

PROGRAMA CURSO

Machine Learning en Geociencias

Prof.: Edier Aristizábal



UNIVERSIDAD
NACIONAL
DE COLOMBIA

SEDE MEDELLÍN

versión: 15 de julio de 2022

Introducción

El curso *Machine Learning* está orientado para estudiantes de posgrados que deseen adquirir conocimientos sobre modelos multivariados, para establecer patrones espaciales y análisis de inferencia estadística y predicción de una variable dependiente a partir de variables independientes o predictoras, con herramientas de aprendizaje automático (*Machine Learning*), minería de datos (*Data Mining*) y *Big Data*, bajo un ambiente de trabajo en Python. El curso es teórico - práctico. Se dictarán clases teóricas con las técnicas y modelos a utilizar, pero exige que el estudiante seleccione una base de datos espacial para poner en práctica el conocimiento adquirido en el curso. El contenido del curso incluye una introducción al ambiente de trabajo en Python, pero no es un curso avanzado de Python o SIG; por lo tanto, no requiere conocimientos profundos en dichas herramientas. Sin embargo, las personas expertas en cualquiera de estas herramientas son bienvenidas a aportar desde su conocimiento y en el marco del contenido del curso. El curso es una construcción conjunta de conocimiento por parte tanto de los estudiantes como del profesor.

1. PROGRAMA

El contenido del curso comprende los siguientes temas a desarrollar:

Introducción al curso

1.1. Ambiente de trabajo

1.2. Formulación de hipótesis

Presentación 1 (10 %)

1.3. Toma y procesamiento de datos

1.4. Análisis exploratorio de datos

1.5. Selección de variables

1.6. Selección de hiperparámetros

1.7. Selección del modelo

1.8. Modelos no supervisados

Presentación 2 – Implementación de modelos no supervisados (30 %)

1.9. Modelos supervisados paramétricos

1.10. Modelos supervisados no paramétricos

1.11. Modelos ensamblados

Presentación 3 – Implementación de modelos supervisados (30 %)

1.12. *Deep learning*

1.13. *Series de tiempo*

Trabajo final (30 %)

2. EVALUACIÓN DEL CURSO

El curso se evaluará con un trabajo individual donde el estudiante debe presentar los avances parciales y un informe final en formato de artículo. La presentación 1 corresponde a la definición del problema a resolver y la base de datos a utilizar con un valor del 10 % de la nota final. La presentación 2 corresponde a los resultados de la implementación de modelos no supervisados con un valor del 30 %. La presentación 3 corresponde a los resultados de la implementación de modelos supervisados con un valor del 30 %.

Finalmente, se deberá hacer entrega de un trabajo escrito en formato artículo, con los resultados finales, donde se seleccione el mejor modelo y optimicen los resultados en los modelos de validación con un valor del 30 %. Para el formato del artículo se puede seleccionar cualquier revista técnica, donde especifican el formato, tal como: número máximo de hojas, tipo de letra, tamaño, entre otras. Se debe seguir el formato estricto de la revista, esto hace parte de la evaluación.

Para la entrega del trabajo se deberá crear un repositorio en Github, donde se deberá cargar tanto el trabajo escrito en pdf, como los codigos utilizados y las bases de datos. El link del repositorio deberá ser compartido en el chat del Google Classroom.