



UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
PROGRAMA ACADÉMICO DE MATEMÁTICAS
FORMATO SYLLABUS
PLAN DE ESTUDIOS 298
VERSIÓN: 2021

RESOLUCIÓN ACREDITACIÓN DE ALTA CALIDAD No. 007575 DE JULIO DE 2019

FACULTAD: Ciencias Matemáticas y Naturales

NOMBRE DEL DOCENTE: Luis Alejandro Masmela Caita

ÁREA DE FORMACIÓN: Electiva

NOMBRE DEL ESPACIO ACADÉMICO: Machine Learning con R

TIPO DE ESPACIO: Teórico () Práctico () Teo-prac () Obligatorio () Electivo ()

CÓDIGO: SIN CÓDIGO

NÚMERO DE CRÉDITOS: 2

HORARIO: Total Horas Semanales Lectivas: _____

DÍA: _____ HORA: _____ SALÓN: _____

DÍA: _____ HORA: _____ SALÓN: _____

DÍA: _____ HORA: _____ SALÓN: _____

1. JUSTIFICACIÓN DEL ESPACIO ACADÉMICO

En la actualidad, estamos bombardeados por una gran cantidad de información. Estamos en la era de los grandes datos, la avalancha de información requiere métodos automatizados de análisis de datos, que es lo que proporciona el aprendizaje automático. En particular, definimos el aprendizaje automático como un conjunto de métodos que pueden detectar automáticamente patrones en los datos y luego usar los patrones descubiertos para predecir el futuro o para realizar otros tipos de toma de decisiones bajo incertidumbre. En el aprendizaje automático, la incertidumbre se presenta de muchas formas. El enfoque probabilístico del aprendizaje automático está estrechamente relacionado con el campo de la estadística, pero difiere ligeramente en cuanto a su énfasis y terminología.

El lenguaje R es básicamente desarrollado por estadísticos para ayudar a otros estadísticos y desarrolladores de manera más rápida y eficiente con los datos. Como ahora sabemos que el aprendizaje automático básicamente trabaja con una gran cantidad de datos y estadísticas como parte de la ciencia de datos, es recomendable el uso del lenguaje R. Por lo tanto, el lenguaje R se convierte en una herramienta útil para quienes trabajan con aprendizaje automático, lo que hace que las tareas sean más fáciles, rápidas e innovadoras.



2. REQUISITOS (Contenidos)

Se recomienda poseer conocimientos básicos en el software R; y en las áreas de Probabilidad e Inferencia Estadística.

3. DESCRIPCIÓN DEL ESPACIO ACADÉMICO

Este es un curso introductorio de Aprendizaje de Máquina, dirigido a estudiantes que deseen dar sus primeros pasos en la Ciencia de Datos. En este curso aprenderá cómo identificar problemas de Machine Learning e implementar soluciones en R-Studio. En el curso se hará referencia a distintos paquetes y a sus respectivas funciones para implementar las principales técnicas de modelado según los distintos problemas, la evaluación de performance de los modelos, las técnicas de calibración y ajuste, llegando hasta la implementación y comunicación de los resultados, no se trata de un curso del lenguaje usado en R. Por ello, es recomendable que los asistentes tengan un conocimiento básico de R, un curso de Teoría de las Probabilidades y conocimientos de Inferencia Estadística básica (de ser posible haber tomado la electiva de Introducción a la Estadística con R y el curso de Teoría de las Probabilidades).

4. OBJETIVOS

4.1 GENERAL

Aplicar conceptos y métodos fundamentales de las técnicas de Machine Learning o aprendizaje máquina, logrando que el alumno se familiarice con algunos de los principales algoritmos, así como sus aplicaciones en diferentes disciplinas, todo esto haciendo uso del lenguaje de programación R.

4.2 ESPECÍFICOS

- Implementar modelos de aprendizaje de máquina a partir de distintos paquetes usando el lenguaje R.
- Aplicar técnicas de análisis y visualización de datos en un conjunto de datos complejo para problemas de machine learning.
- Desarrollar y analizar proyectos de machine learning, aprendizaje supervisado, como regresión, clasificación y multiclase.
- Desarrollar y analizar proyectos de machine learning de Aprendizaje No Supervisado.



- Comprender los diferentes mecanismos y técnicas para aplicar analítica predictiva en problemas de machine learning e interpretar la salida dada por los modelos.

5. UNIDADES TEMÁTICAS Y/O PROBLEMÁTICAS

- Modelos
- Aprendizaje automático supervisado y no supervisado
- Regresión
- Redes neuronales
- Métodos basados en árboles
- Otros métodos avanzados
- Aprendizaje automático con el paquete caret

6. ESTRATEGIAS METODOLÓGICAS:

Clases magistrales asistidas por computador, en donde, a partir de bases de datos reales se implementan los modelos de aprendizaje de máquina más usados e implementados mediante paquetes implementados sobre el software RStudio. Se justifican teóricamente, de manera superficial, los resultados obtenidos luego de implementar los procesos sobre el programa. Interpretación de salidas y variación de los parámetros de los diferentes procesos permiten que el estudiante asimile los fundamentos de los diferentes modelos implementados. Cada tema se presenta de manera concisa con ejercicios que debe desarrollar el estudiante para aclarar la forma de proceder dependiendo del problema que se busca solucionar.

Basado en el sistema de créditos, la distribución de la dedicación horaria del estudiante para este espacio académico es la siguiente:

HORAS			Horas profesor/s emana	Horas Estudiante/sem ana	Total Horas Estudiante/semestre	Créditos
TD	TC	TA	(TD + TC)	(TD + TC +TA)	X 16 semanas	



Convenciones:

TD: Trabajo Presencial Directo; trabajo de aula con plenaria de todos los estudiantes.

TC: Trabajo Mediado cooperativo; Trabajo de tutoría del docente a pequeños grupos o de forma individual a los estudiantes.

TA: Trabajo Autónomo; Trabajo del estudiante sin presencia del docente, que se puede realizar en distintas instancias: en grupos de trabajo o en forma individual, en casa o en biblioteca, laboratorio, etc.)

7. RECURSOS

7.1 TEXTO GUÍA

- Burger, S. V. (2018). *Introduction to machine learning with R: Rigorous mathematical analysis*. "O'Reilly Media, Inc."
- Deisenroth, M. P., Faisal, A. A., & Ong, C. S. (2020). *Mathematics for machine learning*. Cambridge University Press.

7.2 TEXTOS COMPLEMENTARIOS

- Gareth, J., Daniela, W., Trevor, H., & Robert, T. (2013). *An introduction to statistical learning: with applications in R*. Springer.
- Faraway, J. J. (2004). *Linear models with R*. Chapman and Hall/CRC.
- Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4, No. 4, p. 738). New York: Springer.
- Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

7.3 REVISTAS

7.4 DIRECCIONES DE INTERNET

- <https://www.cienciadedatos.net/machine-learning-r.html>
- <https://www.geeksforgeeks.org/introduction-to-machine-learning-in-r/>
- <https://rstudio.cloud/>

7.5 MULTIMEDIA

7.6 MOODLE O PLATAFORMA ACADÉMICA (Link o enlace web)



- Enlace moodle: <https://aulasciencias.udistrital.edu.co/login/index.php>

7.7 SOFTWARE ESPECIALIZADO

- Software R

8. ORGANIZACIÓN /TIEMPO (Organizar contenidos por semanas)

- **Semana 1:** Limitaciones de modelado, Estadística y Cómputo en Modelado, Entrenamiento de datos, Validación cruzada, Uso de R, R y aprendizaje automático.
- **Semana 2, 3:** Modelos supervisados, regresión, datos de entrenamiento y prueba, clasificación, regresión logística, métodos de agrupamiento supervisado, métodos mixtos, modelos basados en árboles, bosques aleatorios, redes neuronales, máquinas de vectores de soporte, aprendizaje no supervisado, métodos de agrupamiento no supervisado.
- **Semana 4, 5:** Sesgo, Muestreo en R, Entrenamiento y prueba, Funciones de los conjuntos de entrenamiento y prueba, ¿Por qué hacer un conjunto de prueba?, Conjuntos de entrenamiento y prueba: modelado de regresión, Conjuntos de entrenamiento y prueba: modelado de clasificación, validación cruzada, validación cruzada k-Fold.
- **Semana 6, 7, 8:** Regresión lineal, regresión multivariante, regularización, regresión polinomial, bondad de ajuste con datos, error cuadrático medio, simplicidad del modelo y bondad de ajuste, regresión logística, Clasificación Binaria, Clasificación Multiclase, Regresión Logística con Caret, Regresión Lineal, Regresión Logística.
- **Semana 9, 10:** Redes neuronales de una sola capa, creación de una red neuronal simple mediante el uso de R, redes neuronales de múltiples capas, redes neuronales para regresión, redes neuronales para clasificación Redes neuronales con signo de intercalación, regresión, clasificación.
- **Semana 11, 12, 13:** Un modelo de árbol simple, decidir cómo dividir árboles, entropía de árbol y ganancia de información, ventajas y desventajas de los árboles de decisión, sobreajuste de árboles, árboles de poda, árboles de decisión para regresión, árboles de decisión para clasificación, árboles de inferencia condicional, regresión de árbol de inferencia condicional, condicional Clasificación de árbol de inferencia, bosques aleatorios, regresión de bosque aleatorio, clasificación de bosque aleatorio.
- **Semana 14, 15, 16:** Clasificación bayesiana naive, Análisis de Componentes Principales, Análisis Discriminante Lineal, Máquinas de Vectores de Soporte, k-vecinos más cercanos, regresión con kNN, clasificación con kNN.



9. EVALUACIÓN (Especificar porcentajes y formas de evaluación)

La evaluación debe ser coherente con la metodología. Para incentivar el estudio permanente y cultivar la disciplina, se recomiendan pruebas escritas cortas y frecuentes. Los estudiantes deben iniciarse en la aplicación de las formas de lenguaje, expresión y argumentación. La Universidad tiene reglamentado tres cortes:

1er corte: 35% Fecha:

2do corte: 35% Fecha:

3er corte: 30% Fecha: