Universidad Tecnológica de Bolívar
CARTAGENA DE INDIAS

FACULTAD DE CIENCIAS BÁSICAS

AS • DS

International Workshop on

Applied Statistics & Data Science

WWW.UTB.EDU.CO

utboficial

Índice general

Libro de Memorias

V International Workshop on Applied Statistic and Data Science

Proccedings book

Facultad de Ciencias Básicas

Julio 2023

Universidad Tecnológica de Bolívar, UTB.*

*V International Workshop on Applied Statistic and Data Science en la Facultad de Ciencias Básicas. Cartagena de Indias, Colombia. Del 28 al 30 de Junio de 2023. Sede Campus Casa Lemaitre.

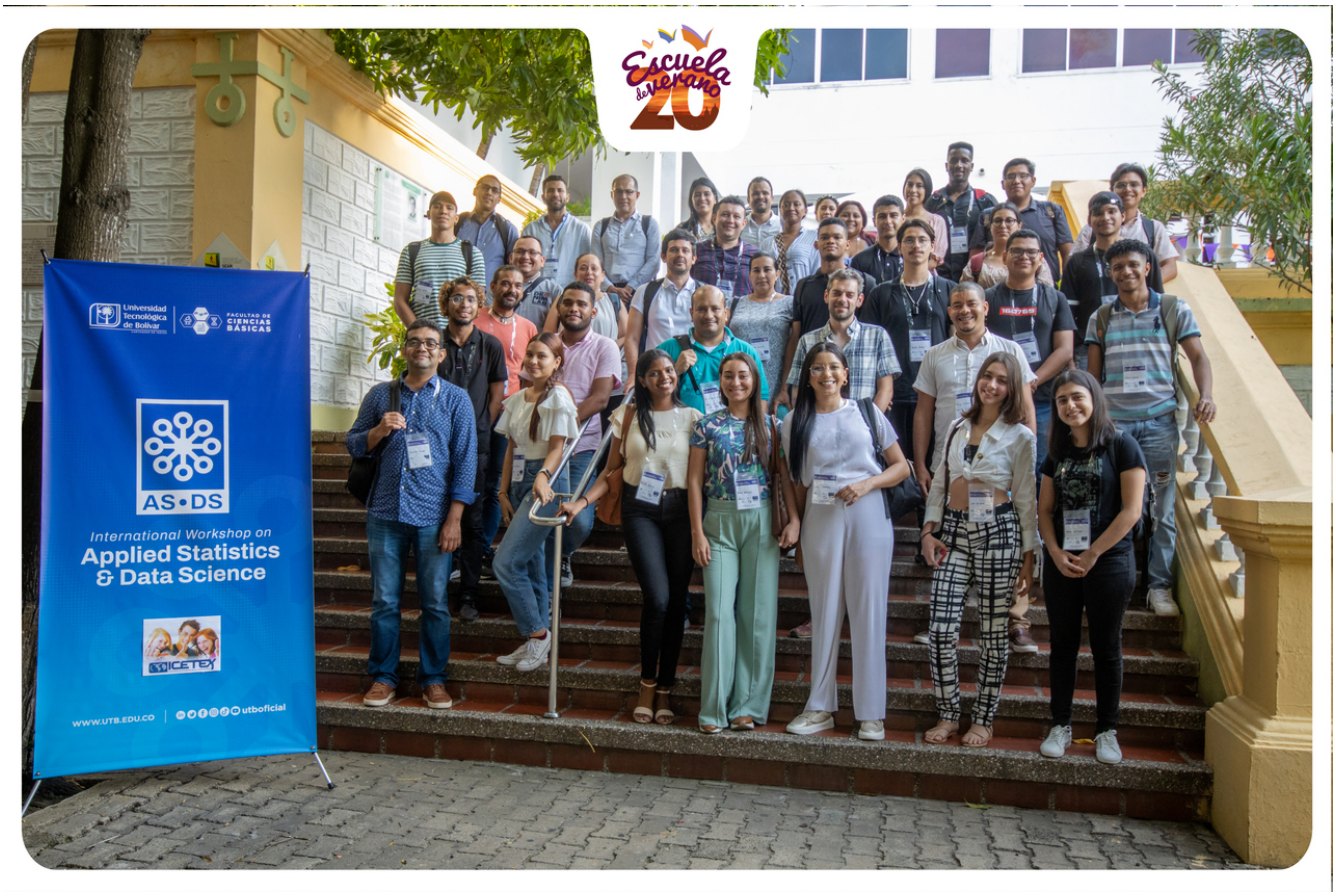
Comité Organizador y Académico:
Yady Tatiana Solano Correa, UTB
Andy Rafael Domínguez Monterroza, UTB
Julio Seferino Hurtado Marquez, UTB
Jorge Luis Villalba Acevedo, UTB
Lenny Alexandra Romero Pérez, UTB
David Sierra Porta, UTB
Asistencia en Organización y Desarrollo:
Yesenia Margarita Rodriguez Melendez, UTB
Mileibys Paola Aycardi Berrio, UTB

Rector
Alberto Roa Varelo
ViceRector Académico
Daniel Toro González
ViceRectora Administrativa
María del Rosario Gutiérrez de Piñeres Perdomo
Secretaria General
Ana María Horrillo Caraballo
Decana de la Facultad de Ciencias Básicas
Lenny Alexandra Romero Pérez
Dirección de Investigación, Innovación y Emprendimiento
Jairo Useche Vivero
UTB Global - Dirección de Internacionalización
Ericka Duncan Ortega
Dirección de Postgrado
Raúl Jose Padrón Carvajal
Maestría en Estadística Aplicada y Ciencia de Datos - Director
David Sierra Porta

Diagramación, Edición, Compilación: David Sierra Porta.
Fotos, Artes Gráficas y Página Web: José David Vergara Saltarín, Anyi Xiomara Giraldo Rivas.
Diseñadora Gráfica: Luisa Fernanda García Yaber.
Producción audiovisual: Carlos David Ortiz Caro.

Ediciones UTB
ISSN: XXXX-XXXX
Universidad Tecnológica de Bolívar. Parque Industrial y Tecnológico Carlos Vélez Pombo Km 1 Vía Turbaco. Cartagena de Indias, D. T. y C., - Colombia
www.utb.edu.co
(c) 2023

V International Workshop on Applied Statistic and Data Science



Ánimos y Motivación

El valor de los datos y de los profesionales con experiencia en datos está creciendo exponencialmente. Los expertos prevén que el mercado mundial de big data crecerá hasta los 103.000 millones de dólares en 2027, duplicando con creces las estimaciones realizadas en 2018. Ese crecimiento expansivo seguirá aumentando la demanda de profesionales capacitados que puedan hacer un uso significativo de enormes volúmenes de datos.

La ciencia de los datos y la estadística aplicada son esenciales para hacer que los grandes datos sean relevantes para una amplia gama de empresas, industrias, instituciones y academia. Aunque estas dos disciplinas operan en algunos de los mismos espacios, no son idénticas. Tanto la ciencia de los datos como la estadística aplicada tienen sus raíces en el campo de la estadística y están relacionadas con él. Gran

parte de los conocimientos y la formación básicos necesarios para una carrera en estos campos se basan en una formación estadística similar. Sin embargo, la principal diferencia entre la ciencia de los datos y la estadística es su enfoque único para dar sentido a los datos y resolver problemas.

Hay matices y excepciones en estos campos que se solapan, pero la ciencia de los datos suele utilizarse para hacer predicciones y optimizar las búsquedas en grandes campos de datos y bases de datos. Aplica técnicas como las herramientas de aprendizaje automático y la inteligencia artificial a problemas que normalmente requerirían inteligencia humana, pero que son demasiado amplios para resolverlos de forma eficiente por esas vías más tradicionales. La ciencia de los datos pretende hacer predicciones precisas de comportamientos y patrones futuros en un mercado o industria determinados.

Sin embargo, la estadística aplicada sigue siendo crucial para resolver muchos problemas del mundo real y sacar conclusiones esenciales para las empresas y organizaciones. Los estadísticos descubren la mejor manera de recopilar datos, realizar mediciones y cuantificar la incertidumbre allí donde las soluciones de la ciencia de los datos basada en las máquinas podrían resultar difíciles de manejar.

"El objetivo final del análisis estadístico suele ser sacar una conclusión sobre qué causa qué, basándose en la cuantificación de la incertidumbre. En cambio, el objetivo final del análisis de la ciencia de datos suele ser más bien una base de datos específica o un modelo predictivo". (https://www.displayr.com/statistics-vs-data-science-whats-the-difference/)

Página web

https://www.utb.edu.co/workshop-applied-statistic-and-data-science/#

Ediciones Anteriores

El Workshop ha sido un evento con historia. Las ediciones anteriores han sido: WorkShop en Estadística Aplicada (2012, Cartagena- Colombia), II Workshop en Estadística Aplicada (2016, Cartagena- Colombia), III International Workshop on Applied Statistics (2019, Cartagena- Colombia) y el último IV International Workshop in Applied Statistics and Data Science (2022, Cartagena- Colombia).

Fechas importantes y lugar del evento

El evento se desarrolló en las instalaciones de la Universidad Tecnológica de Bolivar en la Facultad de Ciencias Básicas. Campus Casa Lemaitre, Calle del Bouquet Cra.21 #25-92, Barrio Manga, Cartagena-Colombia. El evento duró 3 días intensos desde 28 al 30 de junio de 2023.

Dirigido a

Todos los investigadores y profesionales de diferentes disciplinas que deseen conocer de Ciencia de Datos y Estadística Aplicada en el manejo de grandes volúmenes de datos, así como también a estudiantes de pregrados y posgrados de ingenierías, ciencias básicas o de economía y negocios, interesados en desarrollar habilidades en temas relacionados con Estadística Aplicada y Ciencia de datos. Además también es propicio para todos los profesionales del sector empresarial que desempeñan funciones donde el análisis de datos es requerido para la toma de decisiones.

Objetivos y metodología

Este evento reunió a varios profesionales expertos del área que se encargaron de ofrecer a partir de varios cursillos su experiencia y proporcionar así una visión desde cada unas de sus experticias en el contexto de la estadística aplicada y la ciencia de datos. De modo que este evento representó un importante espacio para compartir experiencias y aprender acerca de los usos y casos de uso de la estadística y la ciencia de datos en muchas áreas. Los cursos fueron prácticos y aplicados y a través de la dinámica los participantes explorarán técnicas, metodologías, teorías y aplicaciones particulares. Además se contó con una sesión de posters para que los participantes expongan sus ideas, trabajos en curso, casos particulares de uso y aplicación en un forum que permita la retroalimentación y compartir de experiencias que ayuden, potencien y contribuyan a la generación de sinergias y trabajos colaborativos entre los participantes.

El workshop constó con 2 cursos, de este modo los cursos se realizaron de manera secuencial y se tuvo la oportunidad de seguir el curso completamente. Adicionalmente los cursos fueron diseñados para tener una gran cantidad de horas prácticas en la que los participantes podrán interactuar, probar, experimentar con la ayuda de los talleristas.

Se contó con una sesión de networking y presentación de posters en la que los participantes interactuaron con todos en el workshop y expusieron sus ideas, trabajos, investigaciones, problemas, desafíos, etc. Pudieron contar con la ayuda de los talleristas, expertos en el área y personas invitadas que pudieran ofrecer sus ideas y ayudas en la presentación. Esta fue una oportunidad valiosa para la socializar, estrechar lazos y discutir perspectivas con otros participantes.

Invitados Internacionales y Charlas magistrales

- Andy Domínguez Monterroza, Cand. PhD. Investigador y profesor Timepo Completo de la Universidad Tecnológica de Bolívar y miembro del grupo de investigación de Gravitación Matemática Aplicada de la FAcultad de Ciencias Básicas de la misma universidad. Trabaja en diversos proyectos relacionados con series de tiempo, topología y redes complejas.
    - Charla Magistral: Ciencia de Redes y Sus Aplicaciones (En español)

- Craig Robson, PhD. Profesor de Ingeniería Centrada en Datos en la Escuela de Ingeniería de la Universidad de Newcastle (Reino Unido), su trabajo se centra en el uso de datos y métodos de ciencia de datos para evaluar los impactos del cambio climático. Sus áreas de investigación incluyen la ciencia de datos espaciales, los marcos de evaluación integrados, la complejidad de las redes, la resiliencia de las redes y la visualización de datos.
    - Curso: Análisis de Series de Tiempo (En inglés)

- Jarnishs Beltran, PhD. Investigador y Profesor Asociado de la Universidad de Valparaiso, Chile, experto en ciencia de datos. Sus intereses principales se centran en el manejo de herramientas y habilidades en el procesamiento de lenguaje natural y machine learning para entender coherencia y evolución de emociones en contextos de contenido textual.
    - Curso: Procesamiento de Lenguaje Natural (En español)

Cursos

Análisis de series de tiempo (Craig Robson, UK)

El curso de análisis de series de tiempo que se llevó a cabo en el V International Workshop in Applied Statistics and Data Science se enfocó en proporcionar a los participantes una perspectiva de ciencia de datos sobre la teoría, procesos y métodos para analizar datos temporales. El curso se impartió utilizando ejemplos prácticos con datos abiertos y de libre acceso, y se empleó el lenguaje de programación Python, un lenguaje de código abierto y ampliamente utilizado en el análisis de datos.

A continuación, se presenta una lista de temas y contenido que se abordaron durante el curso:

- Introducción a las series de tiempo: Definición y conceptos básicos de series de tiempo. Componentes de una serie de tiempo: tendencia, estacionalidad y ruido. Aplicaciones en ciencia de datos y campos relacionados.
- Preprocesamiento de datos de series de tiempo: Limpieza y manejo de datos faltantes. Conversión de fechas y frecuencias de muestreo.
- Visualización y exploración de series de tiempo: Gráficos de líneas, gráficos de puntos y gráficos de autocorrelación. Identificación de patrones y comportamientos en los datos.
- Modelado y pronóstico básico: Modelos de media móvil y promedio. Métodos de alisado exponencial. Pronóstico a corto plazo y tendencias.
- Modelos avanzados de series de tiempo: Modelos ARIMA (AutoRegressive Integrated Moving Average). Estacionariedad y diferenciación. Selección de órdenes de modelos.
- Técnicas para series de tiempo estacionales: Modelos SARIMA (Seasonal ARIMA). Decomposición estacional.
- Evaluación y validación de modelos: Métricas de rendimiento para pronósticos. Validación cruzada en series de tiempo.
- Aplicaciones prácticas: Pronóstico de demanda en ventas y marketing. Predicción de datos climáticos y medioambientales.

Cada tema se acompañó de ejemplos prácticos utilizando Python para implementar los métodos y técnicas aprendidas. Los participantes pudieron aplicar estos conocimientos para analizar y pronosticar diferentes tipos de datos de series de tiempo y adquirir habilidades que pueden ser trasladadas a otros entornos de codificación y análisis de datos.

Procesamiento de lenguaje natural (Jarnishs Beltrán, CH)

El curso se enfoca en proporcionar a los participantes herramientas y técnicas para procesar datos de lenguaje natural desde una perspectiva computacional. El objetivo es permitir a los estudiantes utilizar R y RStudio para trabajar con datos de Twitter y realizar análisis de sentimientos y construcción de redes semánticas de hashtags. Se espera que los participantes tengan una palabra clave específica y relevante de carácter político para realizar búsquedas en Twitter durante el taller.

A continuación, se presenta una descripción del contenido y temas que se abordaron en cada día del taller:

- Día 1: Creación de la API de Twitter. Extracción de datos de Twitter. Preprocesamiento de datos y primeras visualizaciones. Introducción al Procesamiento de Lenguaje Natural (NLP) y su importancia en la ciencia de datos. Configuración de la API de Twitter para acceder a datos en tiempo real. Extracción de datos de Twitter utilizando R y la API. Preprocesamiento de los datos extraídos: limpieza, tokenización y eliminación de stop words. Visualizaciones iniciales: nubes de palabras para explorar términos más frecuentes y gráficos de frecuencia de palabras.

- Día 2: Análisis de sentimientos usando diccionarios. Discusión de técnicas de clasificación de sentimientos. Introducción al análisis de sentimientos y su aplicación en el procesamiento de lenguaje natural. Uso de diccionarios de sentimientos para asignar polaridad a palabras y textos. Interpretación de los resultados del análisis de sentimientos y su relevancia en la comprensión de la opinión pública. Comparación de técnicas de clasificación de sentimientos basadas en diccionarios y técnicas de Machine Learning (ML). Posibilidades de investigación mixta en el área de análisis de sentimientos y discurso de odio como un ejemplo destacado.

- Día 3: Introducción al concepto de redes y aplicaciones en redes semánticas de hashtags. Conceptos básicos de las redes y su representación en el análisis de datos. Medidas de centralidad en redes y su importancia en el análisis de datos de lenguaje natural. Aplicación práctica: construcción de redes semánticas de hashtags en Twitter. Análisis de la estructura de las redes semánticas y su interpretación. Discusión sobre el uso de redes semánticas en el análisis de tendencias y opiniones en redes sociales.

Cada día del taller combina conceptos teóricos con ejemplos prácticos utilizando R y RStudio para el análisis de datos de lenguaje natural de Twitter. Al final del taller, los participantes deberían ser capaces de utilizar estas técnicas para analizar datos políticos relevantes en la plataforma de Twitter y explorar su aplicación en la investigación académica en ciencias sociales y computacionales.

Apoyo

El evento contó con el apoyo y soporte de la edición número 20 de la Escuela de Verano UTB 2023 (https://www.utb.edu.co/utb-global/internacionalizacion-en-casa/escuela-de-verano/) de la Dirección de Internacionalización en la cual esta actividad se ha enmarcado, por eso la organización del ASDS expresa su profunda gratitud a la profesora **Ericka Duncan Ortega** (Directora de Internacionalización) y a **Natalia De Jesus Caraballo Noriega** (Coordinadora de Proyectos Internacionales y Visibilidad) por tu esfuerzos en hacer esta edición posible.

También se contó con el apoyo de la Dirección de Mercadeo y Comunicaciones con especiales agradecimientos a **Mavy Catherine Gutierrez Cedeno** (Directora de Mercadeo Y Comunicaciones), **Jose David Vergara Saltarin** (Coordinador de Comunicaciones Institucionales) y a **Anyi Xiomara Giraldo Rivas** (Gestora de Contenidos Web).

Adicionalmente alguna parte de los fondos para cubrir todas las becas otorgadas en en este evento han sido posibles gracias al apoyo del **Instituto Colombiano de Crédito Educativo y Estudios Técnicos en el Exterior - ICETEX**, a los cuales agradecemos y honramos con un reconocimiento.

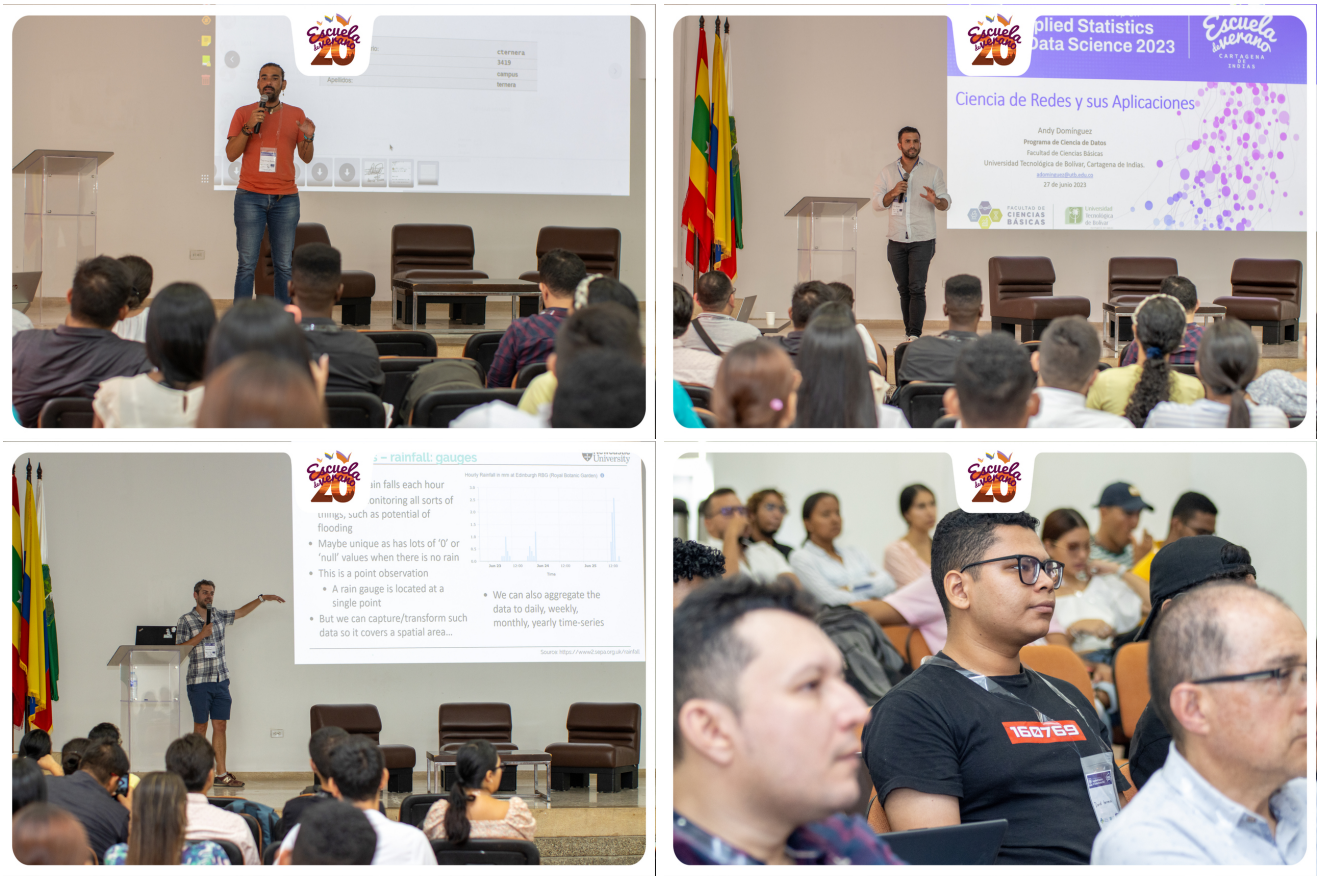
La Sociedad Colombiana de Estadística (SCE-Co) y el Instituto IberoAmericano de Estadísitca (IASI) ayudaron en la difusión y comunicación a través de las redes aliadas.

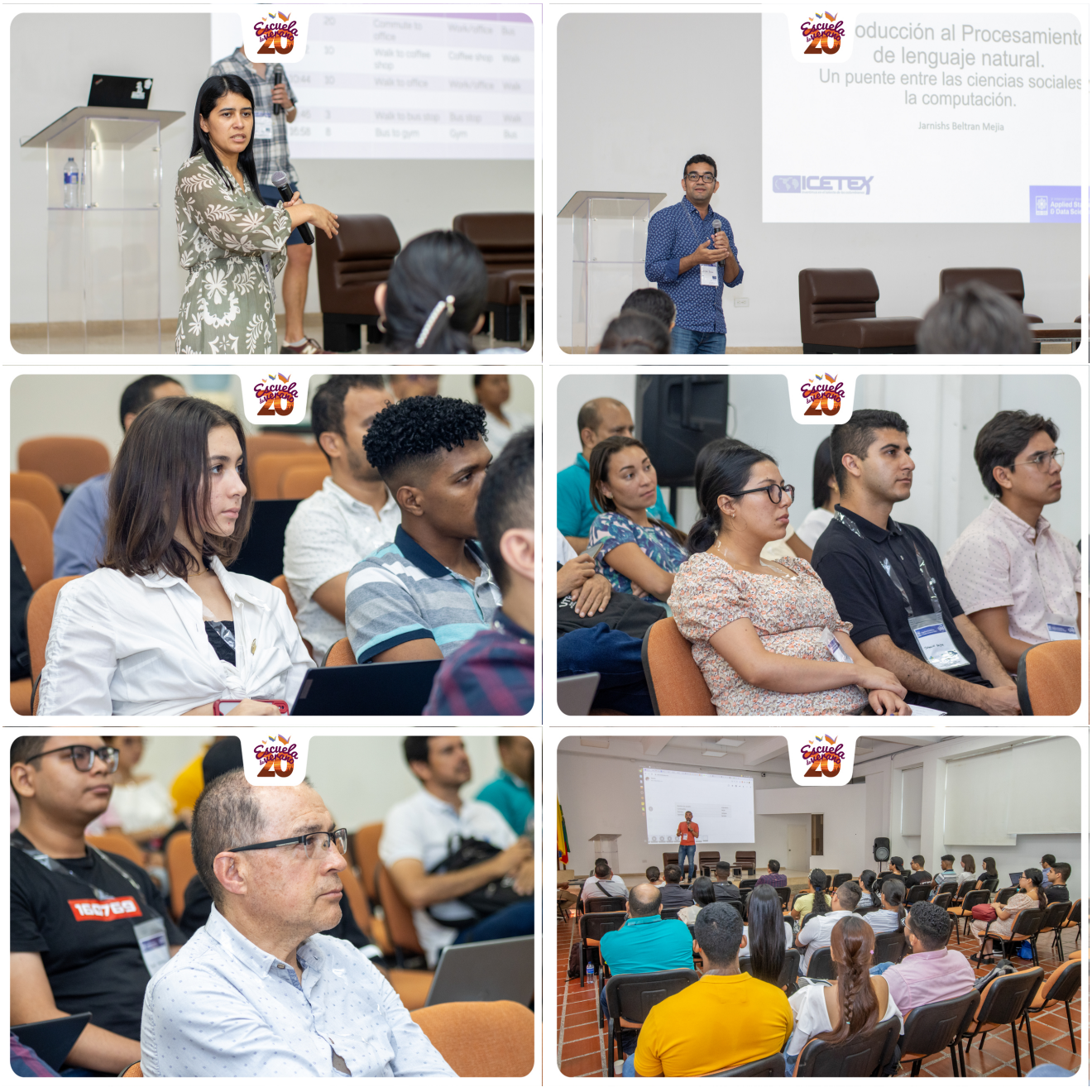
El evento quisiera expresar un profundo y especial agradecimiento a **Yesenia Margarita Rodriguez Melendez** (Auxiliar Administrativo de la Facultad de Ciencias Básicas) por su invaluable entrega y apoyo

en todos los procesos del evento, así como también a **Mileibys Paola Aycardi Berrio** (Profesional de Apoyo Administrativo Y Bienestar Posgrados) y **Astrid Carolina Herrera Uparela** (Profesional de Apoyo Administrativo y Bienestar Pregrado). Sin la ayuda de estas personas nada hubiere podido ser posible.

Nuestros apoyantes:

El Workshop en algunas imágenes




oducción al Procesamiento
de lenguaje natural.
Un puente entre las ciencias sociales
la computación.
Jarnishs Beltran Mejia
ICETEX

Sesión de Posters

Aplicación Shiny para perfiles de verosimilitud

Mario César Jaramillo Elorza*, Valentina Tamayo Guarín, Juan Pablo Martínez
*Departamento de Estadística, Universidad Nacional De Colombia, Bogota, Colombia.
email: mcjarami@unal.edu.co

Resumen: Este proyecto propone una herramienta fundamental para la inferencia estadística, los perfiles de verosimilitud, que permiten calcular un parámetro desconocido. Se desarrollará una aplicación interactiva en R Shiny, basada en técnicas usadas en la literatura de inferencia estadística, con el objetivo de facilitar la comprensión y la importancia de los perfiles de verosimilitud para estudiantes de estadística y la investigación. La herramienta se enfoca en ayudar a los estudiantes a adquirir habilidades prácticas en la construcción y la interpretación de los perfiles de verosimilitud, así como en su uso para seleccionar modelos estadísticos y entender los resultados.

Palabras clave: Perfiles de verosimilitud, Aplicación interactiva, Shiny, Máxima verosimilitud.

A semi-parametric Bayesian extreme value model using a Dirichlet process mixture of gamma densities

Jairo Fuquene-Patino*
*Department of Statistics, University of California, Los Angeles, California, Estados Unidos.
email: jafuquenepatino@ucdavis.edu

Resumen: We propose a model with a Dirichlet process mixture of gamma densities in the bulk part below threshold and a generalized Pareto density in the tail for extreme value estimation. The proposed model is simple and flexible for posterior density estimation and posterior inference for high quantiles. The model works well even for small sample sizes and in the absence of prior information. We evaluate the performance of the proposed model through a simulation study. Finally, the proposed model is applied to a real environmental data.

Palabras clave: generalized Pareto distribution, threshold estimation, Dirichlet process mixture.

Sistema de alertas tempranas para la prevención de la deserción universitaria con el uso de técnicas de machine learning

Miguel Ernesto Velandia Feria*, Oscar Andrés Ramírez Avendaño, Marco Javier Peñaloza Pérez, David Arango Londoño
*Universidad Javeriana, Cali, Colombia.
email: miguel.velandiaf@cecar.edu.co

Resumen: El estudio aborda el problema de la deserción universitaria en la Facultad de Ciencias Básicas e Ingenierías de la Corporación Universitaria del Caribe (Cecar). Se presenta un marco conceptual basado en investigaciones previas que utilizan enfoques cualitativos y cuantitativos, así como la ciencia de datos.

Se realiza un análisis exploratorio descriptivo de los datos de deserción correspondientes a los periodos de 2019A-2022B. Este análisis se centra en comprender y examinar el fenómeno de la deserción en la mencionada facultad. Se utilizan técnicas estadísticas y visualizaciones de datos para identificar factores y tendencias relacionados con la deserción.

Luego, se entrenan varios modelos de machine learning, como la regresión logística, las máquinas de soporte vectorial, los bosques aleatorios de decisión y las redes neuronales simples. Estos modelos permiten predecir y emitir alertas sobre los riesgos de deserción en los programas de ingeniería de sistemas e industrial. Se desarrolla una API y una interfaz gráfica que integran el análisis exploratorio y el modelo predictivo.

El sistema resultante permite predecir la probabilidad de deserción para nuevos estudiantes, configurando un sistema de alertas tempranas. Esto contribuye a la comprensión y mitigación de la deserción universitaria, así como a promover políticas institucionales que buscan la permanencia de los estudiantes.

En resumen, el estudio propone un enfoque integral para abordar la deserción universitaria en la Facultad de Ciencias Básicas e Ingenierías de Cecar. Combina un marco conceptual, análisis exploratorio de datos y modelos de machine learning para comprender el fenómeno, identificar patrones y factores asociados, y desarrollar un sistema de alertas tempranas. Esto tiene como objetivo prevenir la deserción y promover la permanencia de los estudiantes en los programas de ingeniería.

Palabras clave: deserción universitaria, machine learning, ciencia de datos, predicción, sistemas de alertas tempranas.

Métodos estadísticos para el análisis epidemiológico de los factores de riesgo para osteoporosis en mujeres post menopaúsicas

Nidia Milena Babativa Cortes*, Julio Seferino Hurtado Marquez
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: nbabativa@uniguajira.edu.co

Resumen: El diagnóstico de osteoporosis en mujeres Post menopaúsicas, determinado inicialmente por criterios diagnósticos como la medición de la masa ósea a partir del uso de tecnologías basadas en las definiciones de la OMS, que establece la medición de la masa mineral ósea para considerar DMO normal, osteopenia y osteoporosis. El método se realiza a partir de análisis de componentes principales, que permiten la visualización de los grupos y su nivel de representatividad, también se aplicó la técnica de análisis de clúster que permitió observar la distancia, concentración o agrupación de variables a partir de las gráficas en forma de dendogramas, así mismo aplico la técnica de correlación que permitió observar la relación fuerte y débil de las variables analizadas. La data presenta 375 observaciones y 15 variables, sin embargo la presencia de datos faltantes requiere de la implementación de un método de imputación para la optimización de la data, existe correlación fuerte entre las variables prueba de calcio y glicemia, entre las pruebas AST (aspartato amino transferasa) que permite medir la función renal y la prueba ALT (alanino amino transferasa) cuya prueba determina la medición de la función del hígado y el páncreas, y al aplicar el modelo PSA, se observa a partir de la varianza acumulada una reducción de la data que permitió, la observación de las variables más representativas que podrían orientar hacia la presencia de factores de riesgo de osteoporosis, y se determinó que un numero de 10 clúster es suficiente para el análisis del fenómeno de interés.

Palabras clave: Minería de datos, osteoporosis, epidemiologia, clúster, análisis de componentes principales, correlación, modelo estadístico.

Detección de bots en twitter utilizando aprendizaje no supervisado

Lilibeth De Horta Narvaez*, Roberto Trespalacio Alies
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: ldehortana@yahoo.es

Resumen: Este estudio tiene como objetivo detectar bots en Twitter mediante el análisis de sentimientos de un conjunto de tweets. La detección de bots en redes sociales es un desafío importante debido a su capacidad para propagar información falsa y engañar a los usuarios. Para abordar este problema, se utilizará el análisis de sentimientos para identificar patrones característicos en los tweets generados por bots y usuarios reales.

En primer lugar, se recopilará un conjunto de tweets de Twitter que incluya tanto contenido generado por bots como por usuarios auténticos. Estos tweets pueden provenir de diferentes temas y áreas de interés para garantizar la representatividad del conjunto de datos.

Posteriormente, se aplicará el análisis de sentimientos para evaluar las emociones y el tono expresado en cada tweet. El análisis de sentimientos permitirá identificar si los tweets son positivos, negativos o neutrales, lo que proporcionará información valiosa sobre la autenticidad del contenido.

Luego, se empleará aprendizaje no supervisado para encontrar diferencias y patrones distintivos entre los tweets generados por bots y los generados por usuarios reales. Al utilizar técnicas de clustering y reducción de dimensionalidad, se buscará agrupar los tweets en clusters o grupos que compartan características similares.

La identificación de clusters específicos que contengan predominantemente tweets generados por bots podría indicar la existencia de patrones de comportamiento característicos de estas cuentas automatizadas. De manera similar, los clusters con predominancia de tweets de usuarios reales podrían revelar características distintivas de los contenidos auténticos.

Los resultados de este estudio proporcionarán una visión más clara sobre la presencia de bots en Twitter y cómo se comportan en comparación con los usuarios reales. La detección temprana de bots en la plataforma puede ayudar a prevenir la difusión de información falsa y proteger a los usuarios de posibles intentos de manipulación.

En conclusión, este enfoque de detección de bots en Twitter a través del análisis de sentimientos y el aprendizaje no supervisado representa un paso importante para abordar el problema de la propagación de información falsa en las redes sociales. La combinación de estas técnicas puede proporcionar un análisis más profundo y preciso de la autenticidad de los tweets, lo que permitirá tomar medidas preventivas para mantener la integridad de la plataforma y proteger a sus usuarios.

Palabras clave: Análisis de sentimientos, NLTK.

Análisis de redes aplicado al sistema de tránsito de transporte urbano de Cartagena de Indias

Yenifer Matorel Silva*, D. Sierra-Porta
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: ymatorel@utd.edu.co

Resumen: Este estudio se enfoca en realizar un análisis de redes al sistema de transporte urbano de Cartagena de Indias con el propósito de identificar áreas donde el servicio de transporte público es insuficiente. Además, se contrastará la presencia del Mototaxismo en esas zonas para entender su posible relación con la deficiencia del servicio de transporte público.

Para llevar a cabo el análisis, se recopilará información relevante sobre las rutas y el funcionamiento del sistema de transporte urbano, incluyendo datos sobre frecuencia de rutas, capacidad de transporte, horarios, entre otros factores que puedan influir en la eficiencia del servicio.

Se aplicarán técnicas de análisis de redes para mapear y visualizar la conectividad y flujo de tránsito en la ciudad, identificando las áreas donde la oferta de transporte público no satisface adecuadamente la demanda de los ciudadanos.

Además, se investigará la presencia y extensión del Mototaxismo en las zonas donde se ha detectado una deficiencia en el servicio de transporte público. El Mototaxismo es un fenómeno informal de transporte que puede surgir como respuesta a las limitaciones del transporte público formal y puede afectar la movilidad y seguridad de los ciudadanos.

Mediante el contraste de estos datos, se buscarán patrones y relaciones entre la falta de servicio de transporte público y la prevalencia del Mototaxismo en diferentes áreas de la ciudad. Este análisis permitirá proponer estrategias efectivas para mejorar la calidad del transporte urbano y reducir la necesidad de alternativas informales como el Mototaxismo.

Los resultados de este estudio podrían proporcionar información valiosa a las autoridades y planificadores de transporte de Cartagena de Indias para implementar soluciones adecuadas y orientadas a mejorar la movilidad en la ciudad.

En conclusión, este análisis de redes aplicado al sistema de transporte urbano de Cartagena de Indias es una herramienta fundamental para identificar áreas con servicio insuficiente y su relación con el Mototaxismo. La comprensión de estas dinámicas permitirá proponer estrategias efectivas y mejorar la calidad del transporte público en la ciudad, contribuyendo así a una movilidad más eficiente y sostenible para sus habitantes.

Palabras clave: Redes, Transporte, Mototaxismo.

Entrenamiento de modelos de aprendizaje automático para pronosticar la deserción En programas académicos de la universidad de córdoba

Edgar Segundo Ramos Ramirez*, Mario Alfonso Morales Rivera
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: adidactico31@gmail.com

Resumen: En este estudio, nos proponemos ajustar modelos de aprendizaje automático con el objetivo de predecir la intención de los estudiantes de los programas académicos de la Universidad de Córdoba de abandonar sus carreras. El objetivo es proporcionar un mecanismo de reconocimiento temprano de los estudiantes que puedan optar por tomar esa decisión, permitiendo así una alerta temprana para la toma de medidas preventivas.

La deserción estudiantil es un desafío importante en las instituciones académicas, y su identificación temprana es esencial para implementar estrategias efectivas de retención y apoyo a los estudiantes en riesgo de abandonar sus estudios. Para abordar este problema, utilizaremos técnicas avanzadas de aprendizaje automático y análisis de datos.

En primer lugar, recopilaremos datos históricos y relevantes sobre los estudiantes matriculados en los diferentes programas académicos de la Universidad de Córdoba. Estos datos incluirán información demográfica, rendimiento académico, participación en actividades extracurriculares y cualquier otro factor que pueda estar relacionado con la deserción estudiantil.

A continuación, llevaremos a cabo una limpieza y preprocesamiento exhaustivos de los datos para asegurar que estén en un formato adecuado para el entrenamiento del modelo. Posteriormente, dividiremos los datos en conjuntos de entrenamiento y prueba para poder evaluar el rendimiento de los modelos de manera adecuada.

Luego, emplearemos diferentes algoritmos de aprendizaje automático, como árboles de decisión, regresión logística, máquinas de soporte vectorial (SVM) y redes neuronales, para entrenar y ajustar nuestros modelos predictivos. La elección de estos algoritmos se basará en su capacidad para manejar problemas de clasificación y su rendimiento en la predicción de deserción estudiantil.

Finalmente, evaluaremos la precisión y efectividad de nuestros modelos utilizando métricas de evaluación apropiadas, como precisión, recall, F1-score, entre otras. Buscaremos identificar el modelo con el mejor rendimiento para realizar pronósticos precisos sobre la intención de deserción de los estudiantes.

El resultado esperado de este estudio es proporcionar a la Universidad de Córdoba una herramienta efectiva de predicción de deserción estudiantil, lo que permitirá tomar medidas preventivas y de intervención temprana para mejorar la retención y el éxito académico de los estudiantes. Este enfoque puede ser una valiosa contribución para abordar el desafío de la deserción en instituciones académicas y mejorar la calidad de la educación superior en Colombia.

Palabras clave: Entrenamiento, Modelos de aprendizaje automático, pronóstico, deserción, programas académicos, Universidad de Córdoba.

Asociación de las concentraciones dióxido de nitrógeno NO_2 y el índice de vegetación NDVI, en las zonas ucg 11 y 13, Cartagena de indias

Elizabeth Valderrama Serrano*, Yady Tatiana Solano
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: evalderrama@utb.edu.co

Resumen: Este artículo presenta un análisis exhaustivo de la relación entre las concentraciones de dióxido de nitrógeno (NO_2) y el índice de vegetación (NVDI) en dos sectores distintos de Cartagena de Indias: la zona franca UGC 11 y la zona UGC 13. La zona UGC 11 está conformada por varios barrios y se caracteriza por una constante exposición a la contaminación proveniente de empresas y vehículos, mientras que la zona UGC 13 se destaca por su abundante vegetación urbana.

El estudio se llevó a cabo utilizando mapas satelitales de las misiones Sentinel 5 y Sentinel 3, correspondientes al periodo comprendido entre abril de 2022 y marzo de 2023. Estos mapas satelitales permitieron realizar una caracterización temporal de los cambios en los niveles de dióxido de nitrógeno y en el índice de vegetación a lo largo del tiempo.

Mediante el análisis de los datos obtenidos de los satélites, se evaluó la posible asociación entre las concentraciones de dióxido de nitrógeno y el índice de vegetación en ambas zonas. Se buscó identificar patrones y tendencias que pudieran indicar una relación entre la presencia de vegetación y los niveles de contaminación en el aire.

Los resultados de este estudio proporcionan una visión detallada de cómo las concentraciones de dióxido de nitrógeno varían en función de la cantidad de vegetación presente en las áreas de estudio. Se espera que estos hallazgos contribuyan a una mejor comprensión de los factores ambientales que influyen en la calidad del aire en zonas urbanas y en áreas con alta exposición a la contaminación.

Este análisis es relevante para la planificación y gestión ambiental en Cartagena de Indias, ya que brinda información valiosa sobre la relación entre la vegetación urbana y la calidad del aire en distintas zonas de la ciudad. Asimismo, los resultados podrían servir de base para la implementación de estrategias de mitigación y control de la contaminación en áreas urbanas con alta concentración de emisiones de dióxido de nitrógeno.

En conclusión, este artículo proporciona una valiosa contribución al conocimiento científico sobre la relación entre el dióxido de nitrógeno y el índice de vegetación en zonas específicas de Cartagena de Indias. El análisis de datos satelitales y la caracterización temporal de los cambios ambientales permiten una comprensión más profunda de los factores que afectan la calidad del aire en estas áreas urbanas, lo que puede guiar futuras estrategias de gestión ambiental y sostenibilidad.

Palabras clave: Índice de vegetación de diferencia normalizada (NVDI), Dióxido de nitrógeno NO2, Sentinel 2, Sentinel 5P, Análisis Multitemporal.

Analítica aplicada al flujo de visitantes a un supermercado de cadena

Lorena Ibañez Castro*, Andy Dominguez
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: libanez96@hotmail.com

Resumen: En este estudio, se aplica un enfoque de analítica descriptiva y técnicas de machine learning y deep learning para caracterizar los patrones de consumo de visitantes en un supermercado de cadena (MT). El objetivo principal es crear un modelo predictivo que permita pronosticar la demanda de visitas en el tiempo, lo que permitirá optimizar los turnos operativos y mejorar la eficiencia en la atención al cliente.

La metodología se divide en tres fases principales. En primer lugar, se realiza un análisis descriptivo exhaustivo para caracterizar los flujos de visitas en el supermercado MT. Esto implica examinar datos históricos y actuales sobre el número de visitantes, los horarios de mayor afluencia, días de la semana más concurridos, entre otros factores. La finalidad es identificar patrones temporales que ayuden a comprender el comportamiento de los clientes en el establecimiento.

En la segunda fase, se aplican algoritmos supervisados de machine learning, en particular, redes neuronales y Long Short-Term Memory (LSTM), para analizar y modelar la relación entre diferentes variables y el flujo de visitantes. Estos algoritmos permiten capturar relaciones complejas y no lineales, proporcionando una mayor precisión en la predicción de la demanda de visitas al supermercado.

La tercera fase se enfoca en el desarrollo de un modelo de pronóstico del flujo de clientes visitantes utilizando técnicas de inteligencia artificial, en concreto, deep learning. Este modelo busca anticipar la demanda futura con base en datos históricos y tendencias observadas en el comportamiento de los visitantes. El resultado esperado es un modelo que pueda prever la cantidad de visitantes en distintos intervalos de tiempo, lo que permitirá una optimización eficiente de los turnos operativos y recursos disponibles en el supermercado.

La pregunta de investigación clave en este estudio es: "¿Qué patrones temporales exhibe el flujo de clientes visitantes al supermercado MT?"Para responder a esta pregunta, se analizan los datos obtenidos a través de técnicas de visualización y análisis estadístico, lo que proporcionará una comprensión clara y detallada de los patrones y comportamientos de los visitantes.

El resultado esperado de este proyecto es la creación de un modelo preciso y confiable que pueda pronosticar la demanda de visitas al supermercado MT en diferentes momentos del día y de la semana. Este modelo permitirá una toma de decisiones más informada y una gestión eficiente de los recursos humanos y logísticos, mejorando la experiencia del cliente y optimizando la operación general del supermercado. En última instancia, esta investigación busca contribuir al campo del análisis de datos aplicado al comercio minorista y al uso efectivo de técnicas de machine learning y deep learning en la gestión de flujos de visitantes.

Palabras clave: Patrones de consumo, Visitantes, Supermercado, Machine learning, Analítica descriptiva, Algoritmo supervisado, Redes neuronales, LSTM, Patrones temporales, Flujo de clientes, Modelo de pronóstico Inteligencia artificial, Deep learning, Optimización, Turnos operativos.

Optimización del desempeño energético

Luis Fernando Florez Garcia*, Estiven Sánchez Barrera
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: luisfflorezg@gmail.com

Resumen: En este estudio, se aborda el desafío de optimizar el desempeño energético en el proceso de extracción de petróleo. Mediante el monitoreo detallado de dicho proceso, se han identificado escenarios de ahorro y sobre consumo de energía en comparación con un periodo base establecido como referencia.

Nuestra solución se centra en la aplicación de técnicas de Machine Learning para identificar las variables más significativas que afectan el desempeño energético y, posteriormente, determinar los rangos de operación óptimos para estas variables. Este enfoque permite obtener recomendaciones precisas y efectivas para mejorar la eficiencia del proceso.

Primero, se recopilan y analizan datos de diferentes fuentes relacionadas con el proceso de extracción de petróleo. Estos datos incluyen información sobre variables operativas, consumo de energía y otros factores relevantes. Utilizando técnicas de limpieza y preprocesamiento de datos, nos aseguramos de que los datos estén listos para el análisis.

A continuación, aplicamos técnicas avanzadas de Machine Learning, como regresión y algoritmos de clasificación, para identificar las variables más influyentes en el desempeño energético del proceso de extracción de petróleo. Estas variables son clave para determinar los factores que inciden en los escenarios de ahorro y sobre consumo energético.

Una vez identificadas las variables significativas, empleamos técnicas adicionales de Machine Learning, como optimización y algoritmos genéticos, para determinar los rangos de operación recomendados para cada variable. Estos rangos están diseñados para maximizar la eficiencia energética del proceso, asegurando un consumo óptimo de recursos.

Mediante la combinación de estas técnicas de Machine Learning, nuestra solución ofrece una estrategia integral para la optimización del desempeño energético en la extracción de petróleo. Al proporcionar recomendaciones precisas sobre las variables y rangos de operación más eficientes, esperamos lograr reducciones significativas en el consumo de energía y, por ende, en los costos operativos asociados.

Los resultados preliminares muestran una mejora notable en la eficiencia energética del proceso de extracción de petróleo. Esto sugiere que nuestra propuesta tiene un gran potencial para aplicarse en otros contextos industriales, donde la optimización del desempeño energético es fundamental para reducir costos y minimizar el impacto ambiental.

Nuestra investigación demuestra cómo las técnicas de Machine Learning pueden ser utilizadas con éxito para identificar variables significativas y determinar rangos de operación óptimos en el proceso de extracción de petróleo. Estamos seguros de que nuestra solución contribuirá significativamente a la mejora de la eficiencia energética y a la toma de decisiones informadas en el campo de la industria petrolera.

Palabras clave: Optimización, desempeño energético, Variables significativas, Machine Learning.

Predicción de ecuaciones en MathML por medio de procesamiento de lenguaje en base de datos de problemas matemáticos en español

Kevin Andrés Sossa Valencia*, Edwin Alexander Puertas del Castillo
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: ksossa@utb.edu.co

Resumen: En este trabajo, presentamos una contribución innovadora para abordar el desafío de predecir ecuaciones en formato MathML que representen soluciones a problemas matemáticos en español. Nuestro enfoque combina técnicas avanzadas de Procesamiento de Lenguaje Natural (NLP) con modelos de Machine Learning, permitiendo la extracción efectiva de patrones relevantes en un corpus de datos de problemas matemáticos.

El primer paso de nuestro proceso consiste en recopilar y preparar un extenso conjunto de problemas matemáticos junto con sus soluciones en formato MathML. A través de una cuidadosa limpieza y preprocesamiento, aseguramos que los datos estén listos para ser utilizados en el entrenamiento del modelo.

Para representar los problemas de manera numérica, emplearemos técnicas de vectorización como Bag of Words, TF-IDF y Word Embeddings, lo que nos permite trabajar con datos numéricos y facilitar el aprendizaje automático. Asimismo, identificamos características relevantes en los enunciados a través del análisis sintáctico y semántico, lo que contribuye a la precisión de nuestras predicciones.

El diseño de nuestro modelo se basa en arquitecturas de redes neuronales, especialmente aprovechando la potencia de los modelos basados en Transformers, que han demostrado ser altamente efectivos en tareas de NLP. La elección de esta arquitectura nos permite capturar relaciones complejas y sutiles en los problemas matemáticos, mejorando así la calidad de nuestras predicciones.

Mediante un conjunto de entrenamiento y prueba adecuadamente dividido, realizaremos el entrenamiento del modelo y evaluamos su rendimiento utilizando métricas apropiadas para la tarea, tales como precisión, recall y F1-score. Iterativamente, ajustamos y mejoramos el modelo para obtener resultados más precisos y generalizables.

Los resultados en nuestra investigación indican que nuestro modelo es capaz de realizar predicciones precisas de ecuaciones MathML que representan soluciones a problemas matemáticos en español. Esta contribución representa un avance significativo en la aplicación de técnicas de Procesamiento de Lenguaje y Machine Learning para abordar problemas complejos en el ámbito matemático.

Nuestra propuesta tiene un gran potencial de aplicabilidad en el campo educativo y en plataformas de aprendizaje en línea, donde podría facilitar la evaluación automática de respuestas a problemas matemáticos, brindando retroalimentación rápida y precisa a los estudiantes.

Nuestro trabajo ofrece una solución innovadora y prometedora para la predicción de ecuaciones MathML en problemas matemáticos en español. Al unir las capacidades del Procesamiento de Lenguaje Natural y el Machine Learning, hemos demostrado cómo es posible extraer patrones relevantes y mejorar la calidad de las predicciones en este importante dominio matemático.

Palabras clave: NLP, Math Word problem, Spanish, MathML.

Detección de anomalías en señales de LIDAR atmosférico por medio de técnicas de inteligencia artificial

Marco Aurelio Pérez Benítez*, Estiven Sánchez Barrera
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: marcperez@utb.edu.co

Resumen: Esta investigación tiene como objetivo principal caracterizar y evaluar técnicas de inteligencia artificial y aprendizaje automatizado efectivas para la detección de anomalías en las señales de LIDAR atmosférico. El LIDAR es una técnica que utiliza pulsos de luz láser para sondar la atmósfera y recopilar información sobre sus características y composición. Las señales detectadas son almacenadas en diferentes canales que dependen de diversas propiedades de los fotones capturados, como su longitud de onda, polarización o pureza espectral.

El estudio se enfoca en la manipulación y etiquetado de los datos generados por el LIDAR para entrenar y evaluar modelos de inteligencia artificial. Se busca desarrollar un sistema capaz de identificar anomalías y extraer características relevantes de estas señales atmosféricas.

Para lograr este propósito, se recopilarán y preprocesarán conjuntos de datos de señales de LIDAR provenientes de diferentes condiciones atmosféricas. Estos datos servirán para el entrenamiento de algoritmos de inteligencia artificial, incluyendo técnicas de aprendizaje supervisado y no supervisado, como redes neuronales, algoritmos de clustering y sistemas de detección de anomalías.

La manipulación y etiquetado de los datos permitirá la creación de un conjunto de entrenamiento adecuado para cada técnica de inteligencia artificial evaluada. Esto proporcionará la capacidad de identificar patrones y comportamientos anómalos en las señales de LIDAR, lo que a su vez contribuirá a la caracterización de la atmósfera y la detección de eventos inusuales o fenómenos atípicos.

Los resultados esperados de este estudio ofrecerán una visión más profunda de la eficacia y el rendimiento de las técnicas de inteligencia artificial aplicadas a la detección de anomalías en señales de LIDAR atmosférico. La aplicación de estas técnicas podría mejorar significativamente la capacidad de monitorear y comprender cambios atmosféricos inesperados, lo que resulta de vital importancia en aplicaciones relacionadas con la meteorología, la climatología y la investigación atmosférica en general.

En conclusión, esta investigación representa un avance significativo en la detección de anomalías en señales de LIDAR atmosférico mediante el uso de técnicas de inteligencia artificial. La combinación de la capacidad de procesamiento de datos de estos algoritmos con la precisión y sensibilidad del LIDAR permitirá una mejor comprensión de la atmósfera y una mayor capacidad de detección y predicción de eventos atmosféricos anómalos.

Palabras clave: LiDAR atmosférico, inteligencia artificial, detección, anomalías, machine learning.

Análisis bibliométrico de enfermedades transmitidas por alimentos en Colombia

Andrea Carolina Menco Tovar*, Melba Liliana Vertel Morinson
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: amenco@utb.edu.co

Resumen: La presente investigación tiene como propósito delimitar los agentes etiológicos de interés según información disponible en los últimos 10 años en la Colombia, mediante el uso de herramientas y técnicas de analítica de datos (Meta análisis), además del uso de software estadísticos de acceso libre. Se realizó primeramente una acumulación de evidencias a través de la búsqueda en bases de datos (scopus, pudmed, web of science, google académico) haciendo uso de una ecuación de búsqueda que permitiese truncar según lo deseado. Luego se hizo un registro, almacenamiento y cribado de la información obtenida, escogiendo a partir de la extracción de datos los agentes etiológicos de interés tales como el Cólera, Antrax, Toxoplasmosis, Salmollenosis, Escherichia Coli, Campylobacter, Estafilococos, Listeria sp, Epstein-Barr y Leptospirosis. Los cuales son organismos biológicos bacterias, hongos, virus o parásitos capaces de causar enfermedades en los individuos que lo ingieren sea a través de alimentos, agua o superficies llevando al huésped a presentar enfermedades transmitidas por alimentos (ETA) esto último dependiendo de la cantidad de agente consumido.

Palabras clave: Microorganismos, patógenos, alimentos, inocuidad, riesgo, software.

Detección de anomalías con técnicas no supervisadas: impacto en la implementación de modelos de clasificación

Diego Herrera Malambo*, Andy Rafael Dominguez-Monterrosa, Alberto Patiño-Vanegas
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: dherrerambo@gmail.com

Resumen: En esta investigación se aborda el tema de la detección de anomalías, también conocidas como desviaciones, datos extremos u outliers, mediante el uso de técnicas no supervisadas. Las anomalías son puntos que se desvían del comportamiento normal en el contexto analizado y pueden clasificarse de diferentes maneras según sus características y origen.

Enfocándonos en los métodos de detección de anomalías no supervisados, este estudio busca abordar los desafíos que surgen al manejar grandes volúmenes de datos y la interacción compleja entre las diferentes características. Al utilizar técnicas no supervisadas, el objetivo es identificar anomalías sin la necesidad de etiquetas previas o datos de entrenamiento.

Para lograr este propósito, se propondrá una metodología que permita validar las anomalías detectadas de forma no supervisada. Esta metodología se basará en la exploración y análisis detallado de los resultados obtenidos, evaluando la coherencia y relevancia de las anomalías identificadas con respecto al contexto y objetivo del análisis.

La implementación de técnicas no supervisadas para la detección de anomalías es particularmente relevante en escenarios donde no se disponga de etiquetas para el entrenamiento de modelos de clasificación supervisada. Además, permite abordar el desafío de analizar grandes conjuntos de datos con interacciones complejas entre características.

Los resultados de esta investigación pueden tener un impacto significativo en la mejora de modelos de clasificación, ya que la detección precisa de anomalías puede llevar a una selección más efectiva de características y a una clasificación más confiable. Asimismo, se espera que esta metodología ayude a comprender mejor la naturaleza y la importancia de las anomalías detectadas, facilitando su interpretación y toma de decisiones adecuadas.

En conclusión, la detección de anomalías mediante técnicas no supervisadas representa un enfoque prometedor para abordar el análisis de grandes volúmenes de datos en diversas áreas de aplicación. La metodología propuesta permitirá validar las anomalías detectadas de manera no supervisada, lo que contribuirá al desarrollo de modelos de clasificación más eficientes y precisos, así como al mejor entendimiento de los datos y su comportamiento anómalo.

Palabras clave: Machine Learning, Anomalias, No supervisado, Outliers.

Examinado la influencia de los factores socioeconomicos en la incidencia del cáncer infantil en Colombia: un enfoque de aprendizaje automático

Rosa Yamiles Martinez Bello*, D, Sierra-Porta
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: rosimarti95@gmail.com

Resumen: El cáncer infantil es un importante problema de salud pública en Colombia y su incidencia está influenciada por diversos factores, incluyendo las condiciones socioeconómicas. Sin embargo, hasta el momento, ha habido una falta de estudios integrales que exploren el impacto de estos factores en las tasas de cáncer infantil en el país. Por lo tanto, el objetivo de esta investigación es investigar la asociación entre los factores socio económicos y la incidencia de cáncer infantil en Colombia utilizando técnicas de aprendizaje automático.

Se ha demostrado que el cáncer infantil tiene consecuencias económicas adversas para las familias, especialmente aquellas de bajos recursos. En Colombia, los tipos de cáncer más comunes en niños incluyen la leucemia, los tumores cerebrales, el neuroblastoma, el retinoblastoma y el linfoma de Hodgkin. Además, se observa una alta incidencia de cáncer en niños menores de cinco años, lo que subraya la vulnerabilidad de este grupo de edad.

Un análisis preliminar basado en los datos del Sistema de Vigilancia en Salud Pública revela un total de 1.061 notificaciones de casos de cáncer infantil en Colombia. Se observa un ligero predominio de casos en niños varones y una mayor carga de cáncer en los grupos socio económicos más bajos. Esto requiere una atención especial e intervenciones adaptadas para abordar los desafíos que enfrentan estos niños.

Esta investigación se propone abordar varias preguntas clave, incluyendo la correlación entre descriptores socio económicos y la aparición de casos de cáncer infantil en diferentes regiones de Colombia, la capacidad de los algoritmos de aprendizaje automático para predecir la probabilidad de cáncer infantil basándose en factores socio económicos, y la influencia de factores socio económicos específicos en tipos particulares de cáncer infantil.

Se espera que los resultados de esta investigación proporcionen una base sólida de conocimiento sobre el cáncer infantil en Colombia, permitiendo el desarrollo de políticas y estrategias específicas para mejorar los resultados y reducir la carga de esta enfermedad en los niños y sus familias. En última instancia, el objetivo es mejorar las tasas de supervivencia, la calidad de vida y fomentar un entorno de apoyo en el sistema de salud colombiano.

Palabras clave: cáncer infantil, Colombia, factores socio económicos, incidencia, aprendizaje automático, leucemia, tumores cerebrales, neuroblastoma, retinoblastoma, linfoma de Hodgkin, análisis epidemiológico, vulnerabilidad, intervenciones, desafíos, resultados, políticas, estrategias, supervivencia, calidad de vida, sistema de salud.

Developing an EEG, DMN-based tool for mental illnes diagnosis and measurement

Jose Serna Lopez*
*Ingenieria Biomedica, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: jsnlpz1402@gmail.com

Resumen: La red neuronal de estado por defecto (DMN, por sus siglas en inglés) es un sistema específico de regiones cerebrales interconectadas que se activa generalmente cuando el individuo no está enfocado en el ambiente externo o en una actividad específica. Este proyecto de investigación tiene como objetivo verificar la diferencia en la actividad de la DMN entre individuos de control y aquellos con síntomas y diagnósticos relacionados con estados depresivos. Se utilizará el análisis de señales de electroencefalografía (EEG), una técnica no invasiva y portátil, para establecer un criterio de clasificación cuantitativo para estados depresivos. El fin es evaluar y mejorar el desempeño en cada tipo de intervención y terapia para el tratamiento de enfermedades mentales.

Palabras clave: EEG, DMN, DEPRESIÓN.

Detección de fraudes en el sistema de acueducto de la ciudad de Cartagena, usando técnicas de machine learning

Juseff Salim Jalal Luna*, Yady T. Solano
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: jjalal@utb.edu.co

Resumen: Esta investigación se enfoca en abordar uno de los desafíos más complejos para los gestores de servicios públicos en el sector de acueductos y alcantarillados: la detección de fraudes en el sistema de acueducto de la ciudad de Cartagena. Para enfrentar esta problemática, se propone el uso de técnicas de machine learning para analizar los datos relacionados con el comportamiento de consumo de los usuarios y otras variables disponibles que puedan explicar la ocurrencia de fraudes.

El fraude en el sistema de acueducto se refiere a prácticas ilegales o no autorizadas que afectan negativamente la integridad del sistema y la equidad en el suministro de agua. Estas prácticas incluyen la manipulación de medidores, conexiones clandestinas y otros métodos para evitar el pago justo por el consumo de agua.

El estudio se basará en la recopilación de datos relevantes, como historiales de consumo, patrones de uso, ubicación geográfica de los usuarios y otras variables que puedan estar asociadas con comportamientos fraudulentos. Estos datos serán utilizados para el entrenamiento de modelos de machine learning, incluyendo algoritmos de clasificación y detección de anomalías.

La aplicación de técnicas de machine learning permitirá identificar patrones y comportamientos sospechosos que podrían indicar la presencia de fraudes en el sistema de acueducto. Los modelos entrenados serán capaces de detectar anomalías y clasificar usuarios con posibles prácticas fraudulentas, lo que ayudará a los gestores de servicios públicos a tomar medidas preventivas y correctivas de manera más efectiva.

Además, se busca mejorar la eficiencia en la detección de fraudes y reducir los costos asociados a la inspección manual y el proceso de seguimiento de usuarios sospechosos. La implementación de técnicas de machine learning proporcionará una herramienta valiosa para optimizar los recursos y mejorar la gestión del sistema de acueducto.

Los resultados de esta investigación tendrán un impacto significativo en la mejora de la gestión de los servicios de acueducto en Cartagena, permitiendo una detección temprana y precisa de fraudes. Esto contribuirá a garantizar la equidad en el suministro de agua, proteger los recursos hídricos y fortalecer la sostenibilidad del sistema en beneficio de la comunidad.

La detección de fraudes en el sistema de acueducto de Cartagena mediante técnicas de machine learning representa una solución innovadora y eficiente para abordar esta problemática compleja. La aplicación de estas técnicas proporcionará una herramienta poderosa para mejorar la gestión de los servicios de acueducto y garantizar un suministro de agua justo y sostenible para todos los usuarios de la ciudad.

Palabras clave: Fraude, acueducto, machine learning.

Elaboración de un modelo matemático que calcule los parámetros adecuados para mejorar la calidad en la metalización de rollos de biobase de polipropileno en Taghleef industries

Cristian Hernandez*, Victor Pretelt
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: criss.hernandezpadilla17@outlook.com

Resumen: La principal problemática que ocurre en la metalización de estos rollos, son los parámetros que se colocan en las metalizadoras. ya que estos no son demasiados óptimos para el proceso productivo, ya que muchas veces el material no sale de la mejor manera. generando desperdicios y retrocesos en la metalización de este. Afectando de manera continua la eficiencia en las metalizadora. También muchas veces genera roturas en el proceso, generando tiempo muertos que atrasan la producción

La principal problemática que ocurre en la metalización de estos rollos, son los parámetros que se colocan en las metalizadoras. ya que estos no son demasiados óptimos para el proceso productivo, ya que muchas veces el material no sale de la mejor manera. generando desperdicios y retrocesos en la metalización de este. Afectando de manera continua la eficiencia en las metalizadora. También muchas veces genera roturas en el proceso, generando tiempo muertos que atrasan la producción.

Se registró en una base de datos en Excel cada uno de los rollos que fueron metalizados para la producción. Esta base de datos contiene las características de cada rollo, como el ancho, peso y longitud, así como los diferentes parámetros utilizados para la metalización. Posteriormente, solicitamos al analista de calidad los resultados obtenidos en el laboratorio para poder tomar decisiones más informadas en relación a los datos recopilados.

Palabras clave: Metalizacion, parametros, mejoras.

Optimización en el tratamiento para la dislipidemia mediante estrategias de machine learning

Deiby Boneu Yepez*, D. Sierra-Porta
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: djboneu@hotmail.com

Resumen: La dislipidemia es un factor de riesgo importante para el desarrollo de enfermedades cardiovasculares, por lo que mantener niveles adecuados de lipoproteínas de baja densidad (LDLc) es crucial para la atención de los pacientes. A pesar de esto, un porcentaje significativo de pacientes con alto riesgo cardiovascular no logra alcanzar un control adecuado de sus niveles de LDLc. Esto puede deberse a diversas causas, como factores genéticos, ambientales, adherencia al tratamiento, dieta y ejercicio, siendo la mala utilización de la medicación uno de los principales problemas.

En este contexto, se propone una estrategia predictiva basada en machine learning para sugerir un tratamiento farmacológico adecuado para cada paciente. El objetivo es mejorar el control de la dislipidemia y reducir el riesgo cardiovascular mediante una selección personalizada de medicación.

Para probar esta estrategia, se utiliza una base de datos de la práctica clínica con pacientes que presentan algún riesgo cardiovascular. Esta base de datos contiene información relevante sobre los pacientes, como su historial médico, factores de riesgo, resultados de exámenes y tratamientos previos.

Se emplean técnicas de machine learning, como algoritmos de clasificación y regresión, para analizar y modelar la relación entre las características de los pacientes y el éxito del tratamiento farmacológico en el control del LDLc. Esto permite identificar patrones y factores predictivos que influyan en la eficacia del tratamiento.

El uso de machine learning proporciona una ventaja significativa al poder considerar múltiples variables y relaciones complejas entre los datos clínicos, lo que ayuda a mejorar la precisión y la personalización del tratamiento. Con esta estrategia predictiva, se busca identificar qué pacientes pueden responder mejor a un determinado tipo de medicación o enfoque terapéutico.

Los resultados de esta investigación tienen el potencial de mejorar significativamente la calidad de la atención médica para pacientes con dislipidemia y riesgo cardiovascular. Al ofrecer recomendaciones personalizadas de tratamiento, se puede aumentar la adherencia y el control de LDLc, lo que, a su vez, reducirá el riesgo cardiovascular y mejorará la salud cardiovascular general de los pacientes.

En conclusión, la optimización del tratamiento para la dislipidemia mediante estrategias de machine learning representa un enfoque prometedor para mejorar el manejo de esta condición médica. La aplicación de técnicas predictivas en la atención clínica permitirá una selección más precisa de tratamientos farmacológicos, contribuyendo a una atención médica más efectiva y personalizada para pacientes con dislipidemia y riesgo cardiovascular.

Palabras clave: Machine learning, ldl colesterol, hipolipemiantes.

Impacto del covid19 en las características estructurales y de conexión de los componentes del índice S&P Latam 40 del mercado bursátil latinoamericano

Libardo Visbal Ballesteros*, Andy Rafael Dominguez-Monterrosa
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: libavisbal@gmail.com

Resumen: Este estudio tiene como objetivo examinar el efecto de la Pandemia de COVID-19 en las propiedades topológicas del mercado de valores S&P LATAM40 Utilizando el MST entre otras medidas de centralidad y comparar el antes, durante y después del primer año de la Pandemia de COVID-19.

Como pregunta de investigación se plantea: ¿Ha reaccionado el mercado de valores a la pandemia de COVID-19?

El estudio se justifica en virtud del colapso de la economía mundial consecuencia de la pandemia de COVID-19, suspendiendo muchas actividades económicas y creando un declive dramático y abrupto en la demanda y el empleo. Como resultado, los precios del mercado de valores mundial experimentaron la peor caída desde la crisis financiera mundial.

Se plantea la hipótesis alternativa de que la pandemia ha tenido un impacto significativo en la estructura y conexión de los componentes del índice. Para evaluar esta hipótesis, se utiliza una metodología basada en el análisis de datos financieros y técnicas de minería de datos.

Palabras clave: COVID19, S&P LATAM40, PANDEMIA, MERCADOS FINANCIEROS.

Aplicación de modelo de regresión lineal y técnica de machine learning para generar un modelo matemático que permita calcular el déficit habitacional de vivienda en el distrito de Cartagena de indias y proyectarlo mediante mapas de calor, desagregados por localidad, unidad comunera, barrios y manzanas

Steven Calvo Benavides*, Yady Tatiana Solano Correa
*Maestría en Estadística y Ciencia de Datos, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: scalvo@utb.edu.co

Resumen: Esta investigación tiene como objetivo principal aplicar un modelo de regresión lineal y técnicas de Machine Learning para identificar variables estadísticas relevantes y crear un modelo matemático que permita calcular el déficit habitacional en el Distrito de Cartagena de Indias. La proyección de los resultados en mapas de calor desagregados por localidad, unidad comunera, barrios y manzanas proporcionará una visualización clara de la distribución del déficit habitacional, facilitando así la toma de decisiones informadas para abordar esta problemática en el ámbito urbano.

Palabras clave: Déficit, Déficit habitacional, Déficit Cuantitativo, Déficit Cualitativo, Regresión, Regresión Lineal, Machine Learning.

Drone mission and flight planning methodology for drainage water channels reconstruction.

Camilo Naufal*, Laura Paredes, César González, Yady T. Solano, Andrés G. Marrugo
*Facultad de Ingeniería, Universidad Tecnológica de Bolívar, Cartagena de Indias, Colombia.
email: cnaufal@utb.edu.co

Resumen: In the last decade, there has been an increasing use of surveying and mapping applications through remote sensing imagery and UAV platforms. This cutting-edge technology is being utilized for reconstructing drainage water channels, allowing for the analysis of important characteristics. These analyses help in planning, management, and decision-making, particularly in situations involving potential flood risks. To achieve precise reconstruction, it is necessary to develop a good flight protocol. This project aims to acquire aerial images for photogrammetry, maximizing data acquisition efficiency by obtaining high-quality images to performance assseasment of the visión system implemented. The flight methodology involved observing the study area at Universidad Tecnológica de Bolívar, delimitation of the zone using QGIS and mission flight planning software and establishing the positioning of Ground Control Points (GCPs) or targets using a precise GPS and others topogrphic tools. Images were acquired with black and white concentric circles as targets. Spatial Resolution below 5 cm/px was compute for the tested alttitudes to compare with the real experiments once the images have been precessed.

Palabras clave: Surveying, mapping applications, remote sensing imagery, UAV platforms, drainage water channels, photogrammetry, spatial resolution.

Universidad Tecnológica de Bolívar

CARTAGENA DE INDIAS

WWW.UTB.EDU.CO | utboficial