Data Mining

Antiope Project

To extract knowledge from Amazon reviews

Marco Bellia

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

UNIVERSITÀ DI PISA

CROSSLAB Innovation for industry 4.0

Application Highlights

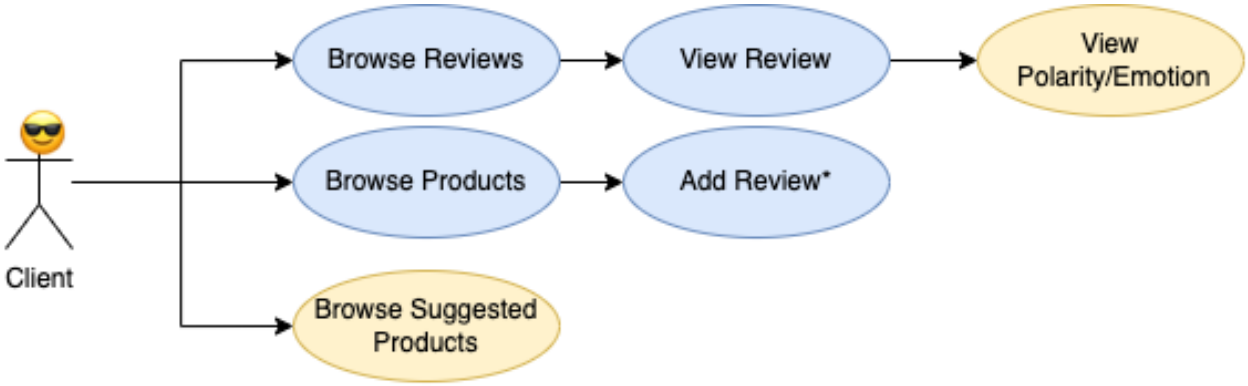
Antiope character in Wonder Woman 2017

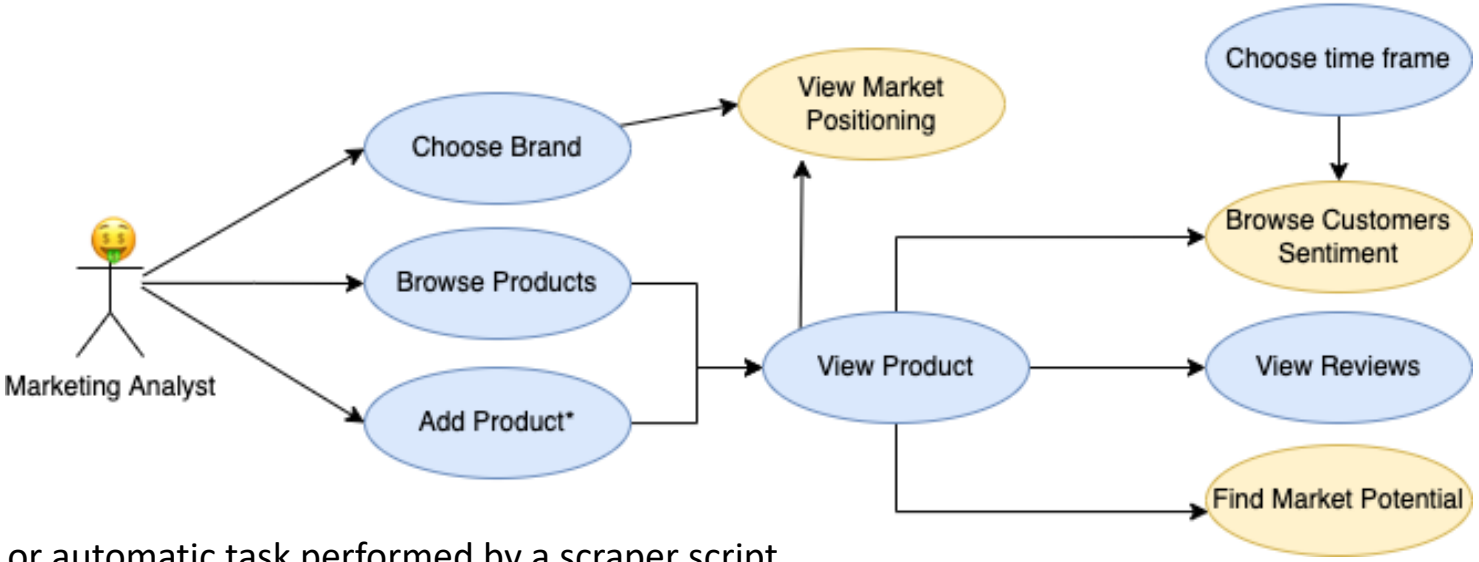
Antiope is a powerful OLAP assistant, queen of the amazons has studied each product and client of the big tech company.

*Antiope use AI to **cluster users** with their linguistic register and buying habits, and suggest new products to the proper target.*
*She's able to **make predictions** about **the success** of a product in the market which is very appreciated from its merchants.*
*Antiope **reveal** to the marketer the real **market segmentation** taking into account the reviews, the price, the category/brand.*

And much more..

Actors and main supported functionalities





*Manual or automatic task performed by a scraper script

Dataset Description

Source: University of California San Diego

Description: This dataset includes **reviews** (ratings, text, helpfulness votes), product **metadata** (descriptions, category information, price, brand, color, size, and image features), and **links** (also viewed/also bought graphs).

29 categories
8,171,199 products
233.1 million reviews in the range May 1996 - Oct 2018.
~50GB of raw data
http://deepyeti.ucsd.edu/jianmo/amazon/index.html

Some consideration

Considering the size of the dataset, and of the statistical sample analyzed in the project, we assume that a **review always corresponds to a purchase.**

So we have that:

- Stars tell us wether a review is positive or negative.
- It is possible to build a network of purchases of all the users.
- We can verify how much the products recommended by amazon are actually purchased.

Techniques Used

Text Mining:
To extract review polarity and user language registry

Clustering:
To find cluster in the space of price, reviews/sentiment, category/brand

Frequent Pattern:
To discover products which are usually bought together