

ANALÍTICA PREDICTIVA DE DATOS

pARTE 2

Prof. José Antonio Taquía Gutiérrez



SEMESTRE 2 - 2018

**Temas:**

[Redes Neuronales],[Support Vector Machine],[Análisis de conglomerados],[Análisis de componentes principales],[Árboles de clasificación],[Aprendizaje supervisado],[Aprendizaje no supervisado],[Redes Neuronales].

Análisis de conglomerados

1. Un importador y distribuidor mayorista desea evaluar la manera de atender a sus respectivos clientes buscando enfocarse especialmente en aquellos que son los que más venta le generan. En el archivo (Segmentacion\_Clientes\_Distribuidoras.xls), se presenta información de las ventas trimestrales de sus principales clientes. Se le solicita un análisis de conglomerados para identificar la cantidad de segmentos que se pueden obtener considerando particionar el dataset por tipo de clientes (Sectorial y Corporativo).
   1. Identifique para cada tipo de clientes si existe un comportamiento normalizado de su variable Ventas\_ML
2. Una empresa que comercializa diversos modelos de autos desea encontrar la manera de agrupar las diversas marcas que tiene para aplicar una campaña de marketing segmentada tomando en cuenta los diversos atributos que podrían ser comunes entre estos modelos. Se le solicita utilizando el dataset ***autos.xls*** :
   1. Identificar los principales atributos para clasificarlos
   2. Proponer una clasificación
3. Una empresa vitivinícola ha desarrollado una labor bastante cuidadosa de registrar las métricas de diversos cultivos que desea comercializar en nuevos mercados fuera de la región siendo bastante competitivos en calidad y precios. Los responsables de estos viñedos cuentan con datos obtenidos en diversas etapas del proceso de cultivo y elaboración.
   1. Se solicita con el dataset **Data\_Vinos\_clasificacion.xls** determinar si es posible encontrar diferencias o variables que influyen más en la clasificación de los vinos y si es posible determinar cluster en los registros entregados.
   2. Se pide presentar gráficamente como se dispersan los grupos identificados.
4. Una empresa de distribución por delivery, desea analizar si existe un patrón en la cantidad de pedidos que distribuye en ciertas zonas de la capital. Para este fin reúne datos del comportamiento diario en las zonas establecidas para su reparto. Si el responsable de operaciones desea analizar aquella zona que es la más relevante en su carga operativa de manera mensual. Se le pide utilizando el dataset ***Distribución Delivery Zonal.xls***, construir la gráfica del reparto semanal por zona que ayude a visualizar si existe un comportamiento para clasificar las zonas de distribución (sugerencia utilice los parámetros de tendencia o estacionalidad en la serie de tiempo). Explique su respuesta

Árbol de clasificación

1. El foco del negocio de la empresa Taxi\_Kal, consiste en el servicio de transporte de pasajeros de diferentes operadores turísticos trasladándolos de diferentes hoteles hacia el aereopuerto. La rentabilidad de la operación de Taxi\_KAL, se sostiene en mantener el flujo constante de servicios y un ciclo de vida de la unidad de transporte que no les lleve a gastos en reparaciones o pérdidas de clientes. En los últimos meses se han presentado estas fallas lo que ha generado malestar en algunos de los hoteles importantes para el área comercial de la empresa por el volumen de turistas que contratan sus servicios. Se le solicita con la data Taxi\_Kal.xls., desarrollar un modelo predictivo de árbol de clasificación, que le permita identificar si la unidad asignada para un traslado tendrá algún tipo de problema durante el servicio.

a) Explique ¿Por qué la gerencia de Taxi\_KAL, debería confiar en su análisis? Justifique su respuesta.

b) Presente 2 reglas de negocio que se pueden obtener de la data considerando una cantidad mínima de 20 casos para que la regla sea válida.

c) Presente una gráfica mostrando la precisión de su modelo considerando 10, 15 y 20 elementos por nodo como cantidad mínima para analizar sus datos

1. La empresa Distribuidora Syntex SAC, tiene presencia en diversos mercados regionales y cuenta con reconocimientos por su gestión de talento. Participa de programas de sobre madurez organizacional y uno de sus grandes logros es retener a sus empleados debido a una cultura de innovación y permanente interacción con proyectos que beneficien al negocio. En la tabla de datos ***Contratacion.csv***, se muestra la historia de los últimos meses con los resultados del proceso de selección de personal de la empresa. Con esta información los directivos responsables de este proceso desean identificar si existe alguna lógica que determine el perfil que tenga mayor posibilidad de ser contratado para orientar hacia ese segmento su publicidad y atraerlos a la empresa. Con esta información se le solicita
2. Identifique que técnica predictiva le sería de utilidad para responder a la interrogante de los responsables del proceso de selección. Justifique su respuesta.
3. Presente su solución considerando la técnica elegida.
4. Presente la matriz de enfrentamiento de su predicción y lo ocurrido en la realidad.

Redes Neuronales

1. El mercado inmobiliario ha tenido diversos cambios debido a la coyuntura macro económica y las expectativas de los compradores de reducciones de precios ante nuevos competidores en el rubro de servicios hipotecarios. En el dataset housing[hoja1].xls, se muestran diversas variables que influyen en el valor comercial de una propiedad para diversos distritos de la capital. Con los datos entregados se le pide:
   1. Muestre gráficamente que variable está más relacionada con el precio de la vivienda.(Sugerencia: Utilice Neural-tools).
   2. Construya una red neuronal que le permita predecir el precio de una vivienda. Considere el modelo de entrenamiento de la red neuronal utilizando una partición de 20%. Utilice el archivo Housing\_Pgta2b.xls
   3. Si se le añade al dataset un atributo adicional, con el archivo Housing\_Pgta2c.xls denominada resultado, que muestra si la propiedad se vendió a es ese precio. Se le solicita hacer los cambios en su red neuronal para predecir si la propiedad se venderá o no. Se le solicita responder si considera que este es el método más adecuado a utilizar. Justifique su respuesta
2. Para la predicción de la edad de una especie de caracol marino, se utilizan mediciones físicas de este molusco. La edad se puede determinar con el caparazón a través del cono, aplicando tinte y contando el número de anillos en un microscopio. La edad se obtiene añadiendo 1.5 al número de anillos. Otra forma es el uso de una red neuronal para predecir la edad a partir de las medidas que se pueden obtener más fácilmente. Por supuesto, la variable Anillos no se debe utilizar en la red neuronal. Diseñe una red neuronal con el método GRN para predecir la edad del molusco. Usando la información del dataset en la hoja “entrenamiento” del archivo ***moluscos.xls***, se le pide:
   1. Obtener el modelo de entrenamiento de la red neuronal utilizando una partición de 30%.
   2. Presente las métricas de la red neuronal y compárelas con las obtenidas con un modelo de regresión multivariable.
   3. Si se le añade al dataset moluscos.xls una variable adicional, como se muestra en la hoja “Variable\_Adicional”, ¿se podría obtener una red neuronal?, ¿qué cambio debería hacer en su diseño?

Ciclo de vida de un producto

1. La empresa Inox S.A está introduciendo al mercado diversos productos electrodomésticos. El mercado retail de electro ha crecido sostenidamente en los últimos cinco años. El Gerente Comercial tiene como referencia una curva de saturación del mercado y espera que las ventas de la nueva línea de refrigeradoras de Inox tengan ese mismo comportamiento. Si las ventas de Inox para esta línea especial de refrigeradora se presentan en la hoja “Inox\_Refrig” del archivo Inox\_Vtas.xls. Se le solicita:
   1. Identifique en cuántos años llegaría a tener un escenario de madurez en el mercado dicha línea de refrigeradoras.
   2. Explique cómo interpreta los valores de M, p y q para el contexto comercial de Inox\_Refrig.
   3. ¿En qué año se lograría el pico máximo de crecimiento de la línea de refrigeradoras?.
2. La empresa de tecnología Xenit, importa equipos de impresión para un mercado regional que comprende varios países en latino-américa. Entre las marcas que trae están sólo las líderes por su prestigio y capacidad de innovación. Entre la información que solicita el gerente general de la empresa está la predicción de la rapidez que el producto ingresará al mercado (con periodicidad mensual, Rango C15:C17, de la hoja “Plantilla”). Si se conoce que la inversión en publicidad de la empresa, acelera crecimiento de las unidades colocadas en el mercado se le pide sugerencia sobre cuáles de las cinco alternativas de marcas a ingresar en su portafolio de impresoras multifuncionales debería elegir Xenit, considerando la función unidades en el mercado por dólares invertidos mostrada en el archivo ***Gestion\_Producto.xls*** y los parámetros de las respectivas curvas de ciclo de vida. Considerar que la empresa en esta categoría tiene una inversión presupuestada de 180,000 dólares para los próximos doce meses. Se le pide responder:
   1. Tomando en cuenta la información sobre el modelo Bass y la inversión publicitaria requerida para insertar unidades en el mercado. ¿Cuál sería su sugerencia de las dos marcas a seleccionar para ingresar al mercado el próximo año? Justifique su respuesta.
   2. Presente una utilidad aproximada factible a obtener con su selección de marcas.