

Minería de Datos y Machine Learning

Módulo I: Fundamentos y Preparación de Datos

software descripción



En los últimos años, los rápidos avances en las tecnologías de recopilación y almacenamiento de datos han dado lugar a conjuntos de datos que son "grandes" en muchos sentidos de la palabra. **La Minería de Datos** está relacionada con el análisis de estos grandes conjuntos de datos con la finalidad de proporcionar ideas, patrones, modelos descriptivos y predictivos que permitan extraer y generar conocimiento para las organizaciones. Algunas de las tareas más comunes son la clasificación, agrupamiento, descubrimiento de reglas de asociación y patrones de respuestas, detección de anomalías, etc.

Para poder realizar estas tareas, la Minería de Datos combina los aspectos teóricos fundamentales de áreas tan diversas como la estadística, las matemáticas, la ingeniería y las ciencias de la información con muchas aplicaciones prácticas y relevantes de la vida real. Dentro de esas disciplinas destaca el aprendizaje automatizado o **machine learning**, que está teniendo profundos efectos en muchas industrias diferentes, desde servicios financieros hasta ventas minoristas y publicidad; convirtiéndose rápidamente en una herramienta fundamental para tomar mejores decisiones en los negocios: decisiones basadas en datos, no en instintos o conjeturas.

Este módulo del curso tiene como objetivo introducir a los participantes en los fundamentos de la minería de datos y el aprendizaje automatizado, así como presentar diversas técnicas de exploración, visualización y preparación de datos previos al modelamiento.

publico objetivo

Analistas de datos. Profesionales relacionados a la inteligencia de negocios, investigación de mercados e interesados en el área de Ciencia de Datos. Académicos e investigadores. . Público en general que requiera extraer conocimiento desde diferentes fuentes de información.

logros de aprendizaje

Al finalizar este curso, el participante conocerá los fundamentos de la Minería de Datos y el aprendizaje automatizado, aplicará las principales técnicas, así como analizará e interpretará los resultados obtenidos. De manera específica el participante estará en capacidad de:

- Reconocer los conceptos fundamentales asociados a Minería de Datos, entendiendo cuáles son las diferentes fuentes de información a utilizar.
- Entender las 4 fases fundamentales del aprendizaje automatizado.
- Comprender el Proceso de Descubrimiento de Conocimiento en Bases de Datos(KDD).
- Realizar un pre procesamiento de datos, con la finalidad de posteriormente entender y aplicar las diferentes técnicas de extracción de conocimiento de datos.
- Entender los grandes desafíos del análisis de bases de datos a nivel del Preprocesamiento necesario.
- Aplicar e interpretar adecuadamente las principales técnicas de Minería de Datos usando R y RStudio.
- Comprender el desarrollo de los algoritmos de las principales técnicas de Minería de Datos
- Presentar de manera efectiva los resultados obtenidos.

contacto



+51970227123



info@perustat.com



<http://www.perustat.com>



fb://perustat



contenidos

UNIDAD I: INTRODUCCIÓN

- Tema 1 Conceptos Básicos** (3 horas)
- Breve historia de la analítica de datos.
 - Definiciones básicas: Estadística. Ciencia de Datos. Minería de Datos. Aprendizaje automatizado. Relación con otras disciplinas
 - Las 4 fases fundamentales del aprendizaje automatizado:
 - Entendimiento de datos: ¿Qué nos están diciendo?
 - Predicción: ¿Que sucederá?
 - Toma de decisiones: ¿Qué hacer?
 - Inferencia causal: ¿Funcionará?
 - Taxonomía de las técnicas de Minería de Datos: Tipos de modelos. Tipos de aprendizaje. Técnicas no supervisadas y supervisadas. Aplicaciones.
 - Fases de la Minería de Datos. Descubrimiento de Conocimiento en Bases de datos (KDD).
 - CRISP-DM: Estructura Básica. Fases.
 - Herramientas de Minería de Datos.
- Tema 2 Entendimiento de datos** (5 horas)
- El proceso de Entendimiento de datos.
 - Tipos de datos.
 - Datos de registro: Atributos y escala de Medición.
 - Exploración y Visualización de datos.
 - Resumen de datos: medidas estadísticas de tendencia, posición (cuantiles), variabilidad y asimetría.
 - Gráficos univariados y bivariados

UNIDAD II: PREPROCESAMIENTO DE DATOS

- Tema 3 Limpieza y transformación de datos** (16 horas)
- Importancia y calidad de datos.
 - Principales tareas.
 - Datos perdidos: Tipos. Detección. Imputación.
 - Outliers (valores atípicos): Outliers univariados. Outliers multivariados. Outliers basados en densidad local. Otros métodos.
 - Transformación de datos: Puntajes Z:
- Tema 4 Reducción de información** (8 horas)
- Discretización.
 - Reducción de la dimensionalidad y extracción de nuevas características: Análisis de Componentes Principales.

contacto



+51970227123



info@perustat.com



<http://www.perustat.com>



fb://perustat



metodología

La metodología del curso se basa en una combinación de clases teóricas y análisis de casos prácticos en la computadora, con la finalidad de que el participante comprenda la metodología, la motivación, los supuestos, las fortalezas y las debilidades de los métodos tratados en el curso. Cada sección del curso está motivada por un conjunto de datos en particular, de tal forma que el participante gane experiencia trabajando con una amplia variedad de fuentes de datos similares a los que usa en la realidad. Los contenidos están estructurados en 8 sesiones con un total de 32 horas académicas (24 horas cronológicas).

evaluación y asistencia

- **Certificado de asistencia:** Para obtener este certificado debe de mantener un porcentaje mínimo de 75% de asistencia a las clases.
- **Certificado de aprobación:** El 100% de la calificación final se obtiene sobre la base de cuestionarios o listas de ejercicios al final de cada sesión. Para recibir el certificado de aprobación el participantes deben obtener al menos un 70% de los puntos posibles y contar con el porcentaje mínimo de asistencia a clases.

Serán otorgados certificados a nombre de PeruStat Analytics S.A.C. que acredita a los participantes del curso. La certificación que se otorga es excluyente.

materiales

Material preparado por el equipo de capacitación con los contenidos del curso el cuál será entregado a los participantes en medios físicos o digitales.

referencias

- Bishop, C. (2011). Pattern Recognition and Machine Learning. Springer.
- Clarke, B., Fokoue, E. y Zhang, H. (2009). Principles and Theory for Data Mining and Machine Learning. Springer Verlag.
- Gareth, J., Witten, D., Hastie, T. y Tibshirani, R. (2013). An Introduction to Statistical Learning: with Applications in R. Springer.
- Géron, A. (2017). Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media.
- Hastie, T., Tibshirani, R. y Friedman, J. (2016). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer.
- Kuhn, M. y Johnson, K. (2013). Applied Predictive Modeling. Springer Verlag.
- Lantz, B. (2015). Machine Learning with R: Expert techniques for predictive modeling to solve all your data analysis problems Packt Publishing; 2 edition .
- Loo, M. V. D, y Jonge, E. (2018). Statistical Data Cleaning with Applications in R. New York: John Wiley & Sons, Incorporated.
- Nisbet, R., Elder IV, J. y Miner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. Academic Press.
- Putler, D. y Krider, R. E. (2012). Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R. Chapman and Hall/CRC.
- Shmuelil, G., Bruce, P., Yahav, I., Patel, N. y Lichtendahl, K. (2017). Data Mining for Business Analytics: Concepts, Techniques, and Applications in R. John Wiley & Sons.
- Véliz, C. (2018). Aprendizaje Automático. Análisis para la Minería de Datos y Big Data. Departamento Académico de Ciencias. Pontificia Universidad Católica del Perú.
- Wu J. y Coggeshall, S. (2012). Foundations of Predictive Anytics. Chapman and Hall/CRC.

contacto



+51970227123



info@perustat.com



<http://www.perustat.com>



<fb://perustat>

