

Proyecto Final

Entrega: Miércoles 9 de Febrero de 2022

Máximo 3 estudiantes por grupo

Prof. Fabio A. González

Intr. Sistemas Inteligentes - 2021-2

Objetivo:

Ejecutar un proyecto de machine learning de forma efectiva usando la metodología y las herramientas presentadas en el curso.

Descripción

Se espera que utilice la metodología de trabajo CRISP-DM y las herramientas de modelamiento para llevar a cabo la planeación y ejecución de un proyecto aplicado. El conjunto de datos sobre el que trabajará puede ser seleccionado por ustedes (entre los conjuntos de datos propuestos) de acuerdo con sus intereses. El objetivo es que a través de un proceso de extensiva experimentación con modelos de machine learning poder llegar a obtener conclusiones con información valiosa que aporte en procesos de toma de decisiones en un dominio de aplicación particular.

El proyecto se desarrollará utilizando el lenguaje de programación Python y su entorno de herramientas para la computación científica, en forma de Notebook en el formato .ipynb. Se debe presentar el proyecto tomando como referencia las etapas previas al despliegue de la metodología CRISP-DM para análisis de datos (IBM, 2012).

Conjuntos de datos

Como se mencionó anteriormente, el planteamiento y desarrollo del proyecto se debe basar en alguno de los siguientes conjuntos de datos:

Google Play Store Apps:

Datos de 10 mil aplicaciones de la App Store obtenidas a través de web scraping con el objetivo de analizar el mercado de Android.

Trip Advisor Hotel Reviews:

20 mil reseñas de hoteles extraídas de Tripadvisor. Se puede usar este conjunto de datos para descubrir cómo son los mejores hoteles o usarla en sus propios viajes.

Avocado Prices:

Datos históricos de los precios del aguacate y volumen de ventas en múltiples mercados de Estados Unidos. Se puede modelar como una serie de tiempo.

Fashion MNIST:

Un conjunto de datos similar a MNIST con 70 mil imágenes con tamaño 28x28 de prendas de ropa. Presenta una tarea de clasificación.

Students Performance in Exams:

Notas obtenidas por estudiantes en varias asignaturas.

IBM HR Analytics Employee Attrition & Performance:

Prediga el desgaste de sus empleados más valiosos. Descubra los factores que conducen al desgaste de los empleados y explora cuestiones importantes como "La relación entre la distancia de la casa al trabajo por puesto de trabajo y el desgaste" o "La relación entre el ingreso mensual promedio por educación y el desgaste". Este es un conjunto de datos ficticio creado por científicos de datos de IBM.

Envío final

Se deb enviar un archivo zip con el informe final y los archivos detallados más adelante a través del siguiente file request de Dropbox (<https://www.dropbox.com/request/Q2sZH6Qj4dbE5xncKLT5>), antes del mediodía de la fecha límite. El archivo debe nombrarse `isi-proj-report-unalusername1-unalusername2-unalusername3.zip`, donde `unalusername` es el nombre de usuario asignado por la universidad (incluir los nombres de usuario de todos los miembros del grupo).

El archivo zip debe incluir los siguientes archivos:

- **Jupyter Notebook:** con todo el código del proyecto. El notebook debe estar debidamente explicado usando celdas de texto. Todos los pasos de carga, preprocesamiento, entrenamiento y prueba deben incluirse. with the code of the model as well as the respective additional files of the model (if there exists), the training, and the evaluation. Asegúrese de que el notebook se visualiza correctamente y está libre de errores antes de enviarlo.
- **Reporte:** un documento pdf que documenta todos los pasos de la metodología CRISP.
- **Poster:** Un archivo pdf con un póster que presenta sus resultados.

Por favor, no incluya imágenes o archivos binarios diferentes a los solicitados.

Adicionalmente, todos los miembros del grupo deben grabar un video de presentación del póster. El video tiene que ser subido a YouTube y el enlace tiene que ser incluido en el documento del reporte. La duración máxima del vídeo es de 6 minutos.

Bibliografía

IBM. "Manual CRISP-DM de IBM SPSS Modeler." *CRISP-DM*, 2012,

<ftp://ftp.software.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf> . Accessed 13 11 2020.