

Instrucciones

- El proyecto será presentado en clases el 22 de Diciembre.

Introducción:

En este proyecto cada grupo trabajará en un dataset de clasificación encontrado en la plataforma Kaggle,

Se le asignará un dataset a cada grupo, los cuales se encuentran al final de este documento.

En cada una de las tareas se trabajará en implementar algunos métodos en el dataset que se le asignó al grupo. En el proyecto, por otro lado, se realizará una presentación sobre el dataset y los diferentes métodos implementados en él.

La presentación del proyecto debe estar dividida en 3 secciones que se presentarán a continuación.

Visualización y Preprocesamiento (30%):

En esta sección debe presentar el dataset que se le fué asignado, de que se trata, cual es la tarea a resolver y las características del dataset. Además debe presentar su trabajo realizado en el procesamiento y visualización inicial del dataset. Se espera se presente al menos 5 gráficos destacando diferentes aspectos del dataset, y en base a este trabajo proponer al menos 3 técnicas de preprocesamiento enfocado en las característica del dataset. Debe explicar claramente la lógica de cada uno de los pasos realizados.

Entrenamiento de Modelos (30%):

En esta sección debe proponer al menos 8 modelos para entrenar en su dataset. Al igual que en la sección anterior, debe explicar detalladamente porque propone cada modelo, además de dar una breve explicación del modelo. Al menos 3 de los modelos propuestos deben ser modelos basados en redes neuronales.

Resultados (40%):

Finalmente, en la sección de resultados debe comparar los resultados obtenidos por cada uno de sus modelos en el conjunto de test que encontrará en Kaggle. Utilice gráficos y tablas para exponer de mejor manera sus conclusiones. Debe presentar un conjunto de conclusiones sobre el dataset y la performance de sus modelos.

Datasets:

Los siguientes son los datasets asignados a cada grupo:

- Grupo 1: <https://www.kaggle.com/c/otto-group-product-classification-challenge/overview>
- Grupo 2: <https://www.kaggle.com/c/otto-group-product-classification-challenge/overview>
- Grupo 3: <https://www.kaggle.com/c/santander-customer-satisfaction/overview>
- Grupo 4: <https://www.kaggle.com/c/santander-customer-satisfaction/overview>
- Grupo 5: <https://www.kaggle.com/c/higgs-boson/overview>
- Grupo 6: <https://www.kaggle.com/c/higgs-boson/overview>
- Grupo 7: <https://www.kaggle.com/uciml/forest-cover-type-dataset>
- Grupo 8: <https://www.kaggle.com/uciml/forest-cover-type-dataset>
- Grupo 9: <https://www.kaggle.com/uciml/forest-cover-type-dataset>

Tips:

- Este es un proyecto abierto, en el sentido es que usted puede tomar las decisiones que le parezcan necesarias. Además puede recopilar más información desde la fuente que desee. Siéntase libre a tomar sus decisiones.

- Los Notebooks de la sección de Code de Kaggle le darán muchos ejemplos por donde empezar. Copias literales a los notebooks serán sancionadas, sin embargo, sienta la libertad para utilizarlos cómo base.
- Los datasets están pensados para ser de tamaño grande, por lo que posiblemente enfrente problemas de memoria al trabajar con el dataset. Para evitarlo se recomienda trabajar con Google Colab (puede revisar el siguiente Tutorial). Si sigue teniendo problemas de memoria puede tomar otras decisiones que le permitan entrenar el modelo.
- Habrá una hora de oficina a la semana para aclarar dudas del proyecto, esta será los Martes a las 12:00, se compartirá el link recurrente por Aula.