



Tecnológico de Monterrey

Inteligencia Artificial Avanzada para la

Ciencia de Datos II

Portafolio de análisis

Adrián Galván Díaz
A01704076

SMA0100 - Modelación matemática: Construye modelos matemáticos deterministas o estocásticos, soportados por herramientas computacionales de vanguardia.

SMA0104C - Análisis de información: Analiza la información proveniente de modelos, considerando la toma de decisiones como objetivo final.

Módulo 1 Big Data

- Trabaja con grandes volúmenes de datos
- Procesa grandes volúmenes de datos de manera eficiente
- Genera tableros útiles y correctos que apoyen a la toma de decisiones

Estas competencias están presentes en

https://github.com/AdrianGalvanDiaz/DeepGazeIIE_ACW/blob/main/documents/Big%20Data/Evidencia_Módulo_1_A01704076.ipynb

Módulo 5 Estadística avanzada para la ciencia de datos

- Realiza los pasos adecuados para calibrar un modelo estadístico
- Mide correctamente el desempeño del modelo y sus métricas son correctas
- Interpreta los resultados de las predicciones de los modelos y los interpreta en el contexto del problema de manera correcta

Estas competencias están presentes en el jupyter notebook en las siguientes secciones:

1. PARTE 2 DEL PROYECTO
2. 2 Part 2 - Forecasting the IGAE index for Mexico

[https://github.com/AdrianGalvanDiaz/DeepGazeIIE_ACW/blob/main/documents/Estad%C3%A9stica%20avanzada%20para%20la%20ciencia%20de%20datos/Proyecto_estad%C3%A9stica%20\(2\).ipynb](https://github.com/AdrianGalvanDiaz/DeepGazeIIE_ACW/blob/main/documents/Estad%C3%A9stica%20avanzada%20para%20la%20ciencia%20de%20datos/Proyecto_estad%C3%A9stica%20(2).ipynb)

SEG0400 - Compromiso ético y ciudadano: Implementa proyectos orientados a la transformación del entorno y el bienestar común, con conciencia ética y responsabilidad social.

SEG0403B - Resuelve situaciones de la vida académica, profesional y social, mediante el cumplimiento de leyes, normas y principios éticos. En este nivel, pone mayor énfasis en las situaciones del campo académico y profesional. Realiza lo anterior, a partir de la comprensión de los valores de la integridad y el uso de herramientas y referentes normativos para actuar con respeto y honestidad.

Módulo Reto

- Verifica que los datos estén anonimizados para no violar normas o leyes de privacidad de la información
- Especifica el proceso o estándar a seguir para validar el manejo de los datos y garantizar que solo el equipo tenga acceso a ellos en cumplimiento con la normativas.
- Presenta registros claros sobre el seguimiento del proceso y las autorizaciones para tener acceso a los datos

Estas competencias están presentes en:

[https://github.com/AdrianGalvanDiaz/DeepGazeIIE_ACW/blob/main/documents/Seguridad%20de%20Datos/Reto%20Privacidad%20y%20Seguridad%20de%20los%20Datos%20\(2\).pdf](https://github.com/AdrianGalvanDiaz/DeepGazeIIE_ACW/blob/main/documents/Seguridad%20de%20Datos/Reto%20Privacidad%20y%20Seguridad%20de%20los%20Datos%20(2).pdf)

Correcciones de módulo 1

La evidencia evaluaba tres competencias principales:

1. Trabajar con grandes volúmenes de datos.
2. Procesar grandes volúmenes de datos de manera eficiente.
3. Generar tableros útiles y correctos que apoyen la toma de decisiones.

La tercera competencia no se cumplió inicialmente debido a que las visualizaciones creadas se incluyeron únicamente en los resultados finales del modelo, sin incluir gráficas ni análisis visuales en las etapas previas y durante la creación del modelo. Esto limitó la capacidad de interpretar los datos y tomar decisiones informadas durante el proceso de desarrollo.

Para corregir esta deficiencia, se realizaron las siguientes acciones:

- Se generaron gráficas de las distribuciones de los datos y visualizaciones de las imágenes en el contexto a analizar desde las etapas iniciales del proyecto. Esto permitió identificar patrones y características clave de los datos antes de entrenar el modelo (EDA).

- Durante el proceso de creación y entrenamiento del modelo, se incluyeron visualizaciones que ayudaron a evaluar los resultados. Estas gráficas facilitaron la interpretación de métricas y el rendimiento del modelo.

Con estas correcciones, el proyecto ahora cumple con la tercera competencia, ya que las visualizaciones no solo se generan en la etapa final, sino también durante todo el ciclo de desarrollo.

Correcciones de módulo Reto

La evidencia evaluaba tres competencias principales relacionadas con el manejo seguro y ético de los datos:

1. Verificar que los datos estén anonimizados para cumplir con normas y leyes de privacidad.
2. Especificar el proceso o estándar para validar el manejo de datos y garantizar que solo el equipo autorizado tenga acceso.
3. Presentar registros claros que documenten el seguimiento del proceso y las autorizaciones para el acceso a los datos.

Inicialmente, la tercera competencia no se cumplía, ya que no se incluían registros claros, como logs o bitácoras de trabajo, que evidenciaran las acciones realizadas para mantener la seguridad de los datos.

Para corregir esta deficiencia, se realizaron las siguientes acciones:

- Se desarrolló una bitácora de seguridad detallada en la cual se documentan las acciones realizadas por el equipo para garantizar la seguridad de los datos, autorizaciones de acceso, fecha y responsable del último cambio y links de acceso.
- También se desarrolló una bitácora de logs para el google drive y el github donde se establecen los cambios más importantes dentro del github o del google drive con hora y fecha y el responsable del cambio. Así mismo se incluyen los links de los cambios en caso de querer ver más información de los logs.

Con esta corrección, el proyecto ahora cumple con la tercera competencia, ya que se han establecido registros claros y estructurados que respaldan el seguimiento del proceso y las medidas de seguridad adoptadas.