
PROGRAMA DE ESTUDIOS: CURSO DE MINERÍA DE DATOS

KOICA

 **HANDONG GLOBAL**
UNIVERSITY



Minería de Datos

Información Básica

Información del curso

1. Título: **Minería de Datos**
2. Año lectivo: **2024**
3. Semestre: **Primero**
4. Departamento: **Centro de Innovación TIC (FP-UNA)**
5. Año/Nivel: **Intermedio**
6. Formato de clase: **Charlas y sesiones prácticas**

Hora y lugar

1. Días: **Martes**
2. Hora: **19:00 a 21:00**
3. Ubicación: **Sesiones síncronas vía plataforma Educa**

Información del instructor

1. Nombre: **Ellen Lujan Méndez Xavier**
2. Oficina: **Plataforma Educa**
3. Contacto: emendez@pol.una.py
4. Perfil profesional:



Ingeniera en Informática, egresada de la Facultad Politécnica de la Universidad Nacional de Asunción. Magister en Tecnologías de la Información y Comunicación con énfasis en Ingeniería de Software. Actualmente es Docente del departamento de enseñanza informática de la FP-UNA. En el sector público, soy Directora del departamento de enseñanza informática en la FP-UNA. En el sector privado, me desempeño como gerente general de una empresa consultora especializada en software. También realizo consultorías al exterior. En el área de la docencia me desenvuelvo como docente en diversas universidades privadas.

Notas adicionales

1. Se anima a los estudiantes a comunicarse con el instructor a través de los datos de contacto proporcionados para cualquier consulta o inquietud relacionada con el curso.
2. Es importante asistir tanto a las sesiones teóricas como a las prácticas para beneficiarse plenamente de la estructura del curso.

Prerrequisito

Programación en Python.

Descripción del Curso

El curso "Minería de datos: técnicas, aplicaciones y ética" está meticulosamente diseñado para satisfacer las necesidades cambiantes de la industria y el mundo académico en la era del big data. Presenta una exploración integral de conceptos, metodologías y herramientas de minería de datos, centrándose en extraer información significativa de vastos conjuntos de datos. El plan de estudios cierra la brecha entre el conocimiento teórico y la aplicación práctica, preparando a los estudiantes para abordar desafíos complejos de análisis de datos en varios sectores. Al integrar principios de la informática, la estadística y el aprendizaje automático, el curso ofrece un enfoque multidisciplinario que enriquece la experiencia de aprendizaje y proporciona a los estudiantes un versátil conjunto de habilidades.

Un elemento central de la filosofía del curso es el énfasis en el aprendizaje práctico. A través de sesiones de laboratorio dedicadas y un extenso trabajo de proyectos, los estudiantes se involucrarán directamente con problemas de minería de datos del mundo real utilizando herramientas de software avanzadas como scikit-learn de Python y la suite de minería de datos Weka. Este enfoque práctico garantiza que los estudiantes no sólo comprendan los fundamentos teóricos de los algoritmos de minería de datos, sino que también adquieran competencia en la aplicación de estas técnicas para resolver problemas prácticos. El componente de proyecto del curso anima a los estudiantes a trabajar con conjuntos de datos reales, navegando por todo el proceso de extracción de datos, desde el preprocesamiento hasta el análisis y la interpretación de los resultados.

Al reconocer la importancia de las aplicaciones de minería de datos en el mundo actual, el plan de estudios del curso incluye una amplia gama de estudios de caso y proyectos extraídos de diversos dominios, como finanzas, atención médica, análisis de redes sociales y más. Estas aplicaciones del mundo real subrayan la relevancia de las técnicas de minería de datos para resolver los desafíos contemporáneos y permiten a los estudiantes explorar el potencial del análisis de datos en diferentes industrias. La inclusión de tendencias y desafíos actuales garantiza que los estudiantes no solo aprendan técnicas de minería de datos sino que también comprendan sus implicaciones y aplicaciones en el mundo moderno.

Las consideraciones éticas en la minería de datos forman una parte crucial del contenido del curso. A medida que la privacidad de los datos, la seguridad y el uso ético de la información se vuelven cada vez más importantes, el curso aborda estas cuestiones de frente. Se anima a los estudiantes a considerar las implicaciones éticas de las prácticas de minería de datos, incluida la responsabilidad de los científicos de datos de garantizar la privacidad y la equidad. Este segmento del curso fomenta un sentido de responsabilidad ética entre los estudiantes, preparándolos para tomar decisiones informadas en su vida profesional que respeten la privacidad individual y la seguridad de los datos.

Por último, el curso "Minería de Datos" está estructurado para dotar a los estudiantes de una comprensión profunda de los principios de la minería de datos y la capacidad de aplicar estas técnicas de manera efectiva y ética. Está diseñado para personas que buscan sobresalir en el campo del análisis de datos y ofrece información sobre las dimensiones técnicas y éticas de la minería de datos. Al finalizar, los estudiantes se convertirán en analistas de datos competentes, listos para aportar conocimientos valiosos en los campos elegidos y navegar por las complejidades del mundo basado en datos con integridad y experiencia.

Características clave:

- **Enfoque interdisciplinario:** el curso integra conceptos de informática, estadística y aprendizaje automático, brindando una visión holística de la minería de datos. Este enfoque interdisciplinario garantiza que los estudiantes puedan abordar los desafíos del análisis de datos desde múltiples perspectivas.
- **Experiencia práctica:** a través de sesiones de laboratorio y proyectos, los estudiantes utilizarán herramientas y software de minería de datos de última generación, como scikit-learn de Python y la suite de minería de datos Weka. Esta experiencia práctica es vital para comprender los matices de los algoritmos de minería de datos y sus aplicaciones prácticas.
- **Aplicaciones del mundo real:** el plan de estudios incluye estudios de casos y proyectos que imitan los desafíos de la minería de datos del mundo real, lo que permite a los estudiantes aplicar sus conocimientos para resolver problemas prácticos. Este enfoque en la aplicación ayuda a cerrar la brecha entre los conceptos teóricos y su utilidad en entornos profesionales.
- **Consideraciones éticas:** con las crecientes preocupaciones en torno a la privacidad y la seguridad de los datos, el curso dedica un segmento a las consideraciones éticas en la minería de datos. Los estudiantes explorarán el equilibrio entre la utilización de datos y la privacidad, aprendiendo a navegar por las implicaciones morales de las prácticas de minería de datos.
- **Contenido de vanguardia:** el contenido del curso se actualiza periódicamente para incluir las últimas tendencias, herramientas y técnicas en minería de datos. Esto garantiza que los estudiantes aprendan las metodologías más actuales y estén preparados para el panorama cambiante del análisis de datos.

Objetivo del Curso

Al completar con éxito este curso, los estudiantes podrán:

1. Comprender los conceptos y metodologías clave de la minería de datos.
 2. Pre-procesar conjuntos de datos para prepararlos para la minería.
 3. Aplicar técnicas de agrupación, clasificación y extracción de reglas de asociación a conjuntos de datos del mundo real.
 4. Evaluar la efectividad de diferentes algoritmos de minería de datos.
 5. Utilizar herramientas y software de minería de datos para el análisis de datos.
 6. Abordar las preocupaciones éticas y de privacidad en la extracción de datos.
- **Dominar los conceptos y técnicas fundamentales:** los estudiantes obtendrán una comprensión sólida de los principios básicos de la minería de datos, incluido el preprocesamiento de datos, la minería de patrones, la clasificación, la agrupación y la detección de anomalías. Este objetivo pretende dotar a los estudiantes de los conocimientos necesarios para identificar, analizar y resolver problemas de minería de datos.
 - **Aplicar algoritmos avanzados de minería de datos:** más allá de las técnicas fundamentales, el curso profundizará en algoritmos avanzados y sus aplicaciones en diversos dominios. Los estudiantes aprenderán sobre estrategias sofisticadas de minería de datos, como métodos de conjunto, aprendizaje profundo en minería de datos y análisis de series de tiempo. Esto preparará a los estudiantes para abordar tareas complejas de análisis de datos en sus futuras carreras.

- **Desarrollar competencia práctica con herramientas de minería de datos:** un objetivo clave es garantizar que los estudiantes dominen el uso de software de minería de datos y lenguajes de programación de última generación, principalmente Python y sus bibliotecas (por ejemplo: scikit-learn, pandas) y el kit de herramientas de minería de datos Weka. A través de sesiones de laboratorio y proyectos, los estudiantes aplicarán estas herramientas a conjuntos de datos reales, obteniendo una valiosa experiencia práctica.
- **Analizar aplicaciones de minería de datos del mundo real:** los estudiantes explorarán cómo se aplican las técnicas de minería de datos en diversos sectores, como la atención médica, las finanzas, las redes sociales y el comercio electrónico. Esto incluye comprender cómo formular problemas de minería de datos, seleccionar algoritmos apropiados e interpretar los resultados. El objetivo es preparar a los estudiantes para aplicar soluciones de minería de datos a desafíos de dominios específicos.
- **Evaluar consideraciones éticas y de privacidad:** reconociendo la creciente importancia de las cuestiones éticas en la minería de datos, un objetivo principal es sensibilizar a los estudiantes sobre las preocupaciones éticas, legales y de privacidad asociadas con las prácticas de minería de datos. Esto incluye debates sobre la recopilación de datos, el consentimiento, la anonimización de los datos y el uso ético de los modelos predictivos. Los estudiantes aprenderán a afrontar estos desafíos y a tomar decisiones responsables en sus proyectos de análisis de datos.
- **Pensamiento crítico y resolución de problemas:** el curso tiene como objetivo fomentar el pensamiento crítico y las habilidades de resolución de problemas. Los estudiantes aprenderán a evaluar críticamente el desempeño de los algoritmos de minería de datos, comprenderán sus limitaciones y abordarán creativamente los desafíos encontrados en el análisis de datos del mundo real.
- **Habilidades colaborativas y gestión de proyectos:** a través de proyectos y colaboraciones grupales, los estudiantes mejorarán su capacidad para trabajar eficazmente en equipos, gestionar proyectos y comunicar conceptos y hallazgos complejos de minería de datos a una audiencia no técnica. Este objetivo pretende preparar a los estudiantes para entornos de trabajo colaborativo y mejorar sus habilidades de gestión de proyectos.

Estos objetivos están cuidadosamente elaborados para garantizar que, al finalizar el curso, los estudiantes no sólo sean técnicamente competentes en minería de datos, sino que también sean capaces de aplicar sus habilidades de manera ética y efectiva en una variedad de entornos. Estarán bien preparados para carreras en ciencia de datos, análisis y campos relacionados, con una base sólida tanto en la ciencia como en el arte de la minería de datos.

Política de Calificación

Calificación absoluta:

La calificación del curso se estructura de la siguiente manera para evaluar la comprensión, el compromiso y la aplicación práctica del material del curso por parte de los estudiantes:

1. Tareas y Pruebas: 40% de la nota final. Este componente integral evalúa la comprensión de los estudiantes del material del curso a través de tareas y pruebas regulares. Estas tareas están diseñadas para reforzar el aprendizaje y garantizar un compromiso continuo con el contenido del curso.
2. Examen Parcial: 25% de la nota final. El examen de mitad de período evalúa la comprensión de los estudiantes de los conceptos y principios clave cubiertos en la primera mitad del curso.

3. Examen Final: 35% de la nota final. El examen final se centra en todo el contenido del curso, poniendo a prueba la comprensión general de los estudiantes y su capacidad para integrar diferentes temas aprendidos a lo largo del curso.

Esta política de calificación está diseñada para evaluar de manera justa el desempeño de los estudiantes en diferentes aspectos del curso, fomentando el esfuerzo constante, la participación activa y una comprensión profunda de la materia.

Libros de Texto y Otros Materiales

Se proporcionarán materiales del curso.

Libro de referencia:

- López, J. M. M., & Herrero, J. G. (2006). Técnicas de análisis de datos. Aplicaciones Prácticas utilizando Microsoft Excel y WEKA.

Tareas y Exámenes

Para un curso integral como "Minería de datos", las tareas, exámenes y proyectos desempeñan un papel crucial en la evaluación de la comprensión y aplicación del material del curso por parte de los estudiantes. A continuación se presenta un enfoque detallado para estructurar estos componentes:

Tareas

Objetivo: Reforzar los conceptos enseñados en conferencias y brindar experiencia práctica con herramientas y técnicas de minería de datos.

Formato:

- Tareas quincenales: Consisten en cuestiones teóricas y ejercicios prácticos.
- Preguntas teóricas: Cubren conceptos clave, algoritmos y consideraciones éticas en la minería de datos.
- Ejercicios prácticos: Involucran preprocesamiento de datos, aplicación de algoritmos usando Python o Weka y análisis de resultados.
- Entrega: Las tareas deben enviarse electrónicamente a través del portal en línea del curso.
- Feedback: Se proporcionarán comentarios detallados, destacando las áreas de fortaleza y mejora.

Exámenes

Objetivo: Evaluar la comprensión de los conceptos centrales, las metodologías y las implicaciones éticas de la minería de datos.

Formato:

- Examen parcial: cubre todo el material impartido en la primera mitad del curso. Incluirá preguntas de opción múltiple, preguntas de respuesta corta y preguntas de resolución de problemas basadas en estudios de caso.
- Examen final: comprensivo, es decir, que cubre todo el contenido del curso. Seguirá un

formato similar al examen parcial, pero incluirá preguntas más complejas sobre resolución de problemas y análisis de casos.

- Preparación: Se anima a los estudiantes a revisar apuntes de clases, tareas y materiales de lectura. Se proporcionarán exámenes de práctica para el estudio.

Proyecto

Objetivo: Brindar a los estudiantes la oportunidad de aplicar técnicas de minería de datos a un problema del mundo real, fomentando habilidades en la gestión de proyectos, el trabajo en equipo y la comunicación.

Formato:

- Proyecto grupal: Los estudiantes formarán pequeños equipos para trabajar en un proyecto de minería de datos. Cada grupo seleccionará un proyecto de una lista proporcionada por el instructor o propondrá su propio proyecto sujeto a aprobación.
- Componentes del proyecto: Incluye propuesta de proyecto, informe provisional, informe final y presentación.
- Propuesta: Breve documento que describe el problema elegido, las fuentes de datos y los métodos propuestos.
- Informe provisorio: Actualización a mitad del proyecto que detalla el progreso, los resultados preliminares y cualquier ajuste al plan inicial.
- Informe final: Documento completo que incluye definición del problema, metodología, resultados, discusión y reflexiones sobre consideraciones éticas.
- Presentación: Cada grupo presentará sus hallazgos a la clase, destacando las ideas clave, los desafíos enfrentados y las implicaciones éticas.
- Criterios de evaluación: Los proyectos se evaluarán en función de la originalidad, la complejidad del análisis, la aplicación de técnicas de extracción de datos, la calidad de los hallazgos, la consideración de cuestiones éticas y la eficacia de la comunicación.

Estos componentes están diseñados para garantizar una evaluación integral del conocimiento teórico, las habilidades prácticas y la capacidad de los estudiantes para aplicar consideraciones éticas en escenarios de minería de datos del mundo real. La combinación de tareas, exámenes y un proyecto integral permite a los estudiantes demostrar su competencia de diversas maneras, preparándolos para carreras en ciencia y análisis de datos.

Actividad del Curso

Para garantizar una experiencia de aprendizaje integral en "Minería de datos: técnicas, aplicaciones y ética", el curso incorpora una variedad de actividades diseñadas para involucrar a los estudiantes tanto en la comprensión teórica como en la aplicación práctica. A continuación se presenta un desglose de las actividades clave del curso y cómo contribuyen a los objetivos de aprendizaje:

Clases

- Formato de contenido: durante las clases se presentan conceptos teóricos, algoritmos y consideraciones éticas en la minería de datos. Las conferencias pueden incluir oradores invitados de la industria para discutir aplicaciones y desafíos del mundo real.
- Componentes interactivos: para mejorar la participación, las clases incorporan sesiones de preguntas y respuestas, encuestas en tiempo real y debates sobre estudios de casos para fomentar el pensamiento crítico y la aplicación de conceptos.

Foros de Discusión

- Plataforma en línea: se utiliza un sistema de gestión de cursos o un foro en línea para facilitar las discusiones entre estudiantes e instructores. Esta plataforma permite compartir recursos, discutir tareas y colaborar en ideas de proyectos.

Proyectos Grupales

- Aprendizaje colaborativo: los proyectos están diseñados para mejorar las habilidades colaborativas, lo que requiere que los estudiantes trabajen en equipos para aplicar técnicas de minería de datos para resolver un problema del mundo real. Esto incluye recopilación de datos, preprocesamiento, análisis y consideraciones éticas.
- Presentaciones de hitos: a lo largo del proyecto, los equipos presentan su progreso, reciben comentarios y perfeccionan su enfoque. Este proceso iterativo culmina con una presentación final a la clase, mostrando sus hallazgos y el proceso de aprendizaje.

Estudios de Caso

- Aplicaciones del mundo real: los estudios de casos seleccionados destacan la aplicación de la minería de datos en diversos ámbitos, como la atención sanitaria, el marketing y el análisis de redes sociales. Los estudiantes analizan estos casos, discuten metodologías y evalúan implicaciones éticas.
- Discusiones grupales: los estudios de caso a menudo se discuten en grupos pequeños o como parte de discusiones en clase, lo que permite a los estudiantes explorar diferentes perspectivas y soluciones a los problemas presentados.

Pruebas

- Verificaciones de conocimiento: Periódicamente se administran pruebas breves y sin calificación para ayudar a los estudiantes a evaluar su comprensión de conceptos clave. Estos cuestionarios brindan retroalimentación inmediata y son una herramienta valiosa para la autoevaluación.

Estas actividades están estructuradas para garantizar un entorno de aprendizaje dinámico e interactivo, que atiende a diferentes estilos y preferencias de aprendizaje. Al participar en estas diversas actividades, los estudiantes obtienen una educación integral en minería de datos, preparándolos para carreras exitosas en este campo en rápida evolución.

Cronograma del Curso

Semana	Clase	Tipo de clases	Materiales
1	Introducción a la Minería de Datos Proceso KDD Preprocesamiento de datos	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
2	Minería por reglas de asociación <ul style="list-style-type: none"> - Algoritmo A Priori - Algoritmo FP-Growth 	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
3	Clasificación <ul style="list-style-type: none"> - Árboles de decisión. - Naive Bayesiano - Random Forest 	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
4	Examen Parcial / Proyecto	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
5	Agrupación <ul style="list-style-type: none"> - k-Means - Hierarchical Clustering - Cobweb 	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
6	Detección de anomalías <ul style="list-style-type: none"> -k-Vecinos Más Cercanos -DBSCAN (Density-Based Spatial Clustering of Applications with Noise) - Otros 	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
7	Temas avanzados <ul style="list-style-type: none"> - Minería de texto - Ética 	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea
8	Examen Final / Proyecto	2 horas de clases virtuales síncronas y 3 horas de prácticas independientes del alumno	<ul style="list-style-type: none"> • Presentaciones • Aula virtual • Banco de ejercicios • Foros en línea

Contenidos del Curso

- Introducción a la Minería de Datos
- Proceso KDD
- Preprocesamiento de datos
 - Limpieza de los datos
 - Integración de datos
 - Transformaciones
- Minería por reglas de asociación
 - Algoritmo A priori
- Clasificación
 - Árboles de decisión.
 - Naive Bayesiano
 - Random Forest
- Agrupación
 - k-Means
 - Hierarchical Clustering
 - Cobweb
- Detección de anomalías
 - k-Vecinos Más Cercanos
 - DBSCAN (Density-Based Spatial Clustering of Applications with Noise)
 - Otros algoritmos
 - Consideraciones para elección de algoritmo
- Temas avanzados
 - Minería de Datos Textuales
 - Ética de los datos
- Proyectos de la vida real.