Using Machine Learning models to Predict Kickstarter success for Technology Projects

Alonso Puente[1] , Marks Calderon[1]

[1]ESAN University

Motivation

More than 5 billions U.S. dollars pledged on Kickstarter projects. They had a **37.86%** success rate. Also, Technology campaigns had only a **20%** success rate in average since 2009.

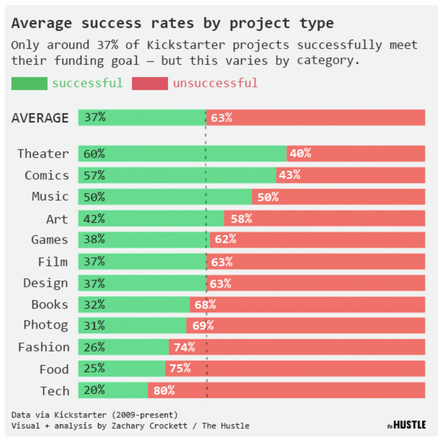

Figure 1:Kickstarter success rate, from 2009 to 2019

Machine Learning models are essential to predict campaign success and reduce uncertainty.

Dataset and Features

27,035 Kickstarter technology projects were obtained from January 2009 until August 2019. The features selected were metadata (backers, goal, pledge and duration), visual content (project image), and textual content (project description). The pre=processing metadata used **Alteryx Designer** software. the image and textual datasets were scraped from Kickstarter site.

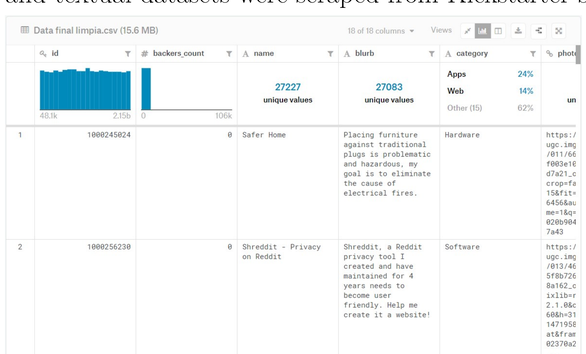

Figure 2:Metadata variables scraped from Kickstarter

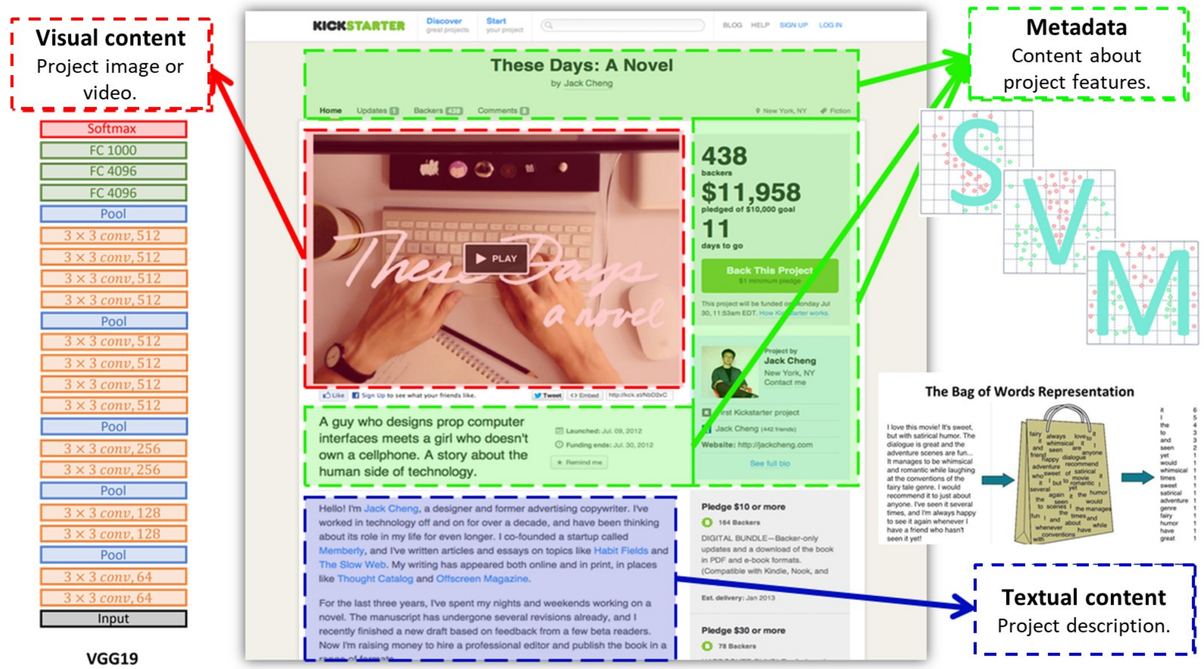

Figure 3:Proposed framework.

Methods

A framework based on the assembly of three predictive models (Fig, 3) 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 one with **BoW**.

Results and Discussion

We split the data with 80% for training set and 20% for test. And the modelos got the following results :

	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

Conclusion and Discussion

We concluded that only the models for metadata and descriptions are acceptable. Image model was a lower result because each project has different images to train, and VGG learned per each project differents patterns. It implies overfitting over our visual model.

References

[1] Chaoran Cheng, Fei Tan, Xiurui Hou, and Zhi Wei, editors.
Success Prediction on Crowdfunding with Multimodal Deep Learning.

[2] Rohith Gandhi.
Support vector machine — introduction to machine learning algorithms.
Towards Data Science, 2018.

[3] Fei-Fei Li, Justin Johnson, and Serena Yeung.
Convolutional neural networks.
5, 2019.