



Analyzing the Successfulness of Crowdfunding Projects on Kickstarter





Index

- Introduction
- Project Flow
- Exploratory Data Analysis
- Results
- Conclusion & Future Work

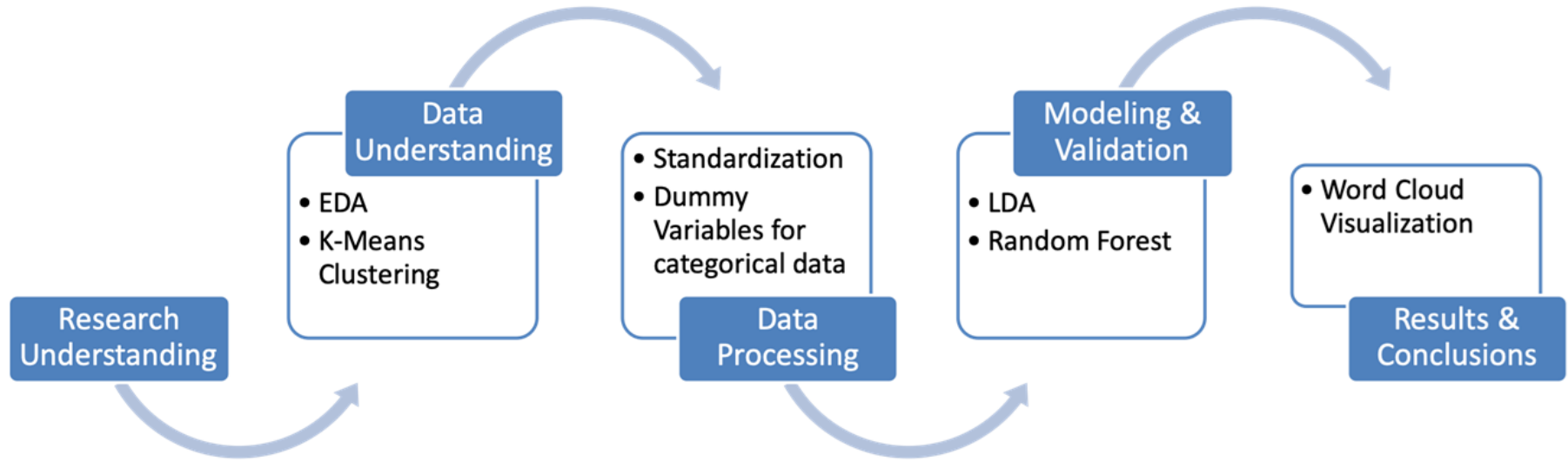


Introduction

- Kickstarter is the world's largest funding platform for creative projects, it works on an all-or-nothing basis i.e if a project doesn't meet its goal by the deadline, the project owner gets nothing.
- Our project digs into which specific characters lead to a successful fundraising for a project, hoping to provide a reference for the Kickstarter community.
- Data Source: Kaggle- kickstarter-projects
- Keywords: Data Mining, Clustering, Classification, R, Tableau



Project Flow





Dataset at a Glance

id	name	category	main_category	currency	deadline	goal	launched	pledged	state	backers	country	usd pledged	usd_pledged	usd_goal_real
1000002330	The Songs of Poetry	Poetry	Publishing	GBP	10/9/15	1000	8/11/15 12:12	0	failed	0	GB	0	0	1533.95
1000003930	Greeting From	Narrative Film	Film & Video	USD	11/1/17	30000	9/2/17 4:43	2421	failed	15	US	100	2421	30000
1000004038	Where is Happiness	Narrative Film	Film & Video	USD	2/26/13	45000	1/12/13 0:20	220	failed	3	US	220	220	45000
1000007540	ToshiCapital	Music	Music	USD	4/16/12	5000	3/17/12 3:24	1	failed	1	US	1	1	5000
1000011046	Community Film	Film & Video	Film & Video	USD	8/29/15	19500	7/4/15 8:35	1283	canceled	14	US	1283	1283	19500
1000014025	Monarch Escape	Restaurants	Food	USD	4/1/16	50000	2/26/16 13:38	52375	successful	224	US	52375	52375	50000
1000023410	Support Solar	Food	Food	USD	12/21/14	1000	12/1/14 18:30	1205	successful	16	US	1205	1205	1000
1000030581	Chaser Strips	Drinks	Food	USD	3/17/16	25000	2/1/16 20:05	453	failed	40	US	453	453	25000
1000034518	SPIN - Premium	Product Design	Design	USD	5/29/14	125000	4/24/14 18:14	8233	canceled	58	US	8233	8233	125000
100004195	STUDIO IN THE	Documentary	Film & Video	USD	8/10/14	65000	7/11/14 21:55	6240.57	canceled	43	US	6240.57	6240.57	65000

378661 Rows X 15 Variables

Target Variable: state (successful, failed, canceled, suspended, live, undefined)

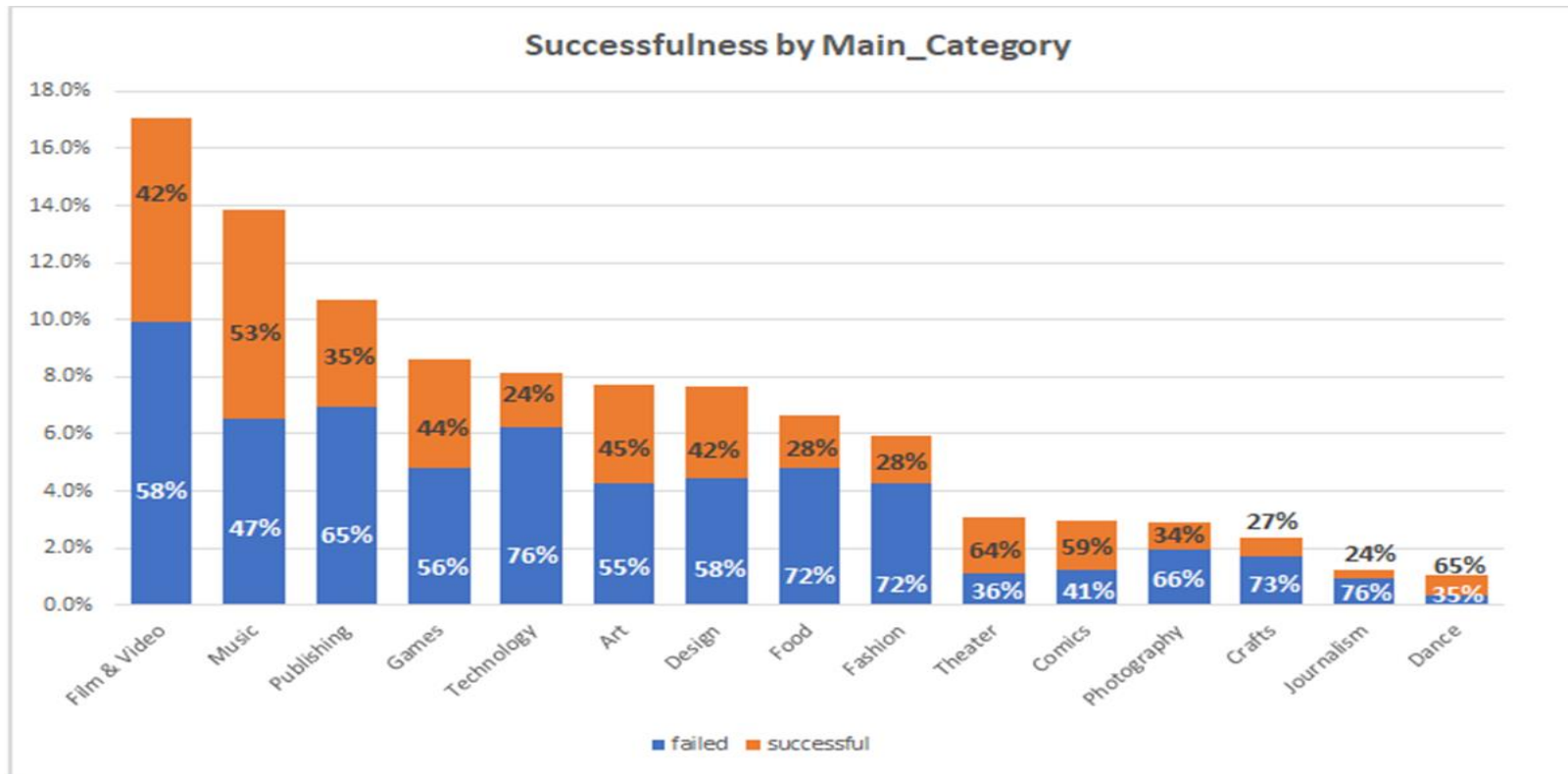


Exploratory Data Analysis

- The projects are divided into 15 main categories and 159 sub- categories. Among all the variables, pledged, goal, backers (supporters) and days of duration are the most important numerical variables.



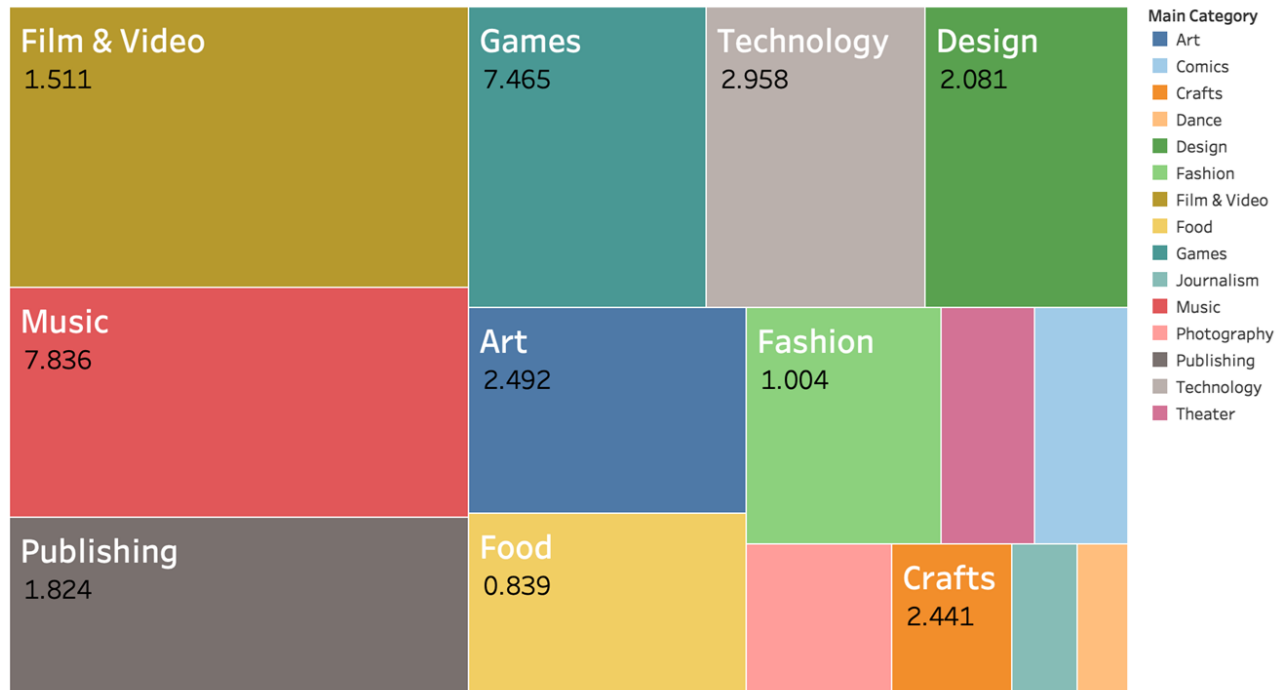
Exploratory Data Analysis





Exploratory Data Analysis

Pledged/Goal Ratio By Size of Main_Category




Main Category and average of Pledged_Goal_Ratio. Color shows details about Main Category. Size shows sum of Number of Records. The marks are labeled by Main Category and average of Pledged_Goal_Ratio.



Cluster Analysis: K - Means (k = 5)

While Cluster#1 & #4 represented the crowded majority, Cluster#2,#3, #5 was perfectly separated.

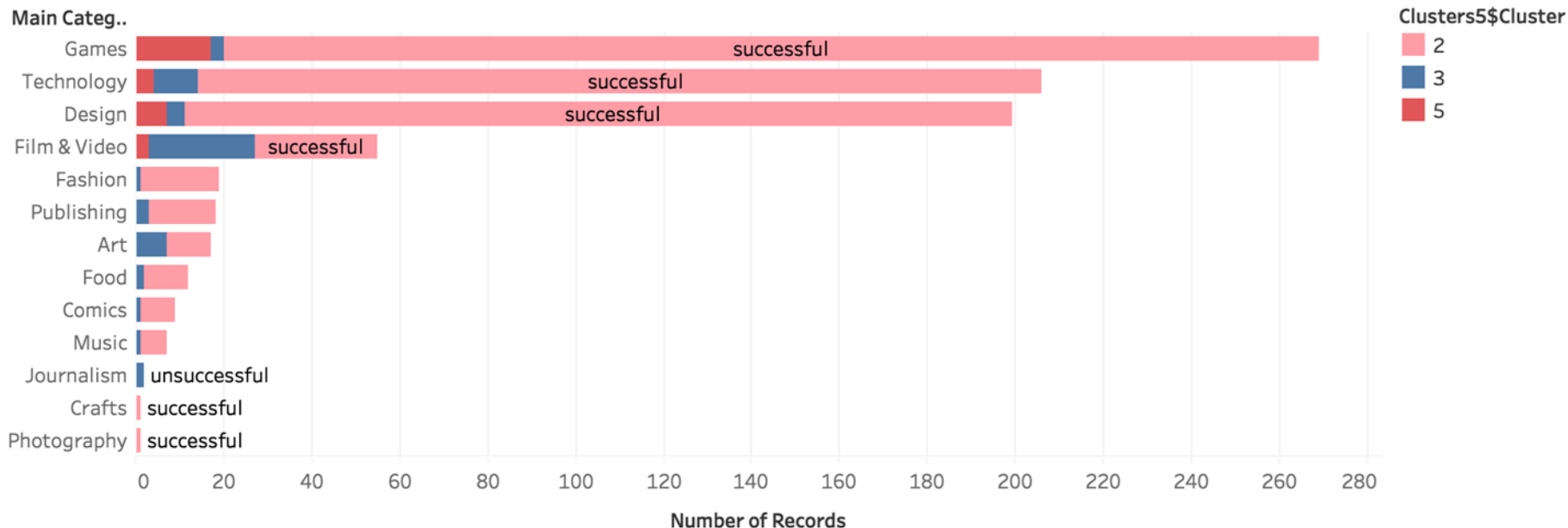
Successfulness By Cluster

state_cal	Clusters5\$Cluster					% of Total Number of R..
	1	2	3	4	5	
successful	15,402 21.53%	707 97.38%		117,816 38.46%	31 100.00%	 2.62%100.00%
unsuccessful	56,123 78.47%	19 2.62%	58 100.00%	188,505 61.54%		



Cluster Analysis: K - Means (k = 5)

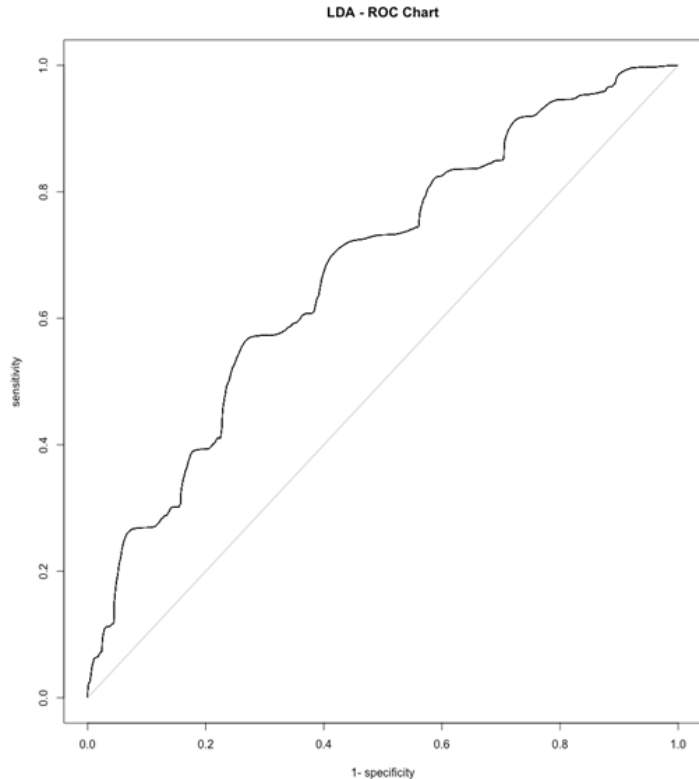
Cluster2, 3, 5 By Main_Category



Sum of Number of Records for each Main Category. Color shows details about Clusters5\$Cluster. The marks are labeled by state_cal. The view is filtered on Clusters5\$Cluster, which keeps 2, 3 and 5.



Results



LDA

The result of LDA yields an accuracy of 66.25%. Among all other variables, the coefficient of Technology, Dance, Theater & Journalism ranked top.



Results

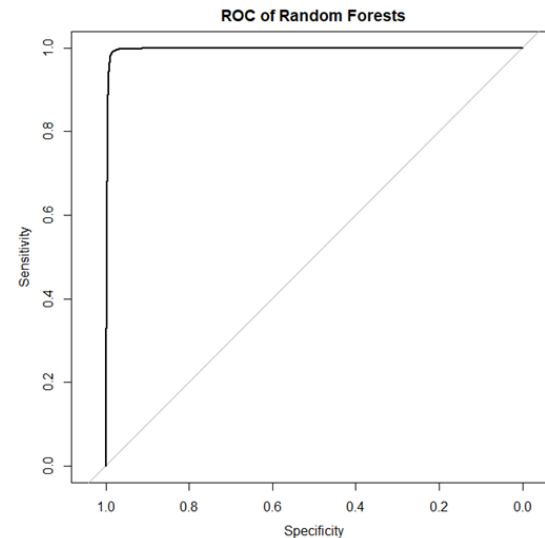
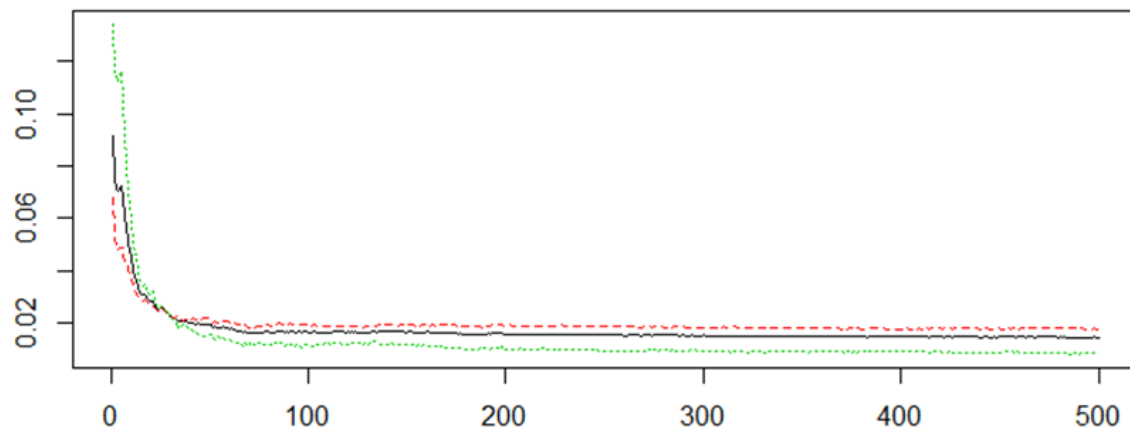
Random Forest

As the amount of trees increased, the error of the model decreased and stabilized at 0.02%. The most important variables are Pledged, Backers, Goal, and Days. Moreover, some categories have larger impact than others. Accuracy = 98.5%

Results

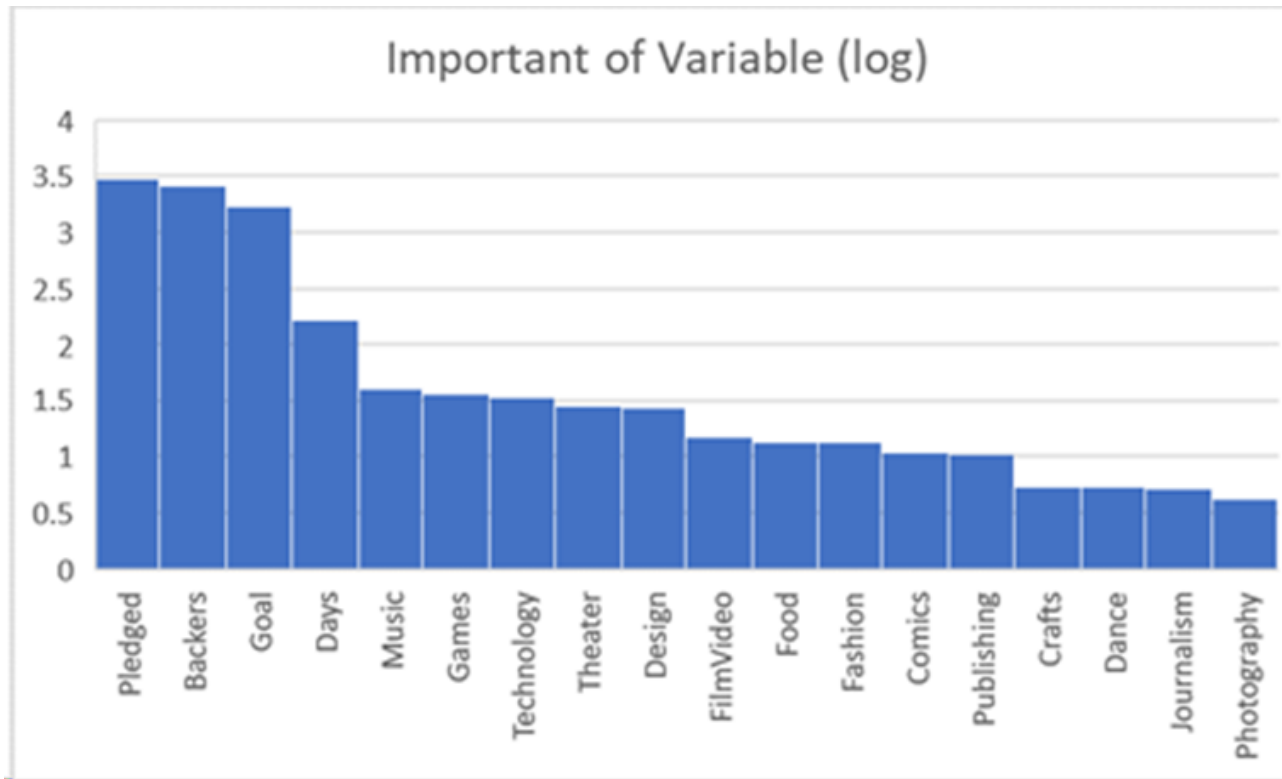


randomforestM





Results



A word cloud of tabletop RPG terms. The word 'card' is the largest and most central. Other prominent words include 'play', 'adventure', 'fantasy', 'miniature', 'dice', 'rpg', 'game', 'board', 'dungeon', 'project', 'novel', 'space', 'wargame', 'monster', 'player', 'book', 'dark', 'edit', 'set', 'box', 'pathfind', 'metal', 'sci-fi', 'horror', 'war', 'build', 'world', 'king', 'magic', 'survival', 'visual', 'tabletop', 'expans', 'cthulhu', 'batt', 'strategy', 'art', 'roleplay', 'deck', 'dragon', 'hero', 'terrain', 'part', 'relaunch', 'zombi', 'citi', 'new', 'battl', 'zombi', 'wargame', 'project', 'novel', 'space', 'wargame', 'monster', 'player', 'book', 'dark', 'edit', 'set', 'box', 'pathfind', 'metal', 'sci-fi', 'horror', 'war', 'build', 'world', 'king', 'magic', 'survival', 'visual', 'tabletop', 'expans', 'cthulhu', 'batt', 'strategy', 'art', 'roleplay', 'deck', 'dragon', 'hero', 'terrain', 'part', 'relaunch', 'zombi', 'citi', 'new', 'battl', 'zombi'.

A word cloud featuring various terms related to writing and media. The most prominent words are 'magazin', 'novel', 'children', 'story', 'art', 'poetry', 'guide', 'publish', 'adventure', 'poem', 'collect', 'journey', 'science', 'history', 'calendar', 'illustration', 'fantasy', 'live', 'memoir', 'two', 'time', 'first', 'make', 'one', 'little', 'world', 'help', 'journal', 'dream', 'year', 'series', 'color', 'day', 'podcast', 'anthology', 'write', 'love', 'edit', 'zine', 'girl', 'print', 'issue', 'world', 'life', 'book', 'year', 'dream', 'series', 'color', 'day', 'podcast', 'anthology', 'write', 'love', 'edit', 'zine', 'girl', 'print', 'issue'. The words are in various colors (blue, green, yellow, orange, red, purple) and sizes, arranged in a circular pattern.



Conclusion & Future Work

The result for EDA, Cluster Analysis, LDA and Random Forest are similar. They revealed the same group of categories that attracted a majority of the backers in the Kickstarter Community.

Games, Technology, Design, Film & Video are the majority industries among the most mature projects. These industries have higher barrier-to-entry and require more technology inputs. On one hand, this resulted in many well-prepared projects and their successful fundraising; on the other hand, immature project also exists, which could be a reason balancing out the successfulness in such categories.



Conclusion & Future Work

The World Cloud helps summarized some featured keywords for the top 6 categories, implying majority of the topics within each category. It could be developed to a future research in NLP.

Fun Fact: Cluster#2 & #5 are mainly made up of “outliers” with high goal & pledged. These data are in fact our focus – the ones that “ought to be successful”.