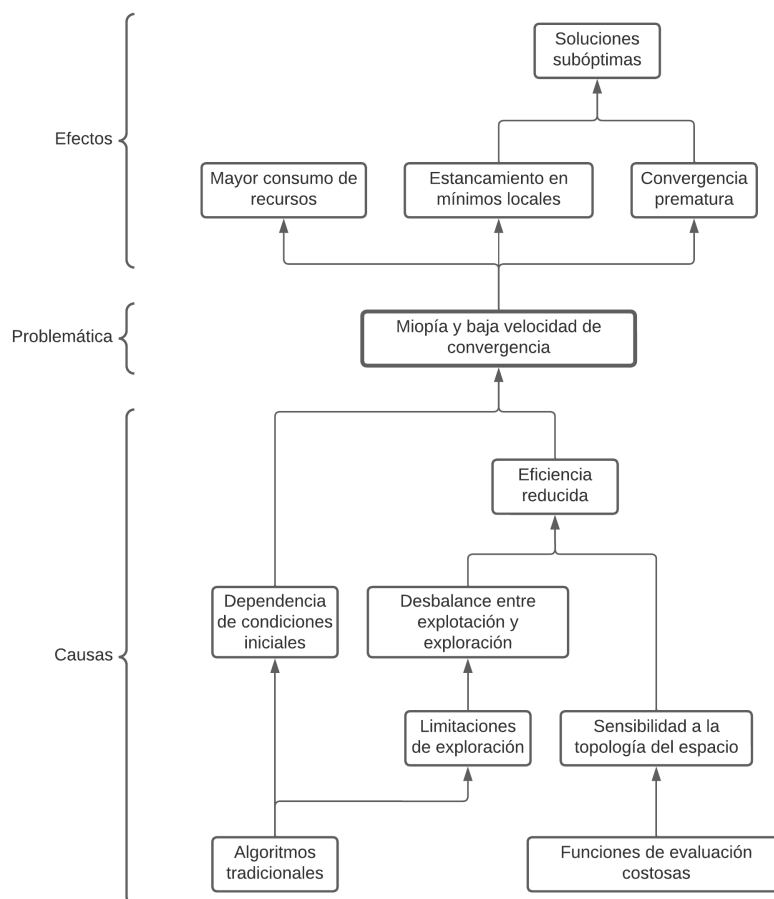


Fig. 1: Árbol de formulación del problema (optimización)

Adicionalmente, reforzando los objetivos de las investigaciones en este campo, el Ministerio de Minas y Energía en Guzmán and Sánchez (2022) plantea la importancia de la

generación de energía eléctrica renovable, la optimización y su papel en el cumplimiento de los objetivos de desarrollo sostenible propuestos por la ONU para el año 2030 en Colombia, surge la necesidad de enfocar los avances de la industria a las energías renovables.

En los últimos años, más del 70% de la matriz energética de Colombia se basa en el uso de hidroeléctricas (Pardo & Jiménez, 2021). La dependencia de una sola fuente de energía puede ser un riesgo en términos de seguridad energética y vulnerabilidad a los cambios climáticos. La energía eólica ofrece una forma de diversificar la matriz energética y reducir la dependencia de las hidroeléctricas. Además, con esto nace la exigencia de la optimización de parques eólicos y la predicción de patrones de viento en Colombia apoyada por Cortes-Pérez, Sierra-Vargas, and Arango-Gómez (2016) y Liapis and Kotsiantis (2022).

La implementación de técnicas de optimización a las energías eólicas trae consigo algunas limitaciones, presentes en el árbol de formulación del problema (Fig. 2), trabajadas en anteriores investigaciones. De acuerdo con Keswani, Verma, and Sharma (2022) y Klaiber and van Dinther (2023), las características estocásticas de las fuentes de energía renovables aumentan la complejidad e inestabilidad del problema a tratar. Además, la estimación inexacta de la velocidad del viento para granjas eólicas puede conducir a errores significativos en la predicción de la generación de energía eólica, estipulado por Kerem, Saygin, and Rahmani (2019).

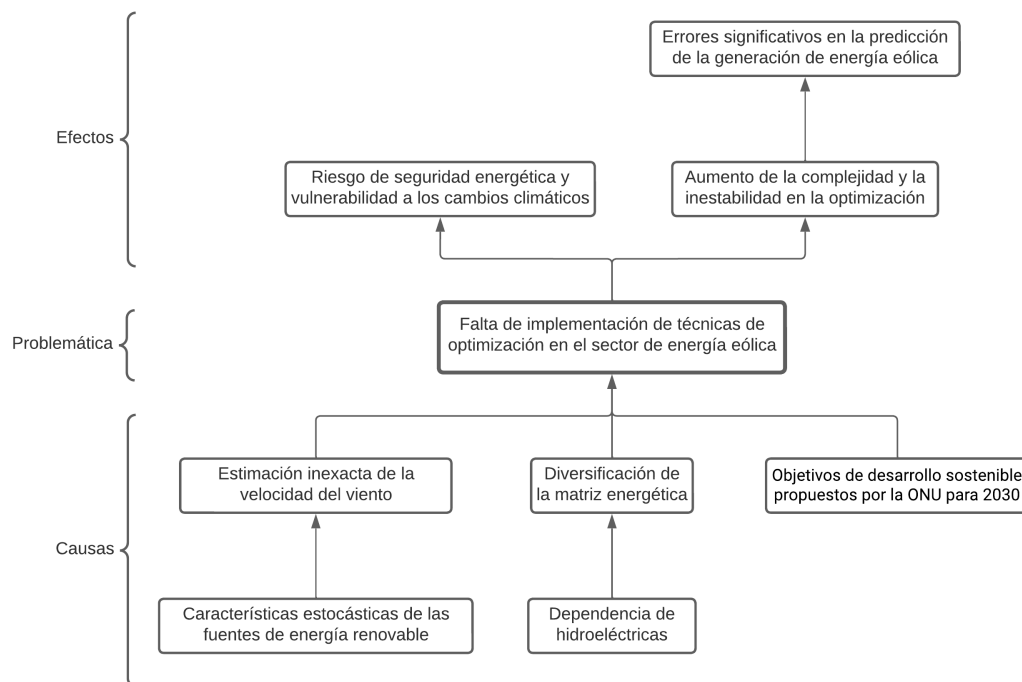


Fig. 2: Árbol de formulación del problema (energía eólica)

Objetivos

Objetivo general

Diseñar, implementar y validar un modelo de optimización metaheurístico que permita mitigar la miopía y mejorar la velocidad de convergencia con respecto a otros métodos de optimización y predicción del viento partiendo de métodos matemáticos y datos históricos del viento.

Objetivos específicos

- Identificar los componentes claves asociados con la optimización metaheurística y predicción del viento a partir de una revisión sistemática de la literatura relacionada.
- Desarrollar una arquitectura de la solución para la implementación de un nuevo método de optimización metaheurística que permita a partir de algoritmos genéticos, espacio de direcciones aleatorias y restricción al cono mitigar la miopía de los métodos tradicionales y mejorar la velocidad de convergencia de los métodos metaheurísticos existentes. Además, predecir la velocidad del viento a partir de datos reales del viento en Colombia.
- Desarrollar el prototipo de la solución del nuevo método de optimización y analítica de datos haciendo uso del lenguaje de programación Python.
- Validar el prototipo de la solución corroborando el funcionamiento en un ambiente controlado.

Metodología

En este proyecto de grado, se aplicará una metodología de investigación-formulación-acción, abordando la creación de un novedoso método de optimización metaheurística y la aplicación del método en la predicción del viento para el uso en energía eólica. Esta metodología tiene como objetivos manejar, inicialmente, los problemas como casos independientes para explorar un enfoque deductivo (de lo general a lo particular) y posteriormente integrarlos para presentar una solución conjunta cumpliendo los requerimientos y objetivos planteados.

Metodología matemática

Esta metodología se plantea como una variante al “proceso incremental” expuesto en Sommerville (2016), como se logra ver en Fig. 3. La discrepancia radica en los usuarios y partes involucradas inexistentes, reemplazados por un sistema de evaluaciones y retroalimentación basado en métricas de precisión. Se opta por esta metodología debido a que se estudia un problema de corta duración con un sistema de una sola funcionalidad. Además, se lleva un sistema de versiones incremental que busca siempre cumplir con uno de los

objetivos. El proceso incremental planteado con basa a la idea de desarrollar una implementación inicial, obtener una evaluación y crear una versión mejorada hasta que se haya desarrollado el sistema requerido. Repitiendo este proceso hasta cumplir con cada objetivo planteado.

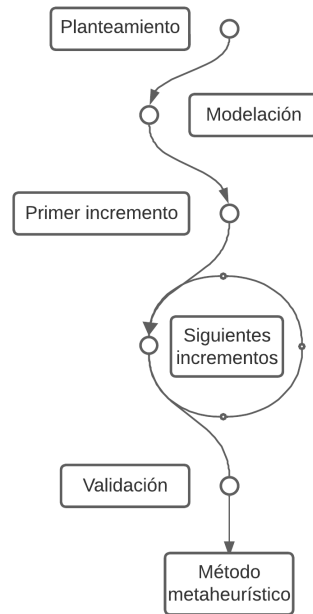


Fig. 3: Metodología matemática

En la primera fase se plantea la problemática con base a una revisión sistemática de la literatura haciendo una búsqueda sobre los métodos de optimización existente y su problema de miopía junto con la necesidad de mejora en su velocidad de convergencia. Se expuso las técnicas matemáticas para la solución de este tipo de problemáticas y se establecieron como requerimientos en la implementación del nuevo método. En adición, se presentaron los costos y pasos a seguir para llegar una solución viable.

Posteriormente, en la fase de modelación y primer incremento, se crea un modelo de la solución matemática como un prototipo matemático del algoritmo, además, se plantea la arquitectura y tecnologías para el desarrollo de la misma. En esta fase se crea un primer prototipo calculando los desafíos y riesgos que se podrán presentar en los incrementos posteriores. Los incrementos o ciclo de versiones de la metodología presenta una mejora del prototipo inicial en base a las métricas de evaluación planteadas, basadas en funciones de prueba en forma de valle las cuales son funciones o conjuntos de datos comunes utilizados para probar algoritmos de optimización (Jamil & Yang, 2013).

Finalmente, se compara el método obtenido con otros métodos en diferentes ambientes para distinguir su comportamiento, precisión y cumplimiento o punto de mejorar de los requerimientos planteados. Esta comparación se realiza tanto con métodos tradicionales de optimización como métodos metaheurísticos.

Metodología analítica

El proyecto consta de un proceso analítico basado en la metodología KDD (Knowledge Discovery and Data Mining), un análisis y modelado exploratorio y automático de grandes repositorios de datos, la cual implica la inferencia de algoritmos que exploran los datos, desarrollan el modelo y descubren patrones previamente desconocidos. El modelo se utiliza para comprender fenómenos a partir de datos, análisis y predicción (Maimon & Rokach, 2010). Esta metodología consta de cinco fases:

- *Búsqueda de información:* se define el problema a resolver y se establecen los objetivos del proyecto. También se identifican las fuentes de datos disponibles y se realiza una exploración preliminar de los datos para determinar su calidad y relevancia.
- *Obtención de datos:* recopilación los datos necesarios para el proyecto teniendo en cuenta la integración de datos de diferentes fuentes y la selección de las variables más relevantes para el caso en particular.
- *Depuración de la base de datos:* limpieza de los datos para eliminar valores atípicos, datos faltantes y errores de registro. También se pueden aplicar técnicas de transformación de datos para mejorar su calidad y relevancia, como la imputación con la mediana (para este caso).
- *Aplicación de técnicas de minería de datos:* descubrir patrones y relaciones en los datos. Esto puede implicar la aplicación de técnicas de clasificación, regresión, agrupamiento, asociación y detección de anomalías.
- *Interpretación y presentación de resultados:* los resultados obtenidos son interpretados y presentados de manera clara y comprensible para los usuarios finales, como un Dashboard. También se pueden realizar pruebas de validación para evaluar la calidad de los resultados y su capacidad para resolver el problema planteado.

Todas estas fases serán trabajadas y evaluadas a lo largo de la investigación al igual que investigaciones similares Torres-Quezada (2021).

Requerimientos

La etapa de requerimientos es crucial para el hilo conductor entre las dos metodologías y su correcta integración. Se planteo el siguiente modelo de requerimientos funcionales

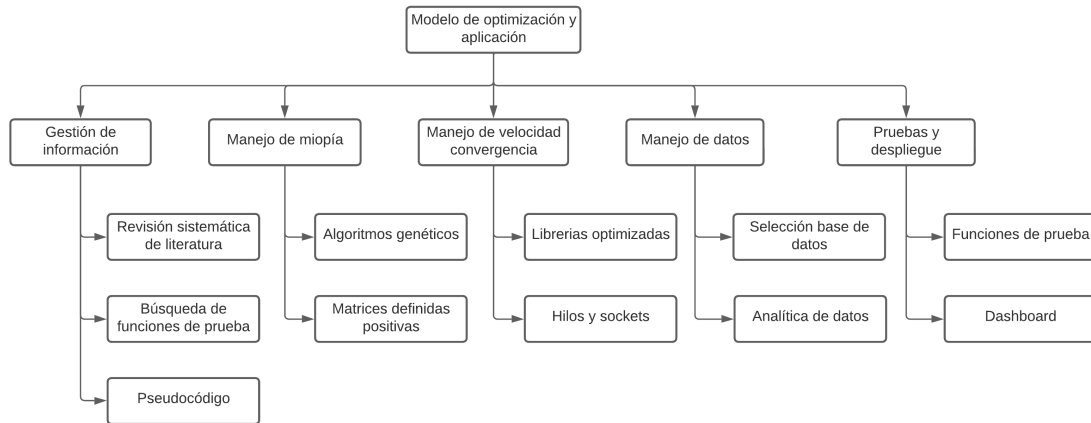


Fig. 4: Requerimientos funcionales

Marco teórico

La optimización metaheurística es un campo que ha evolucionado a lo largo de los años, explorando diferentes enfoques y técnicas para resolver problemas complejos. Inicialmente, se emplearon métodos de búsqueda local iterativa, como el Descenso del Gradiente o el Método de Newton, que partían de una solución inicial estimada y la mejoraban mediante pequeñas modificaciones. Más adelante, se desarrollaron algoritmos basados en población y estrategias evolutivas, que mantenían un conjunto de soluciones candidatas que se combinaban y seleccionaban siguiendo principios inspirados en la genética (Singh et al., 2022), (Schröter, Richter, & Wolter, 2018). También se incorporaron mecanismos de memoria para evitar visitar soluciones ya exploradas, como en la búsqueda tabú y memoria a corto plazo (Kang, Yang, & Youn, 2018).

Asimismo, se propusieron algoritmos basados en la naturaleza e inteligencia colectiva, que imitaban el comportamiento de sistemas biológicos o sociales, como las colonias de hormigas (Sammoud & Alaya, 2022), enjambre de partículas (Kerem et al., 2019). Finalmente, en las últimas décadas se han propuesto variantes, extensiones y combinaciones de estas técnicas, así como nuevos paradigmas inspirados en otros fenómenos naturales o artificiales, como las redes neuronales (Jimenez Navarro, Martinez Ballesteros, Sofia Brito, Martinez Alvarez, & Asencio Cortes, 2023), algoritmos meméticos (Schjølberg, Bekkevold, Sánchez-Díaz, & Mengshoel, 2023), y la propuesta presentada en este proyecto.

En investigaciones previas Santillán et al. (2008), Jiang et al. (2019), se ha observado que los algoritmos tradicionales a menudo sufren de miopía, lo que limita su capacidad para encontrar soluciones óptimas en problemas de optimización complejos. Los métodos tradicionales de optimización a menudo quedan atrapados en mínimos locales y tienen dificultades para explorar el espacio de soluciones de manera eficiente, y, de acuerdo con González et al. (2016), otros métodos solo son capaces de evaluar un puñado de valores

futuros. Esto ha llevado a la necesidad de desarrollar enfoques más avanzados y efectivos para abordar esta limitación y mejorar la velocidad de convergencia de los métodos metaheurísticos existentes.

También se han realizado estudios similares, como González et al. (2016) y Allawi et al. (2019), los cuales plantean nuevos métodos de optimización metaheurística buscando mitigar principalmente la miopía, enfocados en la minimización o maximización global y no local, aplicando diferentes métodos como las direcciones aleatoria o la restricción del espacio. Estos estudios intentan omitir los enfoques de Descenso del Gradiente, Método de Newton debido a que estos enfoques a menudo presentan dificultades para encontrar soluciones óptimas por si solos en problemas complejos. Por otro lado, según Jamil, Yang, and Zepernick (2013), varios proyectos hacen uso de espacios controlados para mirar la validez de los métodos haciendo uso de funciones de prueba comunes como las funciones Dixon-Price y Rosenbrock.

La hipótesis propuesta busca la combinación de generación de direcciones aleatorias, restricción al cono y algoritmos genéticos, para superar la miopía de los métodos tradicionales y mejorará la velocidad de convergencia en problemas de optimización por medio del uso de hilos y sockets. Además, se espera que este enfoque demuestre un mejor rendimiento en términos de encontrar soluciones de mayor calidad en comparación con los métodos existentes en la aplicación de parques eólicos.

Dentro de la revisión sistemática de la literatura se estudiaron diversas bibliotecas, revistas, documentación, principalmente IEEE Xplore y ACM Digital Library con ciertos criterios de búsqueda (Tabla 1), notando el bajo impacto de la metaheurística en los parques eólicos.

Fuente de Datos	IEEE Xplore	ACM Digital Library
Energía Eólica	78, 411	239, 152
Predicción de Viento & Optimización	1, 438	6, 824
Predicción de Viento & Metaheurística	13	194
Predicción de Viento & Metaheurística (≥ 5 años)	8	125

Tab. 1: Criterios de búsqueda

Marco conceptual

Se define la *optimización* como el proceso de ajustar o configurar variables para encontrar mejores propiedades, dentro de un espacio de soluciones, hacia un problema con un objetivo en específico, ya sea hacerlo más eficiente, más rápido, entre otros aspectos. Siguiendo a Wright (2006) y Hernández (2006), los problemas de optimización que trabajaremos en este caso tienen la estructura

$$x^* = \arg \min_x f(x)$$

conocidos como *problemas de minimización*, donde

$$\begin{aligned} f : \mathbb{R}^{n \times 1} &\longrightarrow \mathbb{R} \\ x &\longmapsto f(x) \end{aligned}$$

es la *función objetivo* (única para cada problema en cuestión) respecto al vector de parámetros x y el objetivo final es maximizar o minimizar la función, de acuerdo el interés en particular del problema.

Marti (2003) establece que podemos categorizar los problemas de optimización en tres secciones: métodos exactos, heurísticos y metaheurísticos. Este caso se concentra en los *metaheurísticos*, considerados, según Osman and Kelly (1996), estrategias generales que guían a una heurística subordinada ¹ en la resolución de problemas. Son algoritmos versátiles capaces de abordar situaciones donde no existe una solución específica o la implementación de un método óptimo no es viable. Estas destacan por su capacidad para evitar óptimos locales, centrarse en la búsqueda heurística e introducir elementos estocásticos en sus procesos.

Como se ha mencionado, en la literatura se ha encontrado que los métodos tradicionales presentan *miopía*. Una propiedad que caracteriza a algunos algoritmos por enfocarse en el corto plazo sin considerar las consecuencias a largo plazo (Marti, 2003). Esto se busca evitar, lo máximo posible, debido a que puede conducir a soluciones ineficientes o inviables en el largo plazo, subrayando la necesidad de evaluar el impacto a largo plazo de las soluciones propuestas.

Existen algunas técnicas ya utilizadas en literatura que han demostrado de manera eficiente la mitigación de la miopía, algunos de estos conceptos son las matrices definidas positivas que son base de direcciones aleatorias y a su vez de restricción al cono y algoritmo genético. Las *matrices definidas positivas* son matrices cuadradas en las que todos los autovalores son positivos útil para definir *direcciones aleatorias* por el vector gradiente como se aplica en el método del gradiente conjugado.

La *restricción al cono* emerge como una técnica matemática de optimización utilizada para encontrar el punto interior de un cono que minimiza una función objetivo. Esta técnica se basa en la premisa de que el punto mínimo de una función objetivo debe encontrarse en la frontera del cono (Boyd & Vandenberghe, 2004). Los *algoritmos genéticos* son técnicas de búsqueda y optimización basadas en principios genéticos y de selección natural. Estos algoritmos permiten que una población evolucione hacia un estado que maximice/minimice su aptitud, siendo una herramienta valiosa en problemas de optimización Santos Rodríguez and Aday Martínez (2014).

¹Algoritmo de búsqueda local que sigue una trayectoria en el espacio de búsqueda

Por otro lado, se resalta la optimización de *sistemas complejos* (caso presente), la cual implica encontrar la mejor solución a problemas que involucran múltiples factores e interacciones. Siendo estos sistemas, caracterizados por su comportamiento altamente no lineal y la interconexión de numerosas variables, lo que presentan un desafío significativo para encontrar soluciones óptimas, dada la diversidad de posibles soluciones y la complejidad para evaluar cuál es la mejor Le Thi, Le, and Dinh (2019).

Conjunto de datos

Como se ha establecido anteriormente, la energía eólica es el campo de aplicación de este proyecto. La mejor forma de estudiar este caso es con una base de datos que maneje la velocidad del viento y otras dimensiones como la ubicación, dirección, tiempo. Dentro de los Datos Abiertos Colombia encontramos una variedad de datos de viento y se ha seleccionado Velocidad Viento y Dirección Viento ofrecidos por el Instituto de Hidrología, Meteorología y Estudios Ambientales.

Estas bases de datos contienen más de 100 millones de registros desde el año 2001 hasta el presente (2023). Toma datos del viento como su dirección y velocidad cada 10 minutos, aproximadamente. Además, brinda la fecha de observación, ubicación geoespacial (latitud, longitud, departamento, municipio) de los sensores de captura. De manera específica la base de datos de velocidad del viento presenta

- *Descripción:* Velocidad del viento cada 10 minutos
- *No. de filas:* 131 millones
- *Columnas* (12): CodigoEstacion, CodigoSensor, FechaObservacion, ValorObservado, NombreEstacion, Departamento, Municipio, ZonaHidrografica, Latitud, Longitud, DescripcionSensor y UnidadMedida
- *Propietario:* Instituto de Hidrología, Meteorología y Estudios Ambientales
- *Sector:* Ambiente y Desarrollo Sostenible

Asimismo, la base de datos de dirección del viento presenta

- *Descripción:* Dirección del viento cada 10 minutos
- *No. de filas:* 101 millones
- *Columnas* (12): CodigoEstacion, CodigoSensor, FechaObservacion, ValorObservado, NombreEstacion, Departamento, Municipio, ZonaHidrografica, Latitud, Longitud, DescripcionSensor y UnidadMedida
- *Propietario:* Instituto de Hidrología, Meteorología y Estudios Ambientales
- *Sector:* Ambiente y Desarrollo Sostenible

Aquellos datos que no son relevantes de estudio en ese caso como son los códigos de sensores y estación, zona hidrográfica, unidad de medida y descripción de la muestra. Adicionalmente, ambas bases de datos especifican que los datos dispuestos son datos crudos instantáneos provenientes de los sensores de las estaciones automáticas de la red propia y/o producto de convenios interadministrativos con terceras entidades, y advierten de datos pueden presentar errores y/o inconsistencias estando incluso por fuera de los límites considerados normales, producto de fallas en los sensores de origen.

Planteamiento del problema de optimización

Para el desarrollo del método se considera cualquiera de las funciones de prueba anteriormente presentadas (Dixon-Price o Rosenbrock), las cuales satisfacen

$$f : \mathbb{R}^{n \times 1} \longrightarrow \mathbb{R}$$

donde $n \in \mathbb{N}$ es el número de variables involucradas. El problema de optimización se plantea con la intención de encontrar el valor x^* que minimiza la función $f(x)$

$$x^* = \arg \min_x f(x)$$

Podríamos hacer uso de métodos tradicionales para resolver este problema de optimización sin embargo, como se ha expuesto, la miopía puede degradar los resultados o reducir la velocidad de convergencia. En cambio, queremos proceder de la siguiente manera

A. Generación de direcciones

Se plantea un procedimiento iterativo que se fundamenta en el método de Newton y emplea la dirección de Newton, donde p^k es el vector que apunta en la dirección del gradiente negativo

$$p^{(k)} = -\nabla^2 f_k^{-1} \cdot \nabla f_k$$

El objetivo es generar un conjunto de direcciones aleatorias $\phi^{(k,i)}$, mediante

$$\phi^{(k,i)} = B_i \cdot p^{(k)}, \quad 1 \leq i \leq I$$

donde I es el número de direcciones con $I \ll n$, y B_i son matrices definidas positivas aleatorias.

B. Restricción al cono

Luego se plantea la restricción del espacio de búsqueda al cono

$$x^{(k+1)} = x^{(k)} + \Phi_k \cdot \alpha$$

donde la i -ésima columna $\Phi_k \in \mathbb{R}^{n \times I}$ es representada por $\phi^{(k,i)}$.

C. Segundo problema de optimización (algoritmo genético)

Posteriormente, en cada iteración k , queremos resolver el siguiente problema de optimización para calcular los pesos óptimos α^* ,

$$\alpha^* = \arg \min_{\alpha} f \left(x^{(k)} + \Phi_k \cdot \alpha \right)$$

Por consiguiente, nuestro problema de optimización general queda planteado de la siguiente manera

$$x^{(k+1)} = x^{(k)} + \Phi_k \cdot \alpha^*$$

Arquitectura de la solución

Arquitectura lógica

Recopilación de datos

Se aborda el uso de la energía eólica como una alternativa atractiva a los combustibles fósiles, centrándose en la optimización y toma de decisiones en granjas eólicas en zonas rurales. Para llevar a cabo este proyecto, es esencial la obtención de datos históricos del viento en Colombia, lo cual incluye información de velocidad y dirección proveniente de estaciones meteorológicas y registros de turbinas eólicas ubicadas en diferentes regiones del país. Los datos son recopilados de Datos Abiertos Colombia y es guardado en la nube en Google Sheets y BigQuery, buscando la seguridad de los datos.

Procesamiento de datos

Una vez recopilados los datos, se procederá a limpiar y prepararlos para su análisis. Esto implica eliminación de datos atípicos, imputación o eliminación de valores nulos, y realizar una exploración inicial para identificar posibles problemas de calidad, tal y como mencionan Riquelme Santos, Ruiz, and Gilbert (2006). Se utilizarán técnicas de limpieza de datos y estadísticas descriptivas, donde usaremos una aplicación BigQuery (Fernandes & Bernardino, 2015) la cual es una herramienta de almacenamiento de datos de Google Cloud que permite ejecutar análisis sobre grandes cantidades de datos en tiempo casi real facilitando así esta tarea .

Desarrollo del método metaheurística

Esta etapa es primordial, se llevará a cabo la creación del método de optimización metaheurística, con el objetivo de mitigar la miopía de los algoritmos tradicionales. Este algoritmo será fundamental en el contexto de la optimización de granjas eólicas, con muchos objetivos como lo es la minimización los efectos de la estela de las turbinas en el rendimiento (Kirchner-Bossi and Porté-Agel (2021)). Este método cuenta con generación de direcciones aleatorias, restricción al cono, algoritmo genético y un método de optimización dentro de otro buscando gran precisión. Este método es validado y comparado con otros dos métodos con funciones de prueba Dixon-price y Rosenbrock.

Para mejorar la capacidad predictiva del algoritmo, se llevará a cabo un proceso de entrenamiento. Esto implica ajustar el algoritmo utilizando los datos históricos de viento previamente recopilados y seleccionados. El algoritmo aprenderá de estos datos para hacer predicciones más precisas.

Evaluación del modelo

Una vez entrenado el algoritmo, se evaluará su precisión utilizando datos de validación. Se emplearán métricas específicas para medir el rendimiento del modelo, como el error cuadrático medio (MSE) la cual representa la pérdida promedio de esta discrepancia debido a la aleatoriedad o falta de información precisa (Lehmann & Casella, 2006) o el coeficiente de correlación de Pearson el cual ayudara a realizar esta tarea especialmente en aquellos escenarios en los que debe ser interpretada correctamente o en los que se tienen que comprobar las suposiciones matemáticas que la sustentan (Lalinde et al., 2018) como lo es este desarrollo. Si es necesario, se realizarán ajustes adicionales en el algoritmo para mejorar su rendimiento.

Generación de predicciones

Con el algoritmo entrenado y evaluado, se utilizará para generar pronósticos de patrones de viento en Colombia. Estos pronósticos serán de gran utilidad para la toma de decisiones en la industria eólica, la planificación energética y la gestión de recursos naturales.

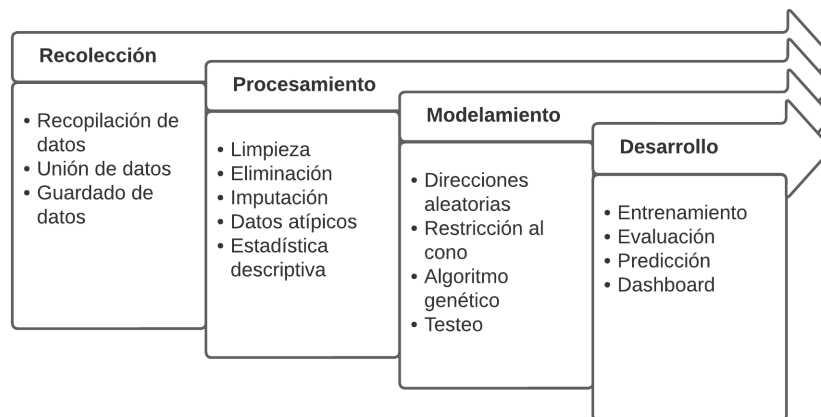


Fig. 5: Arquitectura lógica de la solución

Arquitectura física

Se plantea una arquitectura física donde se realiza la lectura de los datos desde una API que provee archivos .csv con información del viento en Colombia. Estos archivos se almacenan en Google Drive, donde se pueden acceder desde BigQuery, una herramienta de Google Cloud Platform que permite realizar consultas SQL sobre grandes volúmenes de

datos. En BigQuery se realiza el análisis y procesamiento de los datos, aplicando técnicas de limpieza, transformación, agregación y visualización.

Luego, se utiliza Looker Studio, una plataforma de inteligencia de negocios que se integra con BigQuery, para crear un dashboard interactivo que muestra los resultados del análisis. El dashboard se desarrolla con Python, utilizando los frameworks django y graphql para crear una aplicación web que se comunica con Looker Studio mediante una API. Finalmente, se integra el modelo creado para realizar las predicciones del viento en el dashboard. El modelo se entrena con los datos procesados en BigQuery y se evalúa con ciertas métricas de rendimiento.

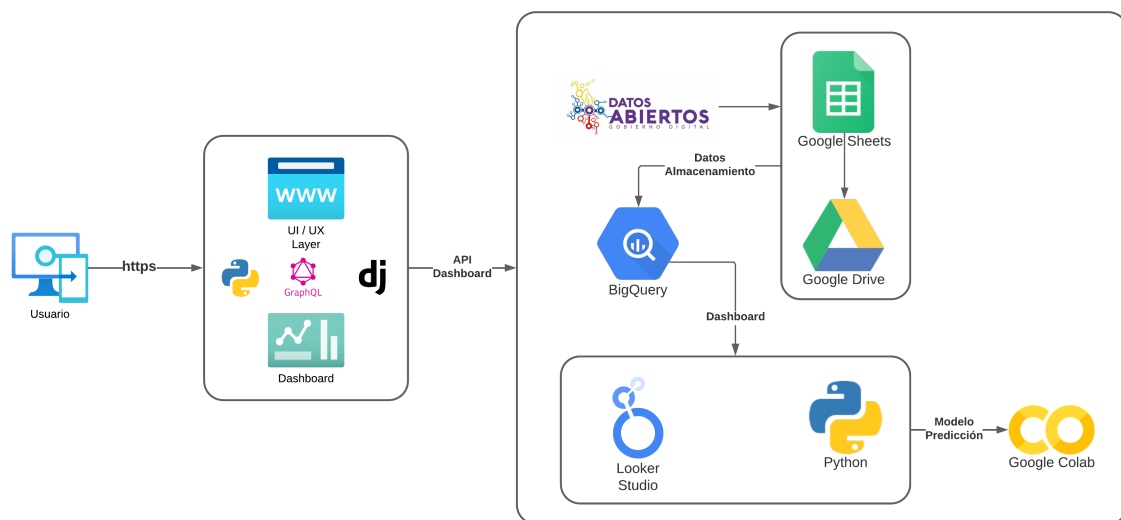


Fig. 6: Arquitectura física de la solución

Prototipo

Prototipo método de optimización

Antes de entrar en contexto del método de optimización creado hacemos mención de las librerías y bibliotecas de Python utilizadas para cumplir con este proceso y se hace una breve explicación de ellas en la siguiente tabla:

Librería / biblioteca	Descripción	Uso
SciPy	Construida sobre NumPy y proporciona funcionalidades adicionales para operaciones científicas y técnicas.	Construcción de matrices dispersas, almacenamiento ligero de matrices y resolver sistemas de ecuaciones lineales.
NumPy	Permite realizar operaciones numéricas en matrices multidimensionales o vectores.	Operaciones matemáticas de manera eficiente sobre matrices y vectores con dimensiones grandes.
itertools	Proporciona herramientas eficientes para trabajar con iteradores, que son objetos que pueden ser iterados uno a la vez.	Repeticiones finitas de variables para subprocesos.
Concurrent	Proporciona una interfaz de alto nivel para la ejecución concurrente de tareas.	Se aprovechan múltiples procesadores o núcleos de CPU para ejecución de tareas.
time	Proporciona diversas funciones relacionadas con el tiempo.	Medir el tiempo de ejecución de un programa.
Matplotlib	Herramienta de trazado 2D que permite crear gráficos de alta calidad y visualizaciones interactivas en diversos formatos.	Gráfico de funciones comparativas del gradiente negativo.

Tab. 2: Librerías y bibliotecas empleadas

Se plantea cumplir con el objetivo de desarrollar un algoritmo eficiente el cual recibe de entrada una función objetivo

$$\begin{aligned}
 f : \mathbb{R}^{n \times 1} &\longrightarrow \mathbb{R} \\
 x &\longmapsto f(x)
 \end{aligned}$$

respecto al vector x , y retorna un vector x^* que lo llamaremos x optimizado. Se aspira encontrar soluciones subóptimas u óptimas (en el mejor de los casos) a problemas con el mínimo conocimiento sobre su función objetivo, es decir, una independencia de los valores iniciales de entrada.

Siendo uno de los objetivos de nuestro método el mejorar la velocidad de convergencia en comparación con otros métodos de optimización, especialmente los algoritmos meta-heurísticos, se plantean ciertas estrategias para cumplirlo. Para ello se procede a hallar, para respectiva función objetivo, el vector gradiente y la matriz Hessiana de forma “manual”, encontrando patrones que simplifiquen su estructura sin afectar los resultados. Tomando esta información se encuentra $p^{(k)}$ (dirección de Newton), valor necesario para

encontrar el conjunto de direcciones aleatorias que se plantea como un algoritmo de generación de matrices definidas positivas enfocado, nuevamente, al objetivo de mejorar la velocidad de convergencia del método.

```
def generate_matrix(n, delta):
    data = rand(n*(n+1)//2 - ((n-(delta-1))*(n-delta)//2))
    I = concatenate([ones(delta)*i for i in range(n-(delta-1))]).astype(int)
    I1 = concatenate([ones(n-i)*i for i in range(n-delta+1, n)]).astype(int)
    J = concatenate([range(i, i+delta) for i in range(n-(delta-1))]).astype(int)
    J1 = concatenate([range(i, n) for i in range(n-delta+1, n)]).astype(int)
    A = coo_matrix((data, (concatenate([I, I1]), concatenate([J, J1]))))
    return A.toarray()
```

En este algoritmo el termino n es el tamaño del vector x y $delta$ (δ), donde $\delta \ll I \ll n$, se toma como el número de diagonales principales de la matriz ($n \times n$) a momento de ser rellenar. Las matrices definidas positivas cuentan con la propiedad de que todos sus autovalores son positivos, importante para asegurar que las direcciones que construimos sean útiles al nuestro problema de optimización. En particular, nos permite garantizar que las direcciones no nos lleven en una minimización equivocada y nos alejen de una solución óptima. En otras palabras, la funcionalidad de las matrices definidas positivas es garantizar que las direcciones generadas estén en una región convexa al espacio de búsqueda.

Continuando en el proceso de mejorar la velocidad de convergencia, dentro del algoritmo principal, se manejan conceptos de Thread y Sockets, como se aprecia en el siguiente código para la generación de la matriz $\Phi_k \in R^{n \times I}$ de dirección aleatorias.

```
with ThreadPoolExecutor(max_workers=I) as executor:
    Phi_k = array([*executor.map(
        directions_set,
        repeat(n),
        repeat(pk),
        repeat(delta, I)
    )]).T
```

Por otra parte, se plantea la restricción al cono para garantizar que la solución a encontrar esté dentro del espacio de búsqueda restringido y para evitar que el método de optimización se salga de este espacio.

```
Phi_k = Phi_k/norm(Phi_k)
```

Posteriormente, se plantea el algoritmo genético, para resolver el segundo problema de optimización. Este método es muy parecido a los algoritmos genéticos comunes, sin embargo, trabaja con una generación de vecinos o candidatos modificando una posición del vector $alpha$ (α) de mejor genética, buscando la creación de candidatos con mejor respuesta al problema de optimización. Como este segundo problema busca minimizar la función inicial entonces se planteo también una función objetivo específica

```
space = linspace(-1.0, 1.0, 10**4)
def generate_neighborhood(alpha):
```

```

pos_alpha = randint(alpha.shape[0])
new_space = choice(space, 1)[0]
alpha[pos_alpha] = new_space
return alpha, pos_alpha
def objective(function, xk, Psi_k, alpha):
    g = norm(-function.gradient(xk + matmul(Psi_k, alpha)))
    return g

```

Recordamos las funciones de prueba en forma de valle (Tabla 3) para el método a desarrollar. Estas serán nuestras funciones objetivo en la fase de desarrollo de este nuevo método de optimización y, también, utilizadas en la comparación con otros métodos de optimización.

	Dixon-Price	Rosenbrock
Función	$(x_1 - 1)^2 + \sum_{i=2}^n [i(2x_i^2 - x_{i-1})^2]$	$\sum_{i=1}^{n-1} 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2$
Mínimo global	$f(x) = 0$ para $x_i = 2^{-\frac{2^i-2}{2^i}}$	$f(x) = 0$ para $x^* = (1, 1, \dots, 1)$

Tab. 3: Funciones de prueba

En adición, se han propuesto dos enfoques distintos para la comparación, siendo el primero el método convencional de Newton y el segundo un método clásico de minimización implementado en Python. Estas estrategias han sido propuestas con la finalidad de exhibir un escenario habitual de optimización, especialmente seleccionado para abordar problemáticas de este tipo. Cabe destacar que la elección de estos métodos obedece a la intención de proporcionar una evaluación representativa, permitiendo así un análisis de casos de usos. Estos métodos requieren la función a minimizar, la dimensión de la variable independiente y una aproximación inicial.

```

def classic_newton_method(function, n, inicial_xk):
    xk = inicial_xk.copy()
    M = 100
    Gk = []
    for k in range(M):
        gk = -function.gradient(xk)
        Hk = function.hessian(xk)
        sk = spsolve(csc_matrix(Hk), gk)
        xk += sk
        gk = -function.gradient(xk)
        Gk.append(norm(gk))
    return Gk, xk

```

Este método busca encontrar el mínimo de una función mediante iteraciones. Durante cada iteración, se calcula el gradiente y la hessiana de la función en el punto actual, se resuelve un sistema de ecuaciones lineales para obtener la dirección de búsqueda, y se actualiza el punto actual sumándole esta dirección.

```

def optimize_alpha2(xk, Psi_k, Alpha, function):
    grad_f = lambda alpha: Psi_k.T @ function.gradient(xk + Psi_k @ alpha)
    hess_f = lambda alpha: Psi_k.T @ function.hessian(xk + Psi_k @ alpha) @ Psi_k

```

```

alpha_star = Alpha.copy()
max_iter = 100
for i in range(max_iter):
    grad = -grad_f(alpha_star)
    hess = hess_f(alpha_star)
    delta_alpha = solve(hess, grad)
    alpha_star += delta_alpha
return alpha_star

```

Por otro lado, el método *optimize_alpha2* es implementado para comparar el segundo problema de optimización creado. En este se utiliza un enfoque iterativo donde se calcula el gradiente y la hessiana de la función objetivo en un punto específico. Con estos cálculos, se resuelve un sistema de ecuaciones lineales para determinar el cambio necesario en el parámetro *alpha*. Este cambio se suma a la estimación actual de *alpha*, y el proceso se repite hasta alcanzar un límite máximo de iteraciones.

Finalmente, se planea un algoritmo de comparación de los tres distintos métodos de optimización donde tienen los mismo parámetros iniciales y la misma cantidad de iteraciones en cada método. Este algoritmo busca hacer una comparación de 3 tipos:

- Error: cálculo de la distancia entre el x optimizado y el valor real del mínimo de la función objetivo.
- Gradiente: menor dirección del gradiente negativo.
- Tiempo: tiempo de ejecución del método.

Estas tres comparaciones indica las ventajas y desventajas de cada método, junto con conocer los ambientes en donde mejor se comportan y las decisiones a tomar sobre estos.

Prototipo analítica

La primera fase plantea la comprensión del problema, definir claramente los objetivos y comprender la pregunta de investigación. Teniendo un modelo de optimización que busca mitigar la miopía un problema estocástico seria recomendable para aplicar sobre este nuevo método, debido al impacto por su gran cantidad de valles sobre la función a definir. El propósito delineado consistió en la identificación de una base de datos que contuviera información de carácter estocástico, tal como los datos relacionados con fenómenos meteorológicos o naturales, específicamente la velocidad del viento. El objetivo principal fue analizar estos datos con el fin de descubrir patrones y tendencias significativas que caracterizan este fenómeno, con el propósito de realizar predicciones lo más precisas sobre la velocidad del viento.

Estos datos fueron obtenidos de Datos Abiertos Colombia y almacenados de primera mano en BigQuery como datos crudos. Después, se generó una Service Account, de los servicios de Google, con permisos de propietario para permitir la lectura y escritura de

información con herramientas de compilación de datos como lo es dbt y posteriormente con Python.

Se planteo un Pipeline (Figura 7) para todo el procesamiento de datos e ingeniería de características (limpieza, eliminación, imputación, datos atípicos y estadística descriptiva), de la siguiente manera:

- Eliminación de acentos, diacrítico y normalización del alfabeto, además, de manejar una sola estructura de frases capitalizadas.
- Unificación de los departamentos donde los departamento con diferentes nombres son unificados para mantener una sola identificación.
- Creación de categoría para la identificación de las regiones en base a los departamentos existentes.
- Eliminación de datos atípicos como lo son municipios (Bello, Puerto Gaitan, Ayapel, Pereira, Purace (Coconuco)) y velocidad mayor que 0.
- Promedio de datos por hora para un mayor control sobre los datos, ya que los datos cada 10 minutos fueron algo inciertos y poco manipulables.

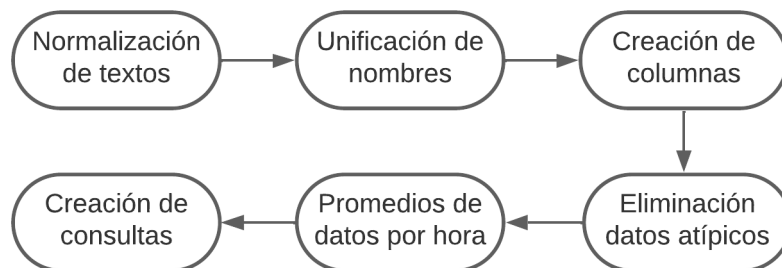


Fig. 7: Pipeline de ingeniería de características

Una vez finalizada la limpieza, se unen las tablas de Velocidad Viento y Dirección Viento y se crean consultas donde almacena la información relevante, lista para ser utilizada en la visualización y modelamiento. Estas consultas ofrecen la siguiente información:

- Información general de Colombia: número de municipios, número de departamentos, número de regiones, número de estaciones, número de registros.
- Promedio de la velocidad del viento por hora del día.
- Promedio de la velocidad del viento por mes del año.

- Información de las regiones, departamentos y municipios: nombre, velocidad media, velocidad mediana, velocidad máxima, velocidad mínima, desviación estándar de la velocidad, dirección media, dirección mediana, desviación estándar de la dirección.
- Información de las regiones, departamentos y municipios por mes del año: nombre, velocidad media, velocidad mediana, velocidad máxima, velocidad mínima, desviación estándar de la velocidad, dirección media, dirección mediana, desviación estándar de la dirección.

Cabe aclarar que todas las imputaciones/promedio sobre la velocidad se hizo a partir de la mediana, debido a que los datos presentan datos atípicos (nuestro caso), por lo que la media puede verse afectada significativamente, mientras que la mediana no. Por lo tanto, la mediana es una medida más robusta y adecuada para trabajar con problemas estocásticos como la velocidad del viento.

La fase de visualización de datos se ha desarrollado con Django, un framework de desarrollo web en Python, y GraphQL por medio de Graphene en Python, un lenguaje de consulta y manipulación de datos para APIs. Estas dos herramientas junto con la Service Account encapsula los datos del BigQuery para realizar sus procesos de lectura sin preocuparnos de acceso o manipulación incorrecta de los datos. Además, permite la solicitud de solo datos que son requeridos, lo que reduce la recuperación excesiva o insuficiente de datos. Todas las consultas creadas en la fase de ingeniería de características son leídas con Python y transformados en objetos de tipo Graphene los cuales guarda una estructura, esto nos permitirá crear consultas en un solo endpoint con GraphQL y realizar una toma de información limpia y consistente. Con esta información se plantean diversos gráficos, tablas, tarjetas y material visual para la facilidad de interpretación de lo usuarios.

De manera simultanea a la creación del dashboard, se procede a la desarrollo del modelo de Machine Learning haciendo uso del nuevo enfoque. Para ello se plantea una función objetivo en base a nuestro problema estocástico de tablas, donde se busca predecir la hora $k + 1$ a partir de una hora k junto con variables que nos permiten describir el espacio como lo son la latitud, longitud, dirección del viento, día del mes, entre otras. Es decir,

$$X_{k+1} = X_k M_{k+1, k} + \varepsilon, \quad 0 \leq k \leq 23$$

donde $X_{k+1} \in \mathbb{R}^{n \times 1}$ son las variables de respuesta (velocidad de la hora $k + 1$), $X_k \in \mathbb{R}^{n \times m}$ son las variables de entrada (velocidad de la hora k , latitud y longitud), $M_{k+1, k} \in \mathbb{R}^{m \times 1}$ representa nuestro x (parámetros) y $\varepsilon : \mathbb{R}^{m \times 1} \rightarrow \mathbb{R}^{n \times 1}$ es una función de errores que puede ser representada de la siguiente manera

$$\varepsilon(M_{k+1, k}) = X_{k+1} - X_k M_{k+1, k}$$

luego, la función objetivo planteada a partir de este problema de optimización es un medio de la norma euclidiana al cuadrado de la diferencia de la variable respuesta y las variables

de entrada por los parámetros

$$\begin{aligned}
f(M_{k+1}, k) &= \frac{1}{2} \|X_{k+1} - X_k M_{k+1, k}\|_2^2 \\
&= \frac{1}{2} (X_{k+1} - X_k M_{k+1, k})^T (X_{k+1} - X_k M_{k+1, k}) \\
&= \frac{1}{2} [X_{k+1}^T X_{k+1} - X_{k+1}^T (M_{k+1, k} X_k) - X_{k+1} (M_{k+1, k} X_k) + (X_k M_{k+1, k})^T (X_k M_{k+1, k})] \\
&= \frac{1}{2} (X_{k+1}^T X_{k+1} - 2X_{k+1}^T X_k M_{k+1, k} + M_{k+1, k}^T X_k^T X_k M_{k+1, k})
\end{aligned}$$

posteriormente se halla el vector gradiente y la matriz hessiana de la función

$$\begin{aligned}
\nabla f(M_{k+1}, k) &= \nabla \left[\frac{1}{2} (X_{k+1}^T X_{k+1} - 2X_{k+1}^T X_k M_{k+1, k} + M_{k+1, k}^T X_k^T X_k M_{k+1, k}) \right] \\
&= \frac{1}{2} \nabla (X_{k+1}^T X_{k+1} - 2X_{k+1}^T X_k M_{k+1, k} + M_{k+1, k}^T X_k^T X_k M_{k+1, k}) \\
&= \frac{1}{2} (0 - 2X_{k+1}^T X_k + 2M_{k+1, k}^T X_k^T X_k) \\
&= -X_{k+1}^T X_k + (X_k M_{k+1, k})^T X_k \\
\nabla^2 f(M_{k+1}, k) &= \nabla [-X_{k+1}^T X_k + (X_k M_{k+1, k})^T X_k] \\
&= X_k^T X_k
\end{aligned}$$

estas funciones son las parámetros de entrada para el nuevo método propuesto con anterioridad. Así, la respuesta de nuestro nuevo método sera la optimización de $M_{k+1, k}$ el cual sera nuestro x optimizado y con esto se han planteado las siguientes funciones

```

def function(X, x):
    return X @ x
def get_gradient(x):
    return matmul(-y.T, X) + matmul(matmul(X, x).T, X)
def get_hessian(x):
    return X.T @ X

```

Resultados

Resultados método de optimización

Para la prueba y comparación de las funciones se hace uso de los parámetros iniciales $n = 10^3$, $I = 10^1$ y un vector inicial $x = 1$. Además, para cada método se estableció un máximo de 100 iteraciones sin límite de tiempo. Como se planteo anteriormente, se consideran las funciones objetivos Dixon-Price y Rosenbrock, se probaran en dos casos: en la suposición de conocimiento de la función inicial, así indicar un x inicial acorde y el caso de no conocer la función por lo tanto nuestro x sera aleatorio.

Dixon-Price con valor inicial conocido

Función	Error	Menor gradiente	Tiempo (seg)
Nuevo enfoque	0.009003385	17.9854061	57.65051746
Método clásico de Newton	$4.577566798 \times 10^{-16}$	9.546575×10^{-14}	0.31700015
Otro método de optimización alpha	1.857657398	178.469461	41.92706418

Tab. 4: Métricas Dixon-Price con x_0 conocido

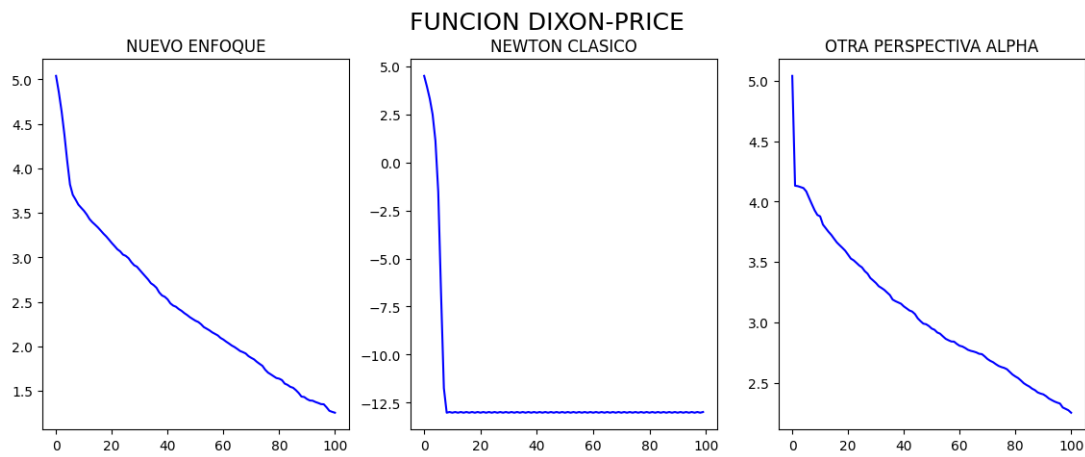


Fig. 8: Comparación Dixon-Price con x_0 conocido

Dixon-Price con valor inicial aleatorio

Función	Error	Menor gradiente	Tiempo (seg)
Nuevo enfoque	12.956994	8924.76281	53.59307
Método clásico de Newton	<i>nan</i>	6816112149.476	0.31699419
Otro método de optimización alpha	13.79384475	6842.91866	42.216299

Tab. 5: Métricas Dixon-Price con x_0 aleatorio

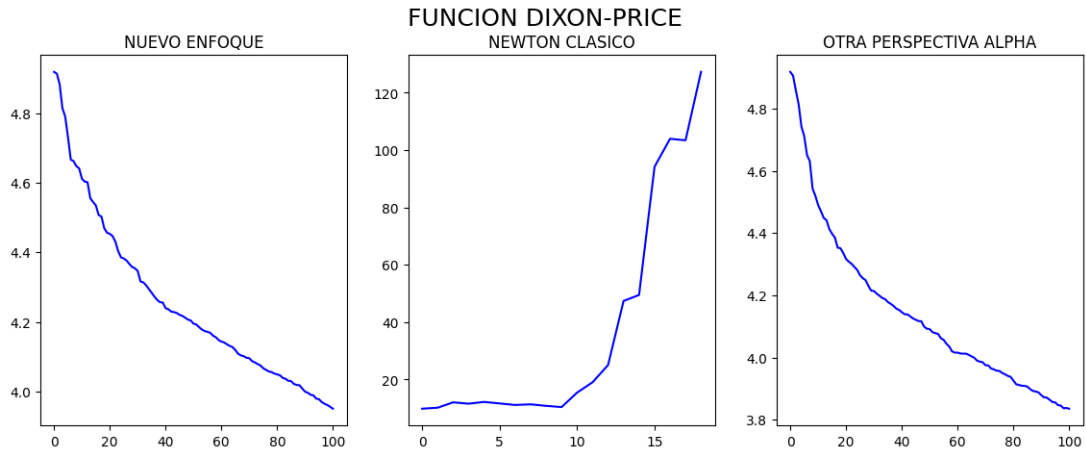


Fig. 9: Comparación Dixon-Price con x_0 aleatorio

Rosenbrock con valor inicial conocido

Función	Error	Menor gradiente	Tiempo (seg)
Nuevo enfoque	0.006027582	8.792572	55.8188989
Método clásico de Newton	$8.72214271 \times 10^{-15}$	$8.52651282 \times 10^{-14}$	1.9009981
Otro método de optimización alpha	0.007227568	12.35089553	92.6577289

Tab. 6: Métricas Rosenbrock con x_0 conocido

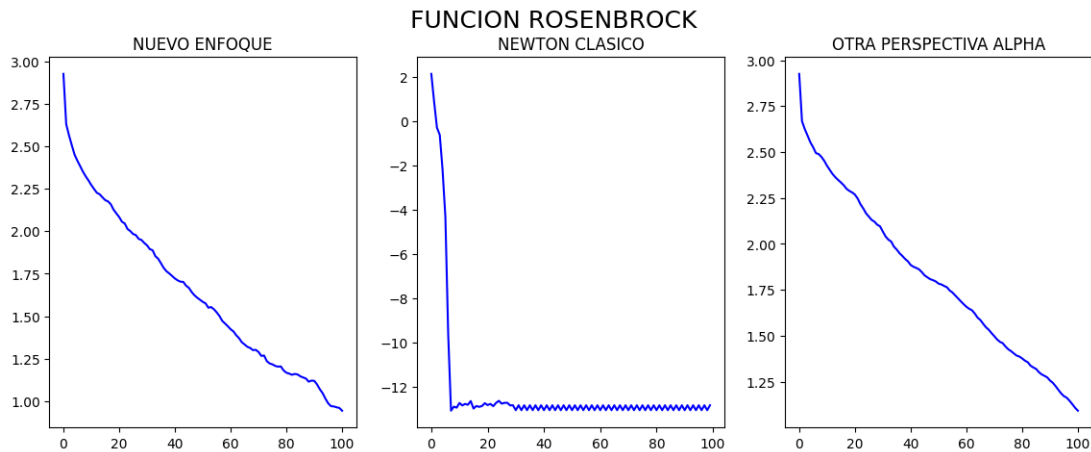


Fig. 10: Comparación Rosenbrock con x_0 conocido

Rosenbrock con valor inicial aleatorio

Función	Error	Menor gradiente	Tiempo (seg)
Nuevo enfoque	23.512680	1520.898389	53.0609996
Método clásico de Newton	139.44521	154327.3710	1.940001
Otro método de optimización alpha	24.710164	1087.86667	93.942814

Tab. 7: Métricas Rosenbock con x_0 aleatorio

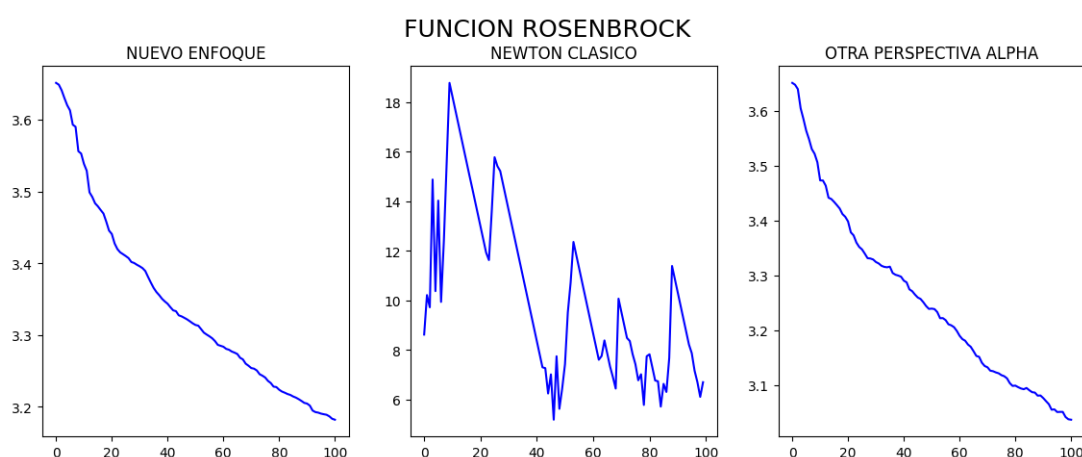


Fig. 11: Comparación Rosenbrock con x_0 aleatorio

A simple vista podemos apreciar como el método el método tradicional (Método de Newton) tiene un excelente comportamiento en los casos de valor inicial conocido, es decir, en los casos donde conocemos lo suficiente la función como para darle un valor inicial adecuado para encontrar su mínimo global. Pero a su vez se aprecia como en los casos donde hacemos la suposición de no tener los conocimientos suficientes de la función y se opta por establecer un valor inicial aleatorio, se aprecia como el método tradicional pierde su rumbo y busca repetidas veces un mínimo pero no logra alcanzarlo de forma global. Vemos directamente como en el caso de la función Dixon-Price con valor inicial aleatorio termina las iteraciones más temprano al quedar en valores infinitos o indeterminados.

Por otro lado, los resultados obtenidos sobre el nuevo método son los esperados, se logra apreciar como nuestro nuevo enfoque siempre conserva la característica de descenso independiente del contexto en el que se encuentre. En cada caso se dirige hacia un mínimo global, en algunos casos mejor que otros. Este método no se ve afectado por el ambiente o entorno en el que se presente la miopía no es un problema para este. A pesar de que el método tradicional cuenta con tiempos excepcionales se ve muy afectado por la miopía, y en comparación con el otro método metaheurístico, se aprecia en la mayoría de casos un mejor comportamiento de su velocidad de convergencia estando sobre $50 \leq \text{tiempo} \leq 60$

segundos, es decir, es independiente de la función y de los valores iniciales, se basa en el número de iteraciones.

Se ve claramente como el método que le ofrece otra perspectiva al segundo método de optimización α^* cuenta con un comportamiento muy similar gráficamente, sin embargo los datos nos ofrecen otra información. Haciendo una comparación de errores y gradientes con nuestro nuevo enfoque, al estar los dos sobre el mismo espacio de restricción al cono, podemos evaluar varios casos donde la norma del gradiente negativo es menor en la otra perspectiva de alpha que en nuestro método, no obstante, el error es mayor en todos los casos. Es decir, la nueva perspectiva de alpha se dirige más rápido hacia un mínimo que nuestro método pero este no es un mínimo global, sino un mínimo local. La miopía prevalece.

Resultados analítica de datos

Primeramente, los datos fueron integrados con el modelo de optimización creado, obteniendo resultados muy importantes y favorables en algunas cosas como se logra ver en la Figura 12 donde se tomo una muestra de 200 elementos y se presenta los valores reales (azul) vs los valores predichos (rojo). Se obtuvieron las siguientes métricas tomando como los valores de entrenamiento al año 2021 y 2022, aproximadamente tres millones de registros, como testeo se tomo el año 2023, aproximadamente un millón de registros:

- Métricas $R^2 \approx 0.71$: indicando un 71% de la variabilidad en la variable dependiente (viento hora $k + 1$) puede explicarse por el modelo. Un valor cercano a 1 es deseable, lo que sugiere que el modelo está capturando una proporción significativa de la variabilidad en los datos.
- Métrica MAE ≈ 0.51 : en promedio, las predicciones tienen un error absoluto de 0.51 unidades en la misma escala que la variable medida. Cuanto más bajo sea este valor, mejor será la precisión del modelo.
- Métrica MedAE ≈ 0.41 : representa la mediana de los errores absolutos. Es menos sensible a valores atípicos en comparación con el MAE.

Por otra parte, los datos obtenidos en la predicción por medio del método de Newton fueron mas acertados. Estos resultados muestran un mejor resultado por parte del gradiente encontrando un mínimo, desconocemos si es un mínimo local o global. Las métricas obtenidas fueron

- Métricas $R^2 \approx 0.83$
- Métrica MAE ≈ 0.34
- Métrica MedAE ≈ 0.23

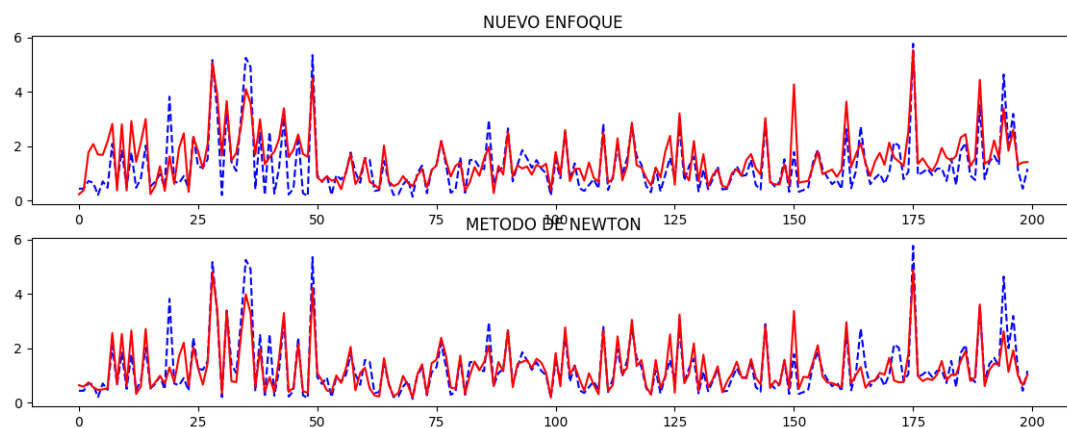


Fig. 12: Valores Verdaderos vs Valores Predichos

Estos resultados son positivos en la integración entre el modelo y este tipo de aplicación enfocada en la predicción del viento. Sin embargo, una gran parte de los datos no logro ser del todo bien predicho. Estos resultados pueden ser causados por diversas cosas, la primera, y más importante, es la restricción computacionales por la gran cantidad de registros con los que cuenta nuestra base de datos lo cual nos lleva a restringir estos datos y los ciclos de iteración de los modelos para no sobre cargar las maquinas o ambientes virtuales, otra posible causa es evidente debido al objetivo del método que ha sido desarrollado para problemas multidimensionales y este tipo de aplicación no ha sido multidimensional y varios parámetros del método dependen de la alta dimensionalidad.

Por otro lado, el dashboard desarrollado esta ha sido dividido en tres secciones con la información de Colombia:

Información general

De los 131 millones de registros con los que cuenta la base de datos se consideran los tres últimos años cuantificados en 29 millones de registros, aproximadamente. Luego del procesos de ingeniería de características el número de registros es 3083630, 5 regiones, 31 departamentos, 284 municipios y 357 estaciones (sensores de toma de datos). Además, se presenta la media y mediana de la velocidad del viento por hora y por mes.

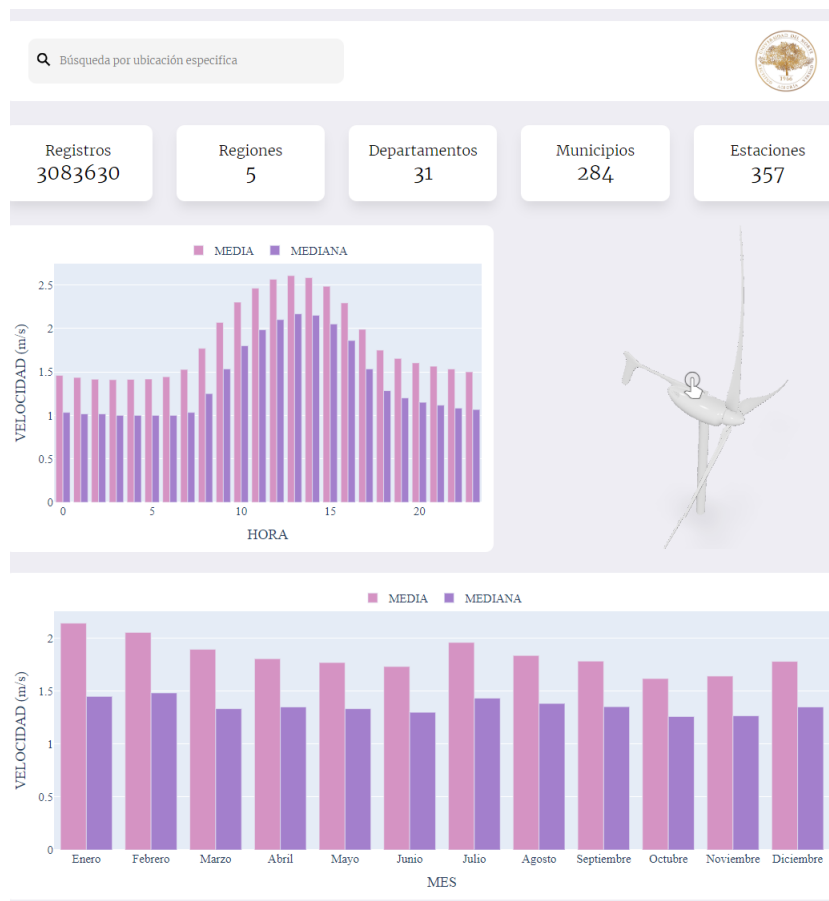


Fig. 13: Dashboard información general Colombia

Información por localidades

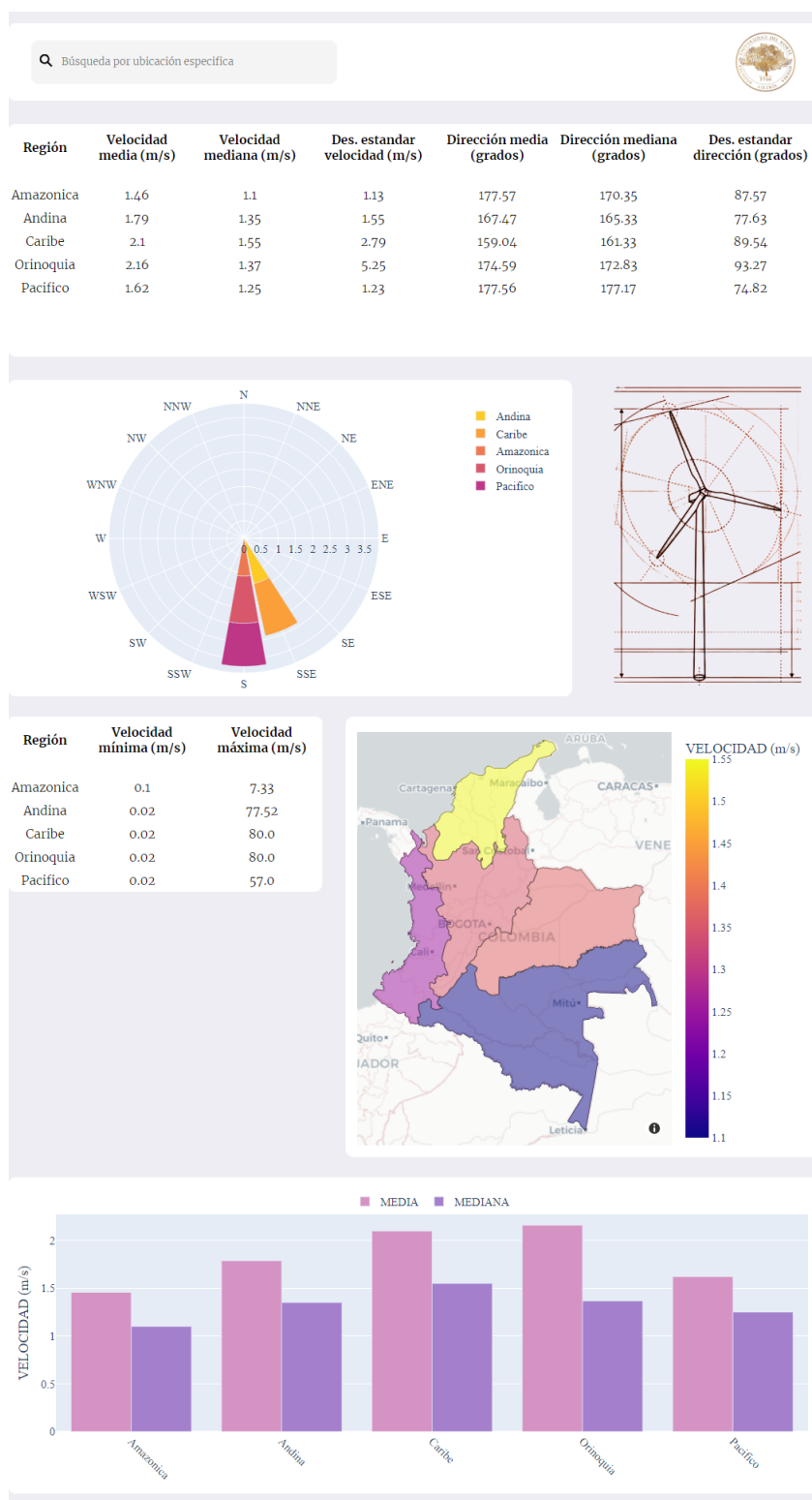


Fig. 14: Dashboard por regiones

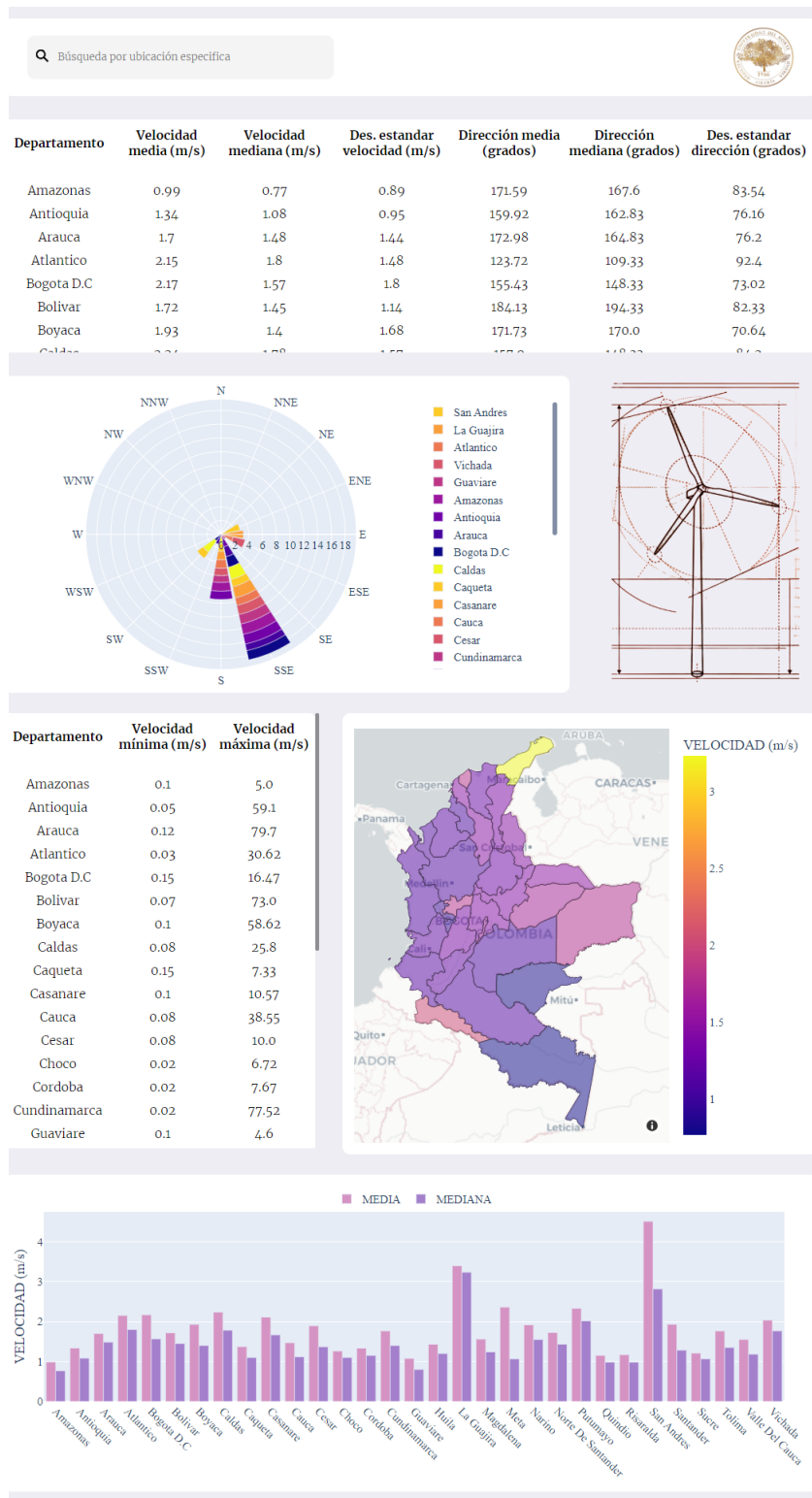


Fig. 15: Dashboard por departamentos

Información por mes y localidades

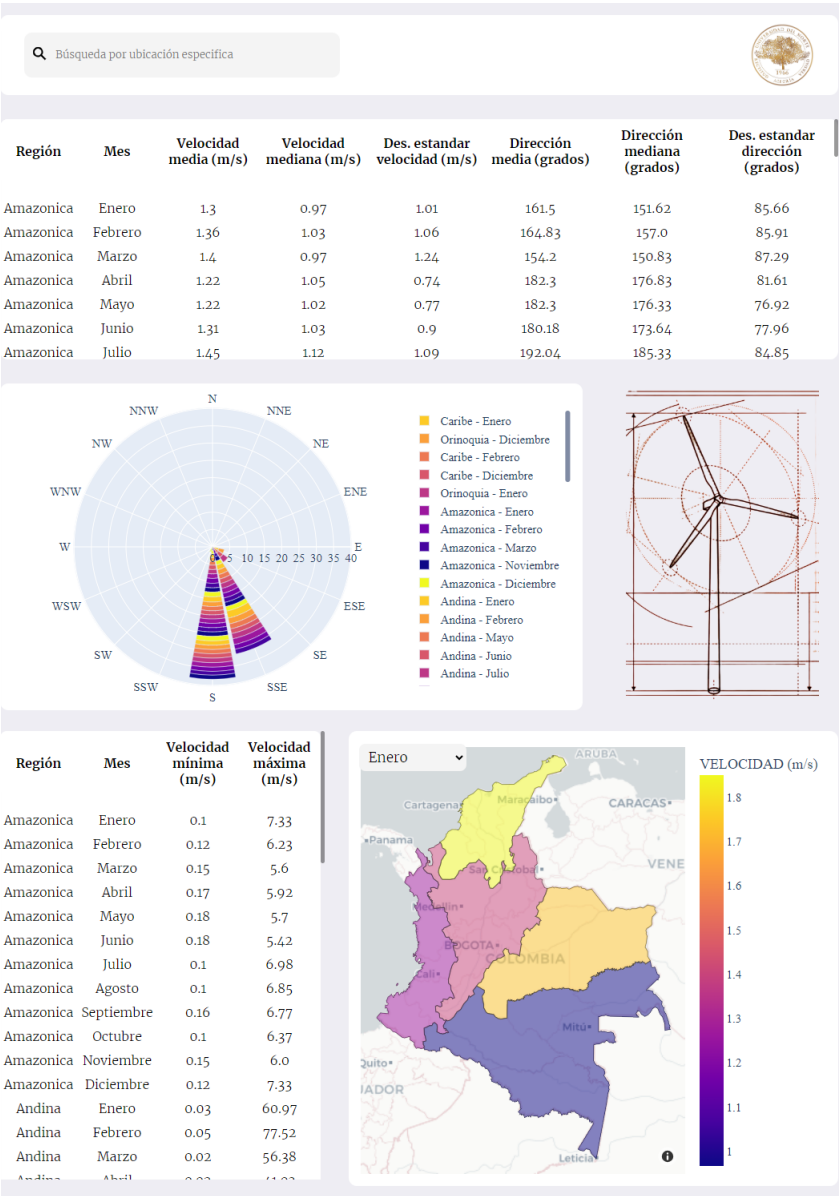


Fig. 17: Dashboard por región y mes

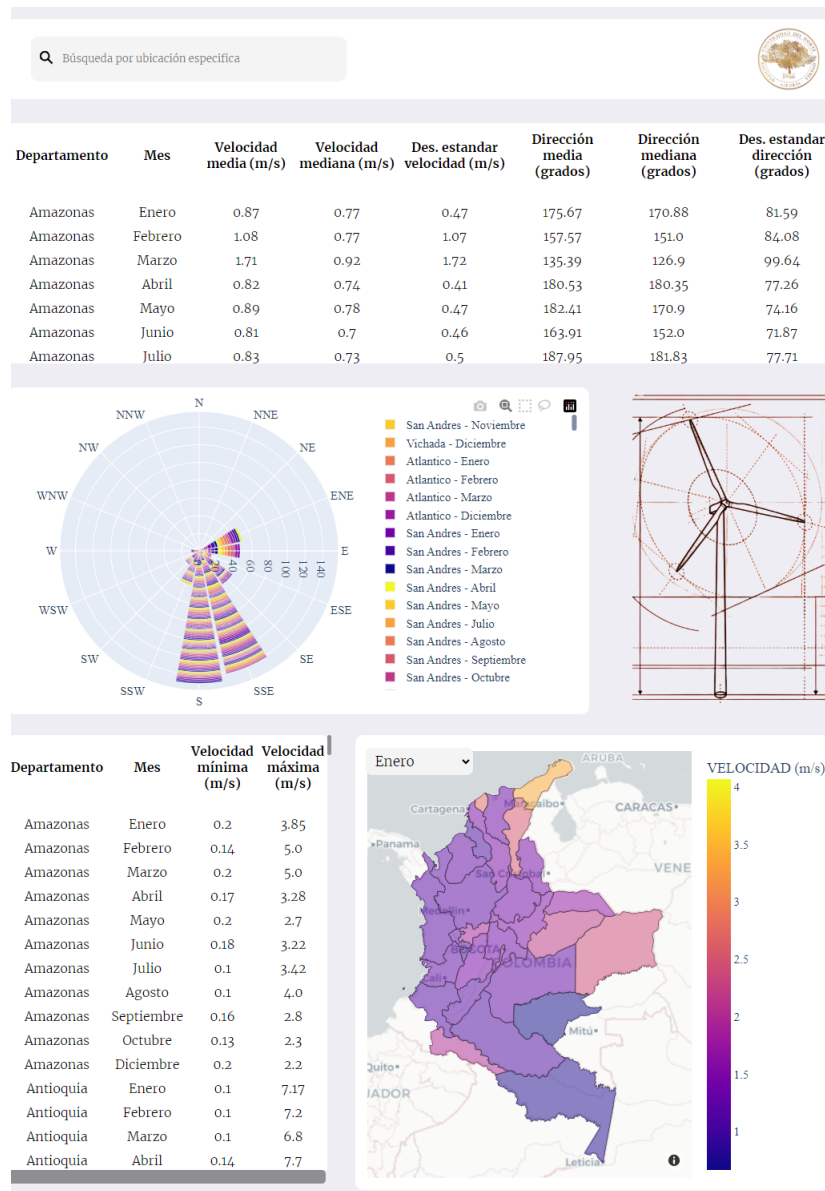


Fig. 18: Dashboard por departamento y mes

Conclusión

Se diseño, implemento y valido un nuevo método de optimización metaheurístico que permita mitigar la miopía, mejorar la velocidad de convergencia y predicción del viento. Este proyecto de grado presenta un pilar en la exploración y aplicación de técnicas de optimización avanzadas en el contexto de la energía eólica para Colombia. Sus hallazgos y contribuciones ofrecen una perspectiva valiosa para el avance tanto científico como tecnológico en la búsqueda de soluciones más eficientes y sostenibles en la generación de energía renovable.

La investigación llevada a cabo ha sido integral y significativa en el campo de la optimización metaheurística aplicada a la predicción del viento, con implicaciones sustanciales

en la eficiencia de las granjas eólicas y el desarrollo sostenible en Colombia. El proyecto se fundamenta en la necesidad imperiosa de superar las limitaciones de los métodos convencionales de optimización, enfocándose en la creación de un nuevo enfoque metaheurístico. Este método, creado a partir de la combinación de algoritmos genéticos, exploración aleatoria del espacio de búsqueda y restricción al cono, ha sido probado y validado con éxito en la predicción de vientos en Colombia.

Las contribuciones de este trabajo no se limitan al ámbito científico, sino que también abarcan esferas sociales y económicas. Se resalta la importancia de la optimización en sectores claves de la industria colombiana, como el consumo energético, el cemento, el carbón y la aviación comercial. La aplicación de estas técnicas no solo puede mejorar la eficiencia de las energías renovables, sino también abordar desafíos en la gestión de recursos y reducir costos, impulsando así la sostenibilidad empresarial y el desarrollo económico.

El enfoque innovador propuesto en este proyecto, probado en la optimización de funciones Dixon-Price o Rosenbrock, presenta un avance significativo al ofrecer una alternativa más robusta y efectiva a los métodos tradicionales en la resolución de problemas complejos de optimización. La arquitectura lógica y física diseñada para este proyecto proporciona un marco detallado que aborda cada etapa clave en el desarrollo y aplicación del método de optimización. Desde la recopilación y procesamiento de datos hasta la aplicación práctica del modelo desarrollado, se ha delineado un camino claro y completo para su implementación y validación.

El uso estratégico de librerías y herramientas de programación, como SciPy, NumPy, y otras, ha sido esencial para el éxito de este proyecto, permitiendo desde la construcción de matrices hasta la visualización y comparación de funciones. Los resultados obtenidos, aunque positivos en cuanto a la capacidad predictiva y la mejora de la velocidad de convergencia en comparación con otros métodos, señalan áreas de refinamiento. La comparación detallada entre métodos muestra la fortaleza del enfoque metaheurístico propuesto en entornos con información limitada sobre la función, destacando su capacidad para buscar mínimos globales de manera consistente. Sin embargo, se identifican desafíos computacionales y limitaciones en la precisión de los resultados, especialmente para entornos multidimensionales.

References

- Allawi, Z. T., Ibraheem, I. K., & Humaidi, A. J. (2019). Fine-tuning meta-heuristic algorithm for global optimization. *Processes*, 7(10), 657.
- Asghar, R., Anwar, M. J., Wadood, H., Saleem, H., Rasul, N., & Ullah, Z. (2023). Promising features of wind energy: A glance overview. In *2023 4th international conference on computing, mathematics and engineering technologies (icomet)* (pp. 1–6).
- Bettinger, A., Brun, A., & Boyer, A. (2022). Independent influence of exploration and exploitation for metaheuristic-based recommendations. In *Proceedings of the genetic and evolutionary computation conference companion* (pp. 475–478).
- Boyd, S. P., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press.
- Carreño Ávila, D. A. (2019). Propuesta de estrategia metodológica para la optimización de los procesos operacionales en las aerolíneas de colombia: caso de estudio avianca.
- Cerveleón, L. J., & Ledesma, J. F. (2022). Optimización de procesos logísticos del sector carbón usando técnicas heurísticas y metaheurísticas. *REVISTA COLOMBIANA DE TECNOLOGIAS DE AVANZADA (RCTA)*, 1(39), 93–99.
- Cortes-Pérez, D. M., Sierra-Vargas, F. E., & Arango-Gómez, J. E. (2016). Evaluación, predicción y modelación del potencial eólico. *Ingeniería Mecánica*, 19(3), 167–175.
- Dietz, T., Klamroth, K., Kraus, K., Ruzika, S., Schäfer, L. E., Schulze, B., ... Wiecek, M. M. (2020). Introducing multiobjective complex systems. *European Journal of Operational Research*, 280(2), 581–596.
- Ezugwu, S. A. N. R., A.E. (2021). Metaheuristics: a comprehensive overview and classification along with bibliometric analysis. *Artif Intell Rev* 54.
- Fernandes, S., & Bernardino, J. (2015). What is bigquery? In *Proceedings of the 19th international database engineering & applications symposium* (pp. 202–203).
- Gendreau, M., & Potvin, J.-Y. (2005). Metaheuristics in combinatorial optimization. *Annals of Operations Research*, 140, 189–213.
- Gonzales, J. E., Ortega, Á. C., & Blanco, R. R. (2022). Metaheurísticas con python: Casos prácticos. *Hatun Yachay Wasi*, 1(2), 43–57.
- González, J., Osborne, M., & Lawrence, N. (2016). Glasses: Relieving the myopia of bayesian optimisation. In *Artificial intelligence and statistics* (pp. 790–799).
- Gualtieri, G. (2020). Comparative analysis and improvement of grid-based wind farm layout optimization. *Energy Conversion and Management*, 208, 112593.
- Gutjahr, W. J. (2009). Convergence analysis of metaheuristics. In *Matheuristics: hybridizing metaheuristics and mathematical programming* (pp. 159–187). Springer.
- Guzmán, I. D. H., & Sánchez, J. W. G. (2022). Integración de métodos de optimización en la generación de energía eléctrica renovable. *Encuentro Internacional de Educación en Ingeniería*.
- Hernández, G. (2006). Métodos clásicos de optimización para problemas no-lineales sin restricciones. *UChile-Departamento de Ingeniería Matemática*, 1–14.
- Herrera-Quintero, L. F., Jalil-Naser, W. D., Banse, K., & Samper-Zapater, J. J. (2015). Smart cities approach for colombian context. learning from its experiences and linking with

- government organization. In *2015 smart cities symposium prague (scsp)* (pp. 1–6).
- Higuita Alzate, D. F. (2018). *Optimización del proceso de abastecimiento de cemento a granel en colombia, aplicación de irp con restricciones particulares (opac)* (Unpublished doctoral dissertation).
- Jamil, M., & Yang, X.-S. (2013). A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(2), 150–194.
- Jamil, M., Yang, X.-S., & Zepernick, H.-J. (2013). Test functions for global optimization: a comprehensive survey. *Swarm intelligence and Bio-inspired Computation*, 193–222.
- Jiang, S., Chai, H., González, J., & Garnett, R. (2019). Efficient nonmyopic bayesian optimization and quadrature. *arXiv preprint arXiv:1909.04568*.
- Jimenez Navarro, M. J., Martinez Ballesteros, M., Sofia Brito, I., Martinez Alvarez, F., & Asencio Cortes, G. (2023). A bioinspired ensemble approach for multi-horizon reference evapotranspiration forecasting in portugal. In *Proceedings of the 38th acm/sigapp symposium on applied computing* (pp. 441–448).
- Jin, W. H. S. C., Y. (2021). Classical optimization algorithms. *Data-Driven Evolutionary Optimization*.
- Kang, D.-K., Yang, E.-J., & Youn, C.-H. (2018). Deep learning-based sustainable data center energy cost minimization with temporal macro/micro scale management. *IEEE Access*, 7, 5477–5491.
- Kerem, A., Saygin, A., & Rahmani, R. (2019). Wind power forecasting using a new and robust hybrid metaheuristic approach: a case study of multiple locations. In *2019 19th international symposium on electromagnetic fields in mechatronics, electrical and electronic engineering (isef)* (pp. 1–2).
- Keswani, R., Verma, H., & Sharma, K. S. (2022). Optimal power flow integrating renewable energy sources in microgrid employing hybrid grey wolf-equilibrium optimizer. In *2022 ieee ias global conference on emerging technologies (globconet)* (pp. 714–720).
- Kirchner-Bossi, N., & Porté-Agel, F. (2021). Wind farm area shape optimization using newly developed multi-objective evolutionary algorithms.
- Klaiber, J., & van Dinther, C. (2023). Deep learning for variable renewable energy: A systematic review. *ACM Computing Surveys*.
- Lalinde, J. D. H., Castro, F. E., Rodríguez, J. E., Rangel, J. G. C., Sierra, C. A. T., Torrado, M. K. A., ... Pirela, V. J. B. (2018). Sobre el uso adecuado del coeficiente de correlación de pearson: definición, propiedades y suposiciones. *Archivos venezolanos de Farmacología y Terapéutica*, 37(5), 587–595.
- Lehmann, E. L., & Casella, G. (2006). *Theory of point estimation*. Springer Science & Business Media.
- Le Thi, H. A., Le, H. M., & Dinh, T. P. (2019). *Optimization of complex systems: theory, models, algorithms and applications* (Vol. 991). Springer.
- Liapis, C. M., & Kotsiantis, S. (2022). Energy balance forecasting: An extensive multivariate regression models comparison. In *Proceedings of the 12th hellenic conference on artificial*

intelligence (pp. 1–7).

- Maimon, O., & Rokach, L. (2010). Introduction to knowledge discovery and data mining. In O. Maimon & L. Rokach (Eds.), *Data mining and knowledge discovery handbook* (pp. 1–15). Boston, MA: Springer US. Retrieved from https://doi.org/10.1007/978-0-387-09823-4_1 doi: 10.1007/978-0-387-09823-4_1
- Marti, R. (2003). Procedimientos metaheurísticos en optimización combinatoria. *Matemáticas, Universidad de Valencia*, 1(1), 3–62.
- Osman, I. H., & Kelly, J. P. (1996). Meta-heuristics: an overview. *Meta-heuristics: Theory and applications*, 1–21.
- Pardo, L. G. A., & Jiménez, J. E. C. (2021). Energía eólica, una alternativa ambientalmente sostenible desde el ejército nacional de Colombia. *Brújula Semilleros de Investigación*, 9(17), 48–66.
- Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. *Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial*, 10 (29), 11–18.
- Sáchica, J. A. (2020). Metodología para la optimización del consumo energético bajo el análisis de eficiencia financiera con un alto impacto en la reducción de emisiones de GEI. caso exitoso de aplicación en el campo más grande de Colombia.
- Sammoud, S., & Alaya, I. (2022). A new ant colony optimization metaheuristic based on pheromone guided local search instead of constructive approach. In *Proceedings of the genetic and evolutionary computation conference* (pp. 13–21).
- Santillán, N. S., López, M. G., Ortiz, W. R., & Ruiz, S. G. (2008). Los límites del pronóstico newtoniano y la búsqueda del orden en el caos. *Ingeniería. Investigación y Tecnología*, 9(2), 171–182.
- Santos Rodríguez, C., & Aday Martínez, A. T. (2014). Algoritmos genéticos: una solución para la optimización del reflector parabólico. *Ingeniería Electrónica, Automática y Comunicaciones*, 35(1), 1–15.
- Schjøelberg, M. E., Bekkevold, N. P., Sánchez-Díaz, X., & Mengshoel, O. J. (2023). Comparing metaheuristic optimization algorithms for ambulance allocation: An experimental simulation study. In *Proceedings of the genetic and evolutionary computation conference* (pp. 1454–1463).
- Schröter, T., Richter, A., & Wolter, M. (2018). Development of methods for an optimized in-feed forecast of renewable energies. In *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS)* (pp. 1–6).
- Singh, M., Brownlee, A. E., & Cairns, D. (2022). Towards explainable metaheuristic: mining surrogate fitness models for importance of variables. In *Proceedings of the genetic and evolutionary computation conference companion* (pp. 1785–1793).
- Sommerville, I. (2016). *Software engineering*. isbn: 978-1-292-09613-1. Pearson, Essex CM20 2JE, England.
- 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.

Wiser, R., Rand, J., Seel, J., Beiter, P., Baker, E., Lantz, E., & Gilman, P. (2021). Expert elicitation survey predicts 37% to 49% declines in wind energy costs by 2050. *Nature Energy*, 6(5), 555–565.

Wright, S. J. (2006). *Numerical optimization*.