

Plan Overview

A Data Management Plan created using DMPTool

Title: Análisis y Predicción del Ciclo de Vida de Productos en una Matriz de Boston para Identificar Transiciones de Estrellas a Interrogantes y Estimar la Probabilidad de Alcanzar el Fin de su Ciclo de Vida

Creator: Christian Saker

Affiliation: Universidad de Los Andes (uniandes.edu.co)

Principal Investigator: Laura Rodriguez

Data Manager: Brayan Torres, Jesica Vique

Project Administrator: Christian Saker

Funder: Digital Curation Centre (dcc.ac.uk)

Template: Digital Curation Centre

Project abstract:

Este proyecto se centra en el análisis del comportamiento y la predicción del ciclo de vida de productos en el contexto de una matriz de Boston preexistente, con un enfoque en la identificación de productos que están transitando de la categoría de "Estrella" a "Interrogante" con el fin de estimar la probabilidad de que los productos estén llegando al final de su ciclo de vida. Utilizando técnicas de análisis de datos y modelado predictivo, este estudio proporciona información valiosa para la toma de decisiones en la gestión de portafolios de productos. La capacidad de anticipar cambios en la viabilidad y longevidad de los productos permite a las empresas tomar medidas proactivas para optimizar su portafolio y recursos. Los resultados de este proyecto tienen el potencial de impulsar estrategias de negocio más efectivas y eficientes, mejorando así la competitividad en el mercado.

Objetivos:

-Estimar la probabilidad de que un producto se encuentre o esté llegando al final de su ciclo de vida.

Posibles transformaciones:

-Limpieza de variables (Corrección de valores nulos y atípicos)

-Estandarización de los datos.

-Ingeniería de variables.

-Codificación de variables categoricas al tipo one-hot.

Start date: 10-15-2023

End date: 11-30-2023

Last modified: 10-22-2023

Análisis y Predicción del Ciclo de Vida de Productos en una Matriz de Boston para Identificar Transiciones de Estrellas a Interrogantes y Estimar la Probabilidad de Alcanzar el Fin de su Ciclo de Vida

Data Collection

What data will you collect or create?

Este proyecto se basa en la utilización de datos reales de ventas en unidades y de la rentabilidad/contribución de muebles de una empresa cuyo nombre se mantiene confidencial, a la cual hemos tenido acceso desde el año 2021. Estos datos están almacenados en un datamart y se recopilan de manera continua. Los datos de ventas se obtienen diariamente a través del sistema de ventas y facturación de la empresa, y también del modelo de rentabilidad de productos. La información relacionada con los productos se encuentra en el repositorio de datos maestros, que incluye detalles como segmentos de productos, códigos, nombres, colores, marcas, entre otros.

How will the data be collected or created?

La información se recopila a partir de archivos en formato CSV proporcionados a los investigadores del proyecto. Estos archivos contienen información encriptada para preservar la confidencialidad de los datos. Existen tres fuentes de información principales:

Ventas_y_Rentabilidad: Un archivo CSV que contiene datos de ventas de productos desde el año 2021 hasta septiembre de 2023. Estos datos se presentan en un desglose mensual y están identificados por códigos de productos.

Datos_Maestros_Productos: Otro archivo CSV que incluye un código único para cada producto, junto con información maestra relacionada.

Cálculo_Matriz: Los datos de clasificación de la matriz se obtienen a través de un cuaderno (notebook) proporcionado por la organización encargada de realizar el cálculo. Este cuaderno utiliza las dos bases de datos mencionadas anteriormente para generar la clasificación de productos en la matriz.

Documentation and Metadata

What documentation and metadata will accompany the data?

Para la primera base de datos, 'Ventas_y_Rentabilidad', no se tiene una documentación formal definida, sin embargo, su interpretación es intuitiva a partir de los nombres de las columnas que conforman el dataset.

Para el caso del dataset 'Datos_Maestros_Productos' tampoco existe documentación formal definida, sin embargo igual al caso anterior, es intuitiva debido a los nombres de las columnas.

Para el caso de 'Calculo_Matriz' existe un pdf que presenta a detalle la información sobre el cálculo y resultados de la matriz de contribución.

Ethics and Legal Compliance

How will you manage any ethical issues?

En líneas generales, los datos que se manejan no son de naturaleza personal. No obstante, debido a su carácter confidencial como información empresarial, se proporcionaron en formato encriptado para salvaguardar su confidencialidad. No se requieren consideraciones éticas adicionales.

Además, todos los análisis planeados en el proyecto tienen un propósito benévolo y no implican cuestiones relacionadas con la privacidad personal. En el improbable caso de que el proyecto implique el manejo de información personal, se seguirá un proceso explícito de solicitud de consentimiento para retener y compartir los datos con las partes involucradas.

How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?

Queremos aclarar que nuestro proyecto no implica ninguna infracción de derechos de autor o violación de derechos de propiedad intelectual (IP/IPR), ya que los datos proporcionados son propiedad exclusiva de la empresa que los suministra. Reconocemos que los datos son de su exclusiva propiedad, y los estamos utilizando con su permiso explícito para los fines de investigación declarados. Estamos comprometidos a garantizar que los datos se manejen de manera responsable y ética, y no participaremos en actividades que comprometan los derechos de propiedad intelectual de la empresa.

Storage and Backup

How will the data be stored and backed up during the research?

Los datos pertinentes se almacenarán en un almacén de datos en la nube proporcionado por la institución universitaria. La conexión a la base de datos en la nube se establece a través de una instancia de AWS (Amazon Web Services) como proveedor de servicios.

How will you manage access and security?

Los datos residen en una instancia de AWS a la cual solo se puede acceder mediante las credenciales proporcionadas por la Universidad y el par de claves utilizadas para la instancia en cuestión, que es la entidad encargada de gestionar el acceso y la colaboración entre los miembros del proyecto. El acceso a la información será abierto y sin restricciones para los cuatro investigadores principales, con el objetivo de no obstaculizar el progreso de la investigación. Cabe destacar que los datos utilizados en los análisis son de carácter privado.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Se considera relevante para su almacenamiento y publicación los datos obtenidos luego de la realización del proceso de ETL en el cual se ejecutó la compresión de los datos contenidos en los 3 datasets para su unificación en una única base de datos.

A partir de estos datos será posible la realización de diferentes análisis que involucren los factores de desplazamiento, periodo de tiempo, condiciones meteorológicas y visitas de puntos de interés en la ciudad de New York

What is the long-term preservation plan for the dataset?

En la actualidad, los datos se almacenan en la instancia de AWS, lo cual conlleva un costo mensual para mantener dicha información. No obstante, en el futuro, los resultados obtenidos serán transferidos y almacenados en el repositorio de datos de la empresa, el cual está alojado en un servidor destinado a este propósito. El servicio de almacenamiento ya está en funcionamiento por parte de la empresa.

Data Sharing

How will you share the data?

La empresa y el equipo de investigación serán los únicos con acceso a los datos originales necesarios para el desarrollo del proyecto. No obstante, los resultados y las conclusiones estarán disponibles en un repositorio público en GitHub. Para la organización, estos datos se encontrarán en un panel de Tableau, la herramienta que la empresa emplea para visualizar datos. Esto incluirá un panel específico para el ciclo de vida de productos en la matriz de Boston.

Are any restrictions on data sharing required?

Como se mencionó anteriormente, la información solo está disponible para el equipo investigador durante la duración del proyecto, y se encuentra encriptada de forma que se proteja la confidencialidad de la información de la empresa que provee los datos.

Responsibilities and Resources

Who will be responsible for data management?

El equipo de investigación será el encargado de la gestión de los datos.

What resources will you require to deliver your plan?

Para llevar a cabo la implementación del proyecto, es esencial contar con capacidad de almacenamiento y

capacidad de procesamiento. La Universidad de los Andes se encargará de satisfacer ambas necesidades mediante la creación de instancias virtuales a través del proveedor AWS. Además, se solicita el respaldo y apoyo del equipo de la Universidad para brindar asistencia en caso de consultas o necesidades relacionadas con la gestión y configuración de estos recursos.

Planned Research Outputs

Software - "Modelo de Pronostico de ciclo de vida para productos con rol en la matriz de Boston"

El resultado de la investigación es un modelo de aprendizaje automático supervisado que tiene como finalidad predecir la probabilidad de que un producto esté próximo a llegar al final de su ciclo de vida en los meses posteriores al actual. Esta probabilidad se mide en un rango de 0% a 100%, donde 0% indica que el producto no tiene posibilidad alguna de llegar al final de su ciclo de vida en el próximo mes, mientras que un valor del 100% señala que el producto definitivamente llegará al final de su ciclo de vida en el próximo mes.

Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Modelo de Pronostico de ciclo de vida para product ...	Software	2023-11-29	Restricted	None specified		None specified	None specified	No	No