


| | | |
|---|--|------------|
|  | UNIVERSIDAD DE CALDAS | |
| | FORMATO PARA CREACIÓN – MODIFICACIÓN DE ACTIVIDADES ACADÉMICAS | |
| | CÓDIGO: R-1202-P-DC-503 | VERSIÓN: 3 |

PLAN INSTITUCIONAL DE ACTIVIDAD ACADÉMICA

I. IDENTIFICACIÓN

| | |
|---|---|
| Facultad que ofrece la Actividad Académica: | |
| Departamento que ofrece la Actividad Académica: | |
| Nombre de la Actividad Académica: | ANALÍTICA DE DATOS Y MACHINE LEARNING PARA LA OPTIMIZACIÓN INDUSTRIAL |
| Código de la Actividad Académica: | |
| Versión del Programa Institucional de la Actividad Académica (PIAA): | 1 |
| Acta y fecha del Consejo de Facultad para: aprobación____ modificación____ | Acta No. ____ Fecha: _____ |
| Programas a los que se le ofrece la Actividad Académica (incluye el componente de formación al cual pertenece): | ESPECIALIZACIÓN EN INDUSTRIA 5.0 Y AUTOMATIZACIÓN INDUSTRIAL |
| Actividad Académica abierta a la comunidad: | Si __ No _X_ |

Tipo de actividad: Teórica ____ Teórico - Práctica
 __X_ Práctica _____

| | | | |
|---|----|------------------------------------|-----|
| Horas teóricas: | 36 | Horas prácticas: | 16 |
| Horas presenciales: | 48 | Horas no presenciales: | 64 |
| Horas presenciales del docente: | 48 | Relación Presencial/No presencial: | 1:1 |
| Horas inasistencia con las que se reprueba: | 5 | Cupo máximo de estudiantes: | 25 |
| Habitable (Si o No): | SI | Nota aprobatoria: | 3 |
| Créditos que otorga: | 3 | Duración en semanas: | 16 |

Requisitos (escribir los códigos y el nombre de las actividades académicas que son requisitos, diferenciados por programas para el caso de una actividad académica polivalente):

1. **JUSTIFICACIÓN:** describe las razones por las cuales es importante la actividad académica desde la perspectiva del conocimiento, el objeto de formación del programa, el perfil profesional del egresado(s), y su lugar en el currículo.

En la era de la Industria 5.0, la capacidad de recolectar, procesar, analizar y extraer valor de los ingentes volúmenes de datos generados por los sistemas industriales (Big Data Industrial) es un diferenciador competitivo clave. La Analítica de Datos y el Machine Learning (ML) proporcionan las herramientas y metodologías para transformar estos datos brutos en conocimiento accionable, permitiendo la optimización de procesos, el mantenimiento predictivo, la mejora de la calidad, la personalización de la producción y la toma de decisiones estratégicas informadas. Esta asignatura es fundamental para el Especialista en Industria 5.0 y Automatización Industrial, ya que le dota de las competencias necesarias para aplicar estas técnicas avanzadas, pasar de un enfoque reactivo a uno proactivo y predictivo, y desbloquear nuevas oportunidades de eficiencia, innovación y creación de valor en el entorno productivo.

2. **OBJETIVOS:** describe en forma clara lo que se pretende con el desarrollo de la actividad académica.

Desarrollar en los estudiantes las competencias para aplicar principios y técnicas de analítica de datos y machine learning en la solución de problemas y la optimización de procesos en entornos industriales, utilizando datos provenientes de sistemas IIoT y de producción para mejorar la toma de decisiones y la eficiencia operativa.

2. Específicos:
 1. Comprender los principios fundamentales de la analítica de datos y el machine learning en el contexto de la industria.
 2. Aplicar técnicas de machine learning para la optimización de procesos y la mejora de la eficiencia productiva.
 3. Utilizar herramientas de analítica de datos para identificar patrones, realizar predicciones y mejorar la toma de decisiones en tiempo real.
 4. Implementar soluciones de mantenimiento predictivo utilizando algoritmos de machine learning.
 5. Desarrollar proyectos que integren analítica de datos y machine learning para la resolución de problemas industriales.

NOTA: en el caso que el Programa Institucional de la Actividad Académica (PIAA) se desarrolle por competencias, es necesario completar los siguientes aspectos, en lugar de objetivos:

-
3. **COMPETENCIAS:** describe actuaciones integrales desde saber ser, el saber hacer y el saber conocer, para identificar, interpretar, argumentar y resolver problemas del contexto con idoneidad y ética.

1. Genéricas

- Pensamiento Analítico y Cuantitativo: Capacidad para abordar problemas industriales desde una perspectiva basada en datos y aplicar métodos estadísticos y computacionales.
- Resolución de Problemas Basada en Evidencia: Habilidad para utilizar los resultados del análisis de datos para fundamentar decisiones y proponer soluciones.
- Manejo de Herramientas Tecnológicas: Destreza en el uso de software y lenguajes de programación para el análisis de datos y el machine learning.
- Interpretación y Comunicación de Resultados Complejos: Capacidad para traducir hallazgos técnicos en información comprensible y accionable para diferentes audiencias.

2. Específicas

C4 (Corresponde al RA4 del programa): Utilizar herramientas de analítica de datos y machine learning para la interpretación de información operativa y la toma de decisiones en tiempo real, orientadas a la mejora de procesos industriales.

(Sub-competencias específicas de la asignatura):

- Aplicar el proceso KDD (Knowledge Discovery in Databases) o CRISP-DM en proyectos de analítica industrial.
- Realizar la limpieza, transformación y preparación de datos industriales para el modelado.
- Seleccionar, entrenar y validar modelos de machine learning apropiados para problemas de predicción, clasificación o segmentación en la industria.
- Visualizar e interpretar los resultados de los análisis para generar insights y recomendaciones de optimización.

COMPETENCIAS GENÉRICAS: describen el conjunto de conocimientos, habilidades, destrezas y actitudes que le permiten al egresado del programa interactuar en diversos contextos de la vida profesional.

COMPETENCIAS ESPECÍFICAS: describen los comportamientos observables que se relacionan directamente con la utilización de conceptos, teorías o habilidades, logrados con el desarrollo del contenido de la Actividad Académica.

4. **RESULTADOS DE APRENDIZAJE (RA):** cada asignatura debe contener resultados de aprendizaje particulares, siempre articulados con los generales de cada programa. Los RA de una asignatura pueden tributar a varios RA generales, y no necesariamente hay una relación uno a uno.

Resultados de Aprendizaje (RA) (Alineados con RA4 del programa y adaptados):

- **RA4.1.** Comprender los fundamentos del ciclo de vida de la analítica de datos industriales (desde la adquisición y preprocesamiento hasta la visualización e interpretación de resultados) y los principales tipos de algoritmos de machine learning (supervisados, no supervisados, por refuerzo) aplicables a la industria.
- **RA4.2.** Seleccionar y aplicar técnicas de preprocesamiento y exploración de datos (limpieza, transformación, reducción de dimensionalidad, análisis exploratorio de datos - EDA) a conjuntos de datos industriales.
- **RA4.3.** Implementar y evaluar modelos básicos de machine learning (ej. regresión, clasificación, clustering) utilizando software y lenguajes de programación especializados para resolver problemas industriales como la predicción de fallos, la optimización de parámetros o la detección de anomalías.
- **RA4.4.** Interpretar los resultados de los modelos de analítica y machine learning en el contexto de un problema industrial, comunicando los hallazgos de manera efectiva para la toma de decisiones y proponiendo acciones de mejora.

5. **CONTENIDO:** describe los temas y subtemas que se desarrollarán en la actividad académica. Estos deben estar en perfecta coherencia con los objetivos, método y evaluación de la asignatura y con los perfiles de formación de los programas a los que se ofrece la actividad académica.

Módulo 1: Introducción a la Analítica de Datos y Machine Learning en la Industria

- El valor de los datos en la Industria 5.0: de Big Data a Smart Data.
- Tipos de analítica: Descriptiva, Diagnóstica, Predictiva y Prescriptiva.
- Ciclo de vida de un proyecto de analítica de datos / machine learning (ej. CRISP-DM, KDD).
- Fuentes de datos en entornos industriales: sensores IIoT, SCADA, PLC, MES, ERP, CMMS.
- Conceptos fundamentales de Machine Learning:
 - Aprendizaje Supervisado (Regresión, Clasificación).
 - Aprendizaje No Supervisado (Clustering, Reducción de Dimensionalidad).
 - Aprendizaje por Refuerzo (Introducción).
- Ética y consideraciones de privacidad en el manejo de datos industriales.

Módulo 2: Preprocesamiento y Exploración de Datos Industriales

- **Adquisición y Calidad de Datos:**
 - Técnicas de recolección de datos de series temporales y datos de eventos.
 - Limpieza de datos: tratamiento de valores atípicos (outliers), datos faltantes, ruido.
- **Transformación de Datos:**
 - Normalización y estandarización.
 - Ingeniería de características (Feature Engineering) para datos industriales.
 - Reducción de dimensionalidad (ej. PCA - Análisis de Componentes Principales).
- **Análisis Exploratorio de Datos (EDA):**

- Estadística descriptiva para datos industriales.
- Técnicas de visualización de datos (histogramas, box plots, scatter plots, series de tiempo).
- Identificación de patrones, correlaciones y tendencias iniciales.
- Herramientas y librerías para preprocesamiento y EDA (ej. Python con Pandas, Matplotlib, Seaborn).

Módulo 3: Modelado con Machine Learning para Aplicaciones Industriales

- **Modelos de Regresión para Predicción:**
 - Regresión Lineal, Regresión Polinómica.
 - Aplicaciones: predicción de consumo energético, estimación de vida útil restante (RUL).
- **Modelos de Clasificación para Detección y Diagnóstico:**
 - Regresión Logística, Árboles de Decisión, Random Forest, Support Vector Machines (SVM), Redes Neuronales (introducción).
 - Aplicaciones: detección de anomalías, clasificación de calidad de producto, diagnóstico de fallos.
- **Modelos de Clustering para Segmentación y Descubrimiento de Patrones:**
 - K-Means, Agrupamiento Jerárquico.
 - Aplicaciones: segmentación de regímenes de operación de máquinas, identificación de patrones de consumo.
- **Evaluación de Modelos de ML:**
 - Métricas de rendimiento para regresión (RMSE, MAE, R^2).
 - Métricas de rendimiento para clasificación (Precisión, Recall, F1-Score, Curva ROC).
 - Validación cruzada.
- Software y librerías para modelado con ML (ej. Python con Scikit-learn, TensorFlow/Keras introductorio).

Módulo 4: Implementación, Interpretación y Casos de Uso en la Optimización Industrial

- Interpretación de los resultados de los modelos de ML en el contexto industrial.
- Comunicación de hallazgos y recomendaciones a equipos técnicos y de gestión.
- Consideraciones para el despliegue de modelos de ML en entornos de producción (MLOps introductorio).
- **Casos de Uso Detallados:**
 - **Mantenimiento Predictivo (PdM):** Uso de ML para predecir fallos en maquinaria.
 - **Optimización de Procesos Productivos:** Ajuste de parámetros de proceso basado en modelos.
 - **Control de Calidad Inteligente:** Detección automática de defectos mediante visión artificial y ML.
 - **Gestión Energética Eficiente:** Predicción y optimización del consumo energético.
 - **Optimización de la Cadena de Suministro:** Predicción de demanda, optimización de inventarios.
- Desafíos en la implementación de analítica y ML en la industria.

- El futuro de la IA y el ML en la Industria 5.0.

6. **METODOLOGÍA:** describe las estrategias educativas, métodos, técnicas, herramientas y medios utilizados para el desarrollo del contenido, en coherencia con los objetivos o competencias.

- **Clases Teórico-Prácticas (Virtual Sincrónico Viernes / Presencial Sábado):** Exposición de los conceptos fundamentales de analítica y ML, y las metodologías de aplicación. Se complementará con la discusión de artículos, demostraciones de software y ejemplos de la industria.
- **Talleres Prácticos de Programación y Análisis de Datos (Presencial Sábado / Entornos Virtualizados con acceso a software):**
 - Ejercicios de preprocesamiento y exploración de datos utilizando Python (Pandas, NumPy, Matplotlib, Seaborn).
 - Desarrollo y evaluación de modelos de regresión, clasificación y clustering utilizando Scikit-learn sobre conjuntos de datos industriales (reales anonimizados o simulados).
- **Aprendizaje Basado en Problemas y Proyectos (Grupales, desarrollo progresivo):** Los estudiantes trabajarán en equipos para abordar un problema industrial utilizando un ciclo completo de analítica de datos/ML, desde la comprensión del problema y los datos hasta la propuesta de un modelo y la interpretación de sus resultados.
- **Análisis de Casos de Estudio y Artículos Científicos (Virtual Sincrónico / Presencial):** Estudio crítico de aplicaciones reales de analítica y ML en la industria, evaluando la metodología, los resultados y el impacto.
- **Uso de Notebooks Interactivos (ej. Jupyter Notebooks, Google Colab):** Facilitar la experimentación con código, datos y la documentación de los análisis.
- **Revisión de Documentación de Librerías y Herramientas:** Fomentar la consulta autónoma de la documentación de las herramientas de software utilizadas.

7. **CRITERIOS GENERALES DE EVALUACIÓN:** describe las diferentes estrategias evaluativas, con valoraciones cuantitativas y reportes cualitativos, si son del caso, que se utilizarán para determinar si el estudiante ha cumplido con lo propuesto como objetivos o como competencias de la Actividad Académica. Ver reglamento estudiantil y política curricular.

- **Participación y Resolución de Ejercicios en Clase (Virtual y Presencial): 15%**
 - Evaluación de la comprensión de conceptos y la participación en discusiones sobre metodologías y aplicación de técnicas.
- **Entregables de Talleres Prácticos (Notebooks de Análisis y Modelado): 35%**

- Calificación de los cuadernos de trabajo donde se evidencie el preprocesamiento de datos, la exploración, la implementación de algoritmos de ML, la evaluación de modelos y la interpretación de resultados.
- **Análisis Crítico de Artículos o Casos de Estudio (Individual/Grupal): 20%**
 - Evaluación de la capacidad para analizar y sintetizar información de investigaciones o implementaciones de analítica/ML en la industria, identificando fortalezas, debilidades y lecciones aprendidas.
- **Proyecto Final Grupal (Aplicación de Analítica/ML a un Problema Industrial): 30%**
 - Desarrollo, documentación y presentación de un proyecto donde se aplique el ciclo de vida de la analítica de datos/ML para resolver un problema industrial específico. Se evaluará la formulación del problema, la preparación de los datos, la selección y aplicación del modelo, la evaluación de resultados y las conclusiones/recomendaciones.

8. **REFERENCIAS BIBLIOGRÁFICAS:** describe los textos guía, manuales, fuentes primarias, páginas de Internet, entre otras, que serán utilizadas para el desarrollo de la Actividad Académica.

- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer. (También existen versiones y ejemplos en Python).
- Murphy, K. P. (2012). Machine Learning: A Probabilistic Perspective. MIT Press. (Más avanzado, para consulta).
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press. (Para introducción a redes neuronales).
- McKinney, W. (2017). Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media. (2nd Edition o más reciente).
- VanderPlas, J. (2016). Python Data Science Handbook: Essential Tools for Working with Data. O'Reilly Media.
- Grus, J. (2019). Data Science from Scratch: First Principles with Python. O'Reilly Media. (2nd Edition).
- Aggarwal, C. C. (2015). Data Mining: The Textbook. Springer.
- Documentación oficial de librerías: Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, TensorFlow, Keras.

- Cursos online y tutoriales de plataformas como Coursera, edX, DataCamp, Kaggle.
- Artículos de conferencias y journals relevantes (ej. IEEE Transactions on Industrial Informatics, Computers in Industry, Journal of Manufacturing Systems)