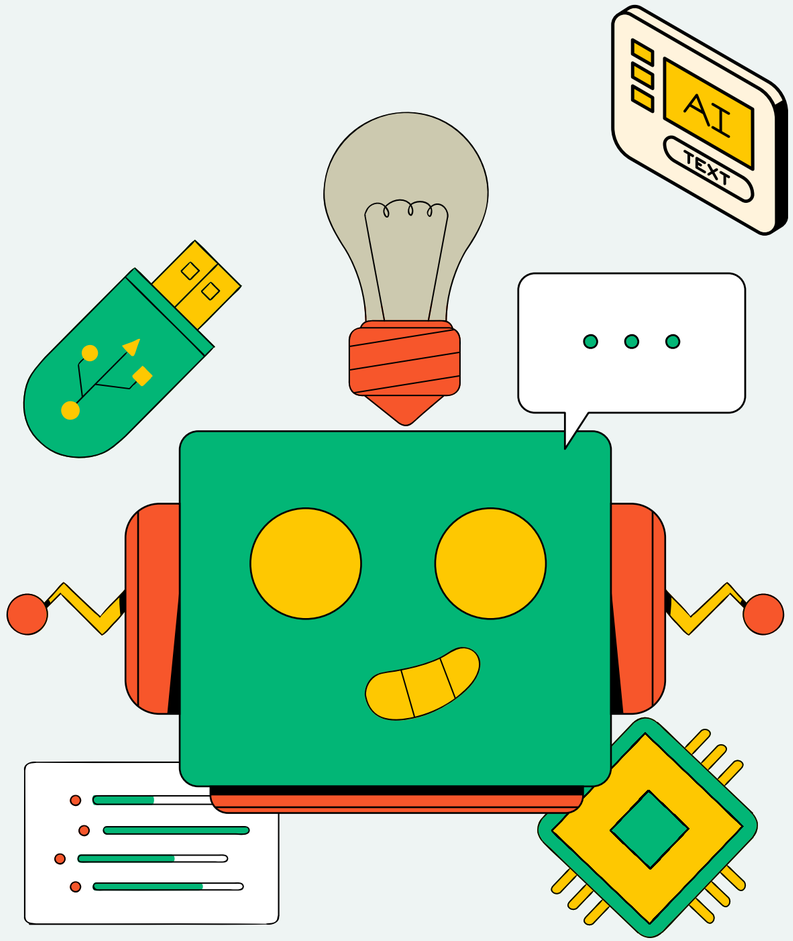
IRONHACK

DATA ANALYTICS OCTOBER 2024



RECOMMENDER SYSTEM FOR LMES PRODUCTS USING ML

PRESENTATION

PRESENTED BY:

ALBERTO GUERRA

PRESENTATION OUTLINE

- Introduction
- Project Overview
- Data Base
- Data Preprocessing & Feature Engineering
- ML Key Concepts used
- Model Building & Web Interface
- Results
- Questions and Answers

INTRODUCTION

OVERVIEW

This project aims to develop a content-based recommender system for Leroy Merlin products.

PURPOSE

Improve the online shopping experience by suggesting similar products based on their characteristics

LEVERAGE

Leverage machine learning to transform raw product data into meaningful recommendations.

PROJECT OVERVIEW

Combine data preprocessing, feature engineering, and machine learning (kNN with cosine similarity) to generate recommendations.

The project uses a large product database with attributes such as dimensions, category, product description, and sales ranking.

The objetive is to build a system that, given a product reference, returns 5 similar products.

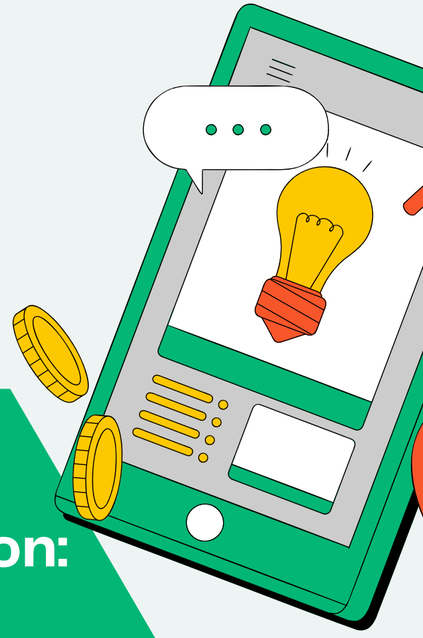
WHAT ABOUT THE DATA?

Dataset with +300.000 different products

Key Attributes:

- **LM Internal Reference: Product identifier**
- **Section & Type: Product categories**
- **Dimensions & Weight**
- **Designación Administrativa: Product description**
- **Letra de Gama: Sales probability ranking**

DATA PREPROCESSING & FEATURE ENGINEERING



Feature Transformation: Categorical Data: One hot encoding

Imputation: Fill missing values in categorical values(place-holders)

Feature Transformation: Convert description to embeddings using Sentence Transformer.

Data Filtering: Exclude 48 &49 and empty rows

Feature Transformation: Numeric Data: Standar Scaling for dimensions

Feature Transformation: Ranking: Encode 'Letra de Gama' with One-Hot Encoding.

ML KEY CONCEPTS USED

CONTENT-BASED FILTERING

Recommending items based on their features.

FEATURE ENGINEERING

Transforming raw data into numerical representations.

SIMILARITY MEASURES

Using cosine similarity to compare product vectors.

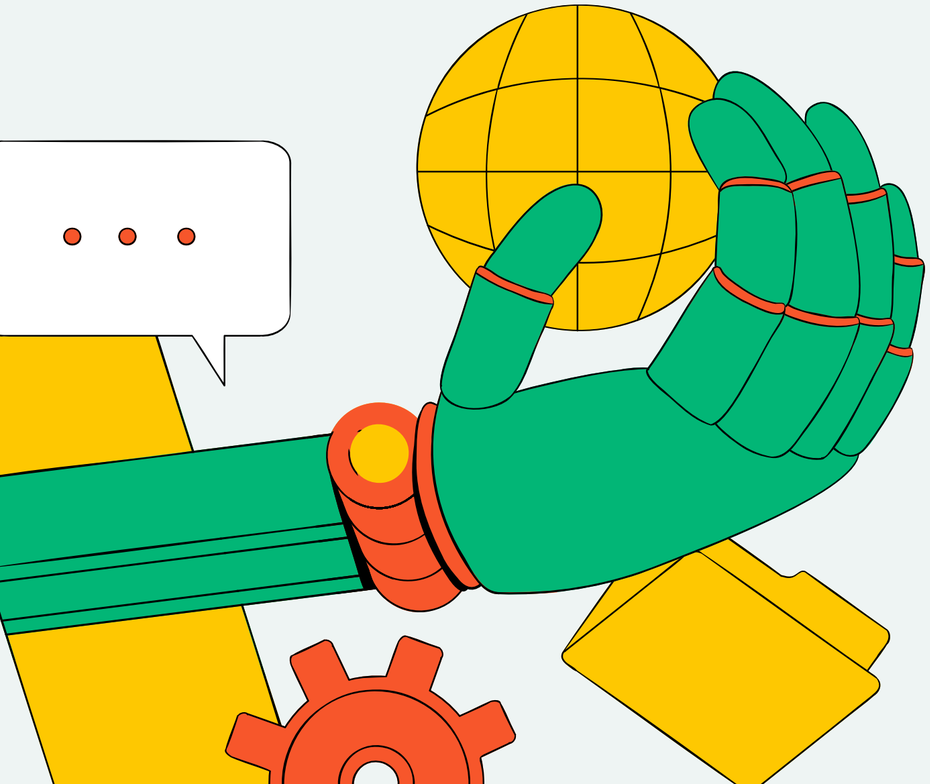
K-NEAREST NEIGHBORS (KNN)

Identifying the 5 most similar products to a given product.

EMBEDDING TECHNIQUES

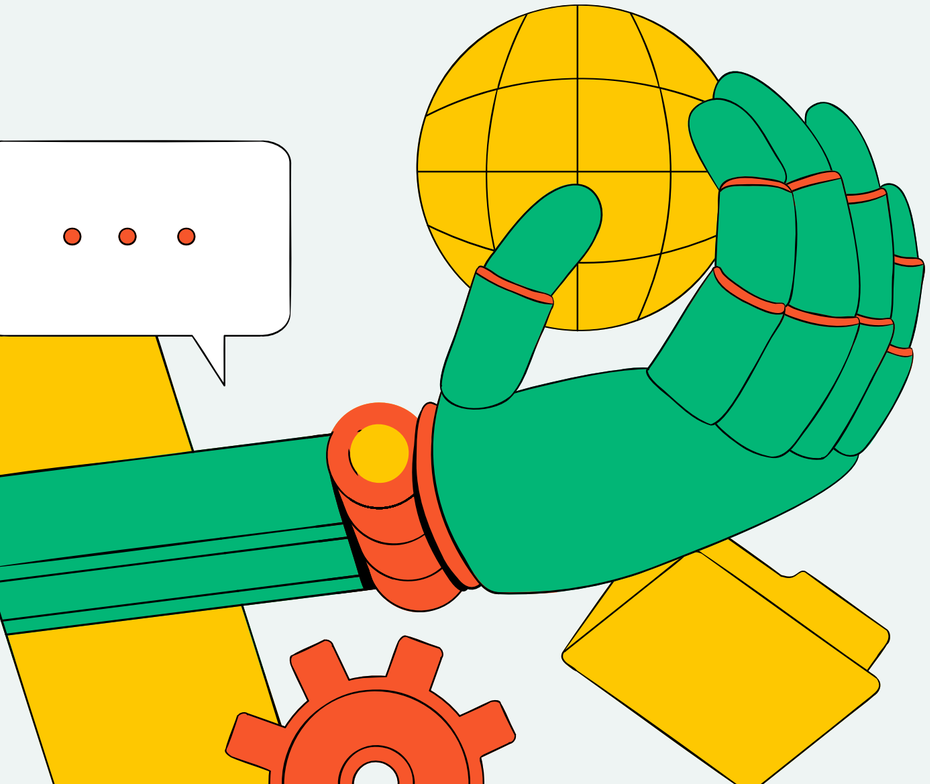
Applying sentence embeddings to capture semantic information from text

MODEL BUILDING



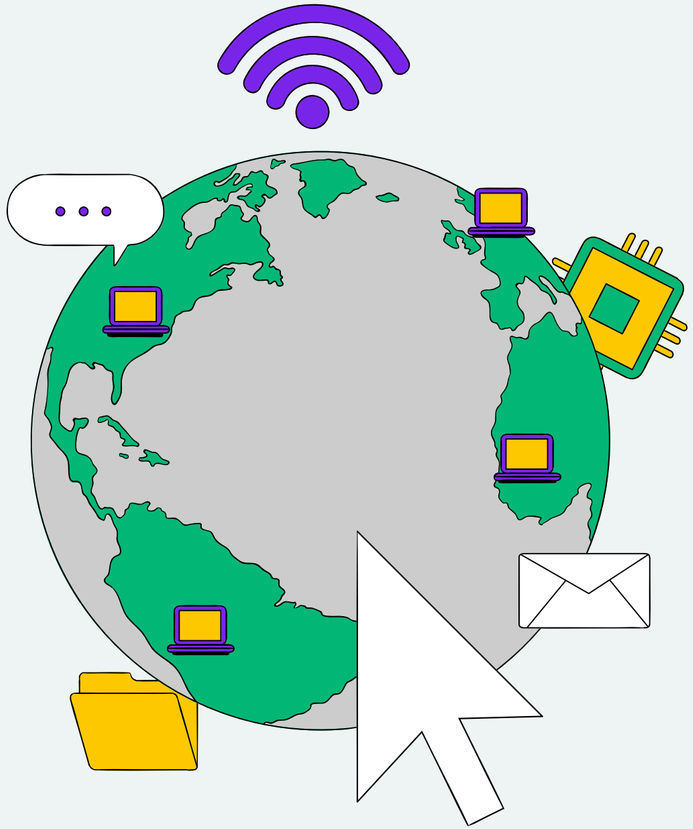
- A k-Nearest Neighbors model is trained using cosine similarity to measure the closeness between product vectors.
- The model combines multiple features (categorical, numerical, textual, and ranking) into a single vector per product.

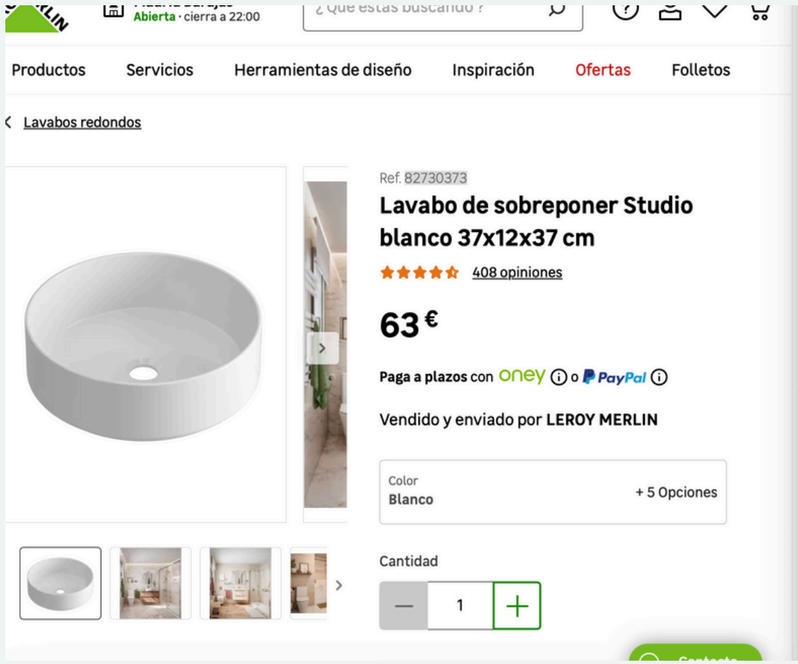
WEB INTERFACE



- A Flask application provides a simple web interface.
- Users enter a product reference, and the system displays 5 recommended products along with key details (LM Internal Reference, Section, Designation).
- This interface demonstrates the practical deployment of the recommendation model

RESULTS



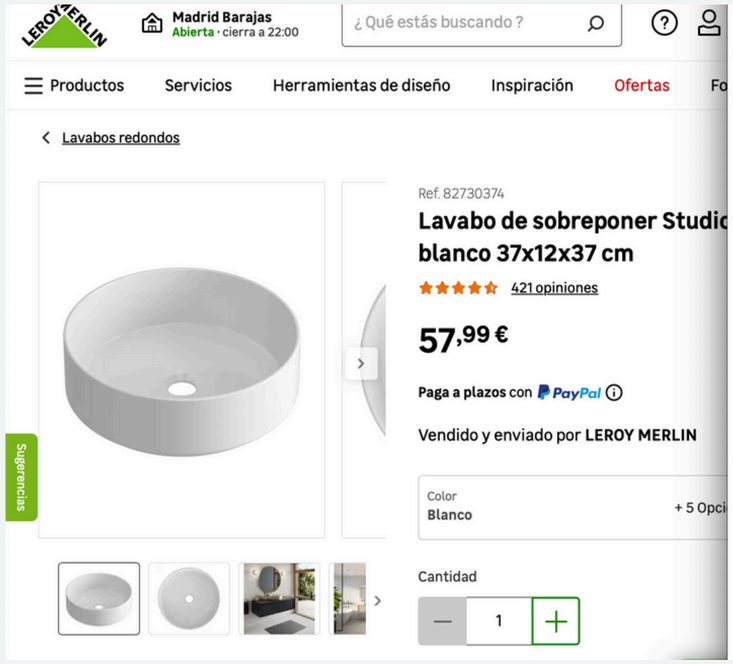


Product Recommender

Enter product reference: Get Recommendations

Recommended Products:

LM Internal Reference	Sección	Designación Administrativa
82730373.0	07 : SANITARIO	LAV STUDIO CERAM REDONDO D37 BLC
82730374.0	07 : SANITARIO	LAV STUDIO CERAM REDONDO D37 BLC
82730377.0	07 : SANITARIO	LAV STUDIO CERAM REDONDO D37 NEG MATT
82644576.0	07 : SANITARIO	LAVABO LUNE CERAMICA RED D40CM
85538218.0	07 : SANITARIO	LAVABO ESSENTIAL BOWL CERAMICA D

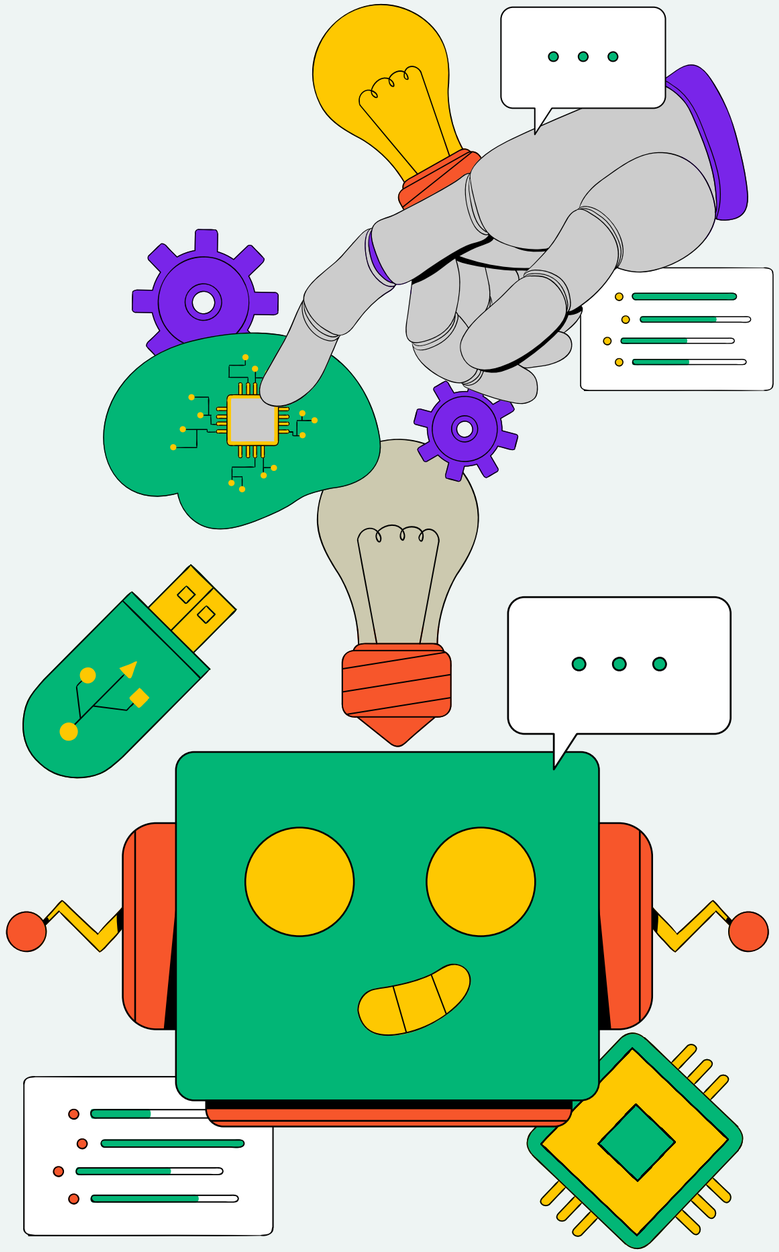


Product Recommender

Enter product reference: Get Recommendations

Recommended Products:

LM Internal Reference	Sección	Designación Administrativa
82730373.0	07 : SANITARIO	LAV STUDIO CERAM REDONDO D37 BLC BR
82730374.0	07 : SANITARIO	LAV STUDIO CERAM REDONDO D37 BLC MATT
82730377.0	07 : SANITARIO	LAV STUDIO CERAM REDONDO D37 NEGRO MATT
82644576.0	07 : SANITARIO	LAVABO LUNE CERAMICA RED D40CM
85538218.0	07 : SANITARIO	LAVABO ESSENTIAL BOWL CERAMICA D33CM

QUESTIONS AND ANSWERS

THANK YOU!