



1. Datos Generales de la asignatura

| | |
|---------------------------------|--|
| Nombre de la asignatura: | Machine Learning |
| Clave de la asignatura: | SVC-2105 |
| SATCA¹: | 2-2-4 |
| Carrera: | Ingeniería en Sistemas Computacionales |

2. Presentación

Caracterización de la asignatura

La asignatura de Machine Learning se encuentra en el VIII semestre y forma parte de un grupo de asignaturas de la especialidad virtualización y cómputo en la nube. Y se relaciona con las competencias del perfil de egreso: Integrar soluciones computacionales con diferentes tecnologías, plataformas o dispositivos y diseñar, desarrollar y administrar bases de datos conforme a requerimientos definidos, normas organizacionales de manejo y seguridad de la información, utilizando tecnologías emergentes.

La importancia de la asignatura se relaciona con la formación profesional de los alumnos en los conceptos y técnicas básicas para identificar millones de patrones a partir de los datos obtenidos de extensas fuentes de datos o big data para incorporar un nuevo conocimiento que le haga reaccionar en diferentes situaciones futuras y agregar valor a la información obtenida. El machine learning se integra en plataformas IoT avanzadas a partir de sensores para predecir patrones a partir del análisis de datos obtenidos de sensores que miden una gran cantidad de variables para tareas como: la monitorización de la calidad del aire en distintos entornos (smart cities, granjas, industria medioambiental, etc), la monitorización de variables meteorológicas en agricultura de precisión (smart agro), la vigilancia de la salud estructural en infraestructuras (edificios, puentes, plantas industriales, etc). así como la monitorización industrial en fábricas (industria 4.0).

La asignatura comprende conocimientos relacionados con diferentes técnicas, las cuales permiten dotar a los computadores de la capacidad de "aprender" modelos tales que, de forma automática, pueden ser usados, por un lado, para resolver problemas nuevos o, por otro lado, para mejorar el rendimiento en problemas ya vistos. El objetivo principal de esta asignatura es dar una visión introductoria de las técnicas y algoritmos de aprendizaje más importantes existentes en la actualidad.

¹ Sistema de Asignación y Transferencia de Créditos Académicos



Intención didáctica

El temario del curso se ha organizado en seis unidades cada una conformada por contenidos que contribuirán al buen aprendizaje de machine learning.

En la primera unidad se abordan los conceptos, características fundamentales y paradigmas de machine learning, así como el diseño de experimentos y métricas de validación que permiten el ecosistema de desarrollo de soluciones de machine learning.

En la segunda unidad se estudiarán los diferentes algoritmos de agrupamiento (clustering) del aprendizaje no supervisado para ajustar un modelo a las observaciones donde no hay un conocimiento a priori creando un código factorial con componentes estadísticamente independientes.

La unidad tres describe los diferentes métodos y algoritmos del aprendizaje supervisado para generar un modelo predictivo basado en un conjunto de datos de entrada y salida previamente etiquetado y clasificado.

Unidad cuatro se pretende que el ing. en sistemas conozca las aplicaciones principales de las redes neuronales artificiales, sus modelos y arquitecturas, así como las clases de neuronas denominadas perceptrón monocapa y multicapa.

En la quinta unidad se tratarán las generalidades y los diferentes operadores de los algoritmos genéticos para codificar soluciones óptimas a problemas de diversa índole en el campo del aprendizaje automático, así como su implementación a través del teorema de esquemas.

Por último, en la unidad seis se verán las diferentes plataformas que soportan la gestión del ciclo de vida del aprendizaje automático.

3. Participantes en el diseño y seguimiento curricular del programa

| Lugar y fecha de elaboración o revisión | Participantes | Observaciones |
|--|--|---|
| TecNM campus Comalcalco. <u>Julio 2021.</u> | Dr. David Ramírez Peralta Mtra. Claudia Morales Barrón. | Diseño y elaboración de la especialidad en virtualización y cómputo en la nube de la carrera de Ingeniería en Sistemas Computacionales. |

4. Competencia(s) a desarrollar

| Competencia(s) específica(s) de la asignatura |
|---|
| <p>Construye modelos analíticos que permitan a las máquinas aprender de manera automática y mejorar a través de la experiencia para identificar patrones y tomar decisiones con mínima intervención humana.</p> |

5. Competencias previas

| |
|--|
| <p><i>Competencias específicas:</i></p> <ul style="list-style-type: none"> ● Aplica los conceptos de la teoría de la probabilidad y estadística para organizar, clasificar, analizar e interpretar datos para la toma de decisiones en aplicaciones de ingeniería biomédica, en computación y comunicaciones ● Desarrolla soluciones de software para resolver problemas en diversos contextos utilizando programación concurrente, acceso a datos, que soporten interfaz gráfica de usuario y consideren dispositivos móviles. <p><i>Competencias genéricas:</i></p> <ul style="list-style-type: none"> ● Capacidad de análisis y síntesis. ● Capacidad de organizar y planificar. ● Comunicación oral y escrita. ● Habilidad para buscar y analizar información proveniente de fuentes diversas. ● Solución de problemas. ● Toma de decisiones. ● Capacidad crítica y autocrítica. ● Capacidad de trabajar en equipo. ● Capacidad de comunicar sus ideas. |
|--|

6. Temario

| No. | Temas | Subtemas |
|-----|----------------------------------|--|
| 1 | Introducción al Machine Learning | 1.1 Introducción al Machine Learning. 1.2 Paradigmas. 1.3 Diseño de Experimentos y Métricas de Validación. 1.4 Uso de herramientas y librerías de machine learning en Python con Jupyter. |
| 2 | Aprendizaje no supervisado | 2.1 Introducción al Clustering. 2.2 Clustering Jerárquico. 2.3 Clustering Particionante. 2.4 Clustering Basado en Densidad. 2.5 Clustering Probabilista. 2.6 Clustering Basado en Grafos. 2.7 Reducción de la Dimensionalidad. 2.8 Aplicaciones |



| | | |
|---|---------------------------------------|---|
| 3 | Aprendizaje supervisado | 3.1 Evaluación 3.2 Árboles de Decisión y Random Forest. 3.3 Regresión y Regresión Logística. 3.4 Vecinos más cercanos 3.5 Modelos Bayesianos 3.6 Máquinas de Vectores de Soporte |
| 4 | Redes neuronales | 4.1 Modelos 4.2 Arquitecturas. 4.3 Perceptrón unicapa a. Sistemas adaptativos 4.4 Perceptrón multicapa. a. Algoritmo de retropropagación |
| 5 | Algoritmos genéticos | 5.1 Generalidades. 5.2 Representación. 5.3 Selección. 5.4 Cruza. 5.5 Mutación. 5.6 Implementación. 5.7 El Teorema de Esquemas. |
| 6 | Plataformas de aprendizaje automático | 6.1 Plataformas de Aprendizaje no Supervisado. 6.2 Plataformas de Aprendizaje Supervisado. |

7. Actividades de aprendizaje de los temas

| 1. Introducción al Machine Learning (ML) | |
|---|--|
| Competencias | Actividades de aprendizaje |
| <p><i>Específica:</i></p> <ul style="list-style-type: none"> ● Comprende los diferentes paradigmas del machine learning y la importancia de este campo de la inteligencia artificial. <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> ● Capacidad de análisis y síntesis ● Capacidad de organizar y planificar ● Comunicación oral y escrita ● Habilidades intelectuales para el desarrollo de un sistema de aprendizaje automático. ● Habilidad para trabajar de manera autónoma y en equipo. ● Habilidad para buscar y analizar información proveniente de fuentes diversas. ● Solución de problemas ● Toma de decisiones. | <ul style="list-style-type: none"> ● Comunicar y presentar la asignatura: encuadre, empleo de recursos educativos abiertos (REA), entrega de productos y evidencias de aprendizaje en la plataforma, netiqueta y tiempo de acompañamiento docente. <ul style="list-style-type: none"> ▪ Organizar al grupo en equipo para identificar los paradigmas de machine Learning, investigar y presentar los resultados del reporte de investigación en plenaria, subirlo a la plataforma en formato pdf. ▪ Organizar al grupo en equipos para realizar la práctica de diseñar experimentos y métricas de validación en clasificadores y agrupamientos. ▪ Valorar las prácticas y retroalimentar los temas en plenaria por parte del facilitador y estudiantes. |

| 2. Aprendizaje no supervisado | |
|--|---|
| Competencias | Actividades de aprendizaje |
| <p><i>Específica:</i></p> <ul style="list-style-type: none"> • Aplica los diferentes métodos y algoritmos de aprendizaje no supervisado para solucionar problemas de agrupamiento donde no hay un conocimiento a priori. <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita • Habilidad para trabajar de manera autónoma y en equipo. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Solución de problemas | <ul style="list-style-type: none"> • Organizar al grupo en equipos para identificar los diferentes métodos y algoritmos de aprendizaje no supervisado, presentar el informe de investigación al grupo, exponer y discutir. Se evalúa con la rúbrica de reporte de investigación. • Organizar al grupo para resolver problemas de agrupamiento aplicando los diferentes métodos y algoritmos de aprendizaje no supervisado, valorar la práctica. • Retroalimentar los temas en plenaria por parte del facilitador y los participantes de los equipos. |
| 3. Aprendizaje supervisado | |
| Competencias | Actividades de aprendizaje |
| <p><i>Específica:</i></p> <ul style="list-style-type: none"> • Aplica los diferentes métodos y algoritmos de aprendizaje supervisado en la solución de problemas que requieran un análisis predictivo de un conjunto complejo de datos previamente etiquetado y clasificado. <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Comunicación oral y escrita • Habilidad para trabajar de manera autónoma y en equipo. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Solución de problemas • Toma de decisiones. | <ul style="list-style-type: none"> • Organizar al grupo en equipos para identificar los diferentes métodos y algoritmos de aprendizaje supervisado, subir el reporte de investigación en formato pdf, presentar y discutir los temas en plenaria. • Organizar al grupo en equipos para resolver problemas de árboles de decisión, regresión lineal, regresión logística, vecinos, modelos bayesianos en la solución de problemas en el campo del aprendizaje automático. Realizar el reporte de prácticas y presentarlo ante el grupo. • Retroalimentación de los temas vistos en plenaria por parte de los participantes y del facilitador. |
| 4. Redes neuronales | |
| Competencias | Actividades de aprendizaje |
| <p><i>Específica:</i></p> <p>Crea y entrena redes neuronales en la solución de problemas difíciles de resolver mediante técnicas algorítmicas convencionales.</p> <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita | <ul style="list-style-type: none"> • Organizar al grupo en equipos para crear redes neuronales para la solución de problemas difíciles de resolver mediante algoritmos convencionales. Subir el producto de aprendizaje en la plataforma y presentarla en plenaria para su retroalimentación por parte del facilitador y los participantes. • Organizar al grupo en equipos para realizar la práctica de entrenar redes |



| | |
|---|---|
| <ul style="list-style-type: none"> • Habilidades intelectuales para el desarrollo de un sistema de aprendizaje automático. • Habilidad para trabajar de manera autónoma y en equipo. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Solución de problemas • Toma de decisiones. | <p>neuronales para su aplicación en análisis financiero, procesamiento de imágenes, procesamiento de voz, diagnóstico y otras aplicaciones. Subir los productos de aprendizaje a la plataforma, evaluar y retroalimentar a los equipos.</p> <ul style="list-style-type: none"> • Retroalimentar los temas, reflexionar sobre el aprendizaje logrado y los avances del proyecto en las prácticas realizadas en cada unidad. |
|---|---|

5. Algoritmos genéticos

| Competencias | Actividades de aprendizaje |
|---|--|
| <p><i>Específica:</i></p> <ul style="list-style-type: none"> • Aplica algoritmos genéticos en la solución de problemas de búsqueda heurísticas y encontrar soluciones buenas u óptimas a problemas que tienen miles de soluciones <p><i>Genéricas:</i></p> <ul style="list-style-type: none"> • Capacidad de análisis y síntesis • Capacidad de organizar y planificar • Comunicación oral y escrita • Habilidades intelectuales para el desarrollo de un sistema de aprendizaje automático. • Habilidad para trabajar de manera autónoma y en equipo. • Habilidad para buscar y analizar información proveniente de fuentes diversas. • Solución de problemas • Toma de decisiones. | <ul style="list-style-type: none"> • Realizar la práctica en equipos de aplicar algoritmos genéticos en la solución de problemas de búsqueda y optimización basados en evolucionar poblaciones de miles de soluciones hacia valores óptimos del problema. Retroalimentar y evaluar la práctica, motivar a los estudiantes para mejorar su producto de aprendizaje. • Organizar al grupo en equipos para desarrollar el proyecto de aplicar técnicas para evitar un ciberataque. 1. Estudio e investigación: Esta fase es crítica debido a que la formulación sustenta el posterior desarrollo del sistema y sin los conocimientos adecuados no se puede obtener un buen resultado. • Una vez obtenida una buena base teórica del proceso de aprendizaje automático, se dedica a investigar y seleccionar las herramientas necesarias para la realización del proyecto. • 2. Desarrollo del sistema: Una vez consolidados los conocimientos necesarios, obtenidas las herramientas de software y los sets de datos, se procede a la realización del sistema de detección como tal. • En primer lugar se realiza un estudio detallado de los datos y se definen los atributos más importantes para la clasificación de las conexiones. • Posterior al estudio del dataset, se procede a probar diversos algoritmos para observar cuál de ellos da los mejores resultados y optimizarlo para obtener la máxima precisión. |



| | |
|--|---|
| | <ul style="list-style-type: none"> • Por último se diseña el script en python que automatizará el proceso. • 3. Documentación: Después de recopilar toda la información necesaria y diseñar el sistema se procede a documentar todo el proyecto mediante la realización del informe o memoria final del trabajo así como un video en forma de presentación en diapositivas mediante el cual se explicará todo el desarrollo del proyecto y la solución propuesta. • Evaluar el proyecto, brindar retroalimentación a los equipos. Realizar la evaluación sumativa y la autoevaluación del trabajo en equipo. |
|--|---|

8. Práctica(s)

| |
|---|
| <ol style="list-style-type: none"> 1. Análisis y modelado de datos con Pandas 2. Regresión lineal con Python numpy 3. Regresión logística con Python scikit-learn y statsmodels 4. Redes neuronales con Tensor Flow |
|---|

9. Proyecto de asignatura

| |
|---|
| <p>Proyecto: Técnicas en la detección de ciberataques</p> <p>Objetivo: Aplicar técnicas de machine learning en la detección de ciberataques.</p> <ul style="list-style-type: none"> • Fundamentación: La revolución digital está produciendo profundos cambios en los hábitos de consumo de los seres humanos, entre otros motivos, por un mayor acceso a datos y un creciente desarrollo de nuevas tecnologías. De acuerdo con algunas estimaciones, a escala global, todos los días se generan alrededor de 2.5 trillones de bytes de información lo que ha favorecido el desarrollo de la disciplina científica de machine learning que es usada en ámbitos tan diversos como los motores de búsquedas, el diagnóstico médico, el análisis de mercado, la robótica entre otros. El machine learning es una disciplina de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender y generalizar comportamientos a partir de una información suministrada en forma de ejemplos. Concretamente puede ser usada en ciberseguridad para la detección de amenazas avanzadas y presentan la ventaja de poder identificar ataques nunca antes vistos, donde los sistemas tradicionales fallan al basar la detección únicamente en firmas o patrones conocidos. • Planeación: El proyecto se divide en tres fases las cuales son: |
|---|

1. **Estudio e investigación:** Esta fase es crítica debido a que la formulación sustenta el posterior desarrollo del sistema y sin los conocimientos adecuados no se puede obtener un buen resultado.

Una vez obtenida una buena base teórica del proceso de aprendizaje automático, se dedica a investigar y seleccionar las herramientas necesarias para la realización del proyecto.

2. **Desarrollo del sistema:** Una vez consolidados los conocimientos necesarios, obtenidas las herramientas de software y los sets de datos, se procede a la realización del sistema de detección como tal.

En primer lugar se realiza un estudio detallado de los datos y se definen los atributos más importantes para la clasificación de las conexiones.

Posterior al estudio del dataset, se procede a probar diversos algoritmos para observar cuál de ellos da los mejores resultados y optimizarlo para obtener la máxima precisión.

Por último se diseña el script en python que automatizará el proceso.

3. **Documentación:** Después de recopilar toda la información necesaria y diseñar el sistema se procede a documentar todo el proyecto mediante la realización del informe o memoria final del trabajo así como un video en forma de presentación en diapositivas mediante el cual se explicará todo el desarrollo del proyecto y la solución propuesta.

- **Ejecución:**

1. Definición del problema que se ha de resolver
2. Adquisición de los datos de aprendizaje y de las pruebas
3. Preparar y limpiar los datos
4. Analizar, explorar los datos
5. Elegir un modelo de aprendizaje
6. Ver los resultados y ajustar o modificar el modelo de aprendizaje

- **Evaluación:**

Se evalúa el proyecto con un instrumento de evaluación que permite que los estudiantes mejoren su propio desempeño, se realiza también una autoevaluación del desempeño en equipo para valorar las competencias genéricas de la asignatura.

10. Evaluación por competencias

| <i>Evaluación formativa</i> | |
|-------------------------------------|---|
| Producto o Evidencia de aprendizaje | Instrumento de evaluación |
| Reporte de investigación | Indicadores de evaluación de un reporte de investigación: <ul style="list-style-type: none"> – Tiene carátula con datos de la institución, nombre del alumno, materia, nombre del autor, título del libro y capítulo, nombre del docente, grupo y fecha. – El documento contiene índice en el que se han incluido todos los subtemas. |

| | |
|--|---|
| | <ul style="list-style-type: none"> - El contenido se encuentra en arial 12, justificados y espacio y medio; los subtítulos en arial 14, centrados y a espacio y medio. - El documento contiene introducción, apartados y conclusión. - El documento emplea citas bibliográficas y referencias ambas emplean las normativas de APA. - El documento se subió a la plataforma en tiene tiempo y formato pdf. |
| Presentación grupal de los temas investigados. | <p>Rúbrica de presentación de temas relacionados a los temas investigados, categorías:</p> <ul style="list-style-type: none"> -Se identifican claramente las ideas primarias de las ideas secundarias. -Se observa la preparación del tema, el uso de referencias empleadas y un orden de ideas. -Se observa seguridad al tratar el tema, buen uso del recurso de apoyo, fluidez de ideas, tono de voz adecuado. -Brinda retroalimentación a sus compañeros. -Explica con claridad los procedimientos y técnicas aplicados en la práctica. |
| Prácticas utilizando los conocimientos aprendidos en los temas de la asignatura. | <p>Rúbrica de prácticas.</p> <ul style="list-style-type: none"> -Analiza, reconoce e interpreta perfectamente los datos, identificando con certeza lo que se busca y demostrando una comprensión del problema. -Utiliza estrategias heurísticas efectivas y eficientes desarrollando los algoritmos matemáticos para la solución del problema. -Detalla los pasos seguidos, relacionando en grado óptimo los conceptos matemáticos necesarios. -Realiza los procedimientos correctos en la solución de la problemática. -Aporta la solución del problema, analiza y discute sobre su fiabilidad, si hay errores revisa el proceso y procede a modificarlo. |
| Autoevaluación del equipo: | <p>Preguntas de reflexión: ¿qué problemáticas se identificaron? ¿Qué necesitamos aprender?, ¿qué temas debo estudiar de manera individual para mejorar mi desempeño en el equipo? ¿qué decisiones tuvo que tomar el equipo para resolver el problema?, ¿en qué áreas de aprendizaje relacionadas con los resultados me considero experto?.</p> |
| Proyecto: Aplicación de técnicas para evitar un ciberataque | <ul style="list-style-type: none"> ▪ Lista de cotejo, indicadores de logro para el desarrollo del proyecto: <ul style="list-style-type: none"> ✓ Definición del problema que se ha de resolver |

- ✓ Adquisición de los datos de aprendizaje y de las pruebas
- ✓ Preparar y limpiar los datos
- ✓ Analizar, explorar los datos
- ✓ Elegir un modelo de aprendizaje
- ✓ Ver los resultados y ajustar o modificar el modelo de aprendizaje

Evaluación sumativa:

1. Participación en clases
2. Evidencias de productos y desempeños en la plataforma
3. Portafolio de aprendizaje personal
4. Autoevaluación

11. Fuentes de información

Referencias:

Bagnato Juan I. (2020). Aprende Machine Learning (Teoría+Práctica Python).
Gerón, A. (2019). Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly
Gerón, A. (2020). Aprende Machine Learning con Scikit-Learn, Keras y Tensorflow. Anaya multimedia
Harrison, M (2019). Machine Learning Pocket. Working with Structured Data in Python. O'REILLY
Mckinney,W. (2017) Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly.
Mirjalili V., y Raschka S. (2019). Python Machine Learning. Marcombo.
Müller, A.C y Guido, S. (2016). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly.

Recursos educativos abiertos:

Aplicación de técnicas de Machine Learning a dispositivos IoT para el control de aforose
Automatización de diagnósticos mediante el uso de técnicas de Machine Learning, disponible en: <https://idus.us.es/handle/11441/114873>
Desarrollo y aplicación de técnicas de Machine Learning para la predicción de contagios por Covid-19, disponible en: <https://idus.us.es/handle/11441/114871>
Diseño de un sistema de gestión de inventarios enfocado en el uso de herramientas de machine learning, disponible en: <http://manqlar.uninorte.edu.co/handle/10584/9537>
Estudio de QSVM, algoritmo de Machine Learning Cuántico, disponible en: <https://uvadoc.uva.es/handle/10324/47344>
<https://repositorio.unican.es/xmlui/handle/10902/21856>
Librerías de Python para Machine Learning, disponible en: <https://www.iartificial.net/librerias-de-python-para-machine-learning/>
Machine Learning para la mejora de la experiencia con MOOC: el caso de la Universitat Politècnica de València, disponible en: <https://revistas.um.es/riite/article/view/466251>
Pandas Python, disponible en: <https://pandas.pydata.org/>
Pronóstico de la demanda del biodiesel mediante la aplicación de técnicas de machine learning, disponible en: <https://repository.libertadores.edu.co/handle/11371/4090>
Scikit-Learn, disponible en: <https://scikit-learn.org/stable/>



Sport-utility vehicle prediction based on machine learning approach, disponible en:
<https://jart.icat.unam.mx/index.php/jart/article/view/1683>

Statsmodels, disponible en: <https://www.statsmodels.org/stable/index.html>

Tecnología de Videojuegos y Machine Learning desde cero:
<http://rua.ua.es/dspace/handle/10045/115983>

Tensor Flow disponible en: <https://www.tensorflow.org/?hl=es-419>

Weka, disponible en: <https://www.cs.waikato.ac.nz/ml/weka/>