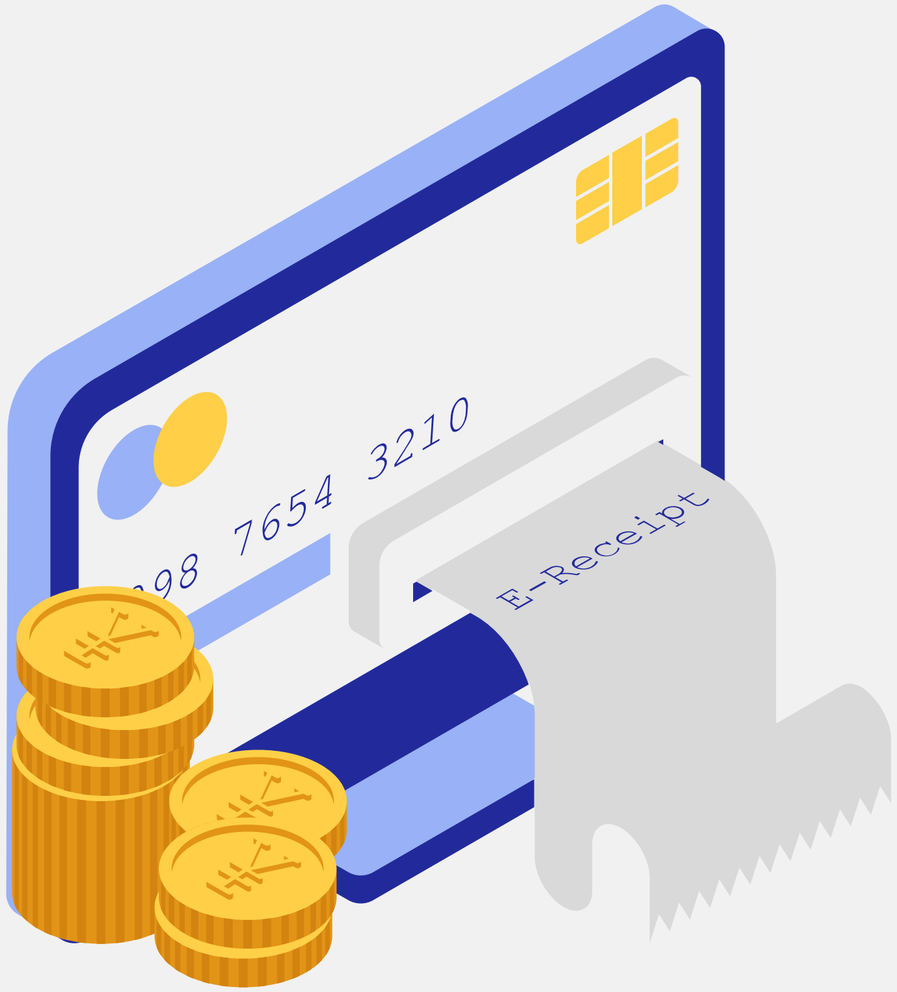
deti universidade de aveiro
departamento de eletrónica,
telecomunicações e informática



The Evolution of Products and Platforms in Portuguese E-commerce

Cristiano Nicolau
108536

Goals

- Analyze the evolution of products offerings over the years
- Analyze price trends and promotional strategies
- Evaluate the evolution of e-commerce platform functionalities

1st Delivery

- Extracted Filter Categories for Each Website
  1. Extracted the main page webistes links
  2. Extracted filter categories across e-commerce platforms

- Extracted Product Links During Black Friday and Non-Black Friday Periods

sites = [fnac, worten, elcorteingles, radiopopular, staples, pcdiga]

2nd Delivery

Data Extraction

1. **Initial Attempt to Extract Product Prices**
   Used *"extractedText"* to identify price patterns but encountered challenges due to lack of consistent patterns.

2. **Switched to Web Scraping for Product Links**
   Implemented web scraping for each product link to retrieve accurate pricing information. For pages with multiple products, captured and stored all products along with their respective prices.

2nd Delivery

Data Processing

1. **Merge Extracted Products Across All Sites**
   Merged all products extracted from each site, whether during Black Friday or regular periods.
2. **Refined Product Titles**
   Removed non-value-adding words and stop words from product titles.
3. **Cleaned Duplicate Entries**
   Identified and removed all duplicate product entries to ensure data consistency.
4. **Created a Comprehensive Dataset**
   Compiled a final dataset containing all products along with their respective details and information.

site	date	title	*extractedData*	promocao

site	date	title	*extractedData*	linkToArchive	linkToExtractedText	snippet	promocao

2nd Delivery

Data Categorization

1. Initial Approach to Categorization Using LDA-NLP

   Attempted to use LDA (Latent Dirichlet Allocation) for automatic categorization of products. Faced challenges as the results did not yield meaningful patterns.

2. Developed a Custom Categorization Model
   - Category (e.g., Tablet, Smartphone, Console, etc.)
   - Brand (e.g., Apple, Samsung, Lenovo, etc.)
   - Model (e.g., iPhone 12, Galaxy S20, etc.)
   - Color (e.g., Black, Silver, etc.)
   - HDD and RAM Specifications

2nd Delivery

Data Categorization

3. Implemented Pattern Matching for Detection

   Used predefined dictionaries and regular expressions for:

   - Categories and Identifiers (e.g., "laptop" for Computers, "galaxy" for Smartphones).
   - Brand Matching (e.g., "macbook" for Apple, "rog" for Asus).
   - Model Identification (e.g., "iPhone 12" for Apple, "Galaxy S20" for Samsung).
   - Color Detection (e.g., "black" for Preto, "silver" for Prata).

4. Performed Groupings and Visualizations

   - Aggregated data by category, brand, and other features.
   - Created visualizations to understand product distributions and trends.

Final Delivery

- Graph Representation of Websites
- Word Cloud for Insights
- Price Prediction Model