

CIENCIAS EXACTAS

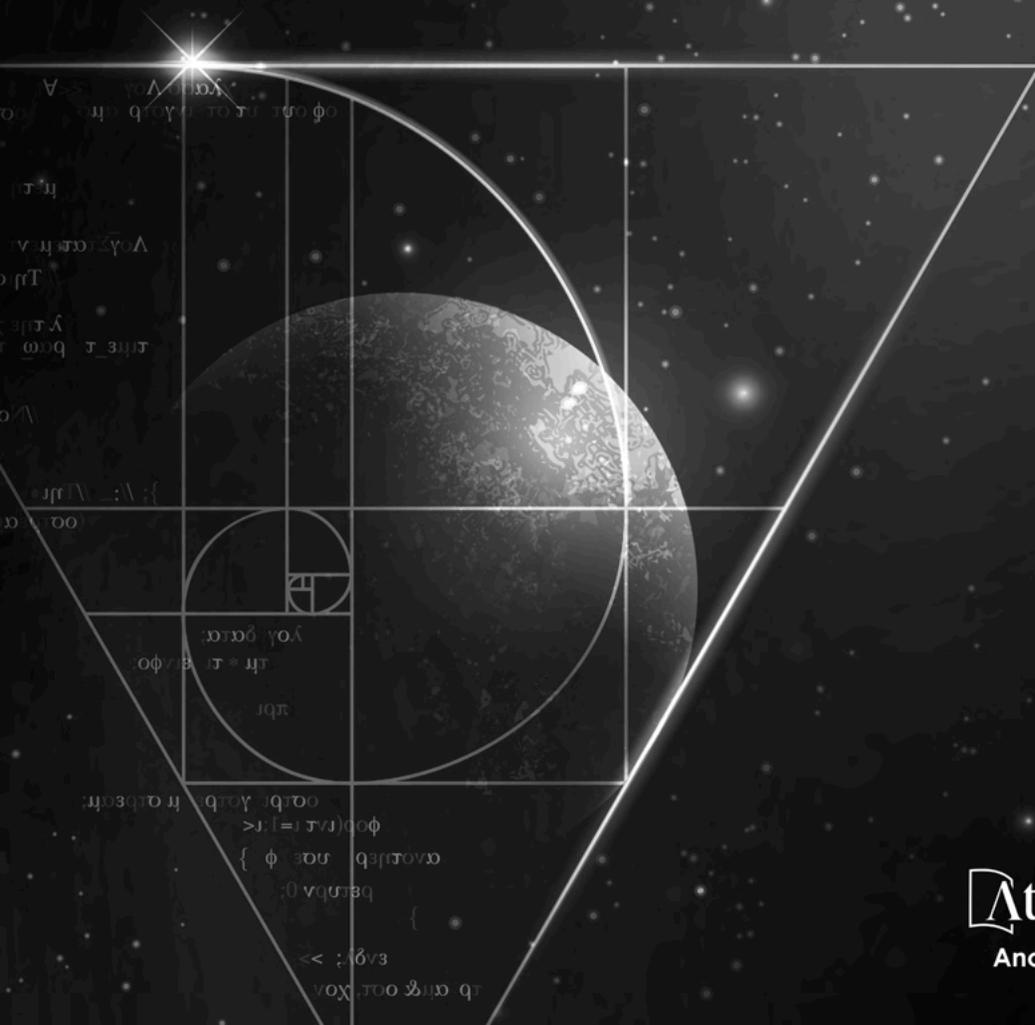
Y DE LA TIERRA:

Observación, formulación y predicción

4

CLEISEANO EMANUEL DA SILVA PANIAGUA

(Organizador)



Editora chefe

Profª Drª Antonella Carvalho de Oliveira

Editora executiva

Natalia Oliveira

Assistente editorial

Flávia Roberta Barão

Bibliotecária

Janaina Ramos

Projeto gráfico

Bruno Oliveira

Camila Alves de Cremo

Luiza Alves Batista

Natália Sandrini de Azevedo

Imagens da capa

iStock

Edição de arte

Luiza Alves Batista

2022 by Atena Editora

Copyright © Atena Editora

Copyright do texto © 2022 Os autores

Copyright da edição © 2022 Atena Editora

Direitos para esta edição cedidos à Atena Editora pelos autores.

Open access publication by Atena Editora



Todo o conteúdo deste livro está licenciado sob uma Licença de Atribuição Creative Commons. Atribuição-Não-Comercial-Não-Derivativos 4.0 Internacional (CC BY-NC-ND 4.0).

O conteúdo dos artigos e seus dados em sua forma, correção e confiabilidade são de responsabilidade exclusiva dos autores, inclusive não representam necessariamente a posição oficial da Atena Editora. Permitido o *download* da obra e o compartilhamento desde que sejam atribuídos créditos aos autores, mas sem a possibilidade de alterá-la de nenhuma forma ou utilizá-la para fins comerciais.

Todos os manuscritos foram previamente submetidos à avaliação cega pelos pares, membros do Conselho Editorial desta Editora, tendo sido aprovados para a publicação com base em critérios de neutralidade e imparcialidade acadêmica.

A Atena Editora é comprometida em garantir a integridade editorial em todas as etapas do processo de publicação, evitando plágio, dados ou resultados fraudulentos e impedindo que interesses financeiros comprometam os padrões éticos da publicação. Situações suspeitas de má conduta científica serão investigadas sob o mais alto padrão de rigor acadêmico e ético.

Conselho Editorial**Ciências Exatas e da Terra e Engenharias**

Prof. Dr. Adélio Alcino Sampaio Castro Machado – Universidade do Porto

Profª Drª Alana Maria Cerqueira de Oliveira – Instituto Federal do Acre

Profª Drª Ana Grasielle Dionísio Corrêa – Universidade Presbiteriana Mackenzie

Profª Drª Ana Paula Florêncio Aires – Universidade de Trás-os-Montes e Alto Douro

Prof. Dr. Carlos Eduardo Sanches de Andrade – Universidade Federal de Goiás

Profª Drª Carmen Lúcia Voigt – Universidade Norte do Paraná



Prof. Dr. Cleiseano Emanuel da Silva Paniagua – Instituto Federal de Educação, Ciência e Tecnologia de Goiás
Prof. Dr. Douglas Gonçalves da Silva – Universidade Estadual do Sudoeste da Bahia
Prof. Dr. Eloi Rufato Junior – Universidade Tecnológica Federal do Paraná
Profª Drª Érica de Melo Azevedo – Instituto Federal do Rio de Janeiro
Prof. Dr. Fabrício Menezes Ramos – Instituto Federal do Pará
Profª Dra. Jéssica Verger Nardeli – Universidade Estadual Paulista Júlio de Mesquita Filho
Prof. Dr. Juliano Bitencourt Campos – Universidade do Extremo Sul Catarinense
Prof. Dr. Juliano Carlo Rufino de Freitas – Universidade Federal de Campina Grande
Profª Drª Luciana do Nascimento Mendes – Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte
Prof. Dr. Marcelo Marques – Universidade Estadual de Maringá
Prof. Dr. Marco Aurélio Kistemann Junior – Universidade Federal de Juiz de Fora
Prof. Dr. Miguel Adriano Inácio – Instituto Nacional de Pesquisas Espaciais
Profª Drª Neiva Maria de Almeida – Universidade Federal da Paraíba
Profª Drª Natiéli Piovesan – Instituto Federal do Rio Grande do Norte
Profª Drª Priscila Tessmer Scaglioni – Universidade Federal de Pelotas
Prof. Dr. Sidney Gonçalo de Lima – Universidade Federal do Piauí
Prof. Dr. Takeshy Tachizawa – Faculdade de Campo Limpo Paulista



Ciências exatas y de la tierra: observación, formulación y predicción 4

Diagramação: Camila Alves de Cremo
Correção: Maiara Ferreira
Indexação: Amanda Kelly da Costa Veiga
Revisão: Os autores
Organizador: Cleiseano Emanuel da Silva Paniagua

Dados Internacionais de Catalogação na Publicação (CIP)

C569 Ciências exatas y de la tierra: observación, formulación y predicción 4 / Organizador Cleiseano Emanuel da Silva Paniagua. – Ponta Grossa - PR: Atena, 2022.

Formato: PDF

Requisitos de sistema: Adobe Acrobat Reader

Modo de acesso: World Wide Web

Inclui bibliografia

ISBN 978-65-258-0622-8

DOI: <https://doi.org/10.22533/at.ed.228221410>

1. Ciências exatas y de la tierra. 2. Matemáticas. I. Paniagua, Cleiseano Emanuel da Silva (Organizador). II. Título.

CDD 507

Elaborado por Bibliotecária Janaina Ramos – CRB-8/9166

Atena Editora

Ponta Grossa – Paraná – Brasil

Telefone: +55 (42) 3323-5493

www.atenaeditora.com.br

contato@atenaeditora.com.br



Atena
Editora
Ano 2022

DECLARAÇÃO DOS AUTORES

Os autores desta obra: 1. Atestam não possuir qualquer interesse comercial que constitua um conflito de interesses em relação ao artigo científico publicado; 2. Declaram que participaram ativamente da construção dos respectivos manuscritos, preferencialmente na: a) Concepção do estudo, e/ou aquisição de dados, e/ou análise e interpretação de dados; b) Elaboração do artigo ou revisão com vistas a tornar o material intelectualmente relevante; c) Aprovação final do manuscrito para submissão.; 3. Certificam que os artigos científicos publicados estão completamente isentos de dados e/ou resultados fraudulentos; 4. Confirmam a citação e a referência correta de todos os dados e de interpretações de dados de outras pesquisas; 5. Reconhecem terem informado todas as fontes de financiamento recebidas para a consecução da pesquisa; 6. Autorizam a edição da obra, que incluem os registros de ficha catalográfica, ISBN, DOI e demais indexadores, projeto visual e criação de capa, diagramação de miolo, assim como lançamento e divulgação da mesma conforme critérios da Atena Editora.



DECLARAÇÃO DA EDITORA

A Atena Editora declara, para os devidos fins de direito, que: 1. A presente publicação constitui apenas transferência temporária dos direitos autorais, direito sobre a publicação, inclusive não constitui responsabilidade solidária na criação dos manuscritos publicados, nos termos previstos na Lei sobre direitos autorais (Lei 9610/98), no art. 184 do Código Penal e no art. 927 do Código Civil; 2. Autoriza e incentiva os autores a assinarem contratos com repositórios institucionais, com fins exclusivos de divulgação da obra, desde que com o devido reconhecimento de autoria e edição e sem qualquer finalidade comercial; 3. Todos os e-book são *open access*, *desta forma* não os comercializa em seu site, sites parceiros, plataformas de *e-commerce*, ou qualquer outro meio virtual ou físico, portanto, está isenta de repasses de direitos autorais aos autores; 4. Todos os membros do conselho editorial são doutores e vinculados a instituições de ensino superior públicas, conforme recomendação da CAPES para obtenção do Qualis livro; 5. Não cede, comercializa ou autoriza a utilização dos nomes e e-mails dos autores, bem como nenhum outro dado dos mesmos, para qualquer finalidade que não o escopo da divulgação desta obra.



PRESENTACIÓN

El e-book titulado: “Ciencias exactas y de la tierra: Observación, formulación y predicción 4” consta de seis capítulos de libros que buscaban investigar: *i)* el proceso de enseñanza- aprendizaje de las ecuaciones en la formación de estudiantes de secundaria y superior curso de ingeniería; *ii)* aplicación de las matemáticas en estudios meteorológicos y desarrollo de aeronaves pilotadas a distancia (RPA); *iii)* análisis de estabilidad coloidal por espectroscopía óptica y voltamperometría; *iv)* evaluación de áreas de preservación permanente (APP's) en la ciudad de Marabá/PA.

El primer capítulo evaluó las numerosas dificultades presentadas en el proceso de enseñanza- aprendizaje de las matemáticas para los estudiantes de secundaria, quienes señalaron una serie de dificultades en la escritura, lectura, interpretación y resolución de problemas en relación a las operaciones matemáticas básicas. El capítulo 2 investigó el uso del algoritmo árbol en aplicaciones de parámetros meteorológicos, lo que resultó en una precisión del 80% en relación al 62% que presenta la regresión bayesiana. El tercer capítulo evaluó los numerosos factores que inciden en el desempeño de los estudiantes de matemáticas en las carreras de ingeniería del Instituto Politécnico Nacional de México, entre los que se encuentran: *i)* las ausencias a clases al final del semestre; *ii)* exceso de interacción social y confianza en los primeros semestres; *iii)* falta de disciplina en los estudios extracurriculares; *iv)* falta de búsqueda de estudios en grupos y la ayuda de medios digitales.

El capítulo 4 presentó un estudio de revisión de literatura que demuestra la amplia aplicación de RPA y la posibilidad de innovación en relación con la recopilación de datos de forma rápida y a bajo costo. El quinto capítulo evaluó el uso de técnicas espectroscópicas (UV-Vis-IR) y electroanalíticas (volamperometría cíclica) en estudios de caracterización de coloides preparados a partir de nanopartículas (NP's) de plata, los resultados mostraron que las técnicas pueden ser utilizadas in loco y que constituyen un instrumentación compacta, simple y de bajo costo. Finalmente, el sexto capítulo investigó el crecimiento urbano de las APP en la región de Cidade Nova en Marabá/PA, estudios realizados entre los años 1990 a 2015 identificaron una disminución del 28% de las APP, lo que resulta en propuestas urgentes de políticas públicas que pueden garantizar tanto la conservación de las APP como la revisión del Plan Director del municipio de Marabá.

En esta perspectiva, Atena Editora viene trabajando para estimular y alentar a cada vez más investigadores de Brasil y de otros países a publicar sus trabajos con garantía de calidad y excelencia en forma de libros, capítulos de libros y artículos científicos.

SUMÁRIO

CAPÍTULO 1..... 1

EQUAÇÕES DE PRIMEIRA SÉRIE PARA O ENSINO MÉDIO NO CONTEXTO DO ENSINO PARA COMPREENSÃO

Edwin Smith Rivera Fernández

Romelio José Gonzales Daza

Gustavo Adolfo Rodriguez

Alcides Paes Soto

 <https://doi.org/10.22533/at.ed.2282214101>

CAPÍTULO 2..... 11

ALGORITMO DE CLASIFICACIÓN MEDIANTE UN ENFOQUE DE MACHINE LEARNING Y SU APLICACIÓN AL ESTUDIO METEOROLÓGICO

Pedro Elizardo Donis del Cid

 <https://doi.org/10.22533/at.ed.2282214102>

CAPÍTULO 3..... 25

ANÁLISIS DE AUTORREGULACIÓN EN FORMACIÓN MATEMÁTICA DE INGENIEROS EN LA UNIDAD PROFESIONAL INTERDISCIPLINARIA DE INGENIERÍA CAMPUS GUANAJUATO (UPIIG)

Gilda Rosa Bolaños Evia

Lenin Augusto Echavarría Cepeda

Luis Rey Díaz Barrón

Yazpik Hernández Vargas

 <https://doi.org/10.22533/at.ed.2282214103>

CAPÍTULO 4..... 33

AERONAVES REMOTAMENTE PILOTADAS: BREVE ABORDAGEM

Dalton Nasser Muhammad Zeidan

Renan Valério Eduvirgem

Maria Eugênia Moreira Costa Ferreira

 <https://doi.org/10.22533/at.ed.2282214104>

CAPÍTULO 5..... 40

ANÁLISIS DE ESTABILIDAD DE COLOIDES DE NANOPARTÍCULAS DE PLATA POR ESPECTROSCOPIA-ÓPTICA Y VOLTAMETRÍA

Margarita Navarrete Montesinos

Rodrigo Mayén-Mondragón

Daniel Aguirre-Aguirre

 <https://doi.org/10.22533/at.ed.2282214105>

CAPÍTULO 6..... 55

CRESCIMENTO URBANO NAS ÁREAS DE PRESERVAÇÃO PERMANENTE (APPS): UM ESTUDO DE CASO DA MARGEM DO RIO ITACAIÚNAS NO NÚCLEO CIDADE NOVA, MARABÁ-PARÁ

Ana Carolina Seabra de Vilhena Linhares

Priscylla Assis Carvalho
Jakeline Oliveira Evangelista
André dos Santos Araújo
Glauber Epifanio Loureiro

 <https://doi.org/10.22533/at.ed.2282214106>

SOBRE EL ORGANIZADOR	68
ÍNDICE REMISSIVO	69

CAPÍTULO 2

ALGORITMO DE CLASIFICACIÓN MEDIANTE UN ENFOQUE DE MACHINE LEARNING Y SU APLICACIÓN AL ESTUDIO METEOROLÓGICO

Data de aceite: 03/10/2022

Data de submissão: 19/07/2022

Pedro Elizardo Donis del Cid

Universidad Mariano Gálvez
Guatemala

<https://orcid.org/0000-0003-0844-9796>

RESUMEN: La clasificación de datos propone una polaridad tanto en el algoritmo de árbol como en el teorema de las causas. Se necesita hacer uso de conjuntos de datos, en este estudio son variables meteorológicas, para construir un modelo predictivo basado en entrenamiento. Los datos son almacenados y procesados haciendo uso de almacenes en grandes volúmenes de datos, estos proponen un sistema basado en entidades con independencia sustancial, de objetos de datos, a partir de una técnica llamada ciencia de datos o Data Science. Técnica que permite adquirir información de valor de los datos, en este caso equipos y sensores de IoT. Inicialmente se capturan en formato no procesado o Raw Data. El objetivo del estudio es hacer uso de las plataformas tecnológicas disponibles para trabajar el pronóstico de datos de meteorología basados en algoritmos de ML (Machine Learning) y datos históricos para el altiplano central, costa del pacífico y valles de oriente de Guatemala. Los métodos utilizados son estadísticos embebidos en algoritmos de tipo predictivos, además de una matriz de confusión para evaluar los resultados obtenidos

en el estudio de estas herramientas y recursos, con un enfoque cuantitativo e instrumentos de IoT, ML y Big Data. El tipo de estudio general es correlacional-predictivo-histórico del lado de la variable y comparativo-explicativo para el algoritmo. Las variables tienen una dimensión anual, mensual y diaria entre 2000 y 2018. En conclusión, se logra el pronóstico del nivel de humedad para las muestras seleccionadas en base a las variables meteorológicas tomadas con dispositivos electrónicos y procesadas por medios informáticos entre dos algoritmos y un dieciocho por ciento de diferencia al comparar la eficiencia.

PALABRAS CLAVE: Clasificación de datos, Internet de las Cosas, Industria 4.0, Modelo Supervisado.

CLASSIFICATION ALGORITHM THROUGH A MACHINE LEARNING APPROACH AND ITS APPLICATION TO THE METEOROLOGICAL STUDY

ABSTRACT: Data classification is a polarity both in the tree algorithm and in the theorem of causes. It is necessary to do use of data sets, in this study they are meteorological variables, to build a predictive model based on training data. The data is stored and processed using big data warehouses, it's a system based on entities with substantial independence from data objects, based on a technique of Data Science. Technique that allows acquiring valuable information from data, in this case IoT equipment and sensors. Initially they are captured in unprocessed format or Raw Data. The objective of the study is to

do use of the available technological platforms to work on the forecast of meteorology data based on Machine Learning Algorithms and historical data for the central highland, pacific coast and valleys of eastern Guatemala. The methods used are statistics embedded in predictive algorithms, in addition to a confusion matrix to evaluate the results obtained in the study of these tools and resources, with a quantitative approach and IoT, Machine Learning and Big Data instruments. The type of general research is correlational-predictive-historical on the side of the variable and comparative-explanatory for the algorithm. The metrics of variables have an annual, monthly and daily dimension between 2000 and 2018. In conclusion, the forecast of the humidity level for the selected samples is accomplished based on the meteorological variables taken with electronic devices and processed by computer between two algorithms and an eighteen percent difference when comparing the efficiency.

KEYWORDS: Data Classification, Internet of Things, Industry 4.0, Supervised Model.

1 | INTRODUCCIÓN

La Industria 4.0 ante todo, la clasificación de datos y el aprendizaje de la máquina por medio de datos, es decir, la generación código que puede convertirse en acciones de un agente de IA sin necesidad de ser previamente programado (Ain, et al., 2015; Nicholson, et al., 2019; Réda, et al., 2020; Turesson et al., 2016; Vu et al., 2018; Zhang, et al., 2017). Es la base de esta revolución industrial (Galvão, et al., 2022; Osmana, Ghirana, 2019). Además, del aumento de las capacidades en las centrales de datos por medio de la tecnología Big Data y la posibilidad de conectividad y recolección de datos automatizado por medio de IoT (Qaffas et al., 2021). Incluye sensores y otras tecnologías especializadas.

Big Data es una tecnología que se ha desarrollado debido a los crecimientos exponenciales en volumen de datos a nivel mundial y las limitaciones de los microprocesadores actuales. Estas técnicas habilitan el cómputo para trabajar una tarea compleja en múltiples nodos, con varios núcleos y memoria agrupada (Franke et al., 2016; Qaffas et al., 2021; Tang et al., 2019; van Evert et al., 2017). Por su parte, IoT recolecta la información, es un conjunto de componentes electrónicos que tienen la capacidad de conectarse y comunicar datos (Qaffas, et al., 2021).

Para este estudio se utilizaron sensores que son implementados y administrados por el Instituto Nacional de Sismología, Vulcanología, Meteorología e Hidrología de la República de Guatemala y es dependiente del Ministerio de Comunicaciones y Obras Públicas. Según acuerdo gubernativo de 26 de marzo de 1976. Estos sensores tienen la capacidad de comunicación con la base de almacenamiento de datos, es decir, en la parte operativa de recolección de los datos de la investigación no es necesario la presencia de entrevistas o encuestas para capturar datos. Pero si es necesario configurar los equipos y parametrizar los sistemas.

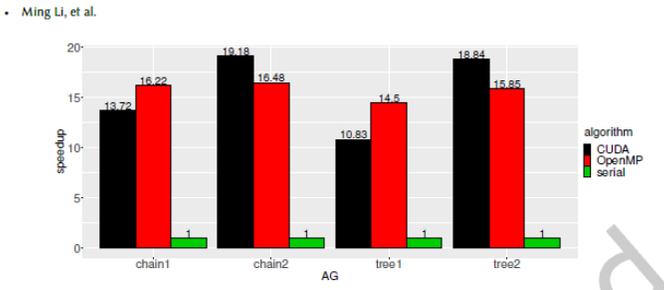
En el método utilizado se hizo un resumen de las variables agrupándolas por día, mes y año. Y las variables operadas de manera continua de temperatura, lluvia, velocidad

del viento, radiación, nubosidad:



Elaboración propia. Origen de datos: Unidad de Información Pública, INSIVUMEH

Por su parte, Galvão et al (2022) utilizando la plataforma Cloudera acompañado de Spark con Python logró presentar un modelo con datos de un proceso industrial. Se usaron funciones de agregación, reportes por medio de tableros. Donde, se determinó un déficit de tornillos apretados del 14.3% para ello se utilizaron DataFrame a partir de objetos JVM y utilizando métodos de mapas y reductores por medio de API's muy semejante a la utilización del lenguaje R. Después de realizar la adquisición de datos, procesarlos y almacenarlos, se generan todos los sistemas de visualización de datos para los temas administrativos correspondientes donde la comunicación entre máquinas es muy importante porque los datos no son digitalizados por personas.



Fuente: Li et al (2022)

Li et al. (2022) con el experimento se realiza comparaciones entre los algoritmos de grafo de ataque utilizando computación de alto rendimiento para algoritmos OpenMP y CUDA AG en un sistema de procesamiento distribuido. El experimento ha demostrado la eficiencia de la estrategia del procesamiento en paralelo. Los resultados fueron utilizados para hacer parche de las vulnerabilidades.

2 | ALMACENES DE DATOS

Los almacenes de datos han estado presentes en la industria desde hace algunos años, sin embargo, gracias a esos datos históricos que se guardan en estos almacenes es

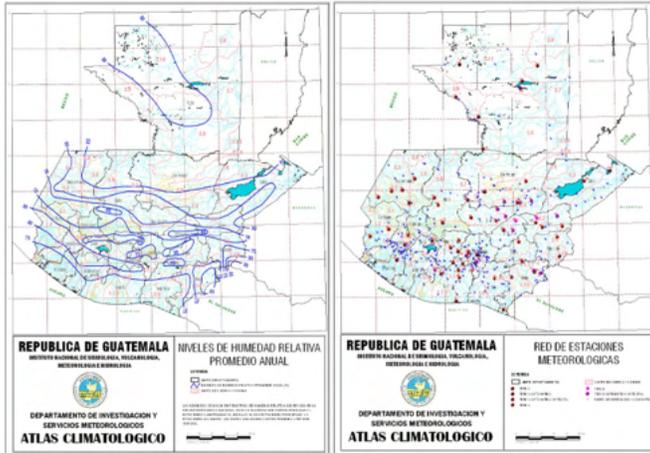
posible predecir. Esto combinado con la variedad de datos que se tienen con las nuevas tecnologías de Big Data y IoT trasciende en tecnologías digitales que permiten mejorar procesos industriales con impactos económicos importantes. La investigación de Galvão et al (2022) hace referencia al uso de estas tecnologías emergentes que se aplican para el tratamiento de la información, tanto la extracción como, el tratamiento y almacenamiento.

En los resultados se puede apreciar una medición a partir de un proceso industrial el cual denota información que se guarda en Big Data, se procesa y se envía para el monitoreo por medio de tableros. Las herramientas estadísticas e informáticas toman un papel relevante. Gracias a la estadística es posible predecir o inferir los sucesos próximos para anticiparse a los escenarios que se presentan. La informática hace posible recolectar, almacenar y procesar la información que llega desde máquinas como es el caso de IoT. Tecnología que permite, por medio de internet, transferir datos de sensores y otros elementos de medición.

Muchos de los gráficos que muestra este caso de estudio está orientado al uso de software especializado para este diseño de información. Donde, se hace presente nuevos métodos para visualizar los datos. Datos que son importantes para la toma de decisiones.

3 | HUMEDAD RELATIVA

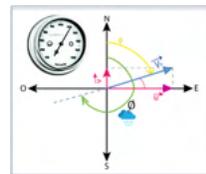
Dentro de la metrología, la humedad relativa viene dada por una relación entre el vapor de agua que tiene el aire y la cantidad que debería tener para saturarse. Este valor viene dado en porcentaje. La fórmula química del agua es H_2O dos moléculas de hidrógeno y uno de oxígeno, este es entonces un compuesto químico. La atmósfera terrestre son mezclas de compuestos en estado gaseoso sobre la superficie de la tierra. Para este estudio se tuvo a bien estudiar variables meteorológicas cerca de la tierra, siendo estas las que se encuentra en interacción directa con las personas, existen otras partes de la atmósfera terrestre que interactúan con la troposfera y que alteran la misma en primer lugar la estratosfera y otras capas que interactúan con esta última. A continuación, se presenta un mapa general:



Fuente: https://insivumeh.gob.gt/hidrologia/ATLAS_HIDROMETEOROLOGICO/Atlas_Climatologico/esmeteo.jpg

La metrología y la meteorología, ambas interactúan con esta variable climática y tiene que ver en gran medida con el sistema socioeconómico en la naturaleza. Dada la naturaleza de ciencias exactas, ocupan diferentes instrumentos para la experiencia y comprensión de los sucesos. Haciendo uso de mapas de información geológica. Para este estudio se ocupó el concepto de Teorema de Bayes (teorema de las causas) para el computo de resultados (Mesa et al., 2021; Tang et al., 2019) enunciado por Thomas Bayes, donde se denota una probabilidad condicional en términos de distribución de probabilidad: $P(A|B) = P(B \cap A) / P(A)$.

$$P(A_i / B) = \frac{P(A_i) P(B / A_i)}{\sum_{i=1}^n P(A_i) P(B / A_i)}$$



Este teorema se caracteriza por representar los resultados en términos de la probabilidad de que un evento ocurra dado que ha sucedido otro. Esta reacción se puede hacer en cadena para poder unir variables. Lo que se hizo en este estudio, pero con la diferencia que se utilizaron herramientas avanzadas de informática. Esta última, se ocupa del tratamiento automático de la información. En el teorema existe un acercamiento a la probabilidad, a 100% o a 0%.

4 | METODOLOGÍA

El estudio hace uso de las plataformas tecnológicas disponibles para trabajar el pronóstico de datos de meteorología basados en algoritmos de ML y datos históricos Big Data Warehouse para el altiplano central, costa del pacífico y valles de oriente para Guatemala.

El método utilizado fue la integración de datos por medio de sensores capaces de tomar las mediciones y transmitir los datos. Se utilizó un modelo supervisado de aprendizaje de máquina con datos cuantitativos de las variables de metrología. Consistente en determinar una variable categórica de baja humedad o alta humedad en base al conjunto de variables independientes del modelo. Además, se utilizó un algoritmo de árbol de decisión y otro método de inferencia bayesiana, muy parecida a la regresión logística que ocupa una variable categórica dicotómica con resultado final. Esto se da por aproximación a un valor ya sea 0 o 1.

El enfoque de la investigación tiene como propósito la indagación sobre estos algoritmos especializados y orientados a datos para que pueda ser aplicado en otras industrias por los profesionales de las ciencias de la computación. Dentro de las herramientas y tareas se llevó acabo la configuración de los equipos para la transferencia de comunicación o transferencia de datos y, una herramienta para la configuración del flujo para el job del procesamiento en ML:



Diseño Conceptual del Proceso. Elaboración propia.

Las variables fueron agrupadas por la dimensión de día, mes y año a partir del año 2000 con 3,690 observaciones completas sin valores atípicos para el altiplano central, muestra de: Santiago Atitlán, Cubulco; 43,464 observación para el pacifico, muestra para Puerto de San Jose; 6,478 observaciones después de la limpieza para valles de oriente, muestra Esquipulas y La Unión. Las variables utilizadas fueron de razón, de intervalo y una variable predictiva con valor dual (dicotómica). Se utilizaron dos algoritmos como métodos

predictivos, además de un matriz de confusión para evaluar los resultados obtenidos en el estudio de las herramientas y recursos, con corte longitudinal, enfoque cuantitativo con instrumentos de IoT, ML y Big Data, y un tipo de estudio predictivo desde el punto de vista de la variable analizada y descriptivo para los algoritmos.

Las mediciones fueron continuas con variables de meteorología: Presión del aire, temperatura dirección promedio del viento, velocidad media del viento, dirección máxima del viento, velocidad máxima del viento, acumulación de lluvia, duración de la lluvia y cantidad de tiempo lloviendo, y procesadas ser registrarse en almacenes de datos de forma periódica en ciclos: diarios, mensuales y anuales para el rango de fechas del año 2000 al 2018. El almacenamiento de estos datos y la calidad juega un papel muy importante para poder generar conjuntos de datos resumidos de las variables.

5 | RESULTADOS

Guatemala es un país agrícola que depende mucho de las situaciones climatológicas para los medios de producción, a continuación, se presenta una imagen de una ciudad (casco urbano de municipio o pueblo) promedio y su situación socioeconómica:

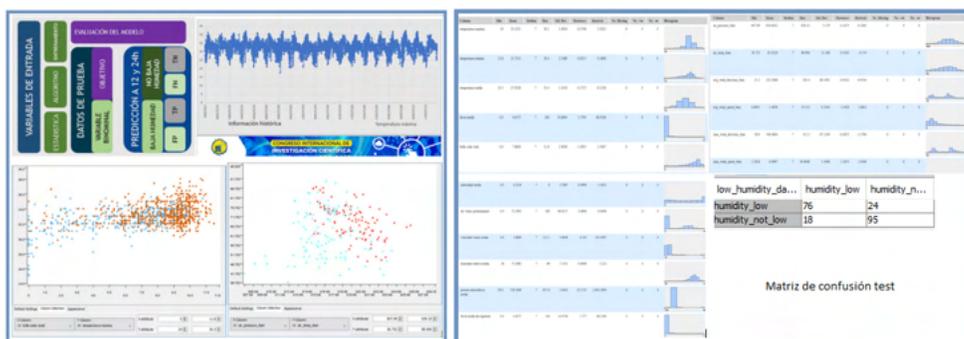


Fuente: Google

Se caracteriza por dos épocas clásicas, una lluviosa y otra seca, la primera va de mayo a octubre y la última de noviembre a abril. La época seca se da por un incremento en la presión atmosférica y movimientos de aire frío. La época lluviosa por lo general da inicios en el mes de mayo. Las temporadas de siembra y cosecha, dependen muchas veces de estas épocas. Las cuales pueden variar en condiciones, lo cual puede afectar las cosechas o mejorar el rendimiento de la producción agrícola.

Se le aplicó un test por medio de la matriz de confusión. En la primera celda se

presenta, 76 observaciones efectivas pronosticando la humedad baja. Y en la última casilla, 95 observaciones fueron efectivas al pronosticar la humedad alta (no baja). El resto de valores presentados corresponden a la tasa de error del modelo cada uno de ellos tiene un tipo: Error Tipo I y Error Tipo II. Afirmar que es baja cuando no lo es, y, por otra parte, que es no baja cuando si lo es. El algoritmo de árbol de decisión ha mostrado una exactitud del 80% frente a un 62% de la regresión bayesiana. Así mismo al comparar el % de error se aprecia una diferencia del 18%. Gráfico de Variables.



En la parte superior izquierda podemos apreciar el esquema del modelo, el cual comprende un conjunto de datos de entrada: Presión del aire, temperatura dirección promedio del viento, velocidad media del viento, dirección máxima del viento, velocidad máxima del viento, acumulación de lluvia, duración de la lluvia y cantidad de tiempo lloviendo. La parte estadística que consiste en las operaciones con datos, el algoritmo, son instrucciones de máquina, para procesar las operaciones. Estas se llevan a cabo de forma ordenada y el entrenamiento.

Posteriormente sigue el modelo, este consiste en tomar la información del entrenamiento y proyectar con datos de prueba, es acá donde se tiene un objetivo que se traduce en una variable. Dentro del modelo también existe una parte de predicción y su parte medular es la matriz de confusión. La cual indica los FP Falsos Positivos TP Verdaderos Positivos FN Falsos Negativos TN Verdaderos Positivos. Cada uno de estos tiene una interpretación en la matriz. Y mide la efectividad del modelo. Para este estudio la variable dependiente es la humedad. La cual tiene un valor dicotómico, alta o baja.

En la parte inferior de la gráfica se puede ver la dispersión entre variables en este caso tenemos dos, el brío solar total y la temperatura máxima, esto puede cambiar dependiendo de cuales variables se comparen. Esto es importante para poder comparar entre sí las variables independientes. A continuación, se presenta una tabla con una muestra de datos, tanto las variables de entrada, la humeada real y la predictiva: Tabla de Muestra.

Row ID	D	temper...	D	temper...	D	temper...	D	fluvia m...	D	evapor...	I	nubosid...	I	dir vien...	D	velocid...	S	humedad di...	S	Prediccion (...)
38111		24.8		15.5		21.8		0		4.6		6		0		15		no_baja_humedad		no_baja_humedad
38113		25.5		14.5		19.8		0		3.1		5		0		17		baja_humedad		baja_humedad
38117		25.5		13.5		19.6		0		4.4		6		0		10		no_baja_humedad		no_baja_humedad
38122		25.3		12.8		18.3		0		1.9		6		0		15		no_baja_humedad		no_baja_humedad
38128		24.3		15.5		20.4		0		3.4		6		0		10		no_baja_humedad		no_baja_humedad
38135		23.8		14.3		19.1		1.4		28.3		5		0		9		no_baja_humedad		no_baja_humedad
38177		22		16		19.4		8.1		1.9		6		0		17		no_baja_humedad		no_baja_humedad
38182		24.5		15.3		19.7		0		3.7		6		0		17		no_baja_humedad		no_baja_humedad
38191		24.8		14.3		19.4		0		3.7		6		0		8		no_baja_humedad		no_baja_humedad
38193		26		15.5		19.8		84.2		3.3		6		0		214		no_baja_humedad		no_baja_humedad
38194		26.8		14		20.5		0		22.9		4		0		10		no_baja_humedad		baja_humedad
38204		24.5		15		20.3		0		1.3		3		0		15		no_baja_humedad		baja_humedad
38206		25		13		19.6		0		4.9		5		0		15		no_baja_humedad		no_baja_humedad
38210		25.5		15.3		20		0		4.9		6		0		20		no_baja_humedad		no_baja_humedad
38211		25.3		13.3		19		0		3.3		4		0		28		no_baja_humedad		no_baja_humedad
38219		27.5		17		20.7		0		5.2		3		0		28		no_baja_humedad		no_baja_humedad
38222		26		15		19.8		0		5.6		6		0		17		no_baja_humedad		baja_humedad
38238		23.5		13.5		17.6		6.3		4.1		6		0		13		no_baja_humedad		no_baja_humedad
38241		24.5		14.5		18.5		0		3		6		0		10		no_baja_humedad		no_baja_humedad
38242		23.8		13.8		17.8		7.2		4.2		5		0		10		no_baja_humedad		no_baja_humedad
38256		24		15.5		18.5		0.6		3.4		6		0		7		no_baja_humedad		no_baja_humedad
38261		25.5		15.5		19.8		0		2.8		4		0		16		no_baja_humedad		no_baja_humedad
38262		26.3		14.5		20.6		0		5.1		6		0		10		no_baja_humedad		no_baja_humedad
38263		24.5		17		19.4		0.1		1.1		5		0		7		no_baja_humedad		no_baja_humedad
38270		24		14.3		17.5		32.2		3.3		6		0		10		no_baja_humedad		no_baja_humedad

Entonces, al comparar los dos algoritmos la diferencia fundamental radica en la interpretación del modelo y de los resultados, donde el árbol de decisión es más fácil de ejemplificar y reconstruir con valores concretos. Por ejemplo, veamos la siguiente imagen: Árbol de decisión.



Podemos apreciar en las hojas del árbol; si la dirección del viento es mayor al parámetro indicado en el diagrama se clasifican las observaciones: Caso primero 15 y 14 respectivamente. Sin embargo, esto puede cambiar según el valor o parámetro del nodo padre anterior en el árbol, que es la variable de temperatura del aire: Si la temperatura del aire es menos o igual que 72.3578 se dirigen las muestras respectivamente, en el primer caso 39 con humedad baja y 240 con humedad alta. Y así el otro nodo hermano en el árbol, si la temperatura del aire es mayor que 72.3578; las clasificaciones son diferentes 86 y 86 respectivamente para la baja y la no baja (alta). Todos estos valores que indican hacia donde se dirige la observación o dato, depende del proceso que genera el algoritmo clasificador o modelo.

Cada mes o segmento de datos, tiene diferentes características climatológicas,

meses fríos, secos y lluviosos. Así, se observa que la temperatura para Santiago Atitlán no es variable comparación con los valles de oriente.



El resultado muestra una eficiencia en el algoritmo del árbol de decisión, el cual se resume a continuación:



En la raíz del árbol inicia los clasificadores como baja humedad con 213 casos. Los siguientes valores paramétricos en orden: 919.4, 72.3578, 66.1774, 89.4, 130.55, 185.95, 175.95; correspondientes a presión del aire, temperatura, dirección del viento promedio, dirección del viento media, respectivamente. Estos valores dependen directamente del modelo de entrenamiento de datos, el cual se puede ir afinando con más datos, es decir más variables o una granularidad menor en la medición cambiando de diario a horaria. Esto para poder pronosticar la humedad correcta.

Cada parte del árbol clasificador tiene un gráfico el cual indica la proporción del valor de la variable de estudio que es clasificada en cada nodo. Estos se van agrupando desde las hojas y los nodos padres hasta llegar a la raíz. Es más fácil visualizar las decisiones

Cuando se compara porcentaje de error se aprecia una diferencia del 18%. Para finalizar, es importante hacer notar que ML (aprendizaje automático o aprendizaje de máquina) es una sub disciplina de la IA, parte fundamental de la informática que se ocupa del tratamiento automático de la información y se apoya de la estadística (Amaya, et al., 2017; Contreras, et al., 2017; Ji, et al., 2018; Tabales, et al., 2017; Wang, et al., 2018). Este estudio se apoyó además de la metrología y meteorología, las cuales fueron descritas en este estudio, en consecuencia, se comprobó la efectividad de un algoritmo basado en árbol de decisión (Contreras, et al., 2017) respecto a la regresión bayesiana (Mesa et al., 2021; Tang et al., 2019) y aplicado a un análisis de correlación causal de variables del clima, de tres regiones en Guatemala.

Para el modelo, cada mes o segmento de datos, tiene diferentes características climatológicas, por ejemplo, existen meses más fríos que otros, dependiendo de la región, todo esto aplica para el resto de variables como dirección del viento, radiación solar, lluvias y temporadas secas. Además, en el caso de la temperatura para Santiago Atitlán no cambia demasiado en comparación con La Unión o Esquipulas.

7 | RECOMENDACIONES

Es importante dar mantenimiento a los equipos de medición para evitar valores atípicos, muchos de estos *missing values* o valores atípicos se deben a ausencia de información, por un incidente en el equipo, en la transmisión de datos o en el almacenamiento.

Utilizar más variables para poder mejorar el modelo que se ha presentado en esta investigación.

REFERENCIAS

- Abd A.; Tawalbeh L.; Maleh Y.; Saldamli G. (2021). Big Scientific Data and Machine Learning in Science and Engineering
- Ain, Q.; Aleksandrova, A.; Roessler, F.; Ballester, P. (2015). Machine-learning scoring functions to improve structure-based binding affinity prediction and virtual screening. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 5(6), 405–424. <https://doi.org/10.1002/wcms.1225>
- Amaya, C.; Magaña, P.; Ochoa, I. (2017). Evaluación de destinos turísticos mediante la tecnología de la ciencia de datos. Universidad de Colima – México.
- Contreras, E.; Ferreira, F.; Valle, M. (2017). Diseño de un Modelo Predictivo de Fuga de Clientes Utilizando Árboles de Decisión. *Revista Ingeniería Industrial*, 16(1), 07–23. <https://doi.org/10.22320/s07179103/2017.01>
- Filiberto, Y.; Caballero, Y.; Bello, R.; Frías, M. (2011). Algorithm to learn classification rules based on the extended rough set theory. *DYNA (Colombia)*, 78(169).

Franke, B.; Plante, J.; Roscher, R.; Lee, E.; Smyth, C.; Hatefi, A.; Chen, F.; Gil, E.; Schwing, A.; Selvitella, A.; Hoffman, M.; Grosse, R.; Hendricks, D.; Reid, N. (2016). Statistical Inference, Learning and Models in Big Data. *International Statistical Review*, 84(3), 371–389. <https://doi.org/10.1111/insr.12176>

Galvão J.; Ribeiro, D.; Machado, I.; Ferreira, F.; Gonçalves, J.; Faria, R.; Moreira, G.; Costa, C.; Cortez, P.; Santos, M. (2022) Bosch's Industry 4.0 Advanced Data Analytics: Historical and Predictive Data Integration for Decision Support.

Jara, M.; Cibertec, I. (2017). *Introducción Machine Learning*. 1–10.

Ji, Y.; Kim, S.; Kim, Y.; Lee, K. (2018). Human-like sign-language learning method using deep learning.

Mesa L.; Rivera M.; Romero J. (2021). Descripción general de la Inferencia Bayesiana y sus aplicaciones en los procesos de gestión.

Mohamed, A.; Berg, W.; Peng, H.; Luo, Y.; Jankowitz, R.; Wu, Sh. (2017). A deep learning method for classifying mammographic breast density categories.

Nicholson, K.; Richardson, R.; van Roden, E.; Quinton, R.; Anzilotti, K.; Richards, J. (2019). Machine learning algorithms for predicting scapular kinematics. *Medical Engineering and Physics*, 65, 39–45. <https://doi.org/10.1016/j.medengphy.2019.01.005>

Osmana, C.; Ghirana, A. (2019). When Industry 4.0 meets Process Mining.

Qaffas A.; Hoque R.; Almazmomi N. (2021). The Internet of Things and Big Data Analytics for Chronic Disease Monitoring in Saudi Arabia

Réda, C.; Kaufmann, E.; Delahaye, D. (2020). Machine Learning applications in drug development. *Computational and Structural Biotechnology Journal*, 18, 241-252. <https://doi.org/10.1016/j.csbj.2019.12.006>.

Roriz, M.; Magalhães, F.; Guedes, Á.; Colcher, S.; Endler, M. (2019). An introduction to data stream processing: a complex event processing approach (DOI: 10.1145/3323503.3345028)

Spark.apache.org: Spark SQL and DataFrames - Spark 1.5.2 Documentation, <https://spark.apache.org/docs/latest/sql-programming-guide.html>, last accessed 2021/07/19.

Tabales, J M Núñez, Carmona, F J Rey, Caridad, J M (2017). Redes neuronales (RN) aplicadas a la valoración de locales comerciales

Tang, Y.; Wang, J.; Nguyen, M.; Altintas, I. (2019). PEnBayes: A multi-layered ensemble approach for learning bayesian network structure from big data. *Sensors (Switzerland)*, 19(20). <https://doi.org/10.3390/s19204400>

Turesson, H.; Ribeiro, S.; Pereira, D.; Papa, J.; de Albuquerque, V. (2016). Machine learning algorithms for automatic classification of marmoset vocalizations. *PLoS ONE*, 11(9), 1–14. <https://doi.org/10.1371/journal.pone.0163041>

van Evert, F.; Fountas, S.; Jakovetic, D.; Crnojevic, V.; Travlos, I.; Kempenaar, C. (2017). Big Data for weed control and crop protection

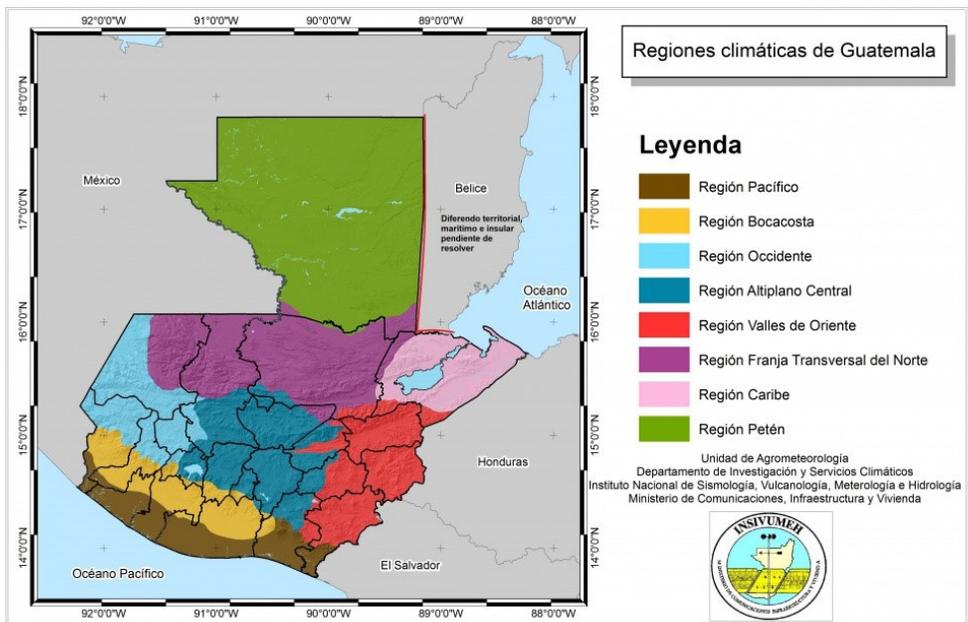
Vu, Ch.; Kim, J. (2018). Human Motion Recognition by Textile Sensors Based on Machine Learning Algorithms

Wang, Sh.; Sun, S.; Xu, J. (2018). Analysis of deep learning methods for blind protein contact prediction in CASP12

Zhang, Q; Wang, L. Xu, Z. (2017). Open source machine-learning algorithms for the prediction of optimal cancer drug therapies.

ANEXOS

Regiones climáticas de Guatemala



Fuente: Unidad de Agrometeorología, INSIVUMEH.

ÍNDICE REMISSIVO

A

- Ações antrópicas 56
- Aços Laminados do Pará (ALPA) 62
- Actividad catalítica 41, 43
- Aeronaves Remotamente Pilotadas (RPA) 33, 34
- Álgebra 1, 2, 3, 28, 29
- Algoritmo 11, 16, 18, 19, 20, 21, 22
- Aprendizaje 12, 16, 22, 25, 26, 27, 28, 29, 31
- Árbol 11, 16, 18, 19, 20, 21, 22
- Área de Preservação Permanente (APP's) 55, 56, 61
- Aritmética 1, 2
- Autorregulación 25, 26, 27, 28, 29

B

- Big Data 11, 12, 14, 16, 17, 23, 24

C

- Coloide 40, 42, 43, 45, 47, 49, 50, 51, 52

D

- Data Science 11

E

- Educación escolar 27, 32
- Electroquímica 40, 42, 43, 44, 45, 46, 47, 49, 52
- Equações 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
- Erodibilidade 59
- Espectroscopia UV-vis-IR 40
- Estudiantes 25, 26, 27, 28, 29, 30, 31, 32

G

- Geoprocessamento 38, 55, 57, 58, 66
- Guatemala 11, 12, 16, 17, 22, 24

H

- Herramientas estadísticas 14

I

Impactos ambientales 56

Industria 4.0 11, 12

Irradiación 40, 44, 45, 46, 49, 50

L

Lluvia 12, 17, 18

Luz solar 40, 42, 43, 44, 45, 50, 51, 52

M

Machine Learning (ML) 11, 12, 22, 23, 24

Microprocesadores 12

Modelo de Zimmerman 25, 29

Modelo pedagógico 1

N

Nanopartículas (NPs) 40, 41

Nubosidad 13

O

Ortomosaico 33, 35, 39

P

Plano diretor 55, 56, 64, 66

Potencial eléctrico 44

Potenciostato 44

Presión atmosférica 17

R

Radiación 13, 22, 42, 49, 68

Resonancia del Plasmon Superficial (SPR) 40

Respuesta electroquímica 40, 43, 44, 45, 52

Revolución industrial 12

S

Software 14, 67

T

Temperatura 12, 17, 18, 19, 20, 22, 42

V

Variables meteorológicas 11, 14, 21

Velocidad del viento 12

Voltamograma 40, 44, 45, 46, 47, 49, 51, 52

Voltamperometría cíclica 40, 43, 52

CIENCIAS EXACTAS

Y DE LA TIERRA:

Observación, formulación y predicción

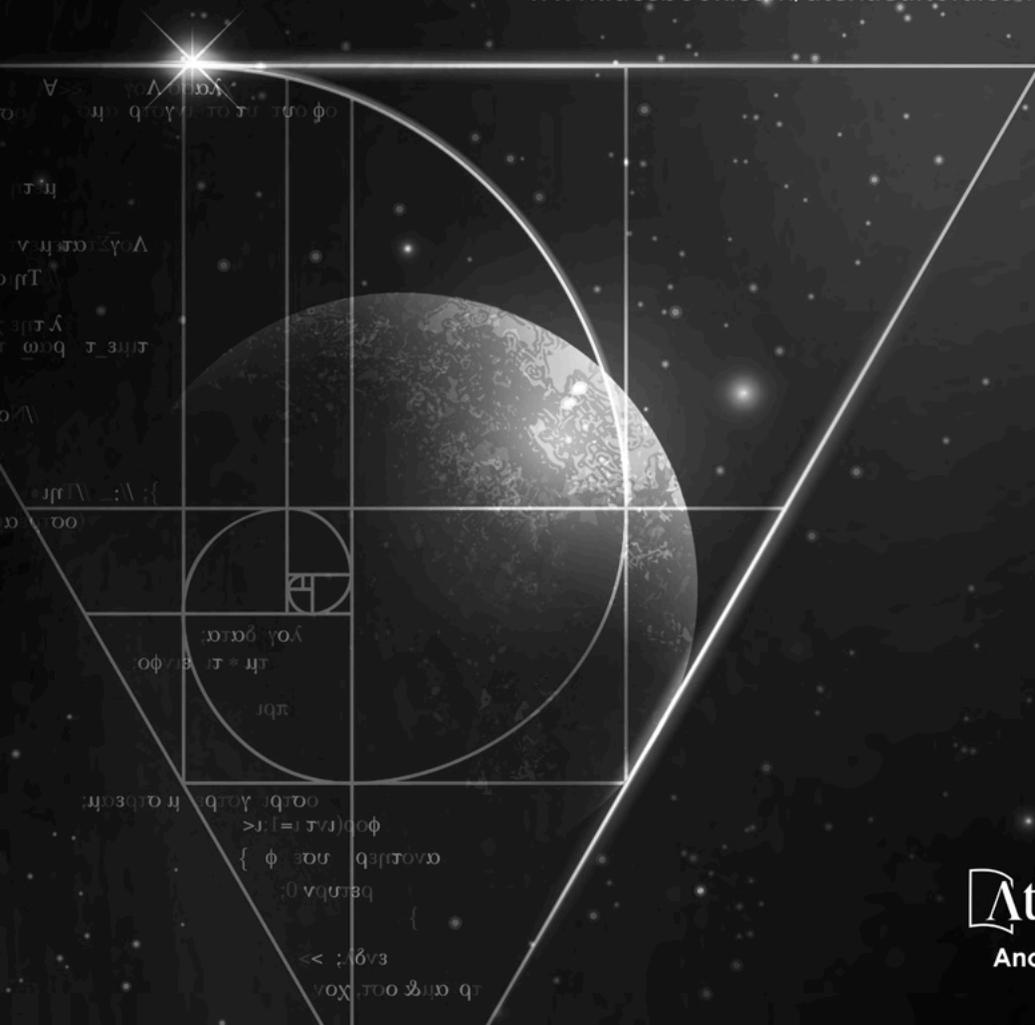
4

www.atenaeditora.com.br 

contato@atenaeditora.com.br 

[@atenaeditora](https://www.instagram.com/atenaeditora) 

www.facebook.com/atenaeditora.com.br 



CIENCIAS EXACTAS

Y DE LA TIERRA:

Observación, formulación y predicción

4

www.atenaeditora.com.br 

contato@atenaeditora.com.br 

[@atenaeditora](https://www.instagram.com/atenaeditora) 

www.facebook.com/atenaeditora.com.br 

