

Instituto Tecnológico y de Estudios Superiores de Occidente

Department of Electronics, Systems, and Informatics

Computing Systems Engineering

Machine Learning Course

Teacher: Edna L. Guevara Rivera

Prediccion del Clima en India

Con ayuda de las herramientas vistas en ML realizaremos predicciones del clima de india, haciendo uso de la base de datos correspondiente al clima registrado en los ultimos años.

Presented by:

Axel Roberto Orozco Hernandez 730820

Luis Eduardo Bruno Selvas 733177

13/Febrero/2024

Tlaquepaque, México

# **Index**

[Index 2](#_Toc158293399)

[Introducción 3](#_Toc158293400)

[Problema para resolver 4](#_Toc158293401)

[Data collection 5](#_Toc158293402)

[Learning type to use 7](#_Toc158293403)

***Limpeza del dataset …………………………………………………………………………………………………….. 8***

[Conclusiones y trabajo pendiente 11](#_Toc158293404)

[References 12](#_Toc158293405)

# **Introducción**

En este trabajo estaremos haciendo uso del aprendizaje máquina, también conocido como machine learning en inglés, el cual es un subconjunto de la inteligencia artificial que permite que un sistema aprenda y mejore de forma autónoma mediante redes neuronales y algoritmos. [1]  
Se espera que el algoritmo a desarrollar se haga en Python, el cual es un lenguaje de programación de alto nivel, orientado a objetos, que se utiliza ampliamente en aplicaciones web, desarrollo de software, ciencia de datos y machine learning. [2] De igual forma usaremos un dataset el cual es un conjunto de datos estructurados y organizados en filas y columnas que se utilizan para analizar patrones, realizar investigaciones y entrenar modelos de inteligencia artificial. [3] en especifico usaremos un dataset sobre el clima de la india para poder realizar el trabajo.

Para el algoritmo se piensa usar el de regresión, los cuales son un subcampo del aprendizaje automático supervisado que se utilizan para modelar la relación entre variables y predecir valores numéricos [4]; junto con nuestro data set que contiene los datos necesarios para poder hacer las predicciones meteorológicas, ya que en general se necesitan datos numéricos como la temperatura, la velocidad del viento, la cantidad de lluvia, la presión del aire, la humedad, la radiación solar, entre otros. [5]

Así mismo esta aplicación que se busca implementar no es algo tan nuevo, ya que el machine learning se utiliza en la predicción del clima y la meteorología. Esta tecnología permite a las máquinas aprender de datos históricos para realizar predicciones y recomendaciones sobre el tiempo. [6]

# **Problema para resolver**

What?:

El Proyecto tiene como objetivo predecir el clima en la india mediante uso de Maching Learning, haciendo uso de la base de datos correspondiente a los datos que usaremos para predecir dichos objetivos.

How?:

Se utilizará Maching Learning y las herramientas vistas en clases para analizar y procesar la base de datos, empleando algoritmos predictivos para generar pronósticos climáticos precisos.

When?:

El Proyecto se llevará a cabo en un marco de tiempo específico, fechas y plazos definidos para la recopilación de datos, entrenamiento del modelo y evaluación de resultados.

Where?:

El enfoque está en la predicción del clima en la India, lo que implica la recopilación de datos específicos y consideraciones geográficas para mejorar la precisión del modelo.

Who?:

El proyecto involucra a nosotros el equipo correspondiente a este proyecto responsables de la implementación del modelo de machín learning, así como la base de datos que fue hecha por expertos los cuales compartieron los datos obtenidos.

# **Data collection**

El dataset del clima de la india lo obtuvimos de la pagina ”kaggle”, el cual cuenta con información meteorológica en tiempo real para las principales ciudades de la India.

Este dataset cuenta con 42 columnas, las cuales estan nombradas como:

* country: País de los datos meteorológicos.
* location\_name: Nombre de la ubicación (ciudad)
* region: Región administrativa de la ubicación
* latitude: Coordenada de latitud de la ubicación.
* longitude: Coordenada de longitud de la ubicación.
* timezone: Zona horaria de la ubicación
* last\_updated\_epoch: Marca de tiempo Unix de la última actualización de datos
* last\_updated: Hora local de la última actualización de datos
* temperature\_celsius: Temperatura en grados Celsius
* temperature\_fahrenheit: Temperatura en grados Fahrenheit
* condition\_text: Descripción de las condiciones climáticas
* wind\_mph: Velocidad del viento en millas por hora
* wind\_kph: Velocidad del viento en kilómetros por hora
* wind\_degree: Dirección del viento en grados
* wind\_direction: Dirección del viento como brújula de 16 puntos.
* pressure\_mb: Presión en milibares
* pressure\_in: Presión en pulgadas
* precip\_mm: Cantidad de precipitación en milímetros
* precip\_in: Cantidad de precipitación en pulgadas
* humidity: Humedad como porcentaje
* cloud: Cobertura de nubes como porcentaje
* feels\_like\_celsius: Temperatura que se siente en grados Celsius
* feels\_like\_fahrenheit: Temperatura que se siente en grados Fahrenheit
* visibility\_km: Visibilidad en kilómetros
* visibility\_miles: Visibilidad en millas
* uv\_index: Índice UV
* gust\_mph: Ráfaga de viento en millas por hora
* gust\_kph: Ráfaga de viento en kilómetros por hora
* air\_quality\_Carbon\_Monoxide: Medición de la calidad del aire: monóxido de carbono
* air\_quality\_Ozone: Medición de la calidad del aire: Ozono
* air\_quality\_Nitrogen\_dioxide: Medición de la calidad del aire: Dióxido de nitrógeno
* air\_quality\_Sulphur\_dioxide: Medición de la calidad del aire: Dióxido de azufre
* air\_quality\_PM2.5: Medición de la calidad del aire: PM2.5
* air\_quality\_PM10: Medición de la calidad del aire: PM10
* air\_quality\_us-epa-index: Medición de la calidad del aire: Índice de la EPA de EE. UU.
* air\_quality\_gb-defra-index: Medición de la calidad del aire: Índice GB DEFRA
* sunrise: Hora local del amanecer
* sunset: Hora local de puesta del sol
* moonrise: Hora local de salida de la luna
* moonset: Hora local de puesta de la luna
* moon\_phase: Fase lunar actual
* moon\_illumination: Porcentaje de iluminación de la luna

Dentro del archivo csv podemos contar con 79,949 registros y ninguna de las celdas esta vacia.

# **Learning type to use**

Tipo de aprendizaje:

Se utilizará aprendizaje supervisado para predecir el clima en la India. En este enfoque, el modelo se estrenará utilizando un conjunto de datos etiquetado que contiene ejemplos históricos de condiciones climáticas y sus correspondientes predicciones.

Algoritmo:

Se empleará algoritmos de regresión, como regresión lineal o regresión logística, para predecir variables climáticas continuas, como la temperatura. Estos algoritmos son comunes en problemas de predicción y son adecuados para el aprendizaje supervisado.

Ventajas:

Precisión en predicciones continuas – Estos algoritmos son eficaces para predecir valores numéricos, como la temperatura, proporcionando resultados cuantitativos.

Facilidad de interpretación – Fácil de entender y visualizar.

Desventajas:

Sensibilidad a datos atípicos – los algoritmos de regresión pueden ser sensibles a datos atípicos, lo que puede afectar la precisión de las predicciones si no se manejan adecuadamente.

Limitaciones en patrones no lineales: En situaciones donde la relación entre las variables climáticas no es lineal, los modelos de regresión pueden no capturar de manera óptima la complejidad de dichos patrones.

**Limpieza del dataset**

En esta parte del proyecto se hace una limpieza del dataset para quitar toda la información que no nos es útil

**Tabla

Descripción generada automáticamente**

Por ejemplo en esta imagen vemos todas la columnas y con esto analizar cuales son las que no nos importan, como podrían ser aquellas que son de tipo object, además de que también nos sirve para ver si hay datos nulos

Una vez limpiado el dataset queda de la siguiente manera

**Tabla

Descripción generada automáticamente**

Estandarización la matriz X

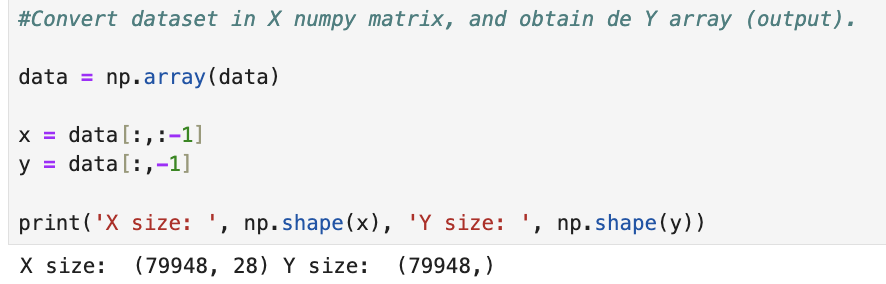
**Texto

Descripción generada automáticamente**

Matriz de correlación**Gráfico

Descripción generada automáticamente**

Obtención de las matrices X y Y



Como podemos observar la matriz X tiene un tamaño de (79948, 28). Esto indica que tenemos esa cantidad de instancias en el conjunto de datos y cada instancia tiene 28 características.

El vector Y tiene un tamaño de (79948). Esto indica los valores correspondientes a la variable objeto, uno para cada instancia en nuestro conjunto de datos.

Por lo tanto, se han separado correctamente los datos en matrices de X e Y, y los tamaños son consistentes con lo que esperaríamos según la estructura de datos.

# **Conclusiones y trabajo pendiente**

En conclusión, Podemos decir que anticipamos obtener predicciones climáticas más precisas y específicas para diferentes regiones de la India mediante la implementación de algoritmos de aprendizaje supervisado y regresión. Esperamos observar mejoras significativas en la precisión de nuestras predicciones en comparación con métodos convencionales. La meta es proporcionar a diversos sectores, como la agricultura y la gestión de recursos, información climática confiable que contribuya a una toma de decisiones más informada y precisa.

Además, buscamos aprender sobre la interacción entre diferentes variables climáticas y como afectan el modelo predictivo. La identificación de relaciones complejas y no lineales entre factores como la temperatura, la humedad y la presión atmosférica será crucial para mejorar la capacidad predictiva del modelo. Este proceso de aprendizaje continuo nos permitirá refinar y ajustar el modelo a medida que se adquiera más comprensión sobre la dinámica climática especifica de la India.

Asimismo, esperamos aprender sobre la influencia de la geografía es las condiciones climáticas, ya que la india presenta una amplia diversidad geográfica. La incorporación de datos geoespaciales y la consideración de la topografía en nuestras predicciones serán factores clave para aumentar la precisión y la aplicabilidad del modelo a lo largo de diversas áreas geográficas.

En cuanto al trabajo pendiente, nos enfocaremos en la expansión del conjunto de datos para incluir información detallada sobre las regiones de India. Esto no solo enriquecerá el modelo, sino que también nos proporcionará una comprensión más profunda. Además, exploraremos la integración de técnicas de machine Learning, para abordar patrones climáticos complejos que podrían no captar completamente modelos más lineales.

En resumen, aspiramos a proporcionar resultados, predictivos mejorados y, a su vez, generar aprendizajes valiosos sobre la interacción de variables climáticas en el contexto especifico de la India.

# **References**

[1] “¿Qué es el aprendizaje automático? | Glosario,” Hpe.com, 2022. https://www.hpe.com/lamerica/es/what-is/machine-learning.html

‌[2] “¿Qué es Python? | Guía de Python para principiantes de la nube | AWS,” Amazon Web Cervices, Inc. https://aws.amazon.com/es/what-is/python/

‌[3] “¿Qué Son Los Datasets Y Dónde Conseguirlos?,” 2020. https://keepcoding.io/blog/que-son-datasets/

‌[4]“Machine Learning: Algoritmos de clasificación y regresión,” The Black Box Lab, May 06, 2022. https://theblackboxlab.com/2022/05/06/machine-learning-diferencias-entre-algoritmos-clasificacion-regresion/

[5]“Base de datos Meteorológica. Red principal Estaciones meteorológicas de AEMET,” datosclima.es. https://datosclima.es

‌[6]“Machine Learning y predicción meteorológica | Basetis now!,” blog.basetis.com. https://blog.basetis.com/es/content/machine-learning-y-prediccion-meteorologica (accessed Feb. 12, 2024).

‌

‌