Prediction of fund raising state of technology projects in Kickstarter using Machine Learning models

Alonso Puente[1] , Marks Calderon[1]

[1]ESAN University

Motivation

Only **20%** of Kickstarter technology projects were successful. It means this category has one of the lowest success rate.

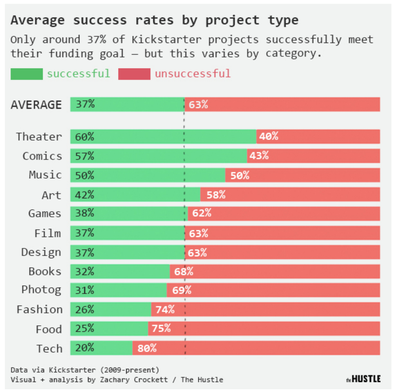

Figure 1:Kickstarter success rate, from 2009 to 2019

In addition to this, previous researches were just focused on metadata variables, such as funding goal, number of backers, pledges, campaign duration, etc.

Dataset and Features

We scraped **27,035** technology projects from Kickstarter, between 2009 and 2019. Those projects contained metadata (backers, goal, pledge and duration), visual content (project image) and textual content (project description).

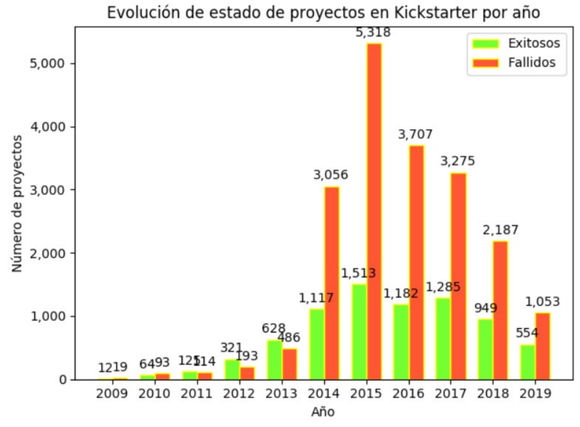

Figure 2:Success rate from Kickstarter's technology scraped projects.

Methods

Metadata start datasets were downloaded from WebRobots.io. Once the files were decompressed, they were pre-processed through **Alteryx Designer** software to build final datasets.

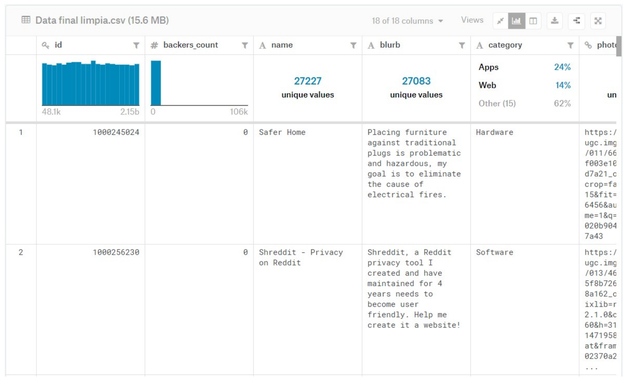

Figure 3:Metadata variables scraped from Kickstarter

With projects URLs, project images and descriptions were scraped using Python libraries such as **Beautiful Soup** and **tqdm**.
A framework based on the assembly of three predictive models for each part of the project was built:

- An **SVM** model for Metadata.
- A **Convolutional Neural Network (CNN)** for visual content.
- Two **SVM** models, the first one with **TF-IDF** and the second one with **BoW**.

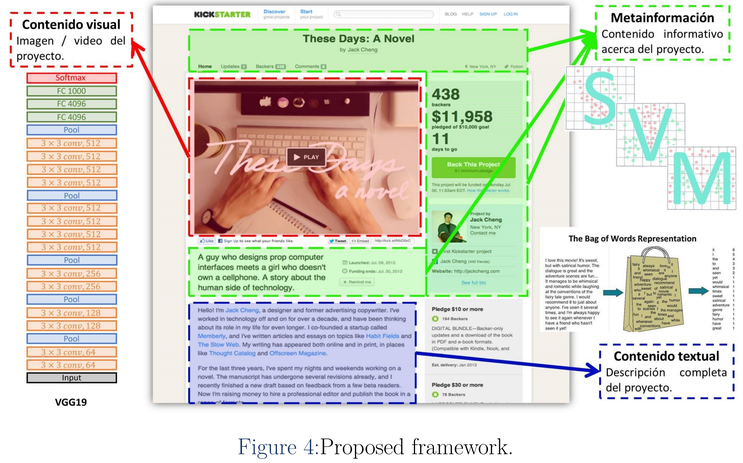

Figure 4:Proposed framework.

Results and Discussion

According to selected evaluation metrics, each model obtained the following results after training stage:

	Metadata		Images	Descriptions	
	SVM	MLP	VGG-19	SVM with TF-IDF	SVM with BoW
Accuracy	0.888108	0.827261	0.700444	0.754023	0.732199
AUC	0.837699	0.704267	0.501000	0.674626	0.670924
Precision	0.860483	0.930636	0.000000	0.584556	0.532942
Recall	0.721569	0.420915	0.000000	0.532349	0.530175
F1-score	0.784927	0.579658	0.000000	0.349252	0.341849
Execution time	00:00:16.22	00:39:30.15	02:43:19.26	03:42:53.30	04:25:49.16

Figure 5:Data training results.

Then, a demo was built in order to test metadata, images and description examples as inputs.

	Metadata		Images	Descriptions	
	SVM	MLP	VGG-19	SVM con TF-IDF	SVM con BoW
Fund state prediction	Failed	Failed	Successful	Successful	Successful
Success probability	0.00%	0.00%	51.19%	50.99%	42.67%

Figure 6:Demo results.

Conclusion and Discussion

After analyzing the models proposed for each part of the project, it was concluded that only the models for metadata and descriptions are acceptable to be used since the image model does not help to predict correctly due to falling into overfitting.

References