

# portfolio\_analysis

December 4, 2022

## 1 Portafolio analisis

### 1.1 SMA0104C Análisis de información

#### 1.1.1 Trabaja con grandes volúmenes de datos

- **¿Porque tengo esta competencia?** Se obtuvo esta subcompetencia ya que para el proyecto se tomaron grandes conjuntos de datos de 42,059 observaciones y una colección de más de 40,000 imágenes que fueron analizadas, procesadas y utilizadas para entrenar y probar la precisión de los modelos de aprendizaje y con base a los resultados obtenidos, se modificaron estos datos y se volvieron a entrenar los modelos.
- **Evidencia de trabajo para competencia**
  - [Estudio de modelo de segmentacion](#)
  - [Estudio del conjunto de datos de rio de la universidad de nebraska](#)

#### 1.1.2 Procesa grandes volúmenes de datos de manera eficiente

- **¿Porque tengo esta competencia?** Se obtuvo esta subcompetencia ya que como parte de la etapa de procesamiento de datos se modificaron la cantidad de observaciones y variables basado en métodos estadísticos para identificar las variables y observaciones más relevantes para obtener el modelo más óptimo. Asimismo, las imágenes fueron escaladas a la mitad de su resolución para disminuir la cantidad de espacio requerido para almacenarlas y reducir drásticamente el tiempo de entrenamiento.
- **Evidencia de trabajo para competencia**
  - [Script para reducir el tamaño de imagenes del conjunto de datos de nebraska para reducir espacio en memoria](#)

#### 1.1.3 Genera tableros útiles y correctos que apoyen a la toma de decisiones

- **¿Porque tengo esta competencia?** Se obtuvo esta subcompetencia ya que en el mismo análisis de la base de datos, se crearon tablas y se comparó con las imágenes para observar si existía alguna discrepancia con la imagen y los datos de esa fotografía, para así poder tomar la decisión de eliminar o conservar estos datos, además de haber hecho análisis de series de tiempo y con base a la información obtenida de las tablas generadas, observar el comportamiento de los datos.
- **Evidencia de trabajo para competencia**
  - [Estudio de modelo de segmentacion](#)
  - [Estudio del conjunto de datos de rio de la universidad de nebraska](#)

#### 1.1.4 Mide correctamente el desempeño del modelo y sus métricas son correctas

- **¿Porque tengo esta competencia?** Se obtuvo esta subcompetencia ya que todos los modelos probados para el proyecto pasaron por una serie de iteraciones hasta llegar a conseguir una serie de pruebas estadísticas para determinar la precisión (accuracy) y error (loss, MSE, RMSE y MAE) de los modelos. Además, para los modelos que utilizan redes neuronales se incluyeron estas métricas en la etapa de entrenamiento para el conjunto de imágenes de validación.
- **Evidencia de trabajo para competencia**
  - [Modelo MLP de regresion para proyecto de la materia](#)

#### 1.1.5 Interpreta los resultados de las predicciones de los modelos y los interpreta en el contexto del problema de manera correcta

- **¿Porque tengo esta competencia?** Se obtuvo esta subcompetencia ya que en los diferentes modelos que se utilizaron para la base de datos numérica se obtienen resultados mediocres, aun después de analizar y modificar los parámetros de los modelos y datos, los resultados seguían siendo malos. Los modelos que utilizaban imágenes también demostraron poca eficiencia aun después de haber adaptado las imágenes seguía sin lograr tener un pronóstico correcto.
- **Evidencia de trabajo para competencia**
  - [Folder que contiene todos los resultados de los modelos utilizados para el proyecto](#)

### 1.2 SEG0403B - Resuelve situaciones mediante el cumplimiento de leyes, normas y principios éticos.

#### 1.2.1 Verifica que los datos estén anonimizados para no violar normas o leyes de privacidad de la información

No aplica

#### 1.2.2 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 normativa.

No aplica

#### 1.2.3 Presenta registros claros sobre el seguimiento del proceso y las autorizaciones para tener acceso a los datos

- **Justificacion** Las bases de datos que se nos brindaron o se consiguieron (ej. [Riwa dataset](#)) eran públicas y todos los cambios o movimientos que se realizaron se compartían entre los equipos para que pudieran eficientar los modelos

### 1.3 Correcciones

A lo largo del proyecto estuvimos haciendo muchas preguntas a los profesores, de como implementar nuestras ideas, (ej. modelo de CNN-LSTM), que ideas servirían y cuales no, porque un modelo esta sirviendo o no esta sirviendo para cambiar nuestras ideas o cambiar la manera en que resolvemos el problema a mano. Cada pregunta que hacíamos con los profesores nos ayudaba a cambiar la

manera en que queriamos resolver el problema y entender que cosas son posibles y si vamos por buen camino.

Tambien en la parte de nuestro paper que cosas llevar, esta bien escrito el orden en donde se explican las cosas es logico o hay una mejor manera y que se tiene que abordar en el paper.