

BIG DATA for BUSINESS

2.9 Modelos analíticos. Problemas propuestos

Conecta Empleo

Contenido desarrollado por
Synergic Partners



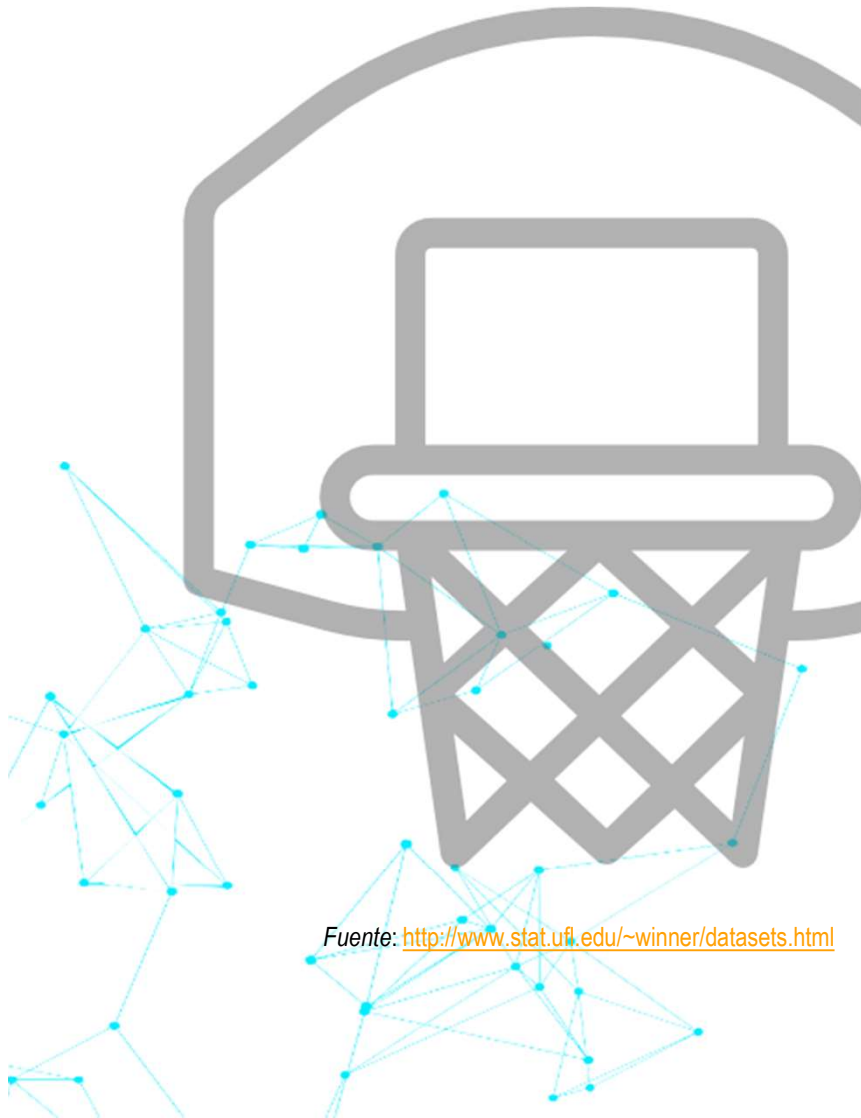


Regresión

Problema I: Puntaje NBA

MODELOS ANALÍTICOS

Problema I: Puntaje NBA



Fuente: <http://www.stat.ufl.edu/~winner/datasets.html>

Descripción

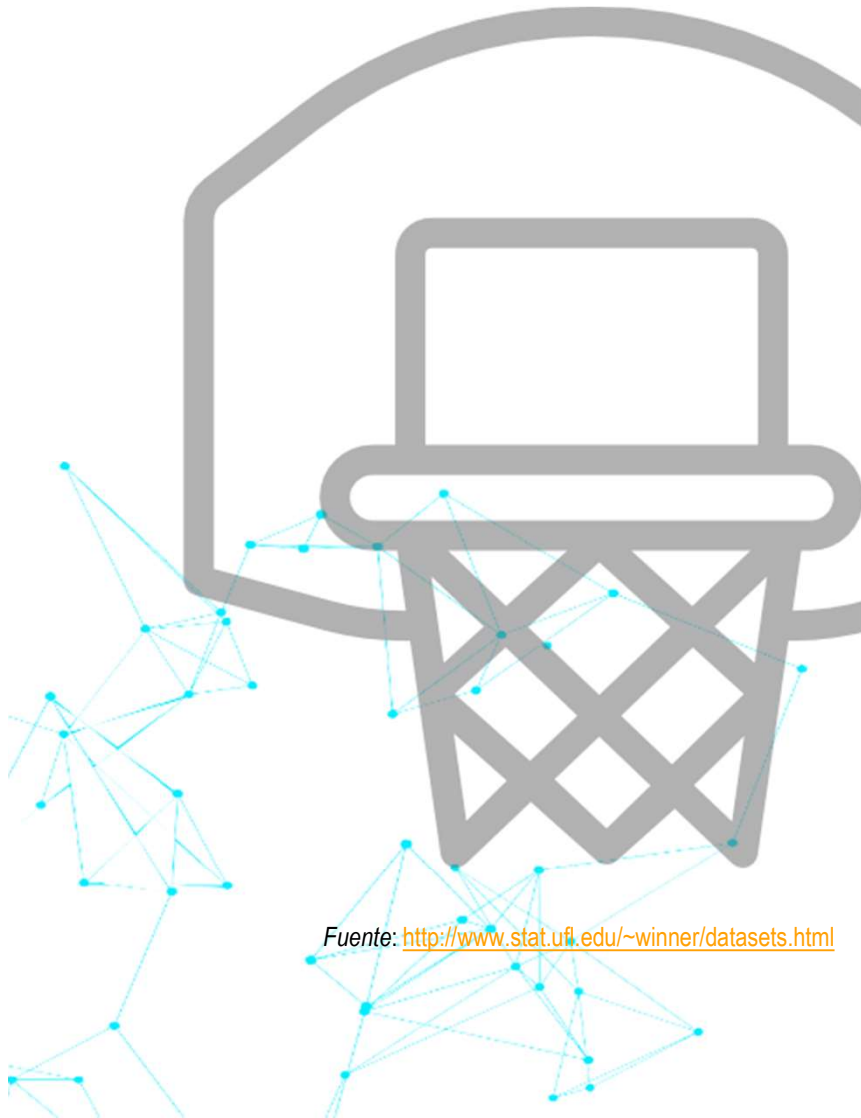
A partir de datos sobre el puntaje en diferentes encuentros Baloncesto de la NBA, nuestro objetivo es predecir el marcador final (Puntos Totales).

URL:

<https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/NBA/nbaodds201415.csv>

MODELOS ANALÍTICOS

Problema I: Puntaje NBA



Fuente: <http://www.stat.ufl.edu/~winner/datasets.html>

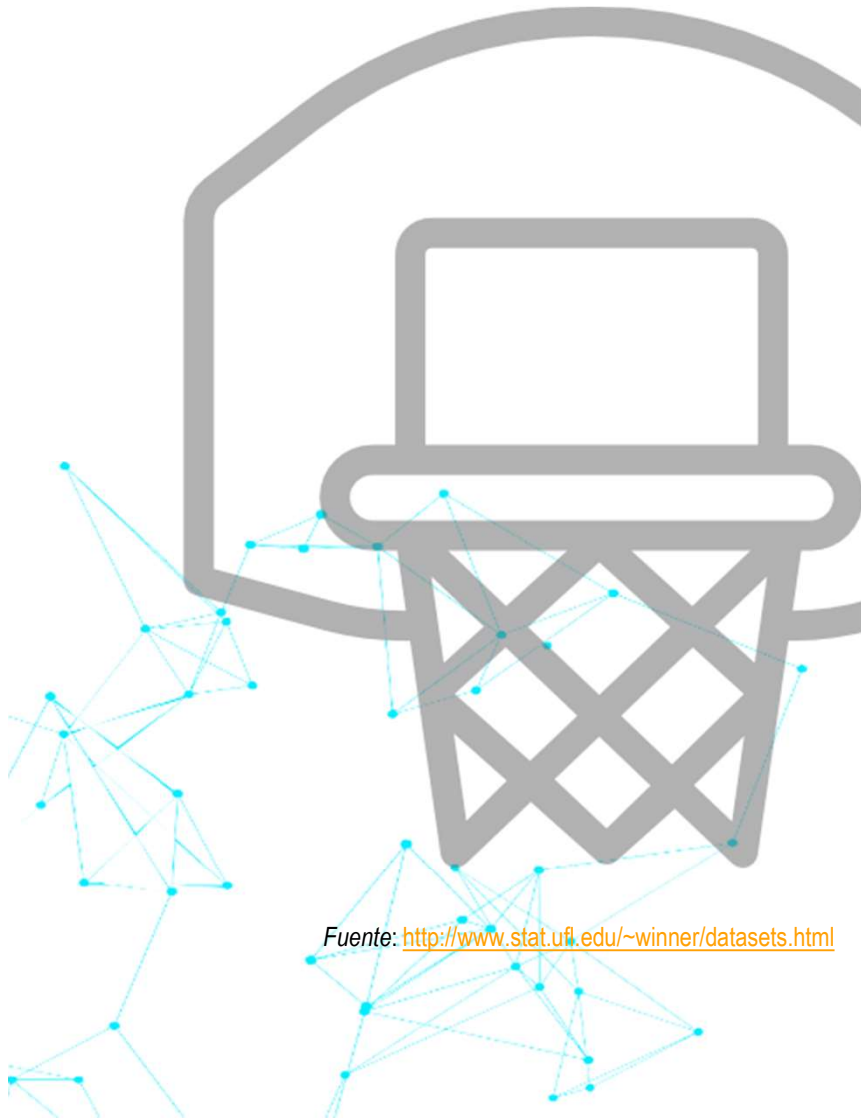
Variables

En primer lugar es importante conocer las variables, para estudiar cuáles de ellas son útiles para nuestro modelo, cuales son redundantes y cuáles no aportan información.

Este dataset contiene 12 variables numéricas, 2 de tipo texto, 1 tipo fecha y 2 de tipo booleano.

MODELOS ANALÍTICOS

Problema I: Puntaje NBA



Fuente: <http://www.stat.ufl.edu/~winner/datasets.html>

Variables

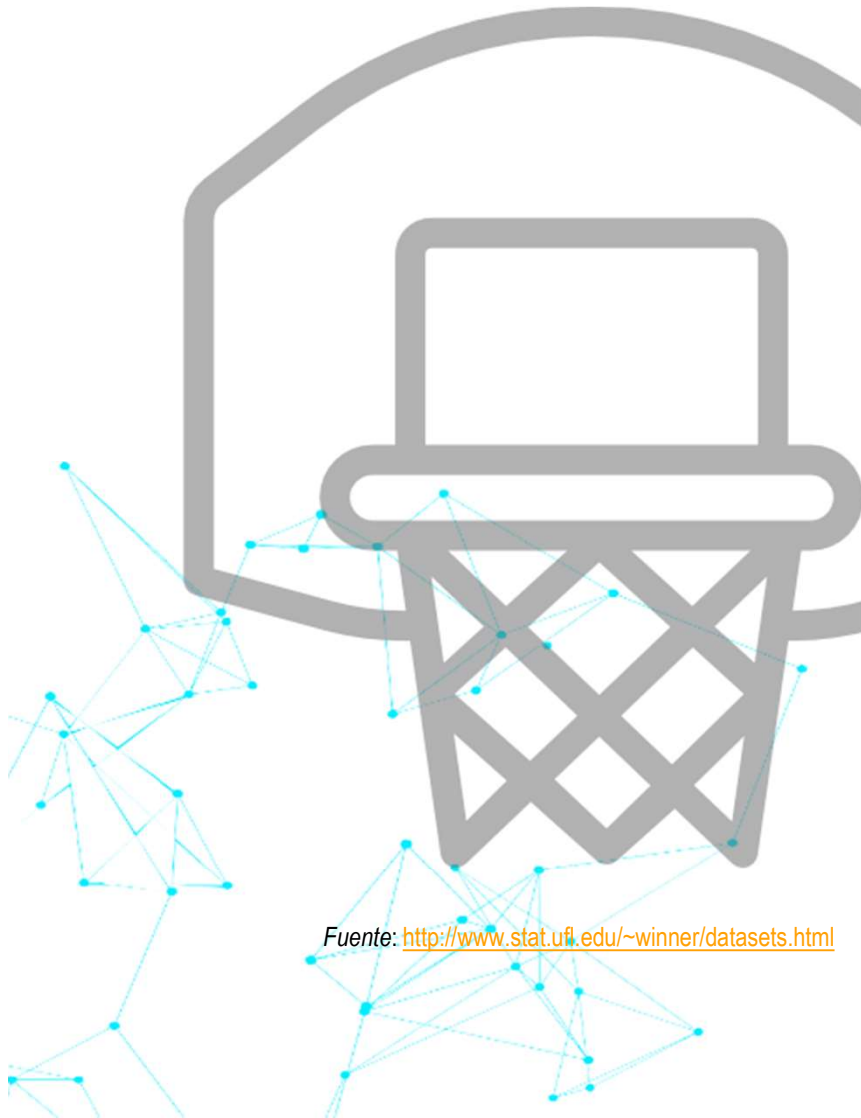
El conjunto de variables contiene la fecha del evento, el equipo local y visitante (tanto nombre como id), los puntos marcados por ambos equipos, el equipo ganador, la diferencia de puntos entre ambos y el total.

Hay una serie de variables que necesitan conocer cómo funciona el reglamento de la NBA para entenderlas. Cuando un equipo se considera mejor que otro, se da un valor mínimo de puntos de diferencia por los que debe ganar al otro equipo, esto se denomina **spread**. En este dataset se da el *spread* dado por la casa de apuestas Las Vegas (variables **TeamSprd**, **TeamCov**), así como la predicción del total de puntos de partido dado por la misma (variable **OvrUndr**).

La descripción completa de las variables [aquí](#).

MODELOS ANALÍTICOS

Problema I: Puntaje NBA



Fuente: <http://www.stat.ufl.edu/~winner/datasets.html>

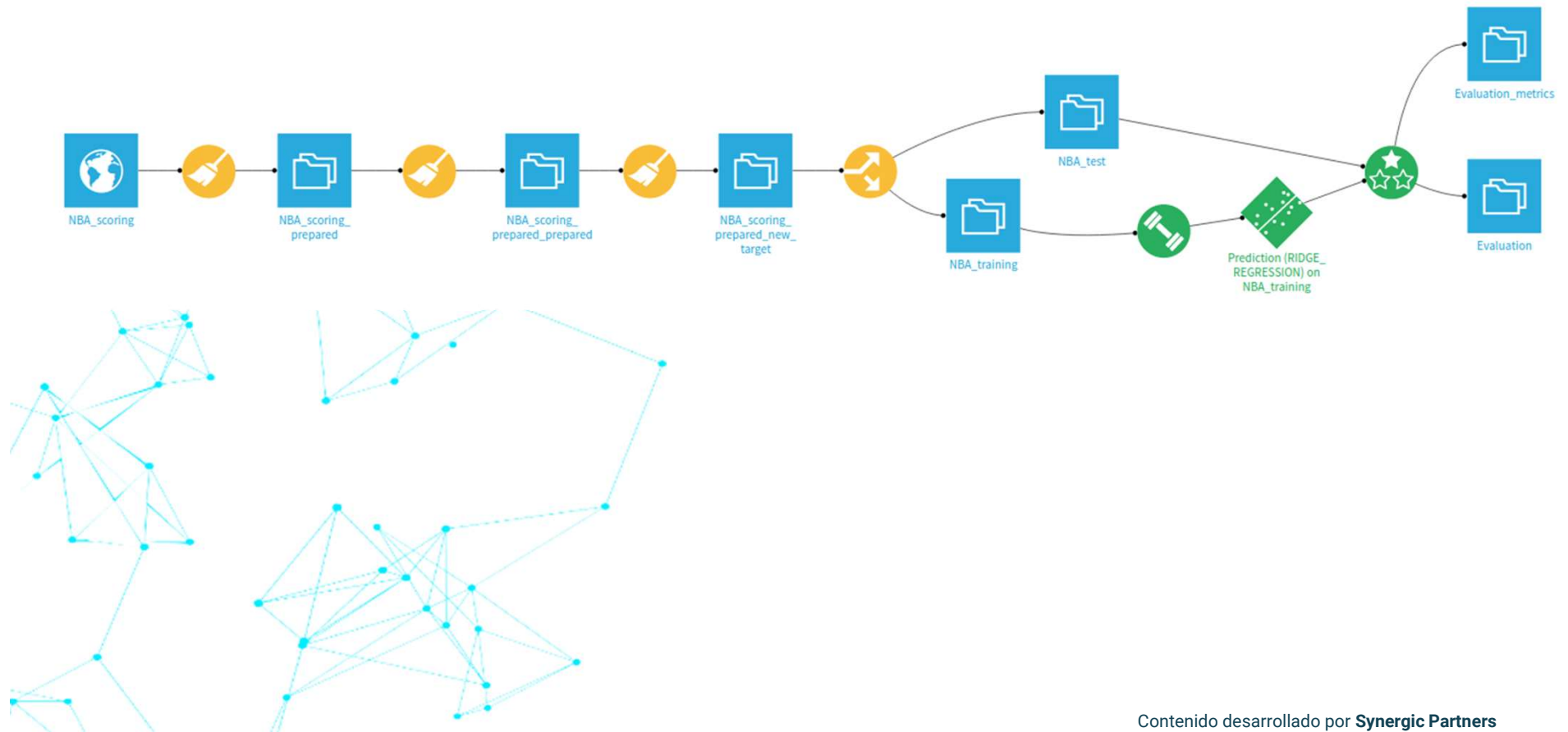
Tareas a realizar

1. Carga de los datos
2. Calidad y limpieza de los datos
3. Análisis exploratorio gráfico y descriptivo
4. Modelo predictivo de regresión
5. Evaluación del modelo
6. Predicción/Scoring

MODELOS ANALÍTICOS

Problema I: Puntaje NBA

Workflow final





Clasificación

Problema II: Vinos Italianos

MODELOS ANALÍTICOS

Problema II: Vinos Italianos



Descripción del problema

Se tienen tres variedades de vino cultivadas en una región de Italia y 13 variables que miden las características químicas de los vinos producidos.

Se pide conocer, para nuevos cultivos y en base a estas características, cuál sería el origen del vino.

MODELOS ANALÍTICOS

Problema II: Vinos Italianos



Datasets

1. **data.csv:** Contiene el valor de los componentes químicos que describen el vino así como el grupo de clasificación al que corresponde, esto es, 1, 2 ó 3.

URL

<https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/wine/wine.data.csv>

MODELOS ANALÍTICOS

Problema II: Vinos Italianos



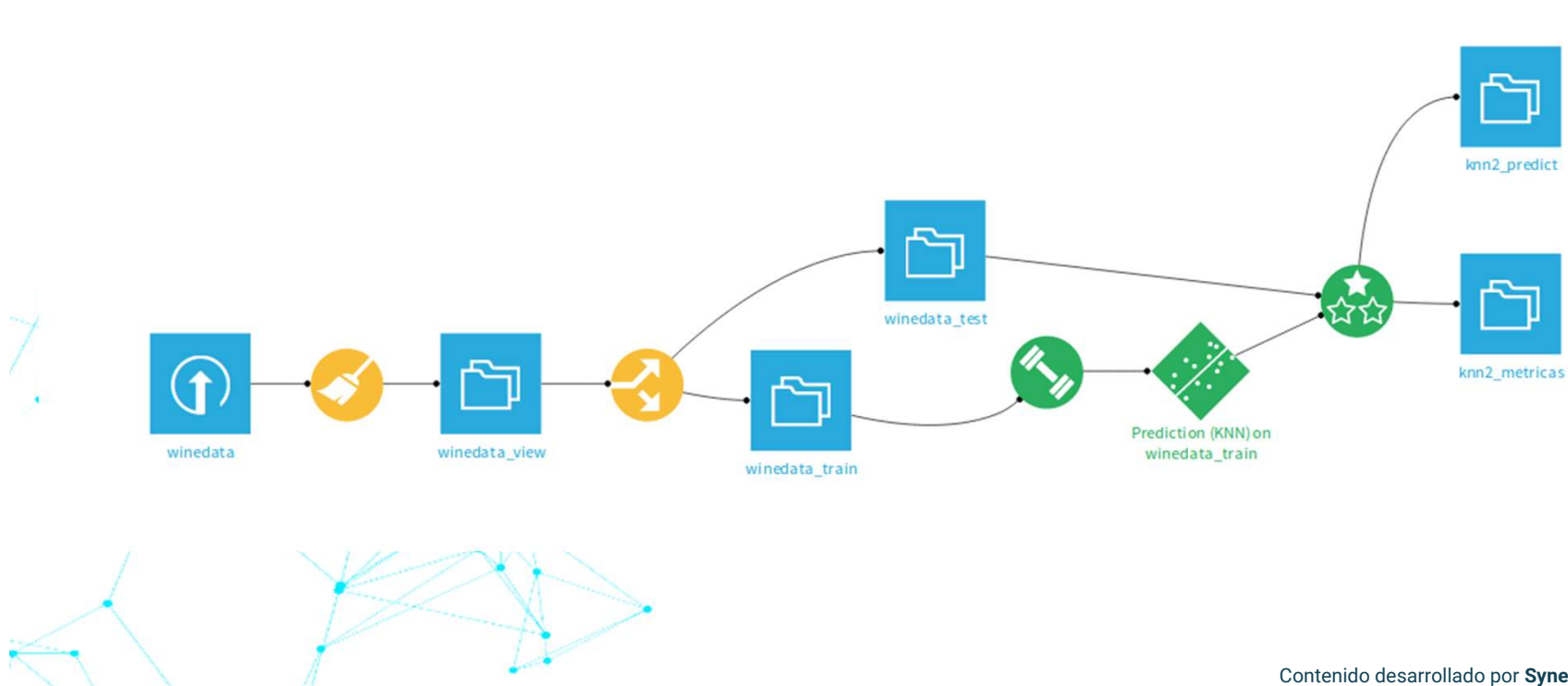
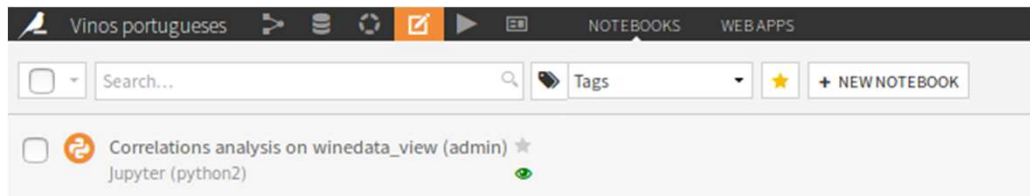
Tareas a realizar

1. Importar datos.
2. Analizar si hay valores nulos o missing y ver la distribución de las variables
3. Analizar si hay variables altamente correlacionadas y eliminarlas si las hubiera
4. Separar los datos en datos de entrenamiento y test
5. Entrenar y analizar el modelo usando los casos conocidos (datos de entrenamiento)
6. Aplicar el modelo a los casos desconocidos (test)
7. Desplegar el modelo
8. Métricas de validación del modelo seleccionado

MODELOS ANALÍTICOS

Problema II: Vinos Italianos

Workflow final



An abstract network diagram consisting of numerous teal-colored nodes (dots) connected by thin teal lines. The nodes are scattered across the left side of the slide, with some forming small, dense clusters and others being isolated or part of larger, more complex structures. The lines represent connections between these nodes, creating a web-like pattern.

Clustering

Problema III: Segmentación y descripción de clientes

MODELOS ANALÍTICOS

Problema III: Segmentación y Descripción de Clientes



Descripción

El cliente es el equipo directivo de un gran centro comercial, quieren entender los tipos de clientes que visitan su centro comercial. Tienen buenas razones para creer que existen algunos segmentos diferentes y están considerando diseñar y posicionar mejor los servicios en el centro comercial para atraer a los segmentos rentables o para diferenciar sus servicios (por ejemplo, invitaciones a eventos, descuentos , Etc).

El dataset cuenta las actitudes de las personas a las compras en un centro comercial.

MODELOS ANALÍTICOS

Problema III: Segmentación y Descripción de Clientes



Datasets

1. **cust_data.csv**: Se trata de una encuesta de investigación de mercado a unos pocos clientes potenciales. En ella cada persona respondió a seis preguntas de actitud y una pregunta sobre la frecuencia con que visitan el centro comercial, todos en una escala 1-7, así como una pregunta con respecto a sus ingresos familiares y la cantidad de visitas al centro comercial.

URL

https://s3-eu-west-1.amazonaws.com/academy-formation/datasets/customers/cust_data.csv

MODELOS ANALÍTICOS

Problema III: Segmentación y Descripción de Clientes



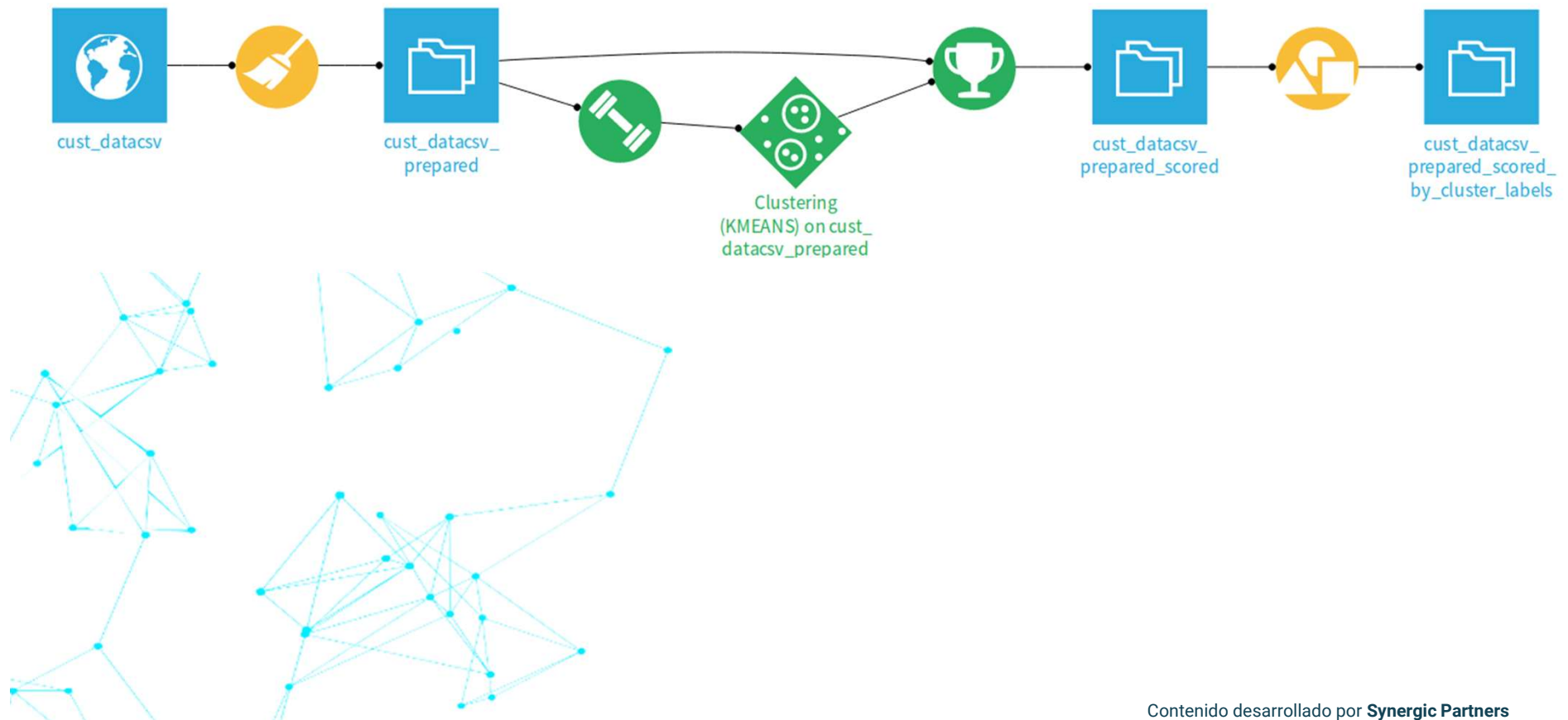
Tareas a realizar

1. Cargar datasets
2. Tratamiento de datos
3. Construcción del modelo de clustering
4. Perfilado y división de clientes
5. Enriquecimiento de datos CRM con los segmentos obtenidos de visitantes conocidos
6. Hacer visualización de los resultados (dashboard)

MODELOS ANALÍTICOS

Problema III: Segmentación y Descripción de Clientes

Workflow final



Telefónica
FUNDACIÓN

Conecta Empleo

