



DIPLOMADO EN CIENCIA DE DATOS: APLICACIONES CON MACHINE LEARNING

MODALIDAD VIRTUAL



VIGILADA MINEDUCACIÓN

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS

RESOLUCIÓN 023655 DE 2021 MEN / 6 AÑOS



Escuela de Posgrados
Facultad de Ciencias



INTRODUCCIÓN

En las últimas dos décadas, las empresas han invertido en tecnología para recopilar información, crucial para la toma de decisiones operativas y estratégicas. Las estrategias tradicionales de informes se han vuelto obsoletas, siendo reemplazadas por modelos descriptivos, predictivos y prescriptivos. Estos modelos extraen patrones de datos, proporcionando una visión completa de la empresa en su estado pasado, presente y futuro, y asegurando su crecimiento continuo.

La Facultad de Ciencias de la UPTC, con el apoyo del CIEC, ofrece el diplomado en **Ciencia de Datos: Aplicaciones con Machine Learning**. Este programa de 16 semanas está dirigido a profesionales y estudiantes que desean adquirir habilidades avanzadas en análisis de datos y machine learning. El curso abarca teoría, metodologías y aplicaciones prácticas, incluyendo clasificación, regresión, clustering y redes neuronales. En un contexto donde los datos son esenciales para la innovación y competitividad, este diplomado busca capacitar a los participantes en la aplicación de algoritmos para descubrir patrones, optimizar procesos y predecir comportamientos futuros.

Al finalizar el diplomado, los participantes estarán capacitados para implementar soluciones de ciencia de datos y machine learning, mejorando la eficiencia y efectividad en sus industrias. El programa ofrece una experiencia intensiva que combina teoría y práctica con casos de estudio reales, preparando a los profesionales para los desafíos de la revolución industrial 4.0 y aumentando la competitividad y desempeño de sus organizaciones.



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

VIGILADA MINEDUCACIÓN

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN 023855 DE 2021 MEN / 6 AÑOS



**Escuela de Posgrados
Facultad de Ciencias**

OBJETIVOS

- **Comprensión del área:** Adquirir una comprensión sólida y detallada de los principios y algoritmos fundamentales de la ciencia de datos y el machine learning.
- **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 lenguajes de programación como Python y SQL, y plataformas de desarrollo de machine learning, tales como TensorFlow, Keras, y Scikit-learn, esenciales para la implementación de modelos de ciencia de datos.
- **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.

Recursos:

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



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN G23655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

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
- Nivel de inglés básico
- Disponibilidad de tiempo horaria

Certificación:

- Asistencia mínima del 75% emitida por CIEC



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN G23655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

ESTRUCTURA GENERAL

Fase 1: Fundamentos de la Ciencia de Datos

Sección dedicada a la Introducción de los conceptos referentes en la ciencia de datos en un contexto organizacional, así como las herramientas computacionales a utilizar a lo largo del diplomado.

Fase 2: Matemáticas para Ciencia de Datos

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

Fase 3: Analítica de datos

Sección dedicada a la Introducción al manejo de bases de datos relacionales, lenguaje de consulta SQL, análisis exploratorio de datos y las librerías fundamentales de visualización en lenguaje de programación Python y en la plataforma de Business Intelligence, PowerBI

Fase 4: Machine Learning: Aplicaciones en la Ciencia de Datos

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 de la Ciencia de Datos (10 horas)

- **Datos, Machine Learning e Inteligencia Artificial**
- **Tendencias en datos e IA**
- **Riesgos, consideraciones legales y normativa nacional**
- **Entornos de Desarrollo: Jupyter Notebooks, Google Colab y VSCode**
- **Software y frameworks comunes**
- **Versionamiento de código con Git/GitHub**
- **Introducción a Python**
 - 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
 - Listas, tuplas y diccionarios
 - NumPy: Arrays
 - Pandas: Procesamiento de estructuras de datos
- **Aplicaciones Prácticas: Casos de uso en diversas industrias 4.0**
- **Database, Datalake y Data warehouse**



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN 023655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

PLAN DE ESTUDIOS

Fase 2: Matemáticas para Ciencia de Datos (25 horas)

- **Álgebra Lineal**
 - 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**
 - Derivadas: Conceptos básicos y reglas de derivación
 - Gradiente y optimización: Gradiente descendente y técnicas de optimización para entrenamiento de modelos
- **Estadística Descriptiva**
 - 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**
 - Conceptos Básicos: Probabilidad, eventos, y espacio muestral
 - Reglas de Probabilidad: Ley de la suma, ley del producto, Teorema de Bayes y Naive-Bayes
 - Variables Aleatorias y Distribuciones: Variables discretas y continuas, esperanza matemática, y varianza
 - Distribuciones de Probabilidad: Binomial, Normal, Poisson y T-Student
- **Estadística Inferencial**
 - Distribuciones de muestra y Teorema de Límite Central
 - Estimación de parámetros
 - Pruebas de hipótesis
 - Regresión y correlación
 - Análisis de varianza (ANOVA)



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN G23655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

PLAN DE ESTUDIOS

Fase 3: Analítica de Datos (25 horas)

- **Bases de datos relacionales y visualización de datos**
 - Bases de datos relacionales
 - Lenguaje SQL
 - Visualización mediante PowerBI
- **Análisis Exploratorio de Datos y Preprocesamiento**
 - **Importación y exploración de datos:**
 - Carga y transformación de datos desde archivos CSV
 - Resumen estadístico: datasets y tipos de características
 - **Análisis descriptivo y diagnóstico**
 - **Ingeniería de características:**
 - **Limpieza de Datos:**
 - Manejo de datos faltantes: Imputación y detección
 - Detección y tratamiento de Outliers
 - Corrección de errores
 - Agrupación de características
 - **Transformación de Datos**
 - Normalización y estandarización
 - Codificación de variables categóricas: One-hot encoding
 - Generación de nuevas variables
 - **Preparación de Datos para modelado**
 - Separación en conjuntos de entrenamiento, validación y prueba
 - Balanceo de conjuntos de datos
 - **Iteración de procesos**

PLAN DE ESTUDIOS

Fase 4: Machine Learning: Aplicaciones en la Ciencia de Datos (36 horas)

- **Introducción al Machine Learning**

- **Modelos Supervisados**

- 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 (6 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**
 - Clustering (5 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
 - Reducción de Dimensionalidad (5 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 (5 horas)
 - Fundamentos y teoría de distribuciones gaussianas.
 - Implementación y ajuste de parámetros.
 - Evaluación de resultados



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN G2.3655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

PLAN DE ESTUDIOS

Fase 5: Redes Neuronales y Deep Learning (14 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 (2 horas)**
 - Perceptrón: Teoría, implementación y aplicaciones
 - Perceptrón Multicapa (MLP): Algoritmos de entrenamiento
 - Funciones de activación: Sigmoide, ReLU, Tanh
- **Entrenamiento y Optimización de Redes Neuronales (3 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) (4 horas)**
 - Arquitectura y entrenamiento de RNN
 - Aplicaciones en series temporales
 - LSTM



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS

RESOLUCIÓN G23655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

PLAN DE ESTUDIOS

Fase 6: Proyecto Final (10 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 hora)**
 - 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 hora)**
 - 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



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN G2.3655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

RECURSOS Y CONTACTO

Costo: \$1.750.000 por persona

Costo Anticipado: \$1.550.000 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



Uptc
Universidad Pedagógica y
Tecnológica de Colombia

ACREDITACIÓN INSTITUCIONAL
DE ALTA CALIDAD
MULTICAMPUS
RESOLUCIÓN 023655 DE 2021 MEN / 6 AÑOS

VIGILADA MINEDUCACIÓN



Escuela de Posgrados
Facultad de Ciencias

