

Científico de Datos y Analista en Inteligencia de Negocios

ALGORITMO DE MACHINE LEARNING PARA OPTIMIZAR
LAS TRANSACCIONES DE COMPRA

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Team



**Luis Andrés
Montoya**

*DATA SCIENTIST
MILLICOM*

Montoyalopera.andres@gmail.com



**Susana
Tilano**

*PRODUCTION
ENGINEER STUDENT
stilanof@eafit.edu.co*



**Mauricio
Villegas**

*ENGINEERING SPECIALIST (last job)
MILLICOM
mauricio.villegas.ramelli@gmail.com*



**Aníbal
Múnera**

*H&G CONSULTING
anibalmuneralopez@gmail.com*

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

4P's

Problema

No se cuenta con modelos descriptivos y predictivos que potencien el análisis en la toma de decisiones de la compañía, para incrementar sus indicadores económicos.

Potencial

Con un modelo de ML de agrupación de datos, se podría diseñar una propuesta de fidelización de clientes (ej. campañas marketing por email, enfocadas por grupos), ofrecer promociones, identificar productos más demandados, anticiparse a la demanda y reconocer proveedores estratégicos. Adicionalmente el modelo serviría de insumo a otros modelos.

Propuesta

Implementación de un modelo de ML de agrupación, usando el algoritmo K Means bajo arquitectura de Azure, que para fechas determinadas permita identificar y describir los grupos encontrados de manera gráfica y fácil de analizar.

Productos

- Modelo descriptivo para agrupar las transacciones de compras efectuadas en el período de tiempo que el usuario seleccione.
- Modelo predictivo para clasificar nuevos pedidos y con base en el grupo asignado, determinar la gestión óptima para la compañía.
- Dashboard con indicadores para visualizar el comportamiento de las transacciones por periodos de tiempo.

Aliados:



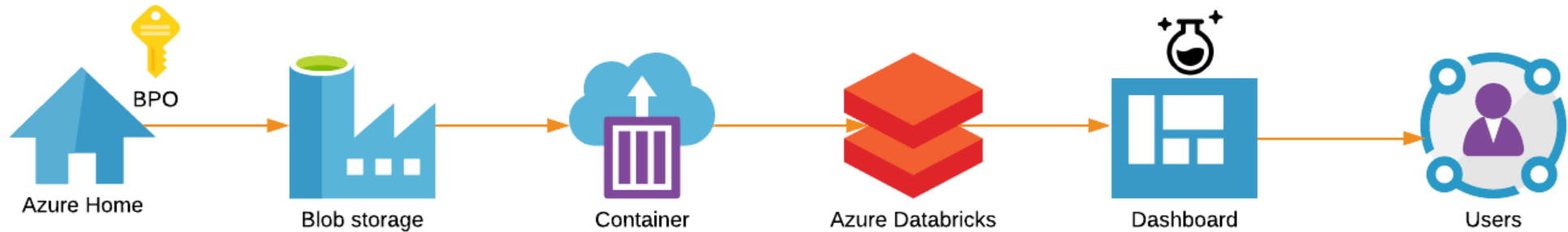
Microsoft

Vigilada Mineducación

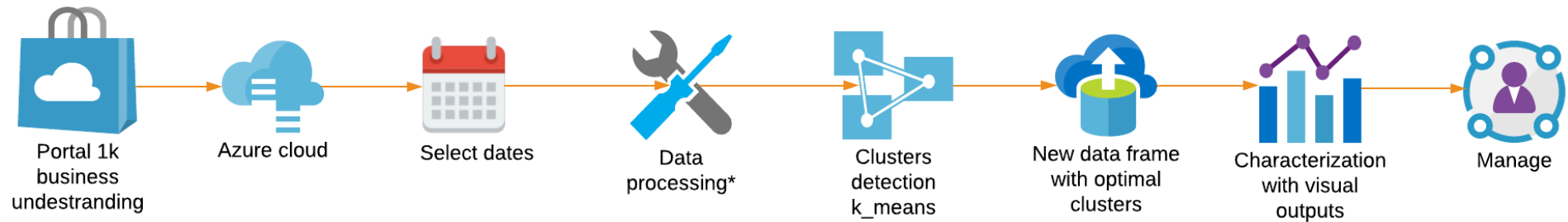


Advanced analytics for business

Architecture



Process flow



* En la etapa de feature engineering se han agrupado como **“Otros”** las entradas menos representativas en cuantía para: categoría, subcategoría, proveedores, compañía compradora.

Cluster: Conjunto de objetos que son similares entre sí.

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business

Proveindustriales_Data_Cleaned_v5 Python

ds-ni-cluster File Edit View: Standard Permissions Run All Clear

Fecha de inicio (YYYY-MM-DD) 2020-04-11 Fecha final (YYYY-MM-DD) 2020-04-18

FILTER BY FECHA_INI AND FECHA_FIN

Cmd 1

```
1 import pandas as pd
2 import matplotlib as plt
3 import seaborn as sns
4 import plotly.express as px
5 import numpy as np
```

Command took 0.03 seconds -- by stilanof@eafit.edu.co at 4/12/2021, 5:53:22 p. m. on ds-ni-cluster

Cmd 2

FILTER BY FECHA_INI AND FECHA_FIN

Cmd 3

```
1 dbutils.widgets.text("Fecha de inicio (YYYY-MM-DD)", "")
```

Command took 0.04 seconds -- by stilanof@eafit.edu.co at 4/12/2021, 5:53:22 p. m. on ds-ni-cluster

Cmd 4

```
1 dbutils.widgets.text("Fecha final (YYYY-MM-DD)", "")
```

Command took 0.03 seconds -- by stilanof@eafit.edu.co at 4/12/2021, 5:53:22 p. m. on ds-ni-cluster

Cmd 5

```
1 Fecha_ini = dbutils.widgets.get("Fecha de inicio (YYYY-MM-DD)")
2 print(Fecha_ini)
```

2020-04-11

Command took 0.03 seconds -- by stilanof@eafit.edu.co at 4/12/2021, 5:53:22 p. m. on ds-ni-cluster

Cmd 6

```
1 Fecha_fin = dbutils.widgets.get("Fecha final (YYYY-MM-DD)")
2 print(Fecha_fin)
```

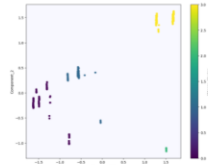
Navigation: Connection to Blob ... 1 DATA ACQUISITION DATA PREPARATION 1. Data integration 2. IRRELEVANT VAR... 2.1 Duplicated rows 2.2 Irrelevant columns 3. STATISTICAL DES... By year By Month Compañía comprad... 4. DATA CLEANING 4.1 Remove outliers 4.2 Fill NA 5. FEATURE ENGINE... SUMARIZE CATEG... Categoría Top CATEGORIA Top SUBCATEG... Top PRODUCTO... Top PROVEEDOR Top PORTAL 6. CORRELATION A... 7. TRANSFORMATION Numeric to Normali... Categorical to One...

Portal stilanof@eafit.edu.co

clideo.com

Dashboard

GRAPHICAL CLUSTERS



GROUPBY CLUSTERS- TOTAL+IVA

KMeans_Cluster	TOTAL_SIN_IVA	
	mean	sum
0	675.298.85	4334967092
1	278.370.68	1924654890
2	789.409.55	4293599050
3	38.868.74	141170245

GENERAL DESCRIPTION

KMeans_Cluster	Total registros	CATEGORIA	SUBCATEGORIA PRINCIPAL	PROVEEDOR PRINCIPAL	COMPRADOR PRINCIPAL	CANTIDAD_PEDIDA (promedio)	CANTIDAD_RECIBIDA (promedio)	PRECIO_UNITARIO (promedio)	TOTAL_SIN_IVA (promedio)
0	6418	Otros	Otros	Otros	Portal comprador 1	186.27	122.19	199,487.65	675,298.85
1	6914	Alimentos y Bebidas	Alimentos preparados y conservados	Otros	Portal comprador 29	389.49	12.19	15,369.47	278,370.68
2	5439	Violencia y seguridad industrial	Seguridad y protección personal	Proveedor 301	Portal comprador 24	42.49	42.49	39,930.22	789,409.55
3	3621	Ases, Cateñerías y Ofertas	Elementos de aseo y limpieza	Proveedor 59	Portal comprador 1	59.47	2.57	15,171.92	38,868.74

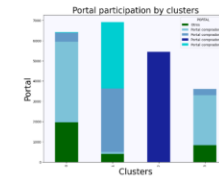
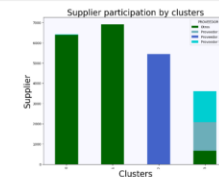
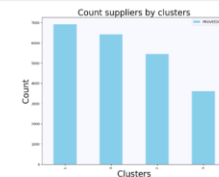
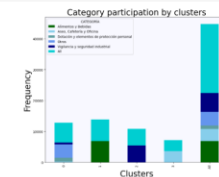
MEAN OF TOTAL_SIN IVA PER CLUSTER



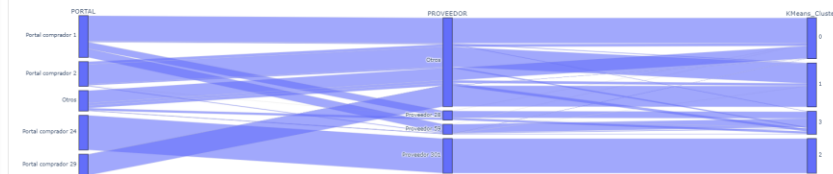
SUM OF TAL SIN IVA PEER CLUSTER



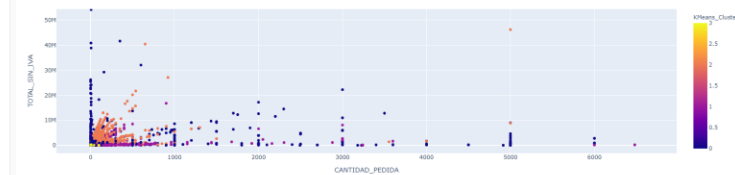
CLUSTER PARTICULAR DESCRIPTION



RELATION PORTAL & SUPPLIER BY CLUSTER

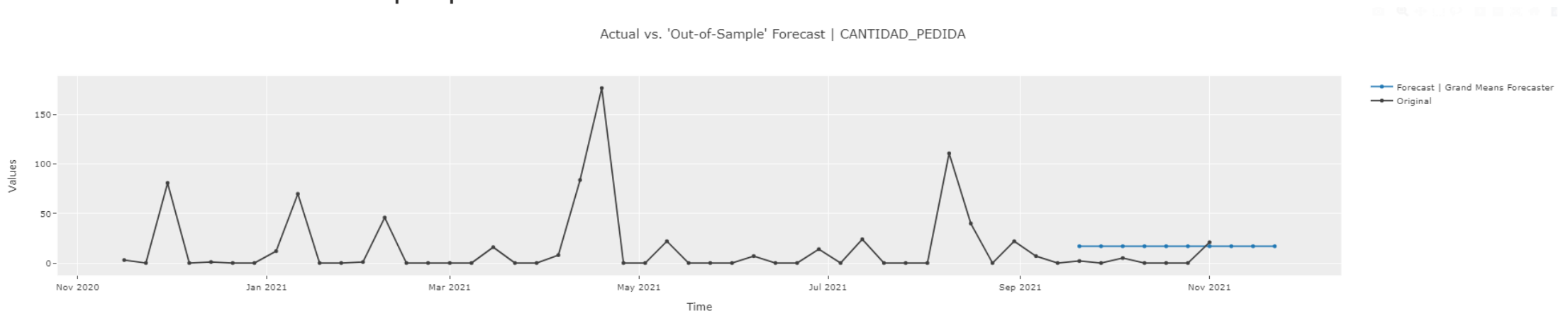


TOTAL SIN IVA VS CANTIDAD PEDIDA



Next Steps


- Despliegue de la app para la predicción de cluster.
- Automatizar la conexión entre los datos generados por el portal y la aplicación para actualizar la agrupación de los datos en el tiempo
- Permitir el acceso a mayor cantidad de datos para crear un modelo de predicción de la demanda futura que permita administrar el abastecimiento.




Aliados:










Repository


main
1 branch
0 tags

Go to file
Add file
Code


Stilano
Update README.md
3e1d1ce 2 minutes ago
33 commits

	business	your textUD2	18 days ago
	model	final	4 days ago
	steps_notebooks	automated	7 days ago
	README.md	Update README.md	2 minutes ago
	architecture.png	Add files via upload	4 days ago
	clusters.png	Add files via upload	14 days ago
	target.png	Add files via upload	4 days ago

https://github.com/Stilanof/ML_BPO_TRANSACCIONES

63 Lines (36 Lines) | 2:45

< > ⌵ ⌶ Save Blame

ML_BPO_TRANSACCIONES

As data scientists we convert data into business values that allow making decisions based on the analysis of behavior in the past, present and future of the operation.

This project is the final work of a eight weeks journey, in which we try to design a Machine Learning product for a local company.

Let's start 🚀

```
graph LR; A[Personal BPO Business understanding] --> B[Azure cloud]; B --> C[Select dates]; C --> D[Data processing]; D --> E[Clusters detection k-means]; E --> F[New data frame with optimal clusters]; F --> G[Characterization with visual outputs]; G --> H[Manage]
```

We implemented a k means model to segment each transaction at the market platform.

Production 🔧⚙️

```
graph LR; A[Azure Home] --> B[Blob storage]; B --> C[Container]; C --> D[Azure Databricks]; D --> E[Dashboard]; E --> F[Users]
```

Implementation 📄

The model consists of two notebooks, "Proveindustriales, Data Cleaned" and "Proveindustriales, Model".

"Proveindustriales, Data Cleaned" loads the csv file and, once the user enters the dates corresponding to the period that he wants to analyze, he proceeds with the preparation and cleaning of the data; the grouping of the categories, subcategories, suppliers and purchasing companies with the fewest occurrences under the class "Others"; correlation analysis; the normalization of the numerical variables and the transformation of the categorical variables.

"Proveindustriales, Model" takes the file generated by "Proveindustriales, Data Cleaned" and executes the clustering algorithm, extracts the most relevant characteristics of each cluster and presents the behavior of the clusters in the Dashboard.

Procedure:

1. Upload the Excel file with all transactions to Azure Blob storage
2. Open the notebook "Proveindustriales, Data Cleaned"
3. Establish Start Date and End Date, corresponding to the period of time that you want to analyze
4. Run the notebook "Proveindustriales, Data Cleaned"
5. Open and run the notebook "Proveindustriales, Model"
6. Refer to the "Dashboard, Proveindustriales"

Builted with 🛠️

- Azure Blob storage
- Containers
- Databricks

Versions 🏠

Take a look at the process! All versions and models tested here (https://github.com/Silanoof/ML_BPO_TRANSACCIONES)

Authors 👤

- Anibal Munera
- Luis Andrés Montoya
- Mauricio Villegas
- Susana Tilano

Thanks to everyone who teach us something along the way!

¡Gracias!

Aliados:



Microsoft

Vigilada Mineducación



Advanced analytics for business