

Análisis exploratorio de datos con RapidMiner

Sílabo

contacto

+51 (1) 282 9524
+51 9 9038 8434

info@perustat.com
<http://www.perustat.com>
fb://perustat



fundamentación

Durante las últimas décadas se ha producido un desarrollo explosivo en las tecnologías de base de datos y la cantidad de datos que es recolectada. Esto ha creado una oportunidad sin precedentes para la Minería de Datos mediante el proceso de descubrimiento, ya sea supervisado o no, de información interesante y útil a partir de repositorios de datos disponibles.

La Minería de Datos está relacionada con el análisis, mayormente estadístico, de 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. Dentro de los distintos programas disponibles, RapidMiner proporciona una poderosa plataforma open source para la aplicación de las técnicas más importantes de análisis de datos, por lo que es actualmente uno de los programas más utilizados.

Este taller tiene como finalidad aplicar las principales técnicas de minería de datos relacionadas con la preparación y el análisis exploratorio de datos usando RapidMiner.

publico objetivo

Analistas e investigadores de mercado. Profesionales de marketing. Público en general que requiera aplicar técnicas analíticas descriptivas.

nivel

Intermedio

conocimientos previos

Conocimientos básicos de estadística descriptiva (recomendable)

logros de aprendizaje

Al finalizar este taller, el participante aplicará las principales técnicas de minería de datos relacionadas con el análisis exploratorio y la preparación de datos usando el programa RapidMiner.

De manera específica el participante estará en capacidad de:

- Interactuar con el ambiente de trabajo de RapidMiner.
- Acceder y preparar datos almacenados en diversos formatos.
- Realizar un análisis exploratorio de datos.
- Realizar representaciones gráficas.
- Aplicar técnicas de limpieza de datos.
- Interpretar los resultados obtenidos luego de realizar un análisis descriptivo.
- Presentar de manera efectiva los resultados obtenidos.

contenidos

- Sesión 1** **Introducción y** manejo de datos con RapidMiner
- Fases de la Minería de Datos. Descubrimiento de Conocimiento en Bases de datos (KDD).
 - Interfaz gráfica de usuario de RapidMiner
 - Terminología básica: Repositorio, atributos y ejemplos, operadores y procesos.
 - Herramientas de importación y exportación de datos.
 - Lectura de archivos de texto: delimitadores alternativos y archivos.
 - Lectura de bases de datos con muchos atributos.
- Sesión 2** **Exploración de** datos
- Resumen de datos: medidas estadísticas
 - Visualización de datos: gráficos univariados y bivariados
 - Detección de valores atípicos (outliers).
 - Detect Outlier (Distances)
 - Detect Outlier (Densities)
 - Detect Outlier (LOF)
 - Detect Outlier (COF)
- Sesión 3** **Limpieza de** datos
- Herramientas de muestreo (selección de instancias).
 - Manejo de datos perdidos.
 - Tipos de datos perdidos.
 - Detección.
 - Imputación.
 - Herramientas de transformación de datos
 - Transformación de atributos.
 - Discretización.
 - Agregación: Tablas dinámicas (pivot tables)
 - Reducción de la Dimensionalidad: Análisis de Componentes Principales.
- Sesión 4** **Aplicaciones**
- Caso de estudio.
 - Discusión final.

metodología

La metodología del taller es completamente práctica. 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 del taller están estructurados en 4 sesiones con un total de 8 horas académicas.

evaluación y asistencia

Se otorgará un certificado a nombre de PeruStat Analytics S.A.C. que acredite la participación en el taller.

materiales

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

referencias

- Chisholm, A. (2013). Exploring Data with RapidMiner. Packt Publishing Ltd.
- Hofmann, M. y Klinkenberg, R. (2013). RapidMiner: Data Mining Use Cases and Business Analytics Applications . Chapman & Hall.
- Kotu, V. y Deshpande, B. (2014). Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner. Morgan Kaufmann.
- Larose, D.T. (2006). Data Mining Methods and Models. Wiley Interscience.
- Nisbet, R., Elder IV, J. y Miner, G. (2009). Handbook of Statistical Analysis and Data Mining Applications. Academic Press.
- Wu J. y Coggeshall, S. (2012). Foundations of Predictive Analytics. Chapman and Hall/CRC.