

Inteligencia Artificial

Proyecto de curso

Name: _____

Profesor: Marco Teran
Deadline: varias

1. Objetivo

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

2. Descripción del proyecto

Se espera que utilice la metodología de trabajo propuesta en el curso 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).

3. 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. [Acceder.](#)
- **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. [Acceder.](#)
- **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. [Acceder.](#)
- **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. [Acceder.](#)
- **Students Performance in Exams:** Notas obtenidas por estudiantes en varias asignaturas. [Acceder.](#)
- **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." "La relación entre el ingreso mensual promedio por educación y desgaste". Este es un conjunto de datos ficticio creado por científicos de datos de IBM. [Acceder.](#)

- **UJIIndoorLoc** Muchas aplicaciones del mundo real necesitan conocer la localización de un usuario para ofrecer sus servicios. La localización de usuarios ha sido un tema de investigación de interés en los últimos años. La localización de usuarios consiste en estimar la posición del usuario (latitud, longitud y altitud) mediante un dispositivo electrónico, normalmente un teléfono móvil. El problema de la localización en exteriores puede resolverse con gran precisión gracias a la tecnología GPS en los dispositivos móviles. Sin embargo, la localización en interiores sigue siendo un problema abierto, principalmente debido a la pérdida de la señal GPS en entornos interiores. Aunque existen algunas tecnologías y metodologías de posicionamiento en interiores, esta base de datos se centra en las basadas en huellas digitales WLAN (también conocidas como WiFi Fingerprinting).

[Acceder.](#)

4. Entregables

4.1. Contenido de la primera entrega del proyecto

Para la primera entrega se requiere la extracción, preprocesamiento, visualización y análisis de los datos. Se deberá encontrar las principales características estadísticas de estos. Estos se deberán representar y visualizar. 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 con el código del modelo, así como los respectivos archivos adicionales del modelo (si existen). Asegúrese de que el notebook se visualiza correctamente y está libre de errores antes de enviarlo.

Reporte: un informe del trabajo en forma de artículo científico en formato PDF generado en LaTeX que documente todos los pasos de la metodología CRISP. El trabajo debe tener al menos estas secciones: una introducción que describa el problema y el trabajo relevante; la descripción del método; visualización y análisis de los datos; y una sección de conclusiones.

Vídeo en YouTube de 5 minutos explicando los resultados del proyecto.

Repositorio Repositorio GIT (el enlace debe estar al final del documento PDF antes de la Bibliografía): el repositorio debe contener carpetas: códigos, \LaTeX , video

4.2. Contenido de la segunda entrega del proyecto

La aplicación de dos técnicas de Machine learning y sus respectivas métricas de evaluación.

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 con el código del modelo, así como los respectivos archivos adicionales del modelo (si existen). Asegúrese de que el notebook se visualiza correctamente y está libre de errores antes de enviarlo.

Reporte: un informe del trabajo en forma de artículo científico en formato PDF generado en LaTeX que documente todos los pasos de la metodología CRISP. El trabajo debe tener al menos estas secciones: una introducción que describa el problema y el trabajo relevante; la descripción del método; la evaluación experimental, incluyendo la descripción de los conjuntos de datos, la configuración experimental, los resultados y la discusión; y una sección de conclusiones.

Vídeo en YouTube de 5 minutos explicando los resultados del proyecto.

Repositorio Repositorio GIT (el enlace debe estar al final del documento PDF antes de la Bibliografía): el repositorio debe contener carpetas: códigos, \LaTeX , video

4.3. Contenido de la tercera entrega del proyecto

Para la tercera entrega se requiere la aplicación de una técnica de *Deep Learning* y sus respectivas métricas.

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 con el código del modelo, así como los respectivos archivos adicionales del modelo (si existen). Asegúrese de que el notebook se visualiza correctamente y está libre de errores antes de enviarlo.

Reporte: un informe del trabajo en forma de artículo científico en formato PDF generado en LaTeX que documente todos los pasos de la metodología CRISP. El trabajo debe tener al menos estas secciones: una introducción que describa el problema y el trabajo relevante; la descripción del método; la evaluación experimental, incluyendo la descripción de los conjuntos de datos, la configuración experimental, los resultados y la discusión; y una sección de conclusiones.

Repositorio Repositorio GIT (el enlace debe estar al final del documento PDF antes de la Bibliografía): el repositorio debe contener carpetas: códigos, \LaTeX , video

Poster: un archivo PDF con un póster que presente sus resultados. El póster debe mostrar de forma visual el problema, el método y los resultados experimentales.

Video: un vídeo de presentación del poster en el cual deben participar todos los miembros del grupo. El vídeo debe subirse a YouTube y el enlace debe incluirse en el documento del reporte. NO incluya el vídeo como parte de su presentación. La duración máxima del vídeo es de 10 minutos.

Por favor, no incluya imágenes o archivos binarios diferentes a los solicitados. Todos los archivos de la presentación deben estar comprimidos en un único archivo ZIP. El archivo debe ser nombrado como **ml-project-username1-username2-username3.zip**, donde nombre de usuario es el nombre de usuario asignado por la universidad en su correo (incluir los nombres de usuario de todos los miembros del grupo). El archivo debe ser enviado a través de la solicitud de Aula, antes de la medianoche de la fecha límite.

5. 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>.