

Planeación del Proyecto

Francisco Castorena Salazar, Juan Pablo Echeagaray González, Emily Rebeca Méndez
Cruz, José Eugenio Morales Ortiz, and Mario Javier Soriano Aguilera

Ing. en Ciencias de Datos y Matemáticas, Tec de Monterrey

3 de junio de 2023

1. Trásfondo del negocio

ClimateAi ha sido capaz de desarrollar algunas de las predicciones de clima más precisas del mundo por medio del uso de Inteligencia Artificial Avanzada y modelos climáticos. Esto con el objetivo de ayudar a distintos grupos, individuos y entidades a crear resiliencia climática.

Con el propósito de apoyar en su misión, el equipo buscará implementar técnicas de TDA (*Topological Data Analysis*) para la realización de un análisis exploratorio de datos del fenómeno El Niño.

2. Alcance

3. Personal

El presente proyecto ha sido desarrollado por parte del Tec de Monterrey como un proyecto para ClimateAi:

■ Tec de Monterrey

- Científico de Datos, Líder de Proyecto: Juan Pablo Echeagaray González
- Científico de Datos: Francisco Castorena Salazar
- Investigador: José Eugenio Morales Ortiz
- Investigador: José Eugenio Morales Ortiz
- Científico de Datos: Mario Javier Soriano Aguilera

■ ClimateAi

- Científico de Datos, CTO: Max Evans
- Científico de Datos: Alitzel Macías

Objetivo	Meta	Justificación
Búsqueda de periodicidad en los datos mediante técnicas de TDA	Sí/No	La información sobre la existencia de la periodicidad en los datos nos permite el poder hacer uso de herramientas de ML para la predicción de cuestiones climatológicas
Reconocimiento de patrones en los datos proporcionados	Sí/No	El análisis de patrones podría revelar anomalías en las mediciones que de otra manera pasarían desapercibidas
Predicción de anomalías Nino3.4 mediante un modelo de ML informado por features de TDA	Coefficiente de determinación superior a 80 %, MAPE menor al 20 %	Se espera que el modelo a implementar pueda predecir fielmente las tendencias de alza o decremento de anomalías, sirviendo como un indicador general del fenómeno
Preparación de la documentación oficial sobre la investigación de los resultados obtenidos	Reporte técnico y presentación ejecutiva	Documentación necesaria para replicar los resultados obtenidos

4. Métricas

El proyecto buscará definir si existe periodicidad en las mediciones de temperaturas de El Niño así como en sus anomalías. Dicha periodicidad puede ser cuantificada y representada mediante técnicas de homología persistente.

Se hará uso del método de clusterización *Mapper* para detectar patrones inherentes en las mediciones de anomalías de El Niño 3.4; dicho análisis es de naturaleza cualitativa, por lo que los patrones encontrados estarán sujetos a una inspección más detallada.

Se implementará un modelo de ML para la predicción de anomalías de El Niño 3.4. Dicho modelo será informado por características obtenidas de TDA, por lo que se espera que el modelo sea capaz de predecir las anomalías con un coeficiente de determinación superior a 80 % y un MAPE menor al 20 %.

5. Planeación

- Limpieza de bases de datos
 - Procesamiento de archivo `pickle` de anomalías de El Niño 3.4 proporcionado por el equipo de ClimateAi
 - Descarga y procesamiento de mediciones de El Niño para distintas regiones del Pacífico del sitio

oficial de la NOAA

- Análisis de Periodicidad: Uso de homología persistente para estimar la presencia de elipticidad en las mediciones de todas las variables en los datos de la NOAA
- Análisis de Patrones: Uso de *Mapper* para la detección de patrones en las mediciones de anomalías de El Niño 3.4
- Predicción de Anomalías: Implementación de un modelo de ML para la predicción de anomalías de El Niño 3.4

6. Arquitectura

- Datos:
 - Archivo `nino34.long.anom.data.txt` obtenido de aquí
 - Datos de NOAA obtenido de aquí
- Herramientas:
 - Google Collaboratory: Entorno de desarrollo en el que se programará la solución
 - GitHub: Sitio donde se alojará un repositorio del código fuente implementado y la documentación técnica
- Entregables:
 - Código fuente del proyecto en formato `.ipynb`
 - Documentación técnica del proyecto en formato pdf
 - Presentación ejecutiva de resultados

7. Comunicación

Los avances del proyecto serán presentados de forma semanal al profesor supervisor designado cada viernes a partir del **19 de mayo del 2023**. El proyecto final será presentado al socio formador el **16 de junio del 2023** durante una sesión de Zoom. En caso de requerir información más detallada del proyecto, contactar a las siguientes personas:

- **Tec de Monterrey**
 - Líder del proyecto: Juan Pablo Echeagaray González, a00830646@tec.mx
 - Profesora supervisora: Lilia Alanís López, lilia.alanislpz@tec.mx
- **ClimateAi**
 - Alitzel Macías, alitzel@climate.ai