



UNIVERSIDAD  
**esan**

---

# FUNDING STATE PREDICTION FOR KICKSTARTER TECHNOLOGY PROJECTS USING MACHINE LEARNING MODELS

---

Student: Alonso Augusto Puente Ríos

Thesis advisor: Marks Calderón

# CONTENT



INTRODUCTION



METHODOLOGY



RESULTS



CONCLUSIONS

# INTRODUCTION

# ABOUT CROWDFUNDING

«All-or-nothing»



«Keep-it-all»



Flexible Funding



Fixed Funding



SINCE APRIL 2009:

→ + 6.3B USD PLEDGED TO PROJECTS

→ + 212K PROJECTS SUCCESSFULLY FUNDED



## Average success rates by project type (2009-2019)



# CURRENT SITUATION



# KICK STARTER



ChopBox: World's First Smart Cutting Board With 10 Features  
Kills germs, weighs your food and more! Sustainable. Replaceable parts. Super easy to clean.  
Watch the video to see all the features.

**\$1,751,252**  
pledged of \$10,000 goal

**12,732**  
backers

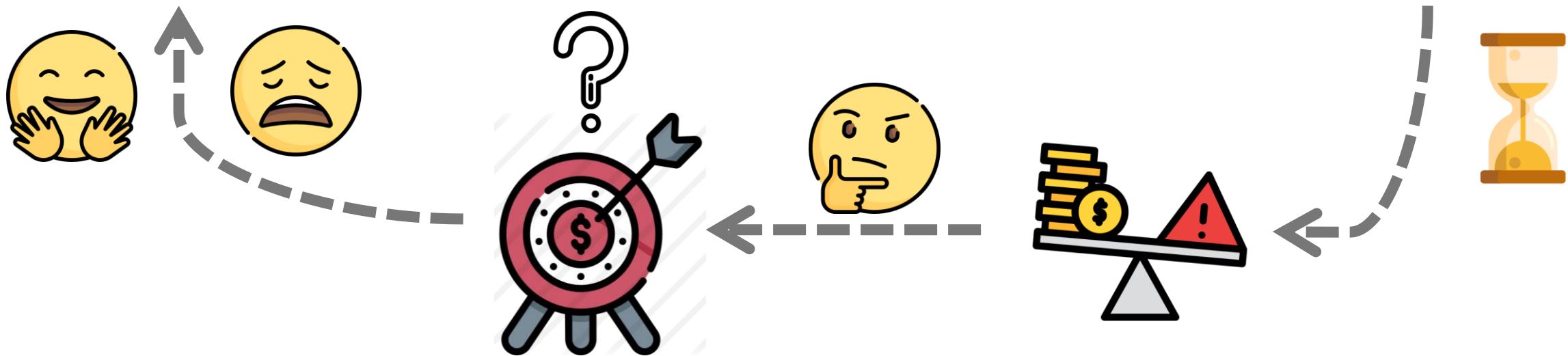
**43**  
hours to go

[Back this project](#)

[Save](#) [Facebook](#) [Twitter](#) [Email](#) [X/](#)

All or nothing. This project will only be funded if it reaches its goal by Fri, October 18 2019 10:55 AM EDT.

The image shows a screenshot of a Kickstarter campaign for the 'ChopBox'. The project has reached over \$1.7 million in pledges towards a \$10,000 goal. It has 12,732 backers and 43 hours left to go. The ChopBox is described as the world's first smart cutting board with 10 features, including sanitizing, weighing food, and being sustainable with replaceable parts. It also includes a timer, scale, and chopping function. A video button is present, along with a review from a superbacker (@brollandboardgames). The project is categorized under Gadgets and located in Dover, DE. A note at the bottom states 'All or nothing. This project will only be funded if it reaches its goal by Fri, October 18 2019 10:55 AM EDT.'



# PROPOSED SOLUTION



KICK  
STARTER

→

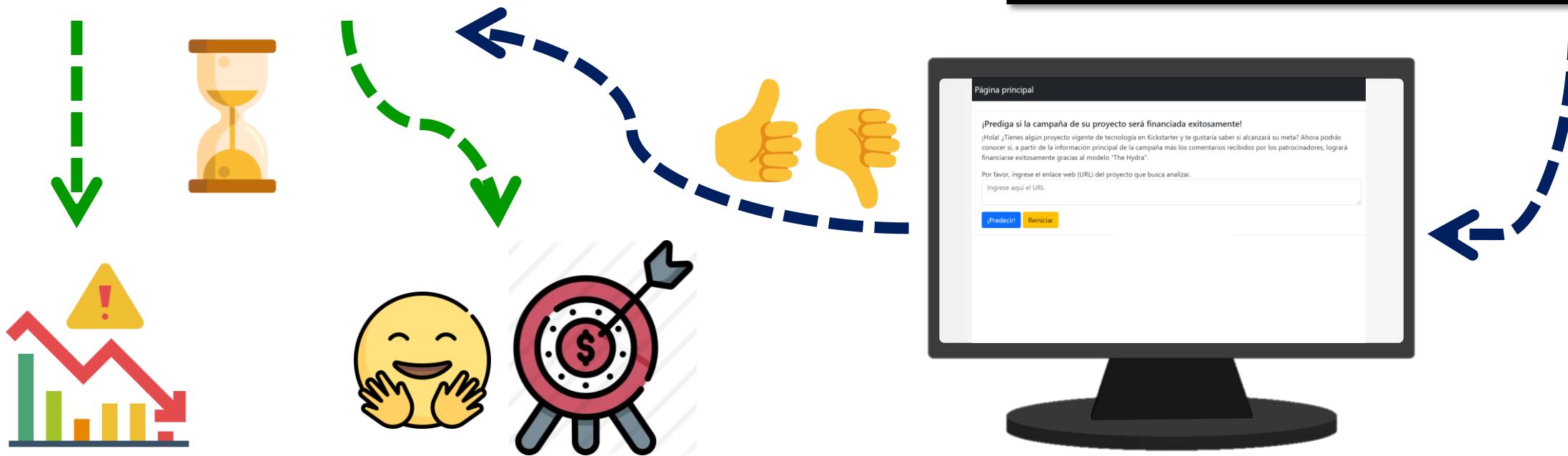
ChopBox: World's First Smart Cutting Board With 10 Features  
Kills germs, weighs your food and more! Sustainable. Replaceable parts. Super easy to clean.  
Watch the video to see all the features.

**CHOPBOX**  
Sanitizer | Sharpeners | Timer | Scale | Chopping

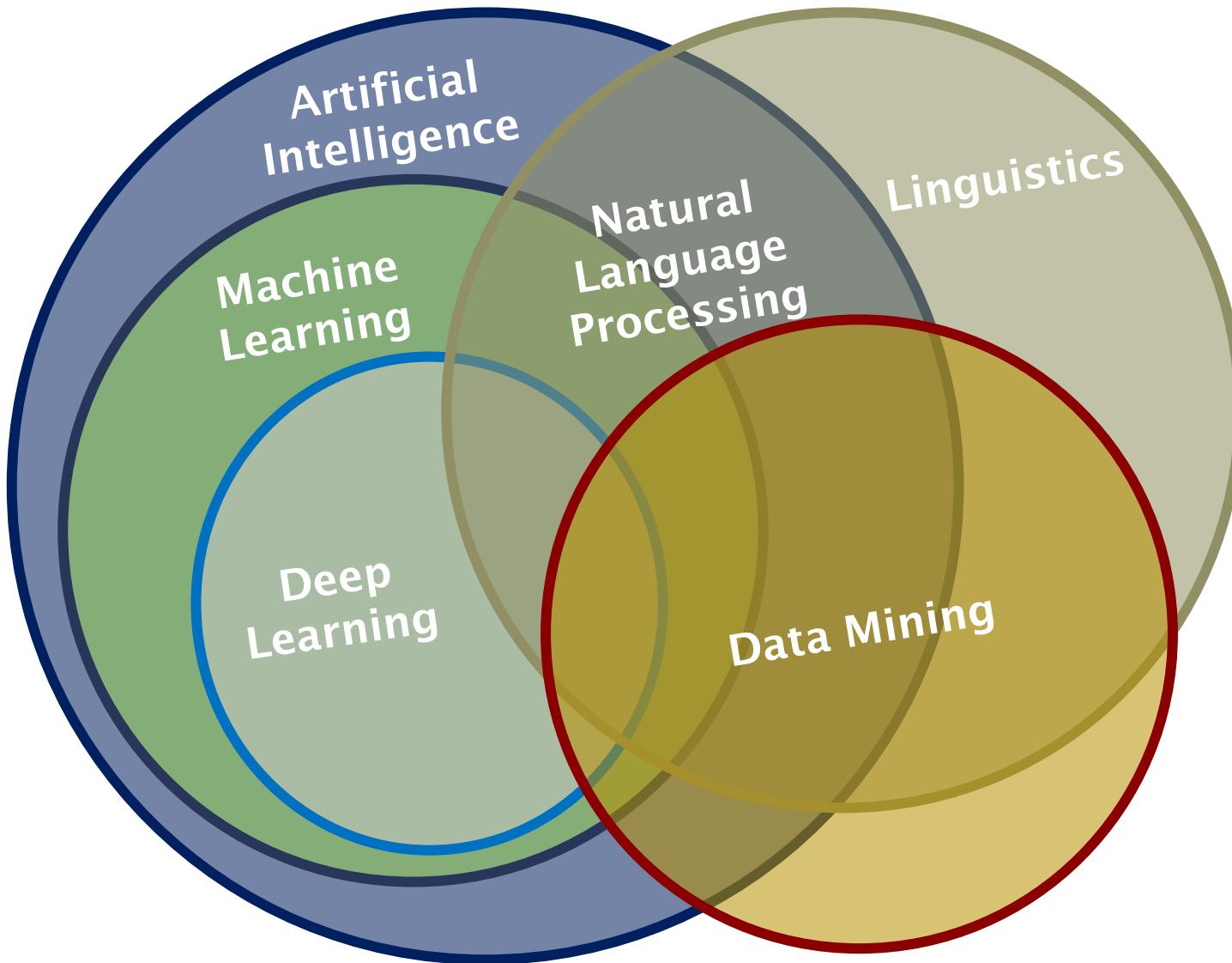
\$1,751,252 pledged of \$10,000 goal  
12,732 backers  
43 hours to go

Back this project

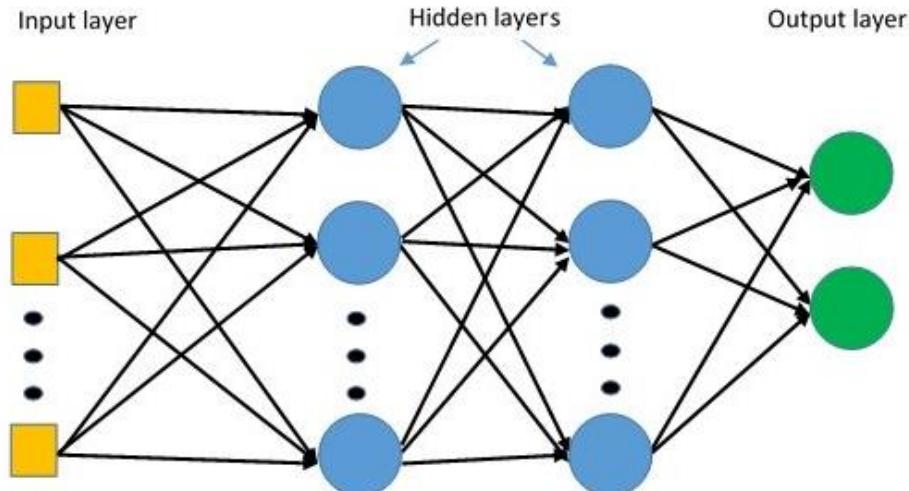
All or nothing. This project will only be funded if it reaches its goal by Fri, October 18 2019 10:55 AM EDT.



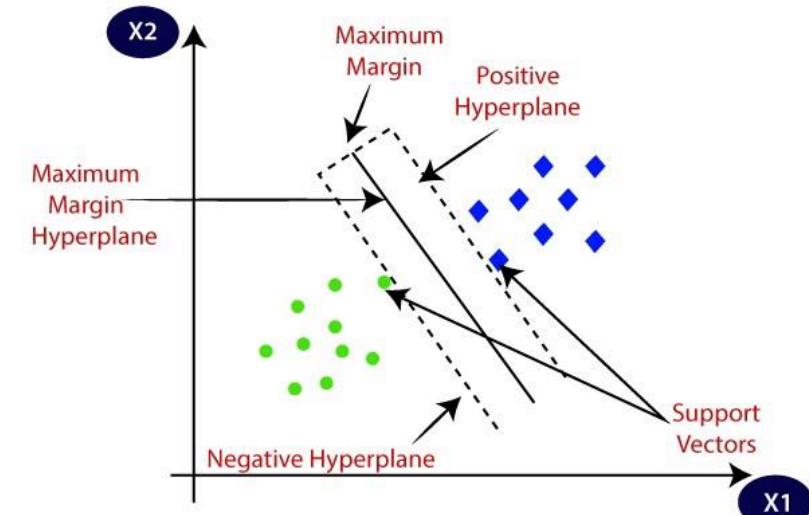
# ARTIFICIAL INTELLIGENCE



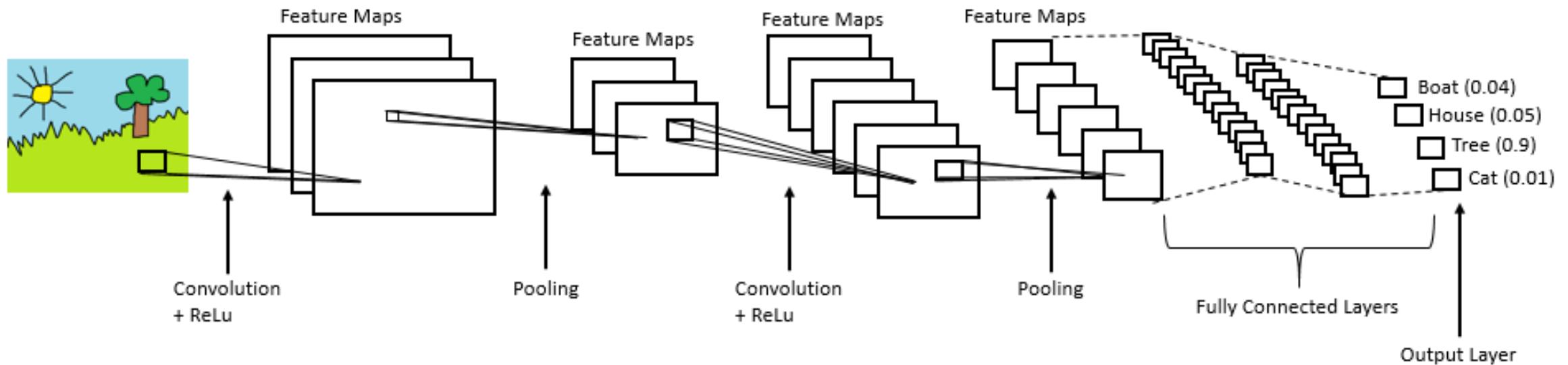
## MULTILAYER PERCEPTRON (MLP)



## SUPPORT VECTOR MACHINE (SVM)



## CONVOLUTIONAL NEURAL NETWORK (CNN)

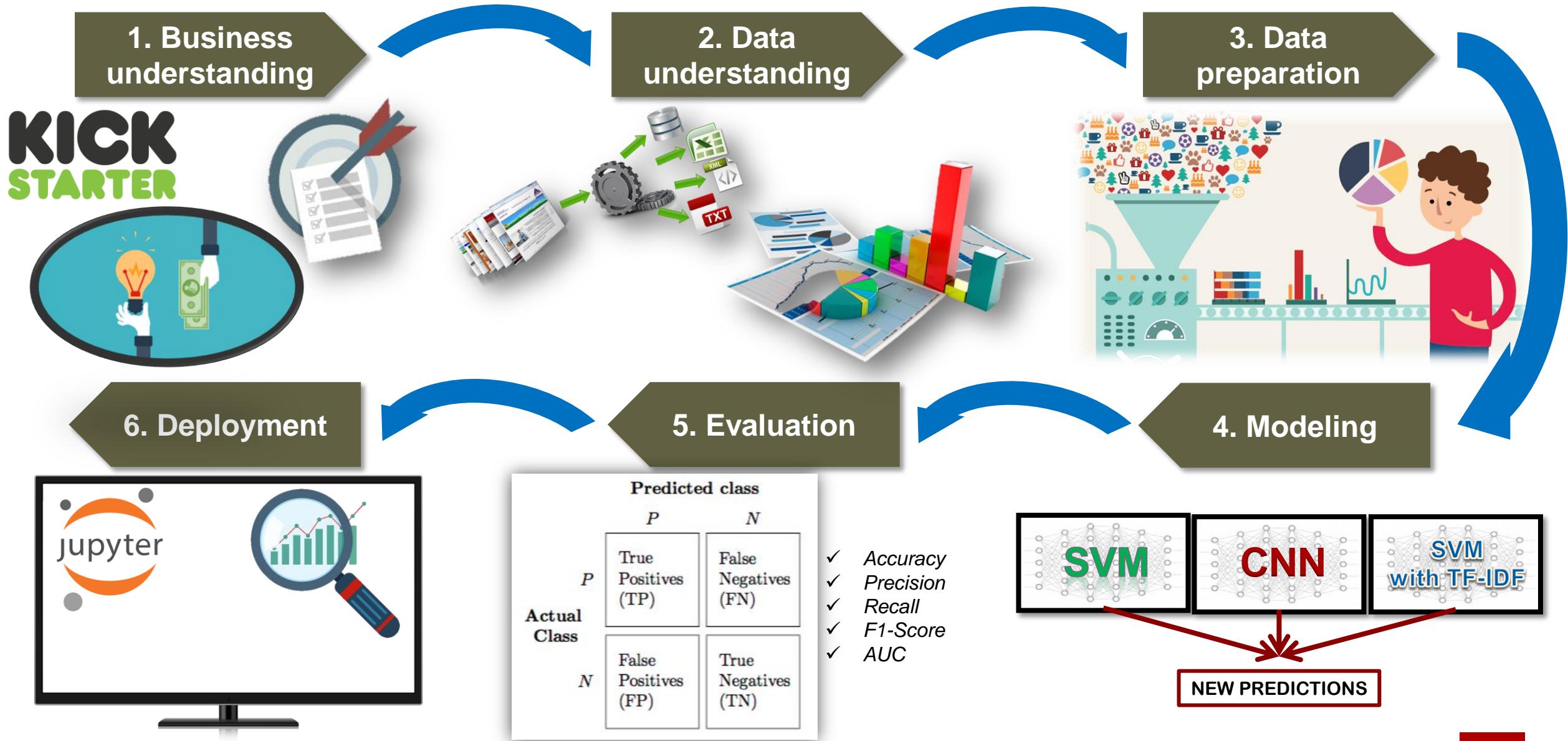


# NATURAL LANGUAGE PROCESSING (NLP)



# METHODOLOGY

# CRISP-DM

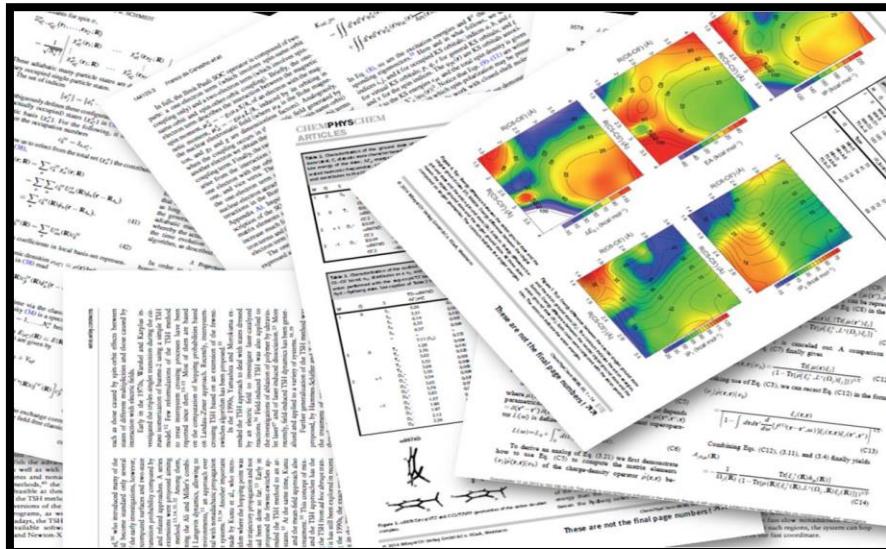


# STAGE 1: BUSINESS UNDERSTANDING

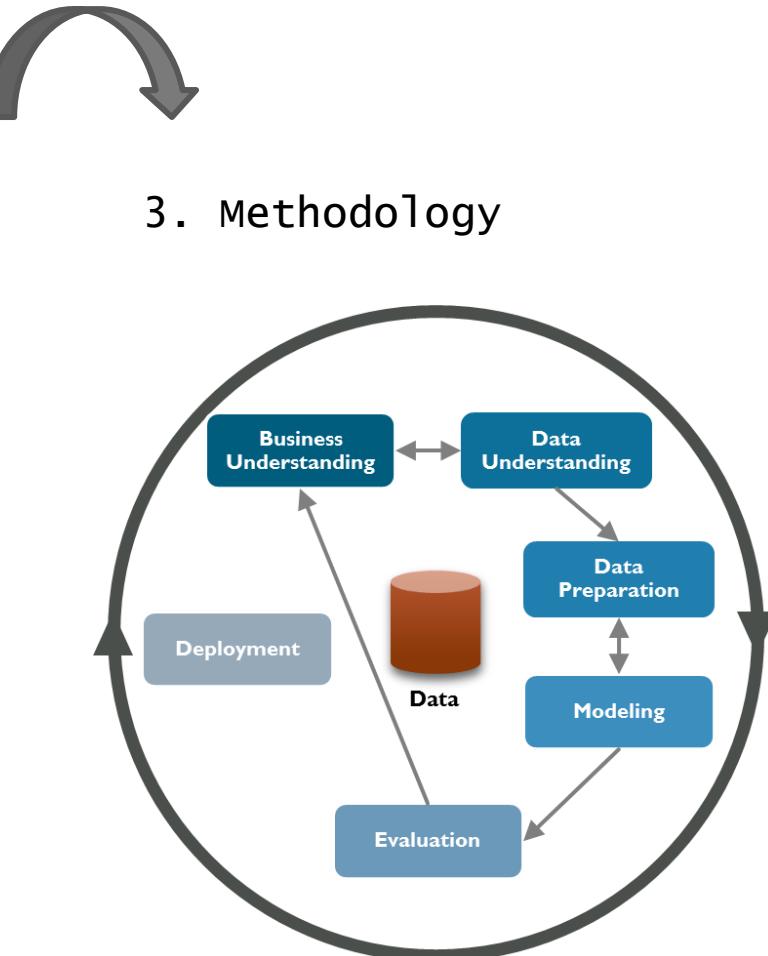
1. Problems, objectives and hypotheses



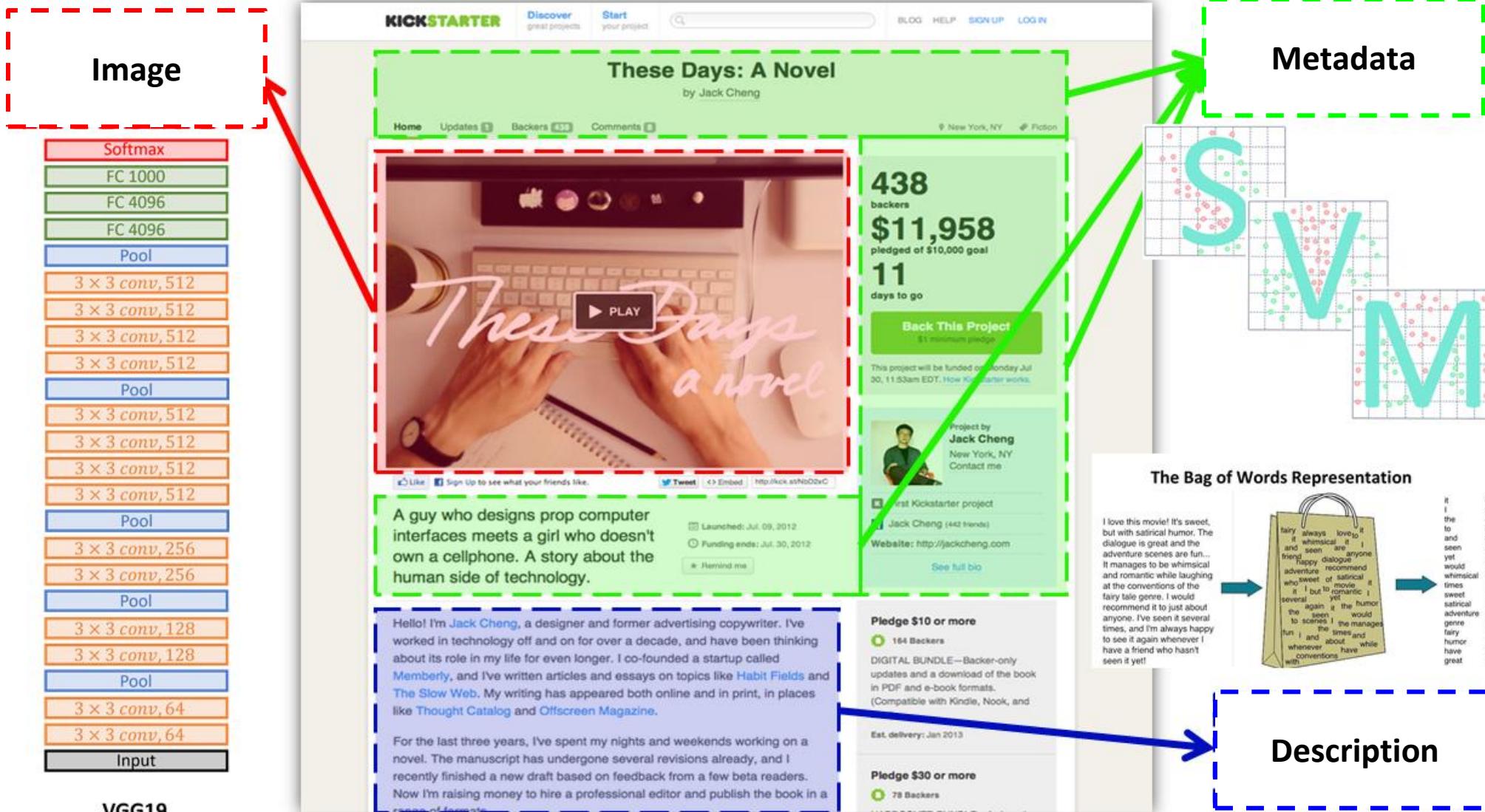
2. State-of-art



3. Methodology



# STAGE 2: DATA UNDERSTANDING



Image

Softmax
FC 1000
FC 4096
FC 4096
Pool
3 × 3 conv, 512
Pool
3 × 3 conv, 512
Pool
Input

VGG19

Metadata

The Bag of Words Representation

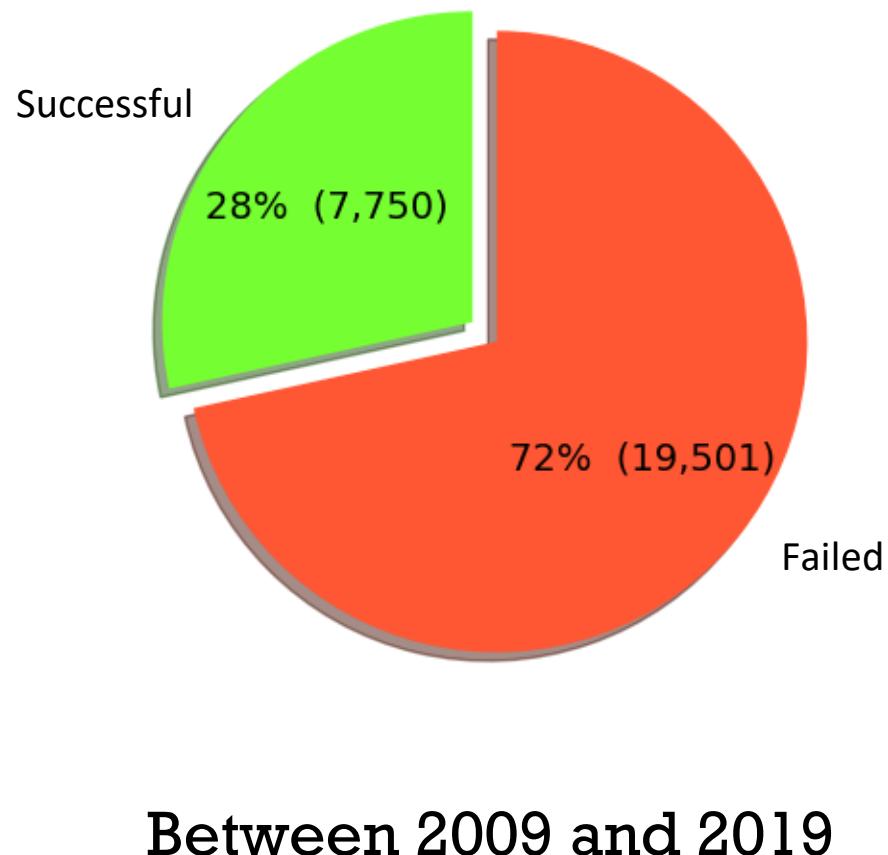


I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun.. It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

it	6
I	4
the	3
to	3
and	3
seen	2
yet	1
would	1
whimsical	1
times	1
sweet	1
satirical	1
adventure	1
genre	1
fairy	1
humor	1
have	1
great	1

Description

## STAGE 2: DATA UNDERSTANDING



### SAMPLE

Metadata			Description		
Dataset	Total	%	Dataset	Total	%
Train	21,628	80	Test	5,407	20

**% Distribution reference: Yu et al. (2018)**

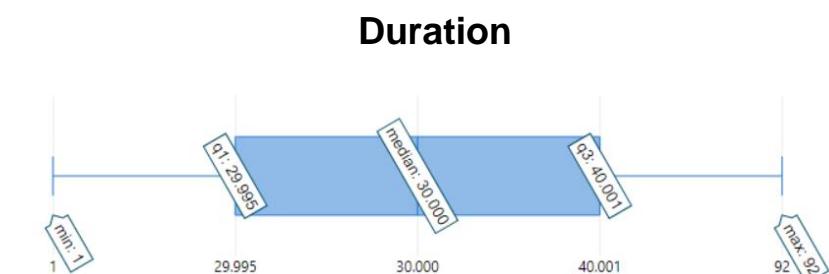
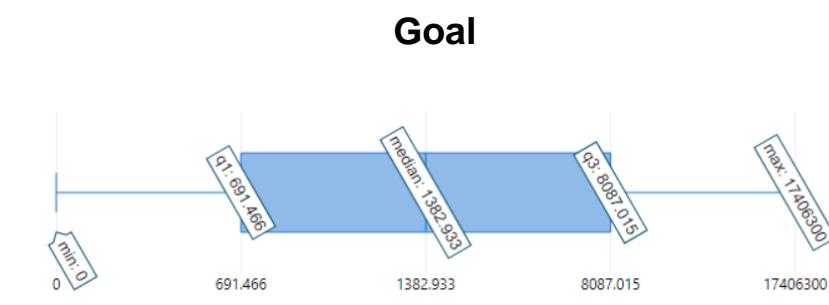
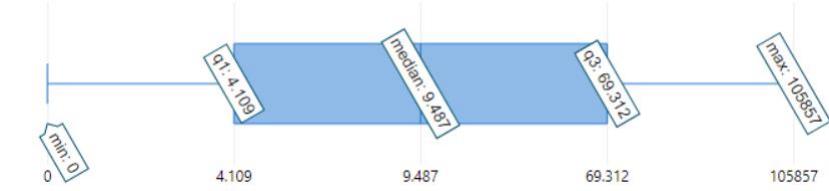
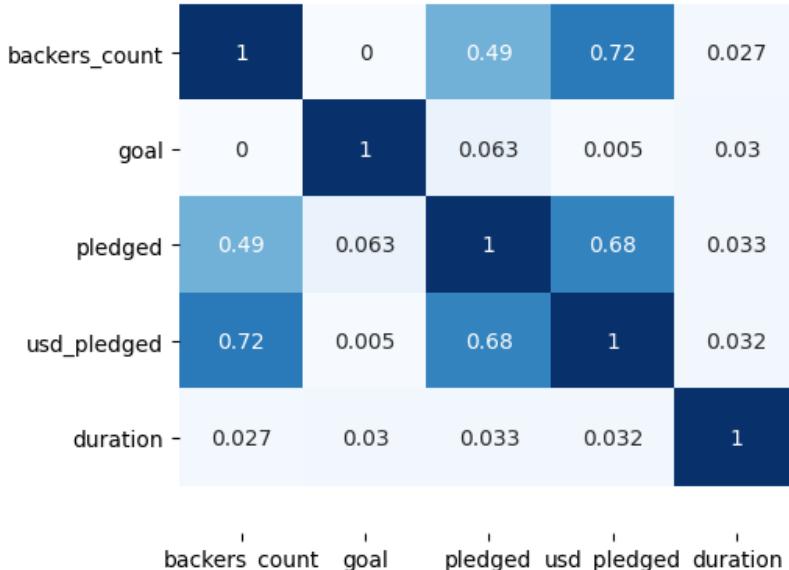
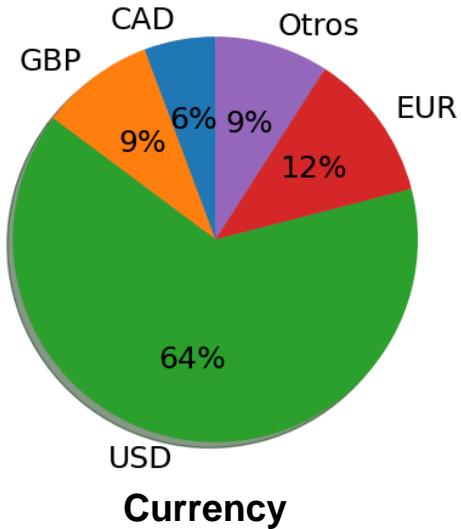
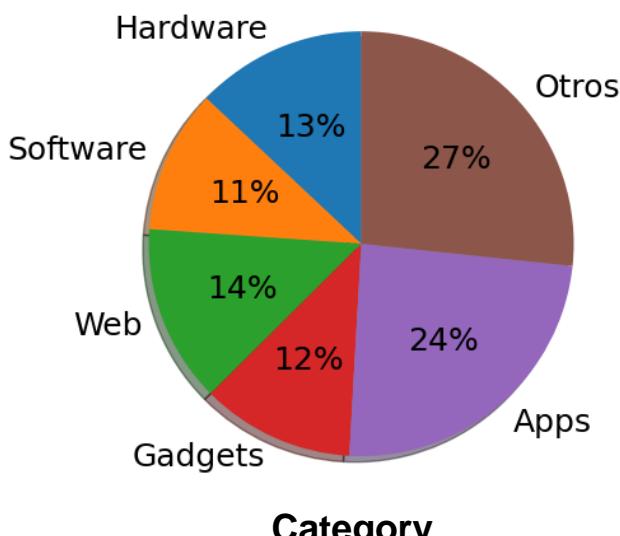
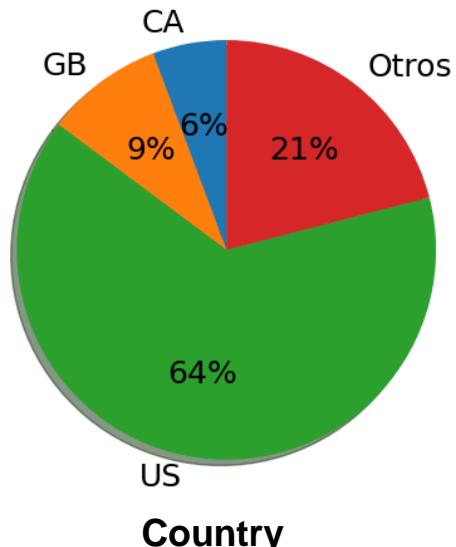
### Images

Dataset	Total	%
Train	21,628	80
Validation	2,703	10
Test	2,704	10

**% Distribution reference: Cheng et al. (2019)**

# STAGE 2: DATA UNDERSTANDING

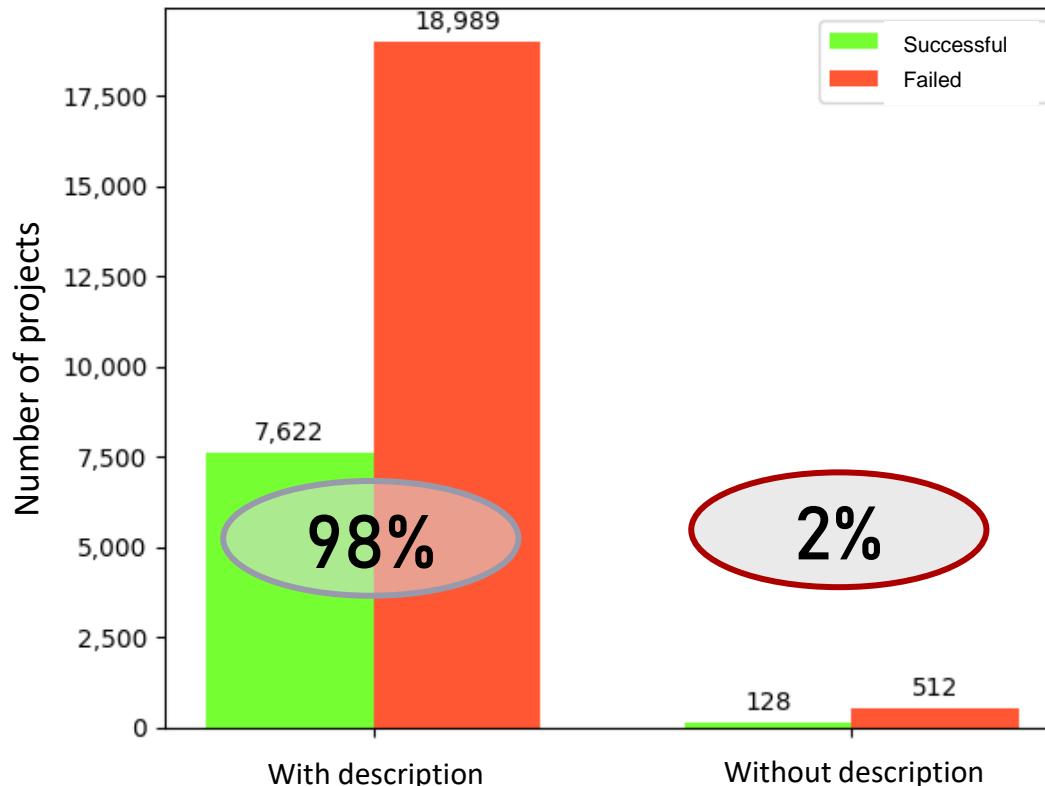
Metadata



## **STAGE 2: DATA UNDERSTANDING**

## Description

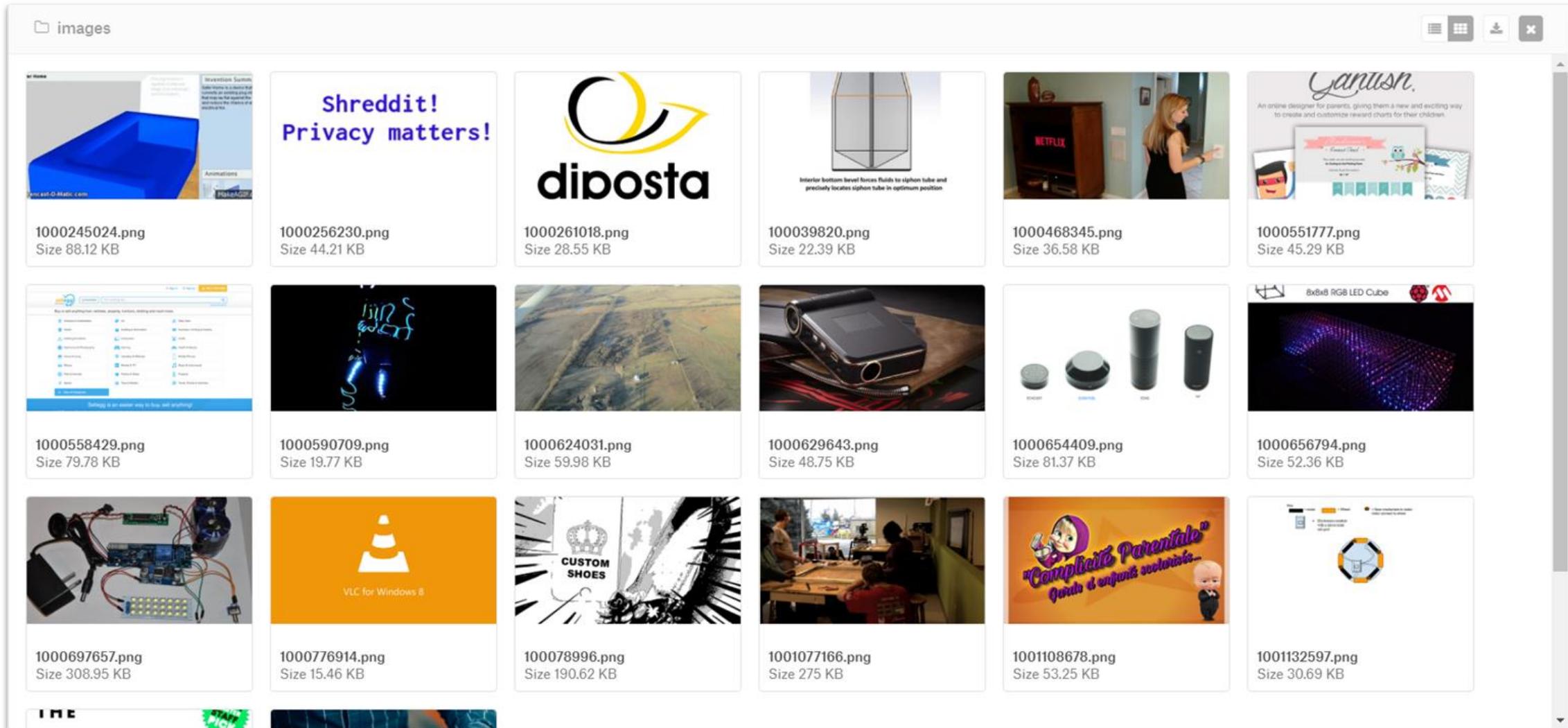
## Projects grouped by description presence and funding state



- Unique words vocabulary size: +165 K
  - Description with the longest length: +5 K

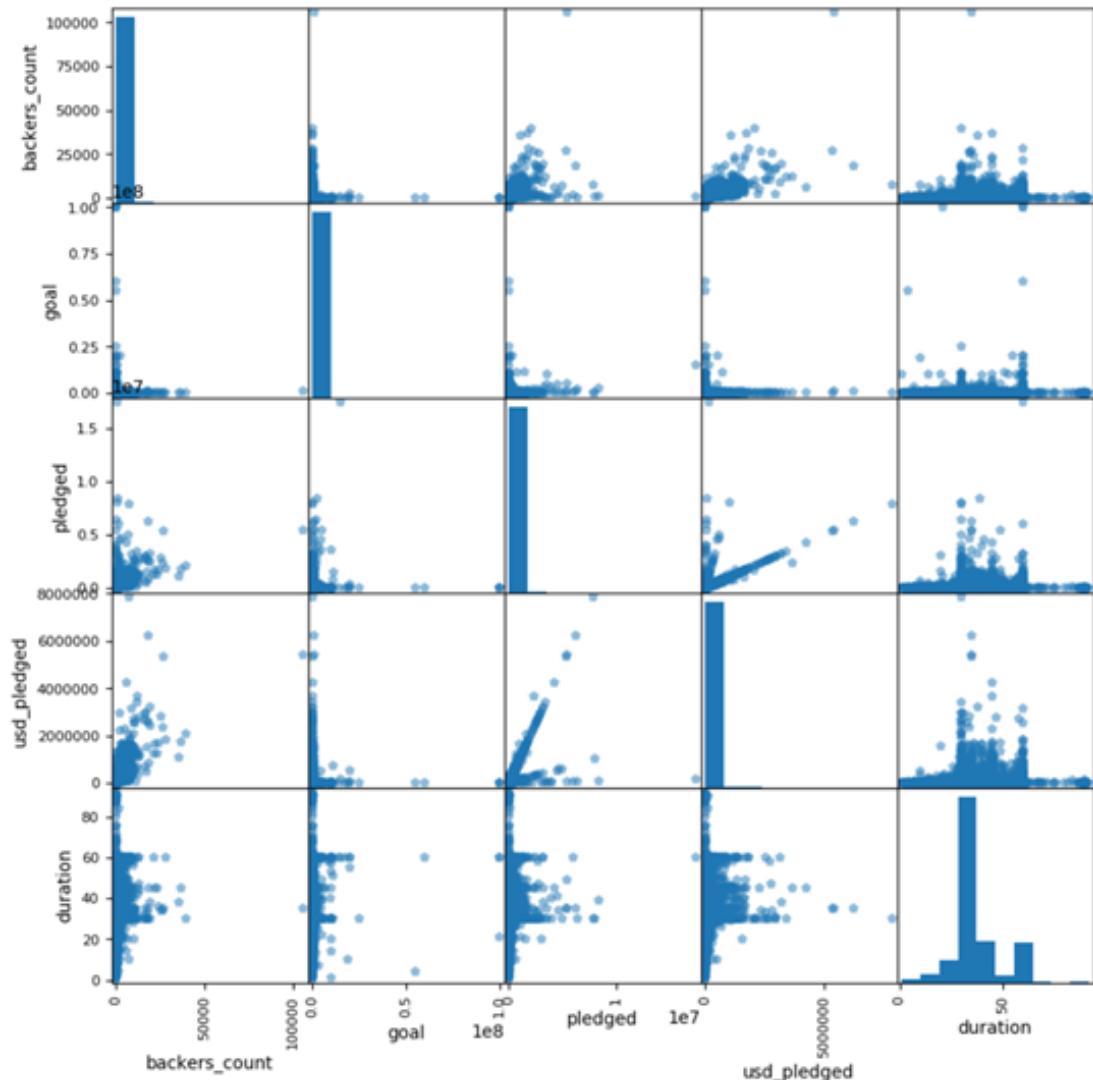
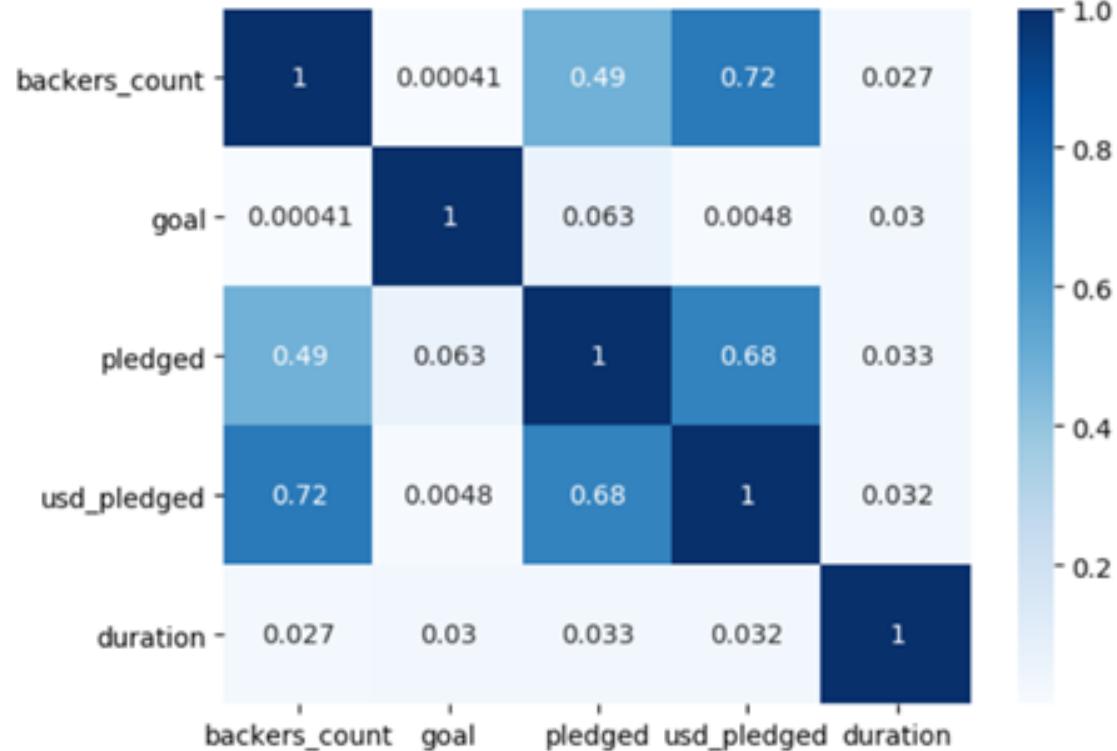
# STAGE 2: DATA UNDERSTANDING

Images



# STAGE 3: DATA PREPARATION

Metadata

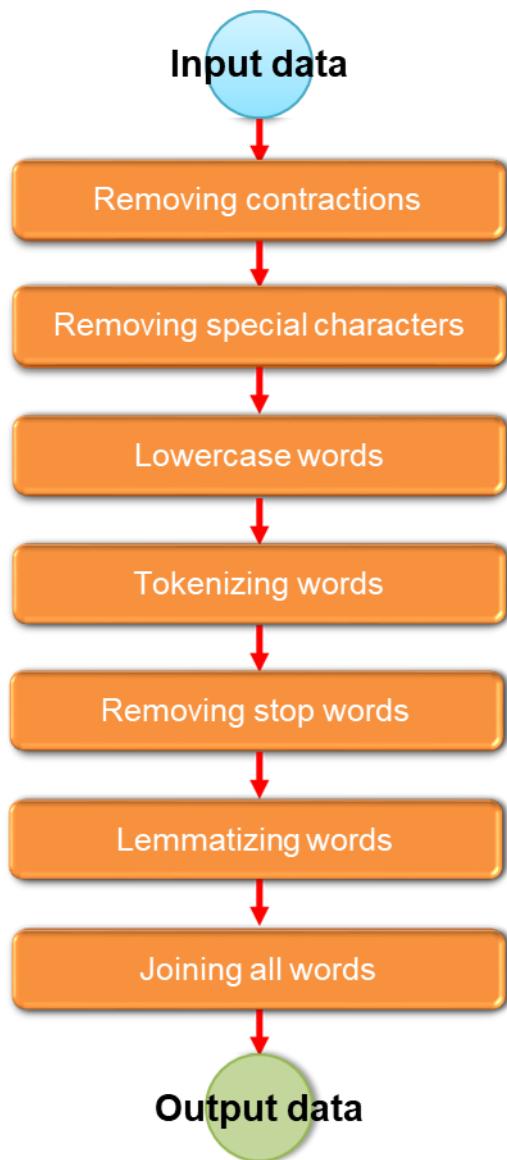


## Min-Max Scaler

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

# STAGE 3: DATA PREPARATION

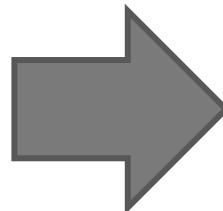
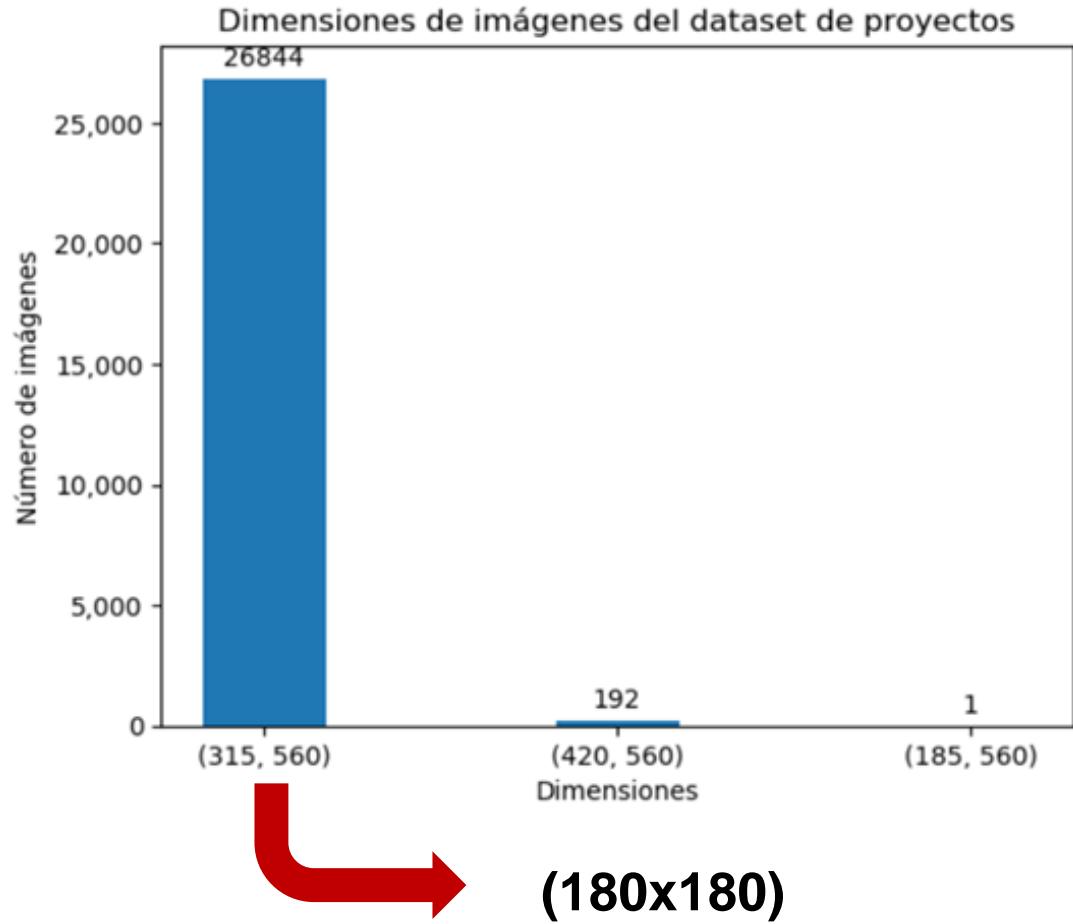
Description



	description	re_contractions	re_special_char	lower_case	tokens	re_stopwords	lemmatization	text_clean
1	I once had a Reddit account that was four years old with thousands of comments forever stored in...	I once had a Reddit account that was four years old with thousands of comments forever stored in...	I once had a Reddit account that was four years old with thousands of comments forever stored in...	i once had a reddit account that was four years old with thousands of comments forever stored in...	[i, once, had, a, reddit, account, that, was, four, years, old, with, thousands, of, comments, f...]	[reddit, account, four, years, old, thousands, comments, forever, stored, reddit, archives, cons...]	[reddit, account, four, year, old, thousand, comment, forever, store, reddit, archive, consider,...]	reddit account four year old thousand comment forever store reddit archive consider see example ...
2	Every day you go home to a mail box filled with junk and even worse, if you are a traveler it is...	Every day you go home to a mail box filled with junk and even worse, if you are a traveler it is...	Every day you go home to a mail box filled with junk and even worse if you are a traveler it is ...	every day you go home to a mail box filled with junk and even worse if you are a traveler it is ...	[every, day, you, go, home, to, a, mail, box, filled, with, junk, and, even, worse, if, you, are...]	[every, day, go, home, mail, box, fill, junk, even, bad, traveler, overflow, quickly, mail...]	[every, day, go, home, mail, box, fill, junk, even, bad, traveler, overflow, quickly, mail, hold...]	every day go home mail box fill junk even bad traveler overflow quickly mail hold mail forward r...
3	Funds needed for tooling. Once tooling is done, approximately 30-45 days, bottle production can ...	Funds needed for tooling. Once tooling is done, approximately 30-45 days, bottle production can ...	Funds needed for tooling Once tooling is done approximately days bottle production can begin P...	funds needed for tooling once tooling is done approximately days bottle production can begin p...	[funds, needed, for, tooling, once, tooling, is, done, approximately, days, bottle, production, begin, patents,...]	[funds, needed, tooling, tooling, done, approximately, days, bottle, production, begin, patent, file, pendin...]	[fund, need, tool, tool, do, approximately, day, bottle, production, begin, patent, file, pendin...]	fund need tool tool do approximately day bottle production begin patent file pending huge market...
4	The iRNinja is a wireless 4 button keypad that controls your entertainment system. It has two co...	The iRNinja is a wireless 4 button keypad that controls your entertainment system. It has two co...	The iRNinja is a wireless button keypad that controls your entertainment system It has two comp...	the irninja is a wireless button keypad controls your entertainment system it has two comp...	[the, irninja, is, a, wireless, button, keypad, controls, entertainment, system, two, components, keypad, mo...]	[irninja, wireless, button, keypad, controls, entertainment, system, two, component, keypad, moun...]	[irninja, wireless, button, keypad, control, entertainment, system, two, component, keypad, mount...]	irninja wireless button keypad control entertainment system two component keypad mount wall base...
5	Gantish is a web-based platform for parents, where they can create beautifully designed reward c...	Gantish is a web-based platform for parents, where they can create beautifully designed reward c...	Gantish is a web based platform for parents where they can create beautifully designed reward ch...	gantish is a web based platform for parents where they can create beautifully designed reward ch...	[gantish, is, a, web, based, platform, for, parents, where, they, can, create, beautifully, desi...]	[gantish, web, based, platform, parents, create, beautifully, designed, reward, charts, children...]	[gantish, web, base, platform, parent, create, beautifully, design, reward, chart, child, simple...]	gantish web base platform parent create beautifully design reward chart child simple step design...

# STAGE 3: DATA PREPARATION

Images



# VARIABLES

Modal	Variable	Detail	Data type
Metadata	Goal	Amount goal of campaign	Numeric
	Duration	Campaign duration (days)	Numeric
	Backers_count	Number of backers during campaign	Numeric
	Pledged	Pledged amount during campaign	Numeric
Description	Description	Project description	String
Images	Comments	Main image of the project campaign	Object (png)
All	State	Campaign funding state	Binary

# STAGE 4: MODELING



## MODEL 1: S.V.M.

### GridSearchCV

C	1	10	100	1000
gamma	0.1	0.01	0.001	0.0001

## MODEL 2: M.L.P

Sample weights balanced:

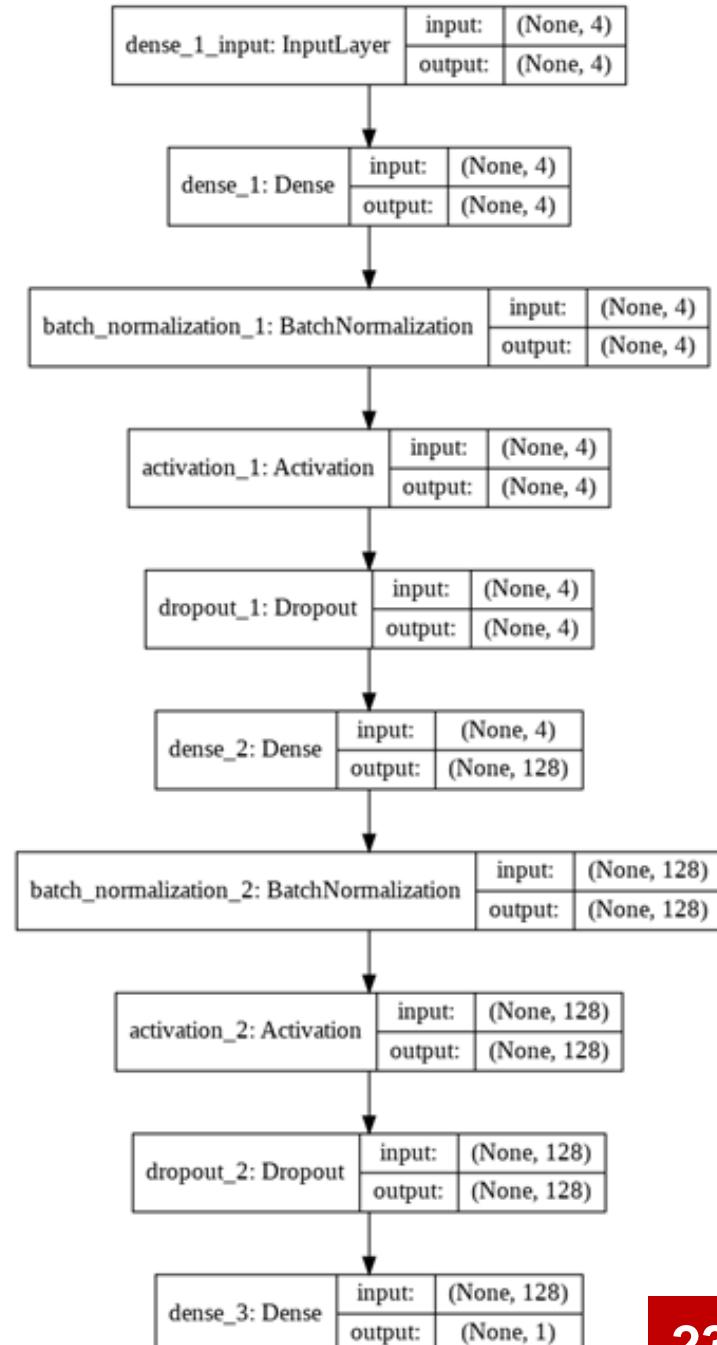
- Failed: 0.69912077
- Successful: 1.75551948

Loss = «binary\_crossentropy»

Optimizer = RMSprop

- Learning rate = 0.00001
- Rho = 0.8

Batch size = 32



# STAGE 4: MODELING

## Images

## Sample weights balanced:

- Failed: 0.69808276
  - Successful: 1.76209875

Loss = «binary\_crossentropy»

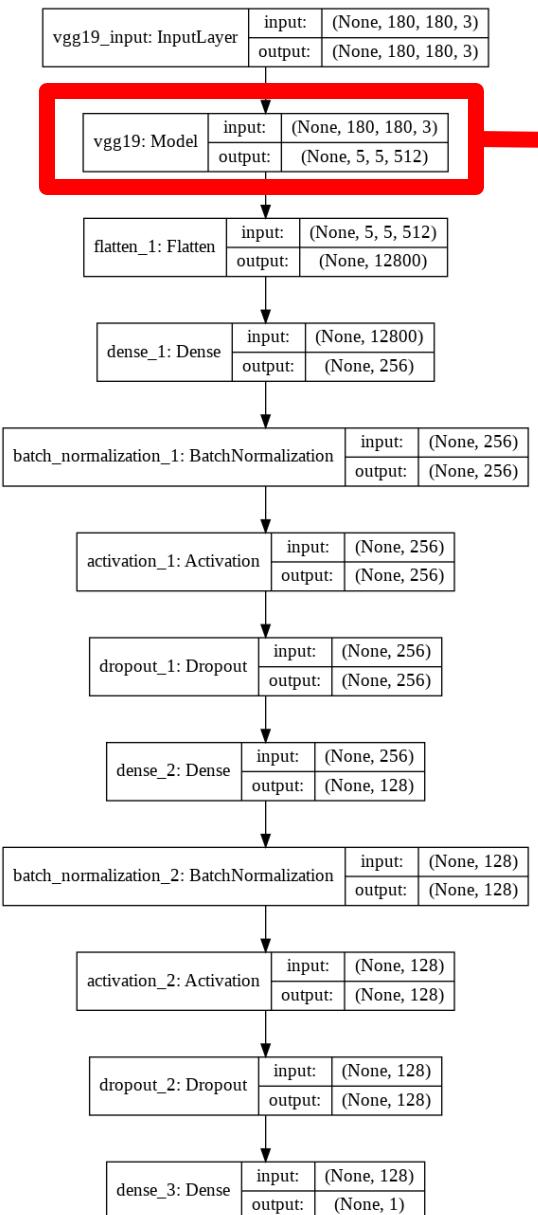
Optimizer = RMSprop

- Learning rate = 0.000001
  - Rho = 0.8

Batch size (núm. de muestras que propaga cada vez) = 64

## Épocas:

- 1st test: 200
  - 2nd test: 150



# Inception model



# STAGE 4: MODELING

## Description

I love this movie! It's sweet, but with satirical humor. The dialogue is great and the adventure scenes are fun... It manages to be whimsical and romantic while laughing at the conventions of the fairy tale genre. I would recommend it to just about anyone. I've seen it several times, and I'm always happy to see it again whenever I have a friend who hasn't seen it yet!

15

	I	love	dogs	hate	and	knitting	is	my	hobby	passion
Doc 1	1	1	1							
Doc 2	1			1	1	1				
Doc 3					1	1	1	2	1	1

**Harris (1954)**

## Bag of Words



## Term Frequency - Inverse Document Frequency

# TF-IDF

TF-IDF is a measure of originality of a word by comparing the number of times a word appears in a doc with the number of docs the word appears in.

$$\text{TF-IDF} = \text{TF}(t, d) \times \text{IDF}(t)$$

Term frequency  
Number of times term +  
appears in a doc,  $d$

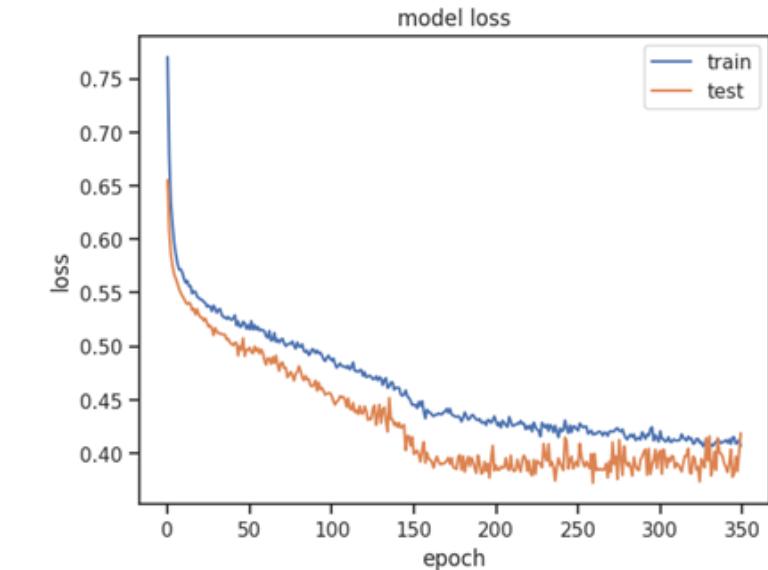
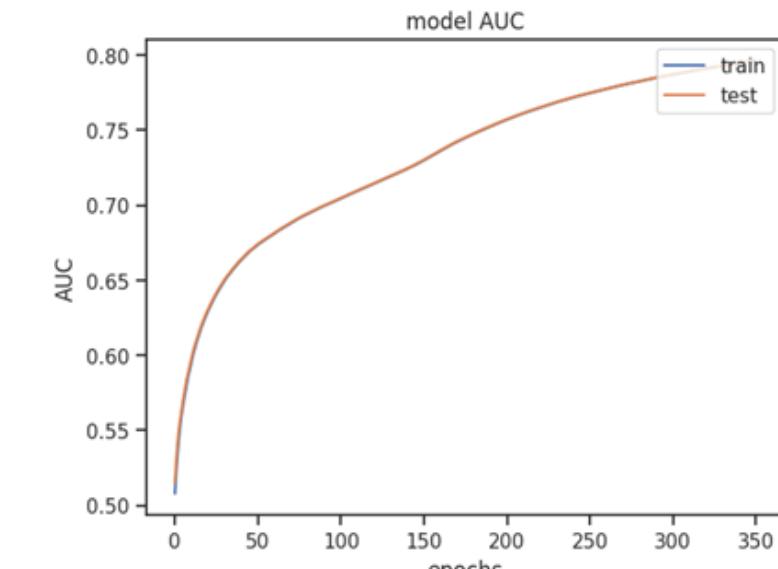
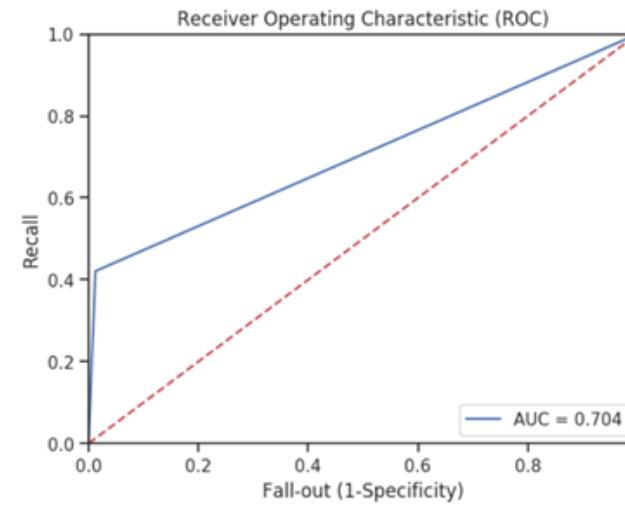
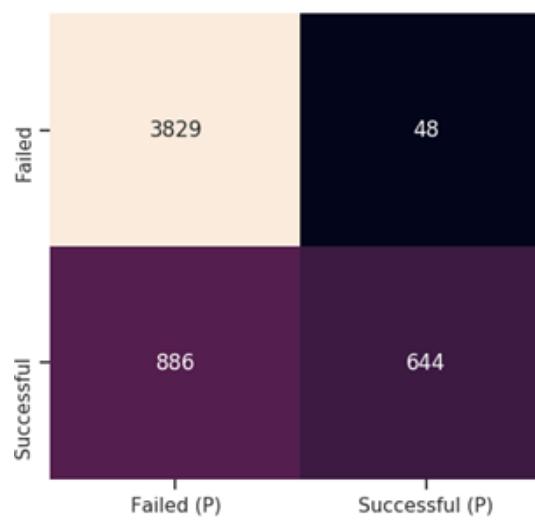
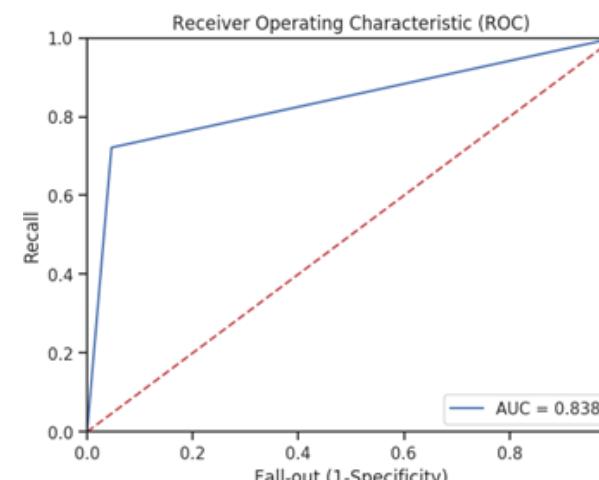
$$\log \frac{1 + \frac{n}{\text{# of documents}}}{1 + \text{df}(d, t)}$$

Document frequency  
of the term  $t$

**Luhn (1957)**

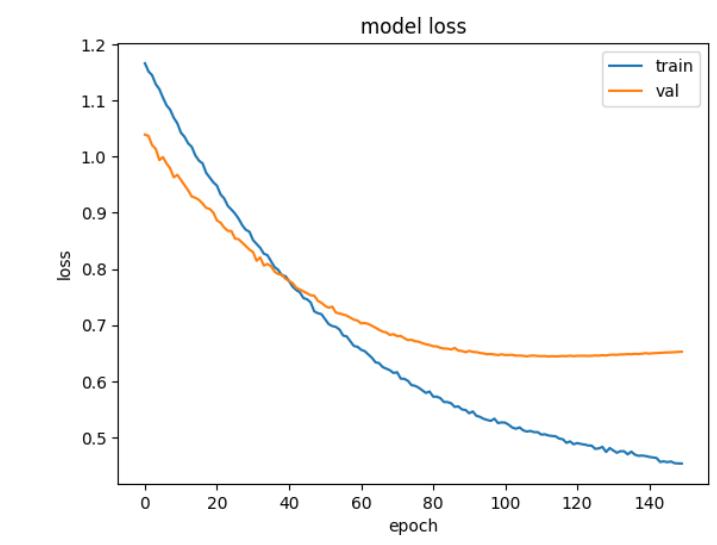
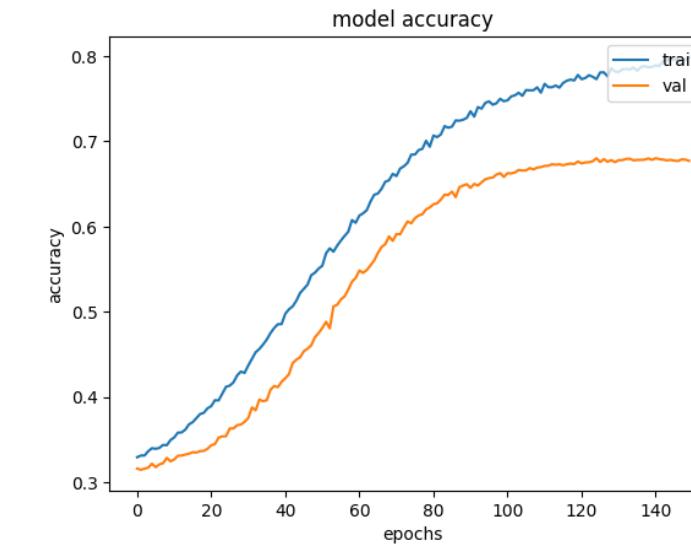
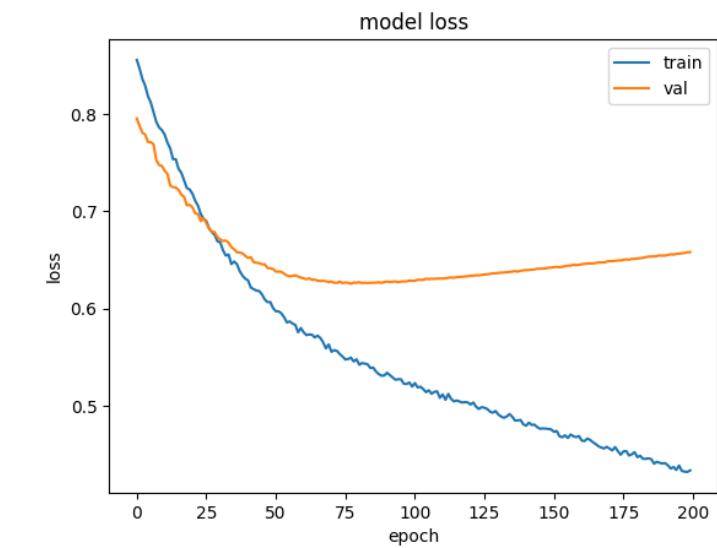
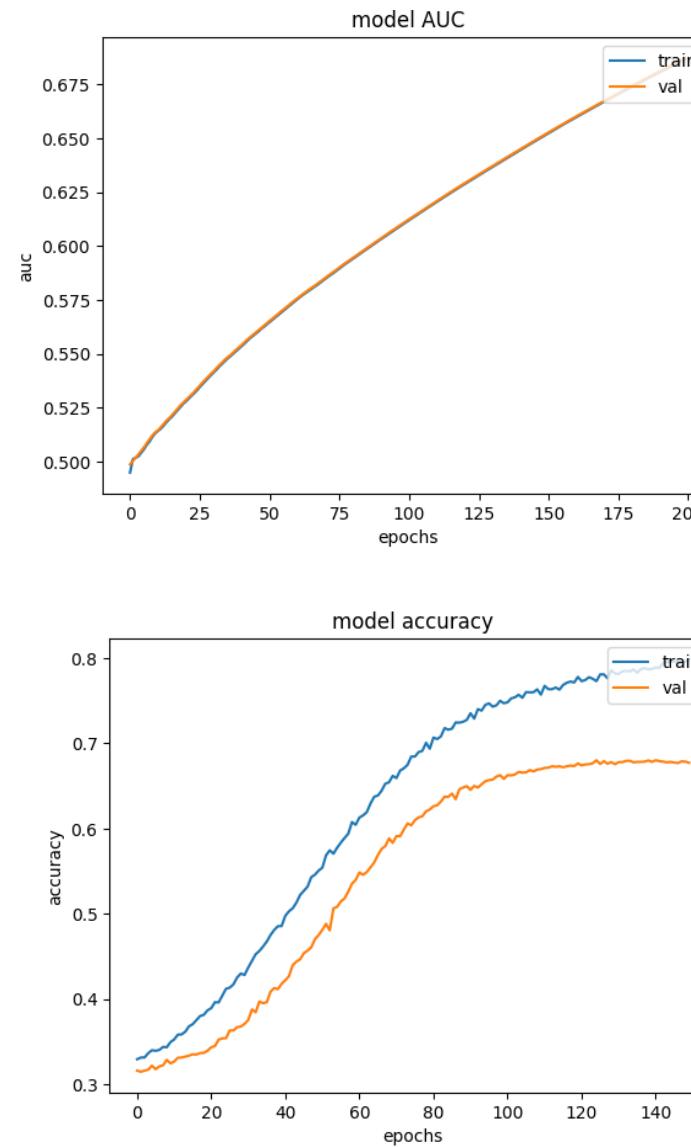
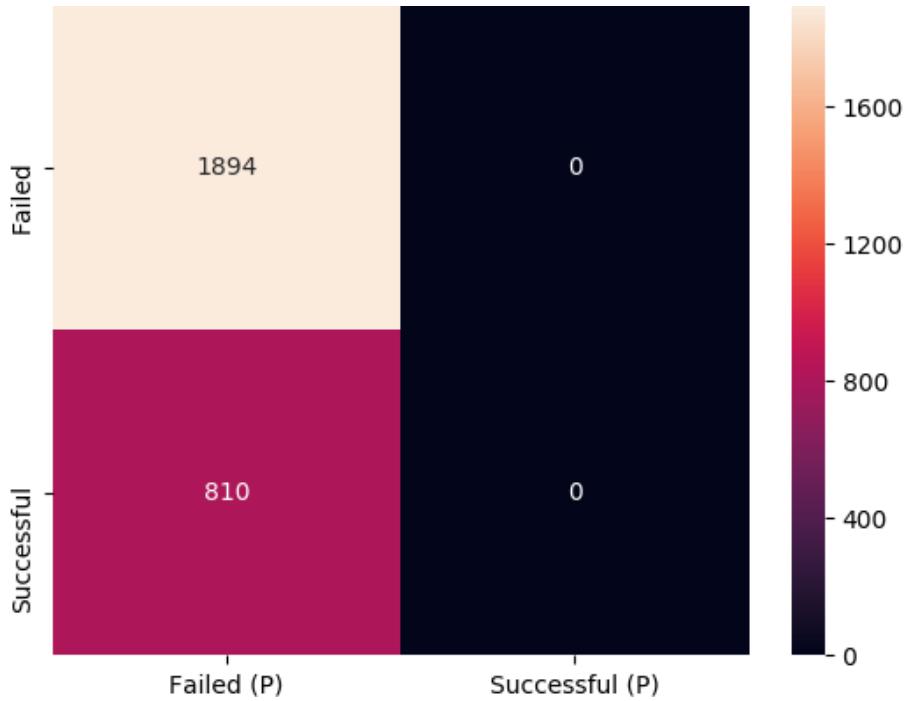
# STAGE 5: EVALUATION

Metadata



# STAGE 5: EVALUATION

Images



# STAGE 6: DEPLOYMENT



```
1/1 [=====] - 1s 875ms/step
Prediction state:      successful
Successful probability: 51.19%
```

```
['want make sell target design use people age walk life go make target want film shoot steel advertise want world see people hit target foal
'group people walk life develop platform real estate agent assist day day activity online presence platform suitable new agent well experienc
'baseline feature launch campaign raise fund project injection app mobile application enhance comic book read experience injection app rewa
'scour internet analysis card performance way crossfire sli configuration notice lack data project aim fix benchmarking various card differen
'pass first stretch goal also preview chapter upload anyone everyone check find update come everybody let u learn along stretch goal add le
```

```
print('Prediction state:\t %s\nSuccessful probability:\t %.2f%%' %\
      (predict_description_tfidf[0],
       (1-description_model_tfidf.predict_proba(tfidf_new)[0][0])*100))

Prediction state:      successful
Successful probability: 50.99%
```

```
print("Número de patrocinadores: ")
backer = input()
backer = float(backer)
print("Meta: ")
goal = input()
goal = float(goal)
print("Monto invertido: ")
pledge = input()
pledge = float(pledge)
print("Duración de la campaña: ")
duration = input()
duration = float(duration)
```

```
Número de patrocinadores:
140
Meta:
200000
Monto invertido:
32728
Duración de la campaña:
40
```

```
print('Prediction state:\t %s\nSuccessful probability:\t %.2f%%' %\
      (predict_metadata_svm[0],
       (1-loaded_model_metadata_svm.predict_proba(array_prueba.reshape(-1,4))[0][0])*100))

Prediction state:      failed
Successful probability: 0.00%
```

```
['want make sell target design use people age walk life go make target want film shoot steel advertise want world see people hit target foal
'group people walk life develop platform real estate agent assist day day activity online presence platform suitable new agent well experienc
'baseline feature launch campaign raise fund project injection app mobile application enhance comic book read experience injection app rewa
'scour internet analysis card performance way crossfire sli configuration notice lack data project aim fix benchmarking various card differen
'pass first stretch goal also preview chapter upload anyone everyone check find update come everybody let u learn along stretch goal add le
```

```
print('Prediction state:\t %s\nSuccessful probability:\t %.2f%%' %\
      (predict_description_bow[0],
       (1-description_model_bow.predict_proba(bow_new)[0][0])*100))

Prediction state:      successful
Successful probability: 42.67%
```

# RESULTS

**Metadata**

	Support Vector Machine	Multilayer Perceptron
<b>Accuracy</b>	0.888108	0.827261
<b>AUC</b>	0.837699	0.704267
<b>Recall</b>	0.721569	0.420915
<b>F1-Score</b>	0.784927	0.579658
<b>Gini</b>	0.675399	0.408534
<b>Execution time</b>	00:00:16.218909	00:39:30.152152

**Images**

Class	Precision	Recall	F1-Score
<b>Failed</b>	0.70	1.00	0.82
<b>Successful</b>	0.00	0.00	0.00

**Description**

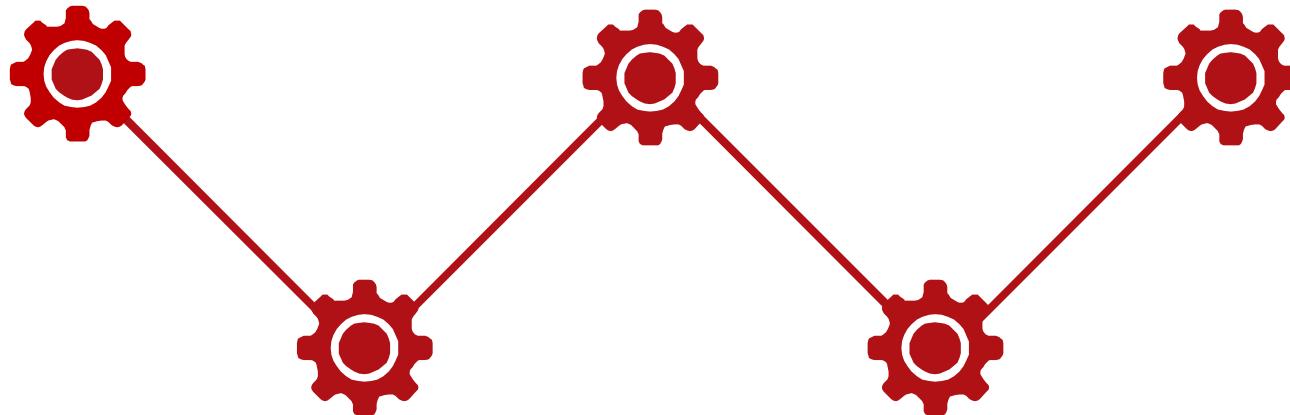
	Accuracy	Training time (seg.)
<b>BoW</b>	0.7321990012946181	7,396.74
<b>TF-IDF</b>	0.7540225633438136	3,933.83

# CONCLUSIONS

2 OF 3 FINAL MODELS  
REACHED ACCURACY  
HIGHER THAN 0.70

MODELS COULD ALSO  
BE USED FOR OTHER  
CATEGORIES

CONTENT FROM SOCIAL  
MEDIA AND NO-SQL  
SHOULD BE CONSIDER



RESULTS WERE HIGHER  
THAN BASELINE

IMAGES MODEL'S  
PARAMETERS MUST BE  
TUNED WITH OTHER  
TECHNIQUES



# QUESTIONS?

## REFERENCES

- Banafa, A. (2019). ¿Qué es el aprendizaje Profundo? *OpenMind*. Obtenido de <https://www.bbvaopenmind.com/tecnologia/mundo-digital/que-es-el-aprendijaze-profundo/>
- Cheng, C., Tan, F., Hou, X., & Wei, Z. (2019). Success Prediction on Crowdfunding with Multimodal Deep Learning. *The Twenty-Eighth International Joint Conference on Artificial Intelligence*, (págs. 2158-2164). Macao. Obtenido de <https://www.ijcai.org/proceedings/2019/0299.pdf>
- Gandhi, R. (7 de junio de 2018). Support Vector Machine — Introduction to Machine Learning Algorithms. *Towards Data Science*. Obtenido de <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- Inzaugarat, E. (30 de octubre de 2018). Understanding Neural Networks: What, How and Why? *Towards Data Science*. Obtenido de <https://towardsdatascience.com/understanding-neural-networks-what-how-and-why-18ec703ebd31>
- Jin, B., Zhao, H., Chen, E., Liu, Q., & Ge, Y. (2019). Estimating the Days to Success of Campaigns in Crowdfunding: A Deep Survival Perspective. *The 33rd AAAI Conference on Artificial Intelligence (AAAI'2019)*. Honolulu: Association for the Advancement of Artificial. Recuperado el Junio de 2019, de [http://staff.ustc.edu.cn/~cheneh/paper\\_pdf/2019/Binbin-Jin-AAAI.pdf](http://staff.ustc.edu.cn/~cheneh/paper_pdf/2019/Binbin-Jin-AAAI.pdf)
- Kickstarter. (s.f.). Acerca de nosotros: Kickstarter. Obtenido de Kickstarter: <https://www.kickstarter.com/about?ref=global-footer>
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook* (Primera ed.). Nueva York: Springer.
- Microsoft. (8 de enero de 2019). Conceptos de minería de datos. Obtenido de Microsoft Docs: <https://docs.microsoft.com/es-es/sql/analysis-services/data-mining/data-mining-concepts?view=sql-server-2017>
- Poole, D., Mackworth, A., & Goebel, R. (1998). Computational Intelligence: A Logical Approach. *Oxford University Press*. Recuperado el 14 de mayo de 2019
- Prabhu, R. (4 de Marzo de 2018). Understanding of Convolutional Neural Network (CNN) — Deep Learning. *Medium*. Obtenido de <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>
- Redacción Gestión. (1 de agosto de 2018). Perú es el tercer país con mayor cantidad de emprendimientos en fase temprana a nivel mundial. *Diario Gestión*. Obtenido de <https://gestion.pe/economia/peru-tercer-pais-mayor-cantidad-emprendimientos-fase-temprana-nivel-mundial-240264>
- Russell, S., & Norvig, P. (2009). *Inteligencia Artificial: Un Enfoque Moderno* (Tercera ed.). Prentice Hall.
- Universo Crowdfunding. (s.f.). ¿Qué es el crowdfunding? Obtenido de Universo Crowdfunding: <https://www.universocrowdfunding.com/que-es-el-crowdfunding/>
- Yu, P.-F., Huang, F.-M., Yang, C., Liu, Y.-H., Li, Z.-Y., & Tsai, C.-H. (Edits.). (2018). Prediction of Crowdfunding Project Success with Deep Learning. *2018 IEEE 15th International Conference on e-Business Engineering (ICEBE)* (págs. 1-8). Xi'an: IEEE. doi:10.1109/ICEBE.2018.00012
- Yuan, H., Lau, R. Y., & Xu, W. (noviembre de 2016). The determinants of crowdfunding success: A semantic text analytics approach. *Decision Support Systems*, 91, 67-76. doi:10.1016/j.dss.2016.08.001
- Zhou, H. (2017). *Predicting the Success of Kickstarter Campaigns*. Proyecto final, University of California, Department of Statistics, Berkeley. Obtenido de [https://www.stat.berkeley.edu/~aldous/157/Old\\_Projects/Haochen\\_Zhou.pdf](https://www.stat.berkeley.edu/~aldous/157/Old_Projects/Haochen_Zhou.pdf)