**2. Introduction:**

With the boom of online marketplaces customers face an overwhelming amount of information when choosing the best products for them, several companies are analyzing these customer choices and which are the main drivers for these decisions.

This project aims to develop an algorithm that takes as input the information from the product in the way it is presented online and tries to forecast its popularity.

Particularly, this project will be focused on predicting Apps popularity. For this purpose, features from different mobile applications will be extracted from Google Play Store including its price, size, category, release date, update date, content rating, in-app purchases and presence of advertising inside the app. More importantly, the main focus of the project is to determine whether the main shown image is relevant to explain the popularity of the application measured as number of Installs. An app will be considered popular when its number of installations is in the top 50 percentile.

Specifically, a PCA algorithm and an advanced Neural Network will be used to preprocess and reduce the dimensionality of the images, a K-means algorithm will be used to cluster similar images and a logistic regression penalized with both L1 and L2 methods will be trained to determine the variables that ‘truly matter’ and to classify the popularity of the apps.

This approach could be very useful to help determine the design of both the marketplace itself but mainly deciding how a product should be presented to maximize its popularity.