



Análisis del Perfil Nutricional, Geográfico y Competitivo de Productos con Chocolate: Un Enfoque de Segmentación, Normalización y Modelado

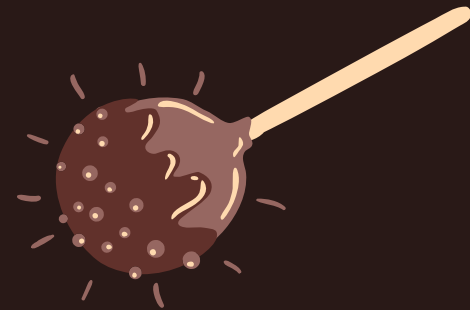


Autora

Wasserman Alejandra



CONTEXTO



El chocolate es uno de los productos más consumidos y comercializados a nivel global, pero su perfil nutricional ha sido motivo de creciente preocupación tanto por consumidores como por organismos de salud. En este contexto, las industrias enfrentan el desafío de innovar sin sacrificar sabor, competitividad o cumplimiento regulatorio.

OBJETIVO



Analizar el perfil nutricional, territorial y comercial de productos con chocolate mediante técnicas de ciencia de datos, con el fin de identificar factores asociados a una mejor calificación nutricional, segmentar el mercado según ingredientes y nutrientes, y evaluar la competitividad entre marcas en distintas regiones del mundo.

AUDIENCIA

Los hallazgos están dirigidos a departamentos de I+D, analistas de mercado y estrategias de producto del sector alimentario.

ETAPAS



1

Objetivos e hipótesis

2

Preparación de los datos

- Data acquisition
- Data wrangling
- Análisis univariado
- Análisis bivariado

3

Tratamiento de outliers

4

Transformaciones categóricas

ETAPAS



5

Entrenamiento de los modelos

H1 Calidad nutricional

- Lineal Regressor
- Random Forest Regression

H2 Segmentación del mercado

- K-Means
- DBSCAN

H3 Competitividad territorial

- Decision Tree Classifier
- Random Forest Classifier

6

Validación de los algoritmos

7

Ajuste de hiperparámetros y modelos de ensamble

8

Conclusiones

Calidad nutricional

Algoritmo
Linear Regression

MAE: 0.3028
MSE: 0.1762
RMSE: 0.4198
 R^2 : 0.5901

Algoritmo
Random Forest Regression

MAE: 0.2311
MSE: 0.1645
RMSE: 0.4057
 R^2 : 0.6172



Segmentación del mercado

Algoritmo
K-Means

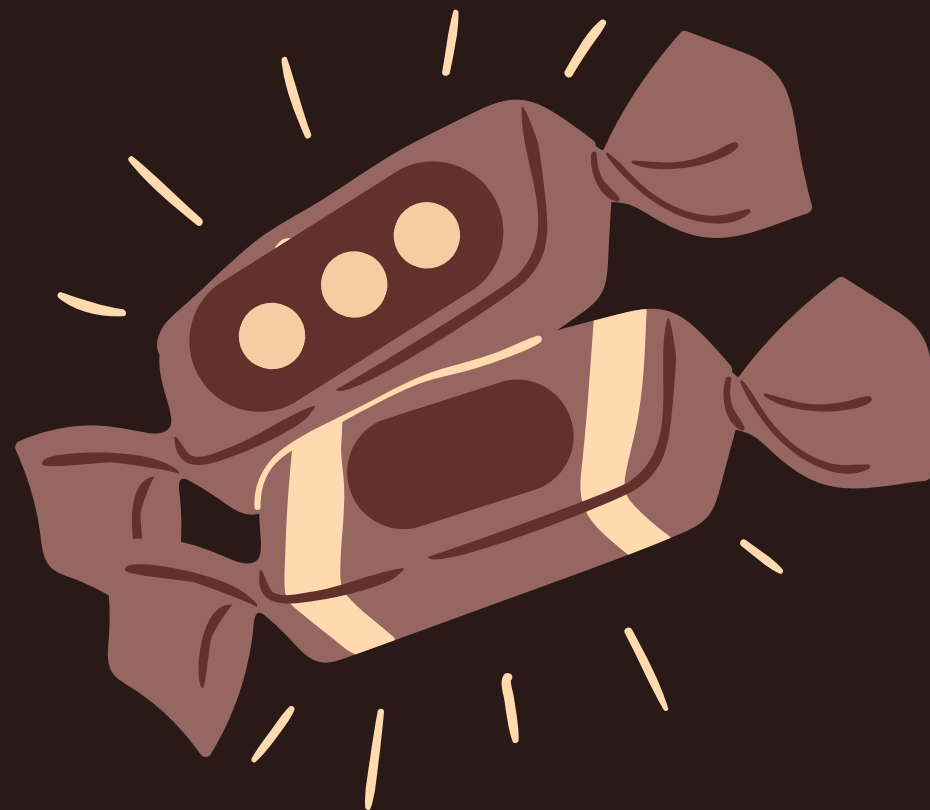
Algoritmo
DBSCAN

Indice de
Silhouette: 0.31

Indice de
Silhouette: 0.02



Competitividad territorial



Algoritmo
Decision Tree Classifier

Accuracy: 0.8772
Precision: 0.8354
Recall: 0.8772
F1-score: 0.8464

Algoritmo
Random Forest Classifier

Accuracy: 0.9756
Precision: 0.9736
Recall: 0.9756
F1-score: 0.9725

Análisis

01

Calidad nutricional:

Al aplicar la validación de los modelos, el algoritmo Random Forest Regressor presenta mejores métricas frente al algoritmo Linear Regression. Es decir, logra predecir mejor el grado de nutrición de los productos según su composición.

02

Segmentación del mercado:

Al aplicar los modelos no supervisados, k-means y DBSCAN, para encontrar grupos de productos con una composición nutricional similar, ninguno de los algoritmos logro formar clusters bien definidos y compactos. Esto requiere de un mayor estudio para lograr segmentar los productos según su composición.

03

Competitividad territorial:

Al clasificar la composición de los productos según los países de origen, se tiene que el algoritmo Random Forest Classifier presenta las mejores métricas al validarlas frente al algoritmo Decision Tree Classifier. Por lo que, dicho modelo logra predecir el país al que pertenece un producto según su composición.

Conclusión

El proyecto demostró que las técnicas de machine learning aplicadas al análisis de alimentos permite evaluar la calidad nutricional, la segmentación de los mercados y analizar la competitividad territorial. Los resultados ofrecen un marco útil para la toma de decisiones en la industria alimentaria, desde el desarrollo de productos hasta la estrategia comercial.

En conjunto, el proyecto demuestra el valor del machine learning en la industria alimentaria, tanto para optimizar la calidad nutricional como para segmentar y analizar la competitividad de marcas en diferentes regiones.

