
Proyecto de aula: Deep Learning

Marco Teran^{1,2}

¹Deep Learning

²Escuela de Ciencias exactas e Ingeniería

This project aims to execute an effective Deep learning project using the methodology and tools presented in the course. The project will involve planning and implementing a practical project, where the data set used will be selected by the student. The ultimate goal is to derive valuable insights through extensive experimentation with Deep learning models that can contribute to the decision-making process in a particular application domain. The project will be implemented using the Python programming language and its environment of scientific computing tools, in the form of a notebook in the iPython format.

1. Introducción

Este proyecto tiene como propósito poner en práctica los conocimientos adquiridos en el curso de Deep Learning mediante la implementación de un proyecto práctico. La realización de un proyecto práctico es fundamental para consolidar y aplicar los conceptos teóricos aprendidos en el curso, permitiendo una comprensión más profunda y significativa del tema. En este proyecto, el estudiante seleccionará un conjunto de datos para planificar e implementar el proyecto. El objetivo final del proyecto es obtener valiosas conclusiones a través de la experimentación con modelos de Deep Learning, lo que permitirá contribuir en procesos de toma de decisiones en un dominio de aplicación particular. La implementación se llevará a cabo utilizando el lenguaje de programación Python y su entorno de herramientas de computación científica en forma de un cuaderno en formato iPython.

2. Objetivos

2.1. Objetivos general

Ejecutar un proyecto de *Deep learning* de forma efectiva usando la metodología y las herramientas presentadas en el curso. Llevar a cabo un proceso de análisis de datos completo utilizando la metodología CRISP-DM y aplicar técnicas y algoritmos de Deep Learning para resolver problemas complejos de clasificación, regresión y predicción en diferentes campos, con énfasis en imágenes, series de tiempo y datos no estructurados, con el fin de obtener conclusiones valiosas a partir de un conjunto de datos seleccionado.

2.2. Objetivos específicos

- Seleccionar y comprender el conjunto de datos a utilizar, identificando las variables relevantes y posibles relaciones entre ellas.
- Realizar una exploración de los datos para detectar anomalías y posibles errores en el conjunto de datos.
- Desarrollar y ajustar un modelo de Deep learning que permita obtener conclusiones valiosas a partir de los datos y evaluar el desempeño del modelo.

3. 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 conjun-

tos de datos propuestos) de acuerdo con sus intereses. El objetivo es que a través de un proceso de extensiva experimentación con modelos de *Deep 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).

4. Conjuntos de datos

Como se mencionó anteriormente, el planteamiento y desarrollo del proyecto se debe basar en el siguiente conjunto de datos:

- **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*). [\[acceder\]](#)

5. Entregables

5.1. 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 utilizando las herramientas vistas en clases. 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, pre-

procesamiento y visualización de los datos, 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 los pasos de la metodología **CRISP** relacionados. Formato *IEEE Conference Template* en inglés. El trabajo debe tener al menos estas secciones: una introducción que describa el problema y trabajo relevante asociado (no menos de 3 fuentes de literatura indexada donde se haya utilizado el repositorio); la descripción del método; visualización y análisis de los datos; 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 , etc.

5.2. Contenido de la segunda entrega del proyecto

La aplicación de dos técnicas de *Deep learning*, sus respectivas métricas de evaluación y comparativa. 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**. Formato *IEEE Conference Template* en inglés. El trabajo debe tener al menos estas secciones: una introducción que describa el problema y trabajo relevante asociado (no menos de 6 fuentes de literatura indexada donde se haya utilizado el repositorio); 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 , etc.

5.3. Contenido de la tercera entrega del proyecto

Para la tercera entrega se requiere la aplicación de una técnica adicional de *Deep Learning*, aplicación de *autoencoders* para filtrado de la señal, y sus respectivas métricas de evaluación y comparativa.

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**. Formato *IEEE Conference Template* en inglés. El trabajo debe tener al menos estas secciones: una introducción que describa el problema y trabajo relevante asociado (no menos de 8 fuentes de literatura indexada donde se haya utilizado el repositorio); 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 , vídeo

- Asegurarse de que el conjunto de datos seleccionado es adecuado para el problema que se busca resolver.
- Documentar todo el proceso de análisis de datos para facilitar la evaluación y comunicación de los resultados.
- Trabajar en equipo y aprovechar las habilidades y conocimientos individuales de cada miembro para lograr los objetivos del proyecto de manera efectiva.

6. Bibliografía

IBM. “Manual CRISP-DM de IBM SPSS Modeler.” CRISP-DM, 2012. [\[Descargar\]](#)

- 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 `d1-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 mi correo antes de la medianoche de la **fecha límite**.

5.4. Recomendaciones para tener éxito en el proyecto:

- Planificar el proyecto con suficiente tiempo para cada una de las etapas del proceso de análisis de datos.
- Familiarizarse con las herramientas de machine learning y Python antes de comenzar el proyecto.