

UNIVERSIDAD ARGENTINA DE LA EMPRESA

Departamento de Tecnología Informática

3.4.096 – INTELIGENCIA ARTIFICIAL

Docentes: Sciolla, Pablo
Ponzoni, Nelson

Trabajo Práctico I – Machine Learning

Machine Learning es parte de la Inteligencia Artificial, la cual utiliza técnicas estadísticas y algoritmos computacionales para inferir patrones y características a partir de un conjunto de datos. Las áreas de aplicación del Machine Learning son muy variadas, salud, detección de fraude, sistemas de recomendación, finanzas, etc.

Objetivo

- Aplicar diferentes arquitecturas de redes neuronales a la clasificación automática de datos reales.
- Implementar algoritmos de entrenamiento para redes neuronales. Evaluar el desempeño.
- Profundizar en los conceptos teóricos relacionados con la retropropagación del error.
- Utilizar diferentes técnicas de validación de modelos y resultados. Valorizar la importancia.

¿Quién no requiere más nuestros servicios?

El presente caso corresponde a una empresa proveedora de servicios de internet para pequeños consumidores. Inicio sus actividades en el 2015 y operan en Argentina, Brasil y Chile. La empresa está en constante crecimiento aumentando su alcance geográfico, calidad de servicio y satisfacción de usuario. En el último período reportó ingresos por más de 200 millones de u\$s según sus asientos contables y estima para el año 2024 triplicar sus ingresos con un plan de extenderse a más países de la región, principalmente Bolivia y Paraguay.

Al inicio del FY21, el contexto impuso limitaciones de trabajo en las oficinas, por lo que se ha incrementado fuertemente la utilización del servicio en pequeños hogares. Esto trajo consigo nuevos clientes, aumentando la demanda significativamente. Luego de unos meses, el área de marketing ha notado que muchos de sus clientes, nuevos o no, dejaron de utilizar el servicio de internet lo que llamó la atención de la junta directiva de la empresa. Iniciaron medidas mitigantes de retención de clientes sin resultados positivos.

Desde el área de Marketing, requieren identificar el causal por el cual no están pudiendo retener clientes. Debido a ello, se le proporciona un set de datos que se debe analizar con las altas de los clientes del último año. Contás con algunos datos característicos y otros que fueron removidos para cumplir con las normativas regionales de protección de datos personales.

El objetivo es generar un modelo predictivo que permita la detección de abandonadores. El modelo, es decir el criterio por el cual un cliente no requiere más de los servicios, debe poder explicarse a organismos reguladores.

VARIABLE	DESCRIPCIÓN
ID	Número único identificador del cliente
NOMBRE	Nombre del cliente
EMAIL	Email del cliente
PAÍS	País de operaciones
GÉNERO	Género del cliente (Femenino, Masculino)
EDAD	Edad
ANTIGÜEDAD	Antigüedad del servicio
FACTURACIONTOTAL	Facturación total realizada al cliente
PUNTIACIÓNDECITICIA	Scoring crediticio del cliente
CANTIDADPRODUCTOS	Cantidad de productos contratados
POSEETARJETA	Tiene registrada una tarjeta de crédito (si, no)
MIEMBROACTIVO	Es miembro activo de la página web – más de un ingreso por semana (si, no)
SALARIOESTIMADO	Salario estimado
VELOCIDADSERVICIO	Velocidad del último servicio de internet contratado
ABANDONO	Estado del cliente (abandono: si, no)
RECORDS := 9.000	Cantidad total de registros etiquetados
COLUMNS := 15	Cantidad total de columnas

Tareas

El equipo debe realizar las siguientes actividades.

1. Identificar el tipo/clase de problema.
2. Identificar las variables y sus características. Describirlas.
3. Realizar la tarea de preprocesamiento en caso de que fuese necesario. Limpieza de variables, normalización, categorización, remuestreo, binarización, categorización, etc.
4. Aplicar al menos cuatro (4) modelos de Machine Learning distintos para el problema identificado.
5. Formalizar una técnica de selección de los datos.
6. Evaluar cada modelo según los resultados obtenidos.
7. Comentar las ventajas y limitación de los modelos elegidos.
8. Presentar los resultados obtenidos.