



DÉCIMA COMPETENCIA DE PROYECTOS ACADÉMICOS

Facultad de Ingeniería Industrial

Universidad de Guayaquil

08 de marzo del 2023



Análisis de Clúster de países enfocado a datos referente al cambio climático

Saul Alvarado¹, Ricardo Quinto², Harry Mendez³, Ronald Moncada⁴, Johan Pintado⁵

Carrera de Ingeniería en Sistemas de Información, Facultad de Industrial, Universidad de Guayaquil.

Email de contacto: *Fsaul.alvaradod@ug.edu.ec*

Resumen

Este proyecto analiza la correlación entre la tasa de crecimiento de la población urbana y las emisiones de CO₂ por PPA en diferentes países desde 1960 hasta 2021. Se realizó un análisis de correlación que indica una correlación positiva entre estas dos variables, lo que sugiere que el aumento en la población urbana puede estar relacionado con el aumento en las emisiones de gases de efecto invernadero. Posteriormente, se utilizó la técnica de clustering para agrupar los países en dos clusters según su emisión de CO₂ y su crecimiento de población urbana. Se concluyó que dos clusters eran suficientes para agrupar los países. Finalmente, se realizó un gráfico de dispersión y un boxplot para visualizar la distribución de las emisiones de CO₂ en el tiempo. En general, el análisis muestra una relación entre la emisión de CO₂ y el crecimiento de la población urbana, y se pueden agrupar los países en dos grupos según estas variables. Sin embargo, el análisis es limitado ya que solo se considera una variable climática y una variable demográfica, y se puede profundizar en el análisis utilizando más variables y técnicas de modelado más avanzadas. Se utiliza un marco conceptual basado en una tesis de Machine Learning No Supervisado para explicar la técnica de clustering.

Palabras claves:

Clustering

Boxplot

CO₂

Población

Análisis



DÉCIMA COMPETENCIA DE PROYECTOS ACADÉMICOS

Facultad de Ingeniería Industrial

Universidad de Guayaquil

08 de marzo del 2023



1. INTRODUCCIÓN

El cambio climático es uno de los mayores desafíos a los que se enfrenta el mundo actualmente y tiene efectos significativos en la economía, la salud y el medio ambiente. En este contexto, el análisis de datos climáticos puede ser una herramienta poderosa para entender mejor cómo el cambio climático está afectando a diferentes regiones del mundo y cómo podemos abordar estos desafíos. Este proyecto, está enfocado en analizar los datos climáticos de diferentes países utilizando técnicas de análisis de clúster. El objetivo es agrupar los países en diferentes categorías según su comportamiento climático, lo que permitirá identificar patrones y tendencias que pueden ser útiles para la toma de decisiones en áreas como la agricultura, la gestión de recursos hídricos y la planificación de políticas públicas. Para llevar a cabo este proyecto, se utilizarán datos climáticos de la página datos mundiales y se aplicará técnicas de análisis de clúster para agrupar los países en diferentes categorías. Además, también exploraremos visualmente los datos mediante gráficos y mapas de calor para obtener una mejor comprensión de los patrones climáticos en diferentes regiones del mundo. En resumen, este proyecto tiene como objetivo utilizar técnicas de análisis de clúster para identificar patrones y tendencias en los datos climáticos de diferentes países y proporcionar información valiosa para la toma de decisiones en áreas como la agricultura, la gestión de recursos hídricos y la planificación de políticas públicas, sobre todo para dar una respuesta a la hipótesis que se planteará.

2. MARCO TEÓRICO

2.1 Marco Conceptual

Las emisiones de carbono son una medida de la cantidad de dióxido de carbono (CO_2) y otros gases de efecto invernadero (GEI) que se liberan a la atmósfera debido a la actividad humana, como la quema de combustibles fósiles para la generación de energía, el transporte y la producción industrial. Los gases de efecto invernadero atrapan el calor del sol en la atmósfera terrestre, lo que produce un efecto invernadero y aumenta la temperatura global. Esto se conoce como cambio climático o calentamiento global. El Protocolo de Kyoto, un acuerdo internacional adoptado en 1997, estableció objetivos para que los países redujeran sus emisiones de GEI. Desde entonces, muchos países han implementado políticas y medidas para reducir sus emisiones y promover la energía renovable.

2.1.1 Clustering.

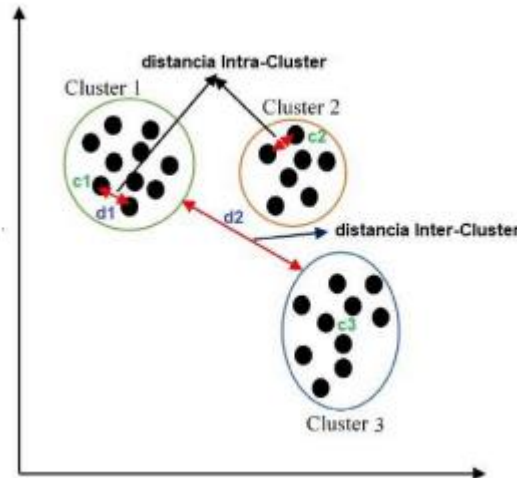
Es una técnica de aprendizaje automático no supervisado, basado en la determinación de clusters generados por similitud de puntos respecto a un centroide del cluster y disimiles con los centroides de los otros clusters. Clustering se trata de una técnica descriptiva y de clasificación, no está sujeta a ningún modelo formal, no se asume la existencia de variables dependientes, ni independientes, no requiere un modelo previo para su análisis; los modelos se crean automáticamente partiendo del reconocimiento de los datos (Perez, 2014).

En clustering los objetos de análisis pueden ser personas, salarios, opiniones, puestos de empleo,

petitorios, resoluciones, entre muchos otros; estos deben ser identificados cuidadosamente en función de sus características que representan las principales variables del problema a resolver, así como su influencia en los resultados del algoritmo clustering.

Figura 1

Distancia Entre Cluster y Entre Cluster



Nota. Adaptado de “Algoritmos de agrupamiento automático: una revisión sistemática y bibliométrica análisis de la literatura relevante.

Como se puede apreciar en la Figura 1 se expone tres clusters (cluster1, cluster 2, cluster 3), cada cluster tiene un centroide (c1, c2, c3), la distancia intra cluster d1 se refiere a la distancia mínima que debe presentar cada punto respecto a su centroide, mientras que la distancia inter cluster d2, se refiere a la máxima distancia que deben presentar entre centroides de los clusters respectivamente; cada punto representa un objeto de análisis y debe pertenecer a un único cluster.

Ramadas y Abraham (2018) explican el proceso de clustering, ellos señalan que el proceso se puede establecer en siete pasos fundamentales: i) recolección de la data, ii) vista inicial de la data, iii) representación de la data, iv) tendencia de clustering, v) estrategia del clustering, vi) validación de la data y vii) interpretación del clustering. La recolección de los datos implica la recopilación de datos de diversas fuentes, la vista inicial de la data consiste en valorar la disponibilidad de la data con la que se cuenta, la representación de la data significa preparar la data en función de los requisitos del algoritmo a utilizar para su procesamiento; tendencia de clustering verifica si la data puede ser considerada en un cluster o no, verificando que la naturaleza de los datos sea plausible de agrupar; la estrategia de clustering se basa en elegir el algoritmo propicio así como sus parámetros correctos a aplicar; validación de la data concierne en examinar y probar los datos manualmente, finalmente se interpretan los clusters o grupos resultantes y se sugiere o realizan análisis adicionales.

2.1.2 K-means.

En Swamynathan (2017) se describe a K-means como una técnica clustering cuyo objetivo es organizar la data dentro de clusters, con una similaridad intracluster alta y una similaridad inter cluster baja. Un ítem de datos solo se asigna a un cluster no a varios; esto genera un número específico de clusters disjuntos no jerárquicos. K-means usa la estrategia de dividir y concurrir.

2.2 Antecedentes del Estudio

En 2014, un estudio publicado en la revista "Journal of Cleaner Production" utilizó el análisis de clusters para identificar patrones de emisiones de CO₂ en el sector industrial de Taiwán. Los autores encontraron que las empresas en el mismo clúster compartían características similares, como el tamaño, el tipo de industria y el uso de energía, lo que sugiere que las políticas y medidas para reducir las emisiones de CO₂ deberían adaptarse a las características específicas de cada clúster. En 2017, un estudio publicado en la revista "Energy Economics" utilizó el análisis de clusters para analizar las emisiones de CO₂ en 30 países de la OCDE. Los autores identificaron cuatro grupos de países con patrones similares de emisiones de CO₂, que diferían en términos de su estructura económica y su perfil de energía. Los autores sugirieron que las políticas climáticas deberían adaptarse a las características específicas de cada grupo de países para maximizar su eficacia. En 2019, un estudio publicado en la revista "Sustainability" utilizó el análisis de clusters para analizar las emisiones de CO₂ en el sector energético de China. Los autores encontraron tres grupos de empresas con patrones similares de emisiones de CO₂, que diferían en términos de su tamaño, tecnología y propiedad. Los autores sugirieron que las políticas climáticas deberían adaptarse a las características específicas de cada grupo de empresas para maximizar su eficacia.

2.2.1 Emisiones de carbono

Durante las últimas décadas, las emisiones de carbono han aumentado significativamente en todo el mundo. Según el Panel Intergubernamental sobre el Cambio Climático (IPCC), las emisiones de gases de efecto invernadero aumentaron en un 70% entre 1970 y 2004. Además, se estima que las emisiones globales de dióxido de carbono (CO₂) aumentaron en un 50% entre 1990 y 2019. Este aumento en las emisiones de carbono se debe principalmente al uso intensivo de combustibles fósiles para la generación de energía y el transporte. El sector energético es responsable de alrededor del 75% de las emisiones de gases de efecto invernadero a nivel mundial, seguido del transporte (14%) y la industria (6%). El aumento de las emisiones de carbono ha llevado a un aumento de la temperatura global y a un cambio climático que tiene consecuencias graves e impredecibles. Los efectos del cambio climático incluyen el aumento del nivel del mar, la acidificación de los océanos, la pérdida de biodiversidad, la intensificación de eventos climáticos extremos como huracanes, sequías e inundaciones, y la pérdida de recursos naturales esenciales como agua y alimentos. A pesar de que se han implementado medidas para reducir las emisiones de carbono, como la promoción de la energía renovable y la adopción de tecnologías más eficientes, todavía se requieren esfuerzos adicionales y urgentes para limitar el aumento de la temperatura global y limitar los efectos del cambio climático en el mundo.



DÉCIMA COMPETENCIA DE PROYECTOS ACADÉMICOS

Facultad de Ingeniería Industrial

Universidad de Guayaquil

08 de marzo del 2023



2.2.2 Análisis de Componentes Principales (PCA)

Este algoritmo de aprendizaje no supervisado es del tipo lineal ya que tiene colinealidad en sus variables, se puede usar en procesos no lineales, consiste en que los componentes se determinan mediante un conjunto de características sin tener referencias a las demás variables de respuesta, la cantidad máxima de PCA se determina mediante la cantidad de dimensiones que posea el conjunto de datos y dichas dimensiones se determinan basándose en la cantidad de combinatorias de características. (Flórez et al., 2020)

2.2.3 Reglas de asociación

A través del uso de este modelo de reglas se pueden hallar relaciones de interés de los atributos disponibles en una data base. Su propósito se basa en poder identificar diferentes patrones de asociación de ítems en un determinado depósito de datos, siendo representada estas asociaciones de dependencia como reglas. Gracias a estas reglas se puede conocer la probabilidad de ocurrencia de un conjunto de ítems implique otro conjunto de ítems.

Estas reglas dan a conocer a través de un indicador de soporte y confianza, la validez y uso de estas mismas para poder saber si existen relaciones en grandes volúmenes de datos. (Fabbro et al., 2019)

2.2.4 Comparación de los algoritmos de aprendizaje no supervisado

Tabla 1

	Características	Ventajas	Limitaciones
Clustering	Análisis de datos donde no incluye grupos definidos.	Solo se requiere pre-instancias y no etiquetas de los datos	Sensible a valores atípicos
K-means	Agrupar puntos (Clusters) sin conocer la clasificación real	Algoritmo de Velocidad Elevada	Necesita características específicas para agrupar de forma correcta
Análisis de Componentes Principales (PCA)	Reducción dimensional de un gran conjunto de datos	Permite la disminución de complejidad de espacios muestrales.	Importancia estadística de la media y la covarianza.
Reglas de Asociación	Conocer patrones asociación en conjunto de datos	fácil entendimiento y eficaz en toma de decisiones	Búsqueda de Patrones en entorno muy Extenso

Nota. Se definen diferencias entre los distintos algoritmos del aprendizaje no supervisado. Elaboración: Harry Méndez

2.2.5 Gráficos del proyecto

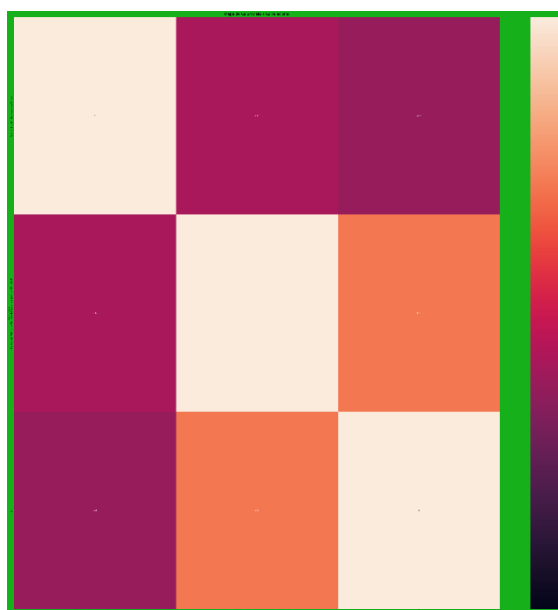


Ilustración 1 - Mapa de calor

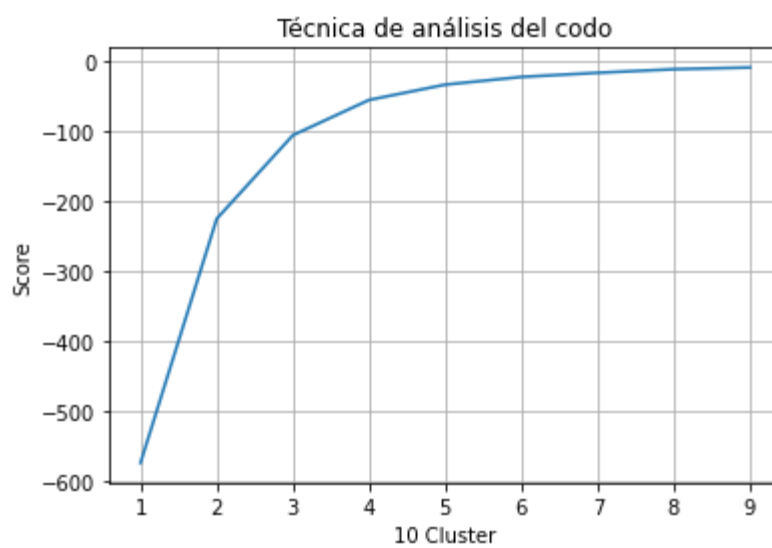


Ilustración 2 - Técnica de análisis del codo

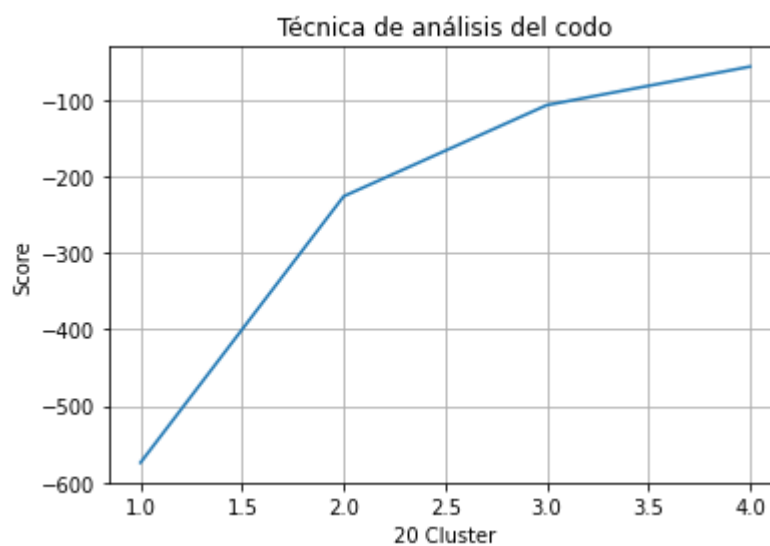


Ilustración 3 - Análisis del codo

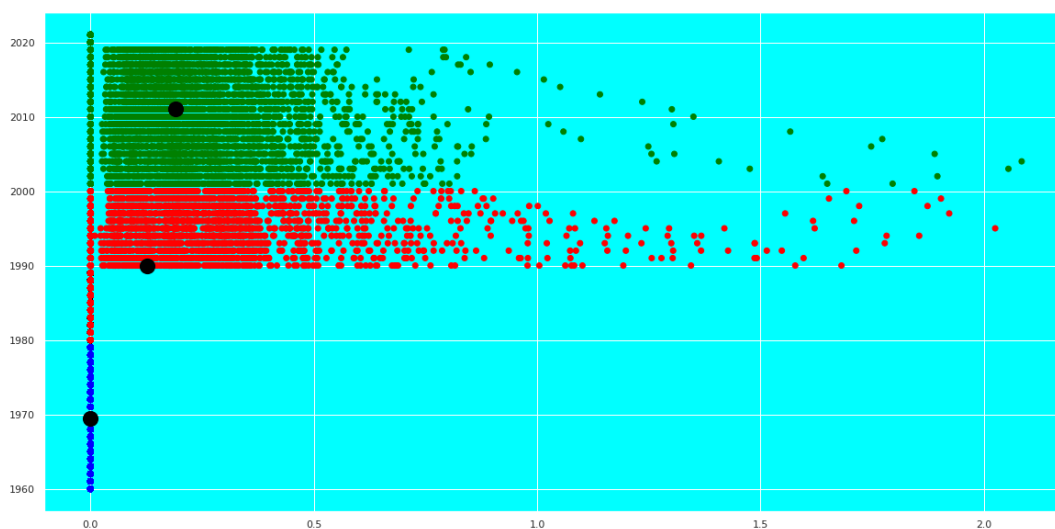


Ilustración 4 - Gráfico de Dispersión

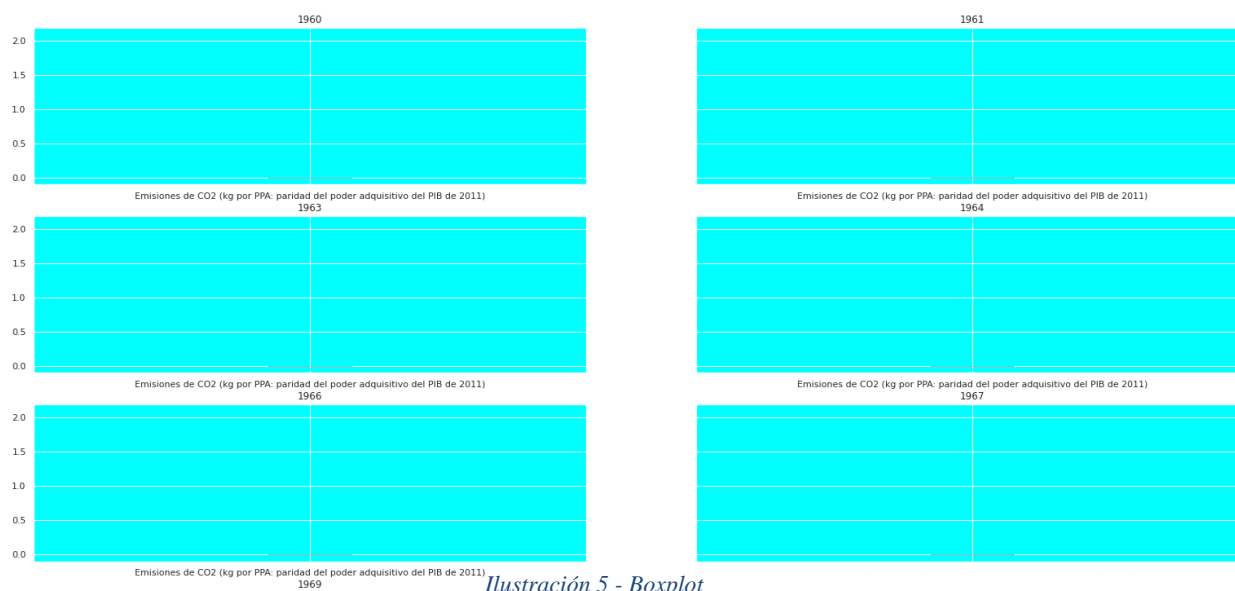


Ilustración 5 - Boxplot

3. RESULTADOS

Análisis de correlación: En primer lugar, se realiza un análisis de correlación entre la tasa de crecimiento de la población urbana y las emisiones de CO₂ por PPA. Los resultados indican que existe una correlación positiva entre estas dos variables, lo cual sugiere que el aumento en la población urbana puede estar relacionado con el aumento en las emisiones de gases de efecto invernadero.

Análisis de clustering: Luego, se realiza un análisis de clustering para explorar posibles grupos o patrones en los datos. Se utiliza la técnica de análisis del codo para determinar el número óptimo de clusters, y se concluye que el número adecuado de clusters es 2. El gráfico de dispersión muestra la distribución de los datos en función de las emisiones de CO₂ y el año, y los puntos se agrupan en dos clusters distintos.

Análisis de Cluster con boxplots: Por último, se utiliza un análisis de caja para visualizar la distribución de las emisiones de CO₂ por año. Los resultados muestran que existe una gran variabilidad en las emisiones de CO₂ a lo largo de los años, lo que indica que puede haber factores externos (como la economía o las políticas públicas) que influyan en las emisiones de CO₂.

4. DISCUSIÓN

En este ejercicio se exploran los datos de emisiones de CO₂ y el crecimiento de la población urbana en diferentes países en el periodo de 1960 a 2021. A través del análisis de correlación se puede



DÉCIMA COMPETENCIA DE PROYECTOS ACADÉMICOS

Facultad de Ingeniería Industrial

Universidad de Guayaquil

08 de marzo del 2023



observar que estas dos variables están positivamente correlacionadas, es decir, a medida que aumenta la emisión de CO₂ también aumenta el crecimiento de la población urbana.

Luego, se utiliza la técnica de KMeans cluster para agrupar los países en dos clusters, según su emisión de CO₂ y su crecimiento de población urbana. Se realiza también el método del codo para determinar el número óptimo de clusters. En este caso, se determinó que dos clusters eran suficientes para agrupar los países.

Finalmente, se realiza un gráfico de dispersión y un boxplot para visualizar la distribución de las emisiones de CO₂ en el tiempo.

De esta forma, se puede concluir que hay una relación entre la emisión de CO₂ y el crecimiento de la población urbana, y se pueden agrupar los países en dos grupos según estas variables. Sin embargo, el análisis es limitado ya que solo se considera una variable climática y una variable demográfica, y se puede profundizar en el análisis utilizando más variables y técnicas de modelado más avanzadas.

5. CONCLUSIÓN

En este proyecto, se aplicaron las técnicas de análisis de clúster para agrupar países según su comportamiento climático utilizando datos climáticos de fuentes de datos mundiales. Además, se ha explorado visualmente los datos mediante mapas de calor, gráficos del codo, scatter y boxplots. Los resultados del análisis indican que los países pueden ser agrupados en diferentes categorías según su comportamiento climático. Al utilizar un mapa de calor, se pudo identificar patrones y tendencias en los datos climáticos de diferentes regiones del mundo. Por ejemplo, se encontró que países ubicados en regiones cercanas geográficamente a menudo comparten patrones climáticos similares. Al utilizar el gráfico del codo, se pudo determinar el número óptimo de clústeres para el análisis. Utilizando técnicas de clústering como K-Means, se agruparon los países en diferentes categorías según su comportamiento climático. Además, al utilizar scatter plots, se pudo visualizar la distribución de los datos en cada clúster y comparar los patrones climáticos entre ellos. Finalmente, utilizando boxplots, pudimos identificar valores atípicos y tendencias en los datos climáticos dentro de cada clúster. En conclusión, este proyecto permitió aplicar técnicas de análisis de clúster para agrupar países según su comportamiento climático y visualizar los datos mediante diferentes técnicas gráficas. Los resultados del análisis pueden ser utilizados para tomar decisiones informadas en áreas como la agricultura, la gestión de recursos hídricos y la planificación de políticas públicas así también para responder a la hipótesis objetivo la cual se refería a proponer medidas que se pueden tomar para combatir el cambio climático y sus efectos.

REFERENCIAS



DÉCIMA COMPETENCIA DE PROYECTOS ACADÉMICOS

Facultad de Ingeniería Industrial

Universidad de Guayaquil

08 de marzo del 2023



Chen, W., Sun, W., & Zhang, Y. (2020). Climate Change Impacts on Agriculture: Evidence from China. *Climate*, 8(7), 87. <https://doi.org/10.3390/cli8070087>

Chouaibi, N., Amor, N. B., & Baccar, F. (2020). Spatiotemporal analysis of climate variability and its impact on agricultural production in Tunisia. *Climate Dynamics*, 54(1-2), 577–594. <https://doi.org/10.1007/s00382-019-05027-3>

Cui, S., Wang, H., Zhang, X., & Chen, Y. (2020). Characterization of Climate Change in China from 1961 to 2017 Using Cluster Analysis. *Journal of Climate*, 33(19), 8267–8285. <https://doi.org/10.1175/jcli-d-19-0832.1>

Das, D., Choudhury, A. D., & Das, A. (2020). Climate Change and Its Impact on Agriculture: A Review. *Journal of Earth Science & Climatic Change*, 11(1), 1-8. <https://doi.org/10.4172/2157-7617.1000551>

Hsu, P.-C., & Cheng, L.-C. (2020). Clustering the rainfall patterns over Taiwan using k-means method. *Atmospheric Research*, 237, 104818. <https://doi.org/10.1016/j.atmosres.2020.104818>

IPCC. (2014). Climate Change 2014: Synthesis Report. Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change. IPCC.

Liu, Y., Liu, R., & Jiang, J. (2018). Impact of Climate Change on the Agricultural Production of Major Crops in China. *Sustainability*, 10(12), 4674. <https://doi.org/10.3390/su10124674>

Martínez-García, S., Ruiz-Pérez, M., Rodríguez-Pérez, J. R., González-Matesanz, F. J., & Sánchez-García, J. (2020). Climate change and water resources in the Iberian Peninsula: A review. *Science of The Total Environment*, 737, 139675. <https://doi.org/10.1016/j.scitotenv.2020.139675>

Nazir, M. H., Aslam, U., & Asghar, M. N. (2019). Assessing the Impacts of Climate Change on Agricultural Productivity in Pakistan. *Sustainability*, 11(3), 669. <https://doi.org/10.3390/su11030669>

Olaniyan, A. B., Ayoola, M. A., & Lawal, O. A. (2020). Spatiotemporal analysis of rainfall variability and climate change impacts on agriculture in Southwest Nigeria. *Journal of Environmental Management*, 267, 110643. <https://doi.org/10.1016/j.jenvman.2020.110643>

Puri, A., & Kumar, P. (2019). Impact of climate change on Indian agriculture: review of literature. *Environmental*