



DIPLOMADO EN MACHINE LEARNING: DECISIONES BASADAS EN DATOS



Uptc[®]
Universidad Pedagógica y
Tecnológica de Colombia



INTRODUCCIÓN

El diplomado ***Machine Learning: Decisiones Basadas en Datos*** está diseñado para profesionales y estudiantes que buscan adquirir competencias avanzadas en el análisis de datos y la toma de decisiones informadas mediante técnicas de machine learning. A lo largo de 16 semanas, los participantes se sumergirán en los fundamentos teóricos, metodologías y aplicaciones prácticas del machine learning, abarcando áreas como la clasificación, regresión, clustering, y redes neuronales.

En un mundo donde los datos son fundamentales para la innovación y la competitividad, el conocimiento en machine learning se ha convertido en una habilidad esencial. Este programa aborda la necesidad crítica de comprender y aplicar algoritmos que permitan descubrir patrones ocultos en los datos, optimizar procesos y predecir comportamientos futuros. Al finalizar el diplomado, los participantes estarán equipados para implementar soluciones de machine learning que transformen diferentes tipos de datos en decisiones estratégicas efectivas, mejorando así la eficiencia y efectividad en sus respectivas industrias.

Este diplomado proporciona una experiencia de aprendizaje intensiva, combinando teoría y práctica con casos de estudio reales, preparando a los profesionales para enfrentar los desafíos de la revolución industrial 4.0.

OBJETIVOS

El diplomado ***Machine Learning: Decisiones Basadas en Datos*** tiene como propósito formar profesionales altamente capacitados en el ámbito del machine learning. A través de este programa, los participantes adquirirán conocimientos y habilidades esenciales para proponer soluciones de modelos predictivos y prescriptivos basados en datos.

- **Comprensión del área:** Adquirir una comprensión sólida y detallada de los principios y algoritmos fundamentales del machine learning, incluyendo métodos supervisados y no supervisados.
- **Aplicación práctica:** Aplicar técnicas de machine learning para resolver problemas complejos y reales, mejorando la capacidad de análisis y predicción en diversas industrias.
- **Habilidades técnicas:** Desarrollar habilidades prácticas en el manejo de herramientas y plataformas de análisis de datos como Python, TensorFlow, Keras, y Scikit-learn, esenciales para la implementación de modelos de machine learning.
- **Decisiones basadas en datos:** Capacitar a los participantes para tomar decisiones estratégicas basadas en datos, utilizando insights derivados del análisis de datos para mejorar la eficiencia y efectividad organizacional.
- **Innovación y competitividad:** Fomentar la innovación y competitividad en los participantes, permitiéndoles integrar soluciones de machine learning en sus respectivos campos profesionales, potenciando su capacidad para liderar proyectos tecnológicos y de análisis avanzado.

ESTRUCTURA GENERAL

Fase 1: Fundamentos computacionales

Sección dedicada a la fundamentación en el lenguaje de programación de Python, así como las herramientas computacionales a utilizar a lo largo del diplomado. Asimismo, se introduce la fundamentación de Machine Learning, sus aplicaciones y su relación con los datos y la inteligencia artificial.

Fase 2: Matemáticas para ML

Sección dedicada a la fundamentación matemática necesaria para trabajar en problemas de machine learning.

Fase 3: Preprocesamiento de datos

Sección dedicada a la fundamentación en limpieza, transformación y preparación de los datos.

Fase 4: Modelos Supervisados, No Supervisados y de Ensamble

Sección dedicada al enfoque en desarrollo de modelos de Machine Learning: supervisados y no supervisados a partir de datos preparados. Así como el uso de modelos de ensamble para acoplar los modelos y optimizar los resultados.

Fase 5: Redes Neuronales y Deep Learning

Sección dedicada a la introducción de las redes neuronales y sus aplicaciones en modelos predictivos, con extensión a aplicaciones de la inteligencia artificial.

Fase 6: Proyecto Final

Sección dedicada al desarrollo del proyecto final de aplicación basado en la fundamentación vista a lo largo del programa

PLAN DE ESTUDIOS

Fase 1: Fundamentos Computacionales (9 horas)

- **Entornos de Desarrollo y Herramientas (0.5 horas)**
 - Jupyter Notebooks, Google Colab y VSCode
 - Versionamiento de código con Git/GitHub
- **Introducción a Python (3 horas)**
 - Sintaxis Básica: Variables, operadores y estructuras de control
 - Funciones: Definición, argumentos, y retorno de valores
 - Manejo de Errores: Try-except y manejo de excepciones
 - Control de flujo
- **Estructuras de datos en Python (2 horas)**
 - Listas, tuplas y diccionarios
- **Librerías esenciales (2 horas)**
 - NumPy: Arrays
 - Pandas: Procesamiento de estructuras de datos
- **Introducción a Machine Learning (1.5 horas)**
 - Definición de machine learning, historia y evolución
 - Tipos de Aprendizaje: Supervisado y no supervisado
 - Software y frameworks comunes utilizados en ML
 - Relación entre ML y IA: Diferencias y conexiones
 - Aplicaciones Prácticas: Casos de uso en diversas industrias
 - Database, Datalake y Data warehouse

PLAN DE ESTUDIOS

Fase 2: Matemáticas para ML (15 horas)

• Álgebra Lineal (3 horas)

- Vectores, matrices y tensores: Definiciones, operaciones básicas, y propiedades
- Descomposición de Matrices: Eigenvalores, eigenvectores, y descomposición en valores singulares (SVD)
- Soluciones de sistemas de ecuaciones lineales

• Cálculo Diferencial (3 horas)

- Derivadas: Conceptos básicos y reglas de derivación
- Gradiente y optimización: Gradiente descendente y técnicas de optimización para entrenamiento de modelos

• Estadística Descriptiva (4 horas)

- Medidas de Tendencia Central: Media, mediana, y moda
- Medidas de Dispersión: Varianza, desviación estándar y cuantiles
- Análisis multivariado
- Covarianza
- Correlación
- Visualización de datos
- Matplotlib y Seaborn
- Visualizaciones más comunes:
 - Barplot
 - Histograma
 - Boxplot
 - Scatterplot
 - Mapa de calor

• Teoría de Probabilidades (5 horas)

- Conceptos Básicos: Probabilidad, eventos, y espacio muestral
- Reglas de Probabilidad: Ley de la suma, ley del producto, y teorema de Bayes
- Variables Aleatorias y Distribuciones: Variables discretas y continuas, esperanza matemática, y varianza
- Distribuciones de Probabilidad: Binomial, Normal, Poisson y T-Student

PLAN DE ESTUDIOS

Fase 3: Preprocesamiento de datos (10 horas)

- **Importación y exploración de datos (2 horas)**
 - Carga y transformación de datos desde archivos CSV
 - Resumen estadístico: datasets y tipos de características
- **Ingeniería de características:**
 - **Limpieza de Datos (3.5 horas)**
 - Manejo de datos faltantes: Imputación y detección
 - Detección y tratamiento de Outliers
 - Corrección de errores
 - Agrupación de características
 - **Transformación de Datos (3.5 horas)**
 - Normalización y estandarización
 - Codificación de variables categóricas: One-hot encoding
 - Generación de nuevas variables
 - **Preparación de Datos para modelado (1 hora)**
 - Separación en conjuntos de entrenamiento, validación y prueba
 - Balanceo de conjuntos de datos

PLAN DE ESTUDIOS

Fase 4: Modelos Supervisados, No Supervisados y de Ensamble (60 horas)

• Modelos Supervisados (30 horas)

- Regresión Lineal (3 horas)
 - Fundamentos, implementación y ajuste del modelo
 - Regularización: L1 (Lasso) y L2 (Ridge)
 - Validación cruzada y K-Fold Cross Validation
 - Métodos de validación: Leave-One-Out, Bootstrap
 - Ajuste de Hiperparámetros: Grid Search y Random Search
 - Evaluación de desempeño: R^2 , error cuadrático medio (MSE), Error Absoluto Medio (MAE)
- Regresión Logística (3 horas)
 - Fundamentos y aplicación en clasificación binaria
 - Implementación y ajuste del modelo
 - Evaluación de desempeño: matriz de confusión, AUC-ROC curve
- Árboles de Decisión (3 horas)
 - Fundamentos y algoritmos de construcción
 - Poda de árboles y prevención de sobreajuste
 - Evaluación de desempeño: Accuracy, precisión-recall, F1-score
 - Overfitting y Underfitting
 - Matriz de Confusión
- Support Vector Machines (SVM) (3 horas)
 - Conceptos básicos y teoría de márgenes
 - Implementación de SVM lineal y no lineal
 - Ajuste de hiperparámetros y evaluación
- K-Nearest Neighbors (KNN) (3 horas)
 - Fundamentos y teoría
 - Implementación y ajuste de K
 - Evaluación de desempeño
- Modelos de Ensamble (15 horas)
 - Teoría de ensambles y beneficios de los modelos de ensamble
 - Métodos de combinación: votación y promedio ponderado
 - Grid Search
 - Bagging: Random Forest
 - Fundamentos y construcción de bosques aleatorios
 - Evaluación de importancia de características
 - Ajuste de hiperparámetros

PLAN DE ESTUDIOS

- Boosting (AdaBoost, Gradient Boosting, XGBoost)
 - Fundamentos y diferencias entre técnicas de boosting
 - Implementación y ajuste de hiperparámetros
 - Stacking y Blending
- **Modelos No Supervisados (30 horas)**
 - Clustering (20 horas)
 - K-means: teoría, implementación y selección de K.
 - Clustering jerárquico: algoritmos aglomerativos y divisivos.
 - Evaluación de resultados: Elbow Method, coeficiente de silhouette, índice de Dunn
 - Medidas de cohesión y separación
 - Basados en densidad: DBSCAN (3 horas)
 - Reducción de Dimensionalidad (7 horas)
 - PCA (Análisis de Componentes Principales)
 - Fundamentos y matemáticas detrás de PCA
 - Implementación y análisis de componentes principales.
 - t-SNE, UMap, TriMap y PacMap
 - Modelos de Mezcla Gaussianos (3 horas)
 - Fundamentos y teoría de distribuciones gaussianas.
 - Implementación y ajuste de parámetros.
 - Evaluación de resultados

PLAN DE ESTUDIOS

Fase 5: Redes Neuronales y Deep Learning (16 horas)

- **Introducción a Redes Neuronales (2 horas)**
 - Fundamentos de las Redes Neuronales Artificiales (ANN)
 - Estructura y funcionamiento de una neurona artificial
 - Arquitectura de una red neuronal: capas de entrada, ocultas y salida
- **Perceptrón y Perceptrón Multicapa (4 horas)**
 - Perceptrón: Teoría, implementación y aplicaciones
 - Perceptrón Multicapa (MLP): Algoritmos de entrenamiento
 - Funciones de activación: Sigmoide, ReLU, Tanh
- **Entrenamiento y Optimización de Redes Neuronales (4 horas)**
 - Proceso de entrenamiento: Forward y Backward Propagation
 - Algoritmos de optimización: GD, GD Estocástico y Adam
 - Regularización, Dropout, Early Stopping
- **Redes Neuronales Convolucionales (CNN) (3 horas)**
 - Fundamentos y arquitectura de las CNN
 - Capas convolucionales y de *pooling*
 - Aplicaciones en visión por computadora: Clasificación de imágenes y detección de objetos
- **Redes Neuronales Recurrentes (RNN) (3 horas)**
 - Arquitectura y entrenamiento de RNN
 - Aplicaciones en series temporales
 - LSTM

PLAN DE ESTUDIOS

Fase 6: Proyecto Final (11 horas)

- **Definición del proyecto y alcance (1 hora)**
 - Selección del problema a resolver
 - Justificación del problema
 - Definición de objetivos y metas del proyecto
 - Delimitación del alcance y expectativas
- **Recolección y preparación de datos (2 horas)**
 - Recolección de datos necesarios para el proyecto
 - Limpieza y preprocesamiento de datos
 - Exploración y análisis inicial de los datos
- **Desarrollo y entrenamiento de modelos (2 horas)**
 - Selección de algoritmos y modelos adecuados
 - Implementación y entrenamiento de modelos
 - Ajuste de hiperparámetros y optimización
- **Evaluación y validación de modelos (1.5 horas)**
 - Aplicación de técnicas de validación cruzada
 - Evaluación de desempeño utilizando métricas adecuadas
 - Ajustes y mejoras basadas en resultados de validación
- **Implementación y despliegue del modelo (1.5 horas)**
 - Implementación final del modelo en un entorno de producción
 - Consideraciones para el despliegue y mantenimiento del modelo
 - Documentación del proceso y resultados
- **Presentación y defensa del proyecto (3 horas)**
 - Preparación de la presentación del proyecto
 - Estructura y contenido de la presentación
 - Defensa del proyecto frente a un comité evaluador o audiencia

INFORMACIÓN ADICIONAL

Duración: 16 semanas

Intensidad: 7.5 horas semanales

Horarios: Martes (07:00 p.m. - 09:00 p.m.)
Jueves (06:00 p.m. - 08:30 p.m.)
Viernes (06:00 p.m. - 09:00 p.m.)

Modalidad: Virtual

Dirigido a profesionales y/o estudiantes de:

- Ingenierías
- Ciencias Básicas
- Economía
- Ciencias de la Salud
- Administración de Empresas

Requisitos:

- Conocimientos básicos de programación y matemáticas
- Disponibilidad de tiempo horaria

Certificación:

- Asistencia mínima del 70% emitida por CIEC



RECURSOS Y CONTACTO

Recursos:

- Repositorio oficial de GitHub
- Acceso a grabaciones
- Acceso a aula Moodle y carpeta de Google Drive

Costo: \$1.750.000 por persona

Realiza el proceso de inscripción mediante el ***Centro de Gestión de Investigación y Extensión de la Facultad de Ciencias***

Contacto: ciec@uptc.edu.co

