

MI353 Project: Predicting Crowdfunding Success

Nolan Beck, Matthew Costello, Spencer Daugherty, Brian
Freiss, & Jake Gadaleta
May, 2021

KICK STARTER



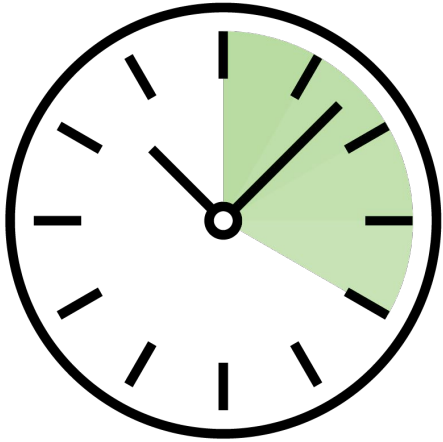
Project Overview

Goal was to find techniques which maximize success on the Kickstarter platform.

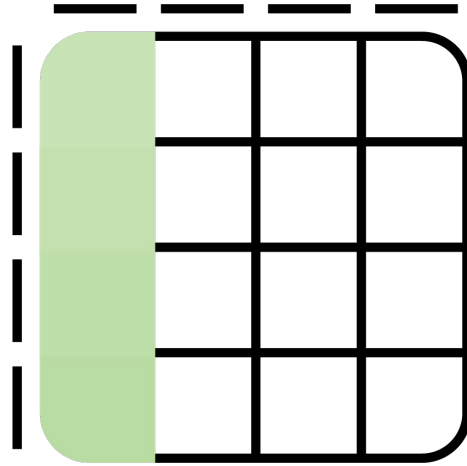
- Classification strategies that lead to successful funding.
- Useful for:
 - Startups seeking project funding.
 - Donors searching for investments.

The initial inspiration behind this project was to help Jake's brother execute a successful KickStarter project.

Data Description



2016-2018



~350K rows



16 variables

Data Dictionary: The Important Bits

Targeted Variables

- “success_factor”:
 - denotes whether or not an individual project met its goal. A value of 1 means that the goal was met, a value of greater than 1 means that the goal was surpassed, and a value of less than 1 means that the goal was not met
- “success”:
 - a boolean denoting if the project had more money pledged to it then it set as its goal containing the values [“TRUE”, “FALSE”]. This data can be represented by “success_factor,” but was purposefully parsed out in order to run decision trees
- “state”:
 - a senary variable denoting the current state of the project consisting of [“CANCELLED,” “FAILED,” “LIVE,” “SUCCESSFUL,” “SUSPENDED,” or “UNDEFINED”]

Non-Targeted Variables

- “Index”, “ID”, “Name”, “Category”, “Main_Category”, “Currency”, “Deadline”, “Goal”, “Launched”, “Pledged”, “Backers”, “Country”, “Usd_pledged”, “Usd_pledged_real”, “Usd_goal_real”

Data Preparation and Processing 1/4

Exploring the Data

- Stat & Graph Explores
- Clustering to see if we have outliers
 - several objects had a “success_factor” of 200.00 (200 x the goal), and some that a “success_factor” of .000001.

Preparing the Data

- filtered out the Kickstarter projects that had the outliers for “success_factor”
 - because these objects led to our first cluster analysis providing a heavily skewed view of the data.
 - we developed a succinct data spread which could be used for further modeling.
- We attempted to run various the models the following slides covers what we tried and how that led to our thinking in the model

Data Preparation and Processing 2/4

Regressions

We decided to use the variable “success_factor” as our target variable.

- After filtering, the initial regression used all variables and obtained an Adjusted R-Squared value of 0.6225
 - “backers” and “usd_pledged” were the greatest indicators.

We followed up the initial regression with one including only the top indicators.

- The new Adjusted R-Squared became 0.0000

We moved forward and attempted to implement Principle Component Analysis instead.

- “success_factor” is essentially too small of a variable to be accurately defined by any model
 - attempting to inflate “success_factor” did not help

Data Preparation and Processing 3/4

Auto Neural Networks

- Attempted to run the Auto Neural Network
- Targeted “success_factor” to start but the network was unable to accomplish this goal.
- Initially thought that perhaps the algorithm was failing due to the close range of the “success_factor” variables, reaching from 0.01 to 2.00.
- We attempted to correct this by manipulating the data and increasing the “success_factor.”
 - This was achieved by multiplying it by a value of 100
- We believe that the variables were largely incapable of predicting the exact value within due to something outside of the collected data. This could be something related to marketability power, social media support, or some other confounding variable. At this point we opted to switch gears, looking to see if we could predict the binary variable success.

Data Preparation and Processing 4/4

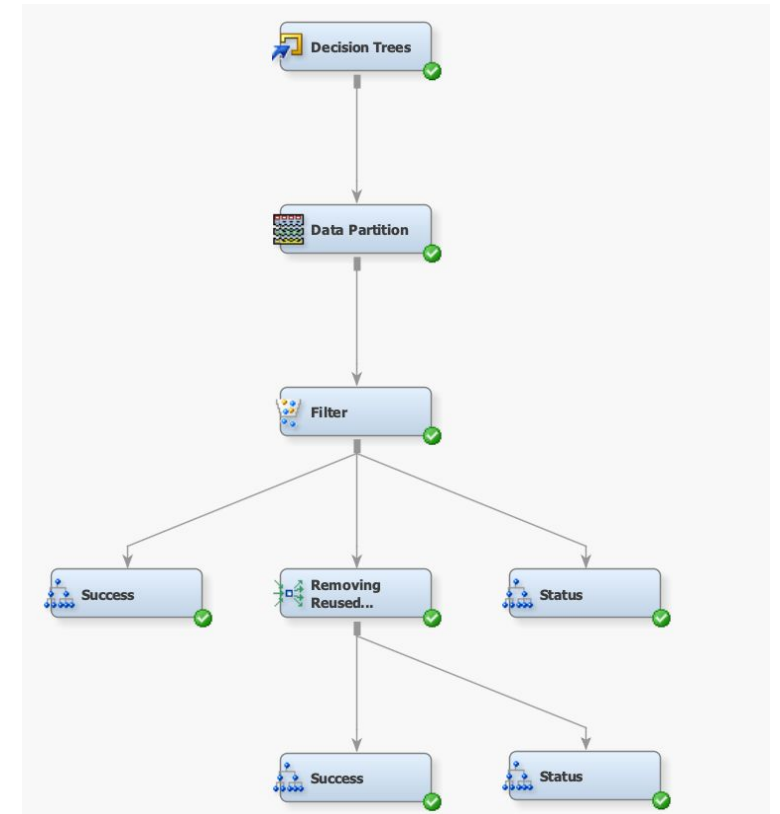
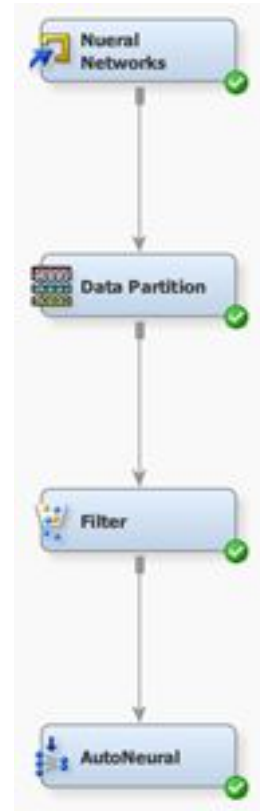
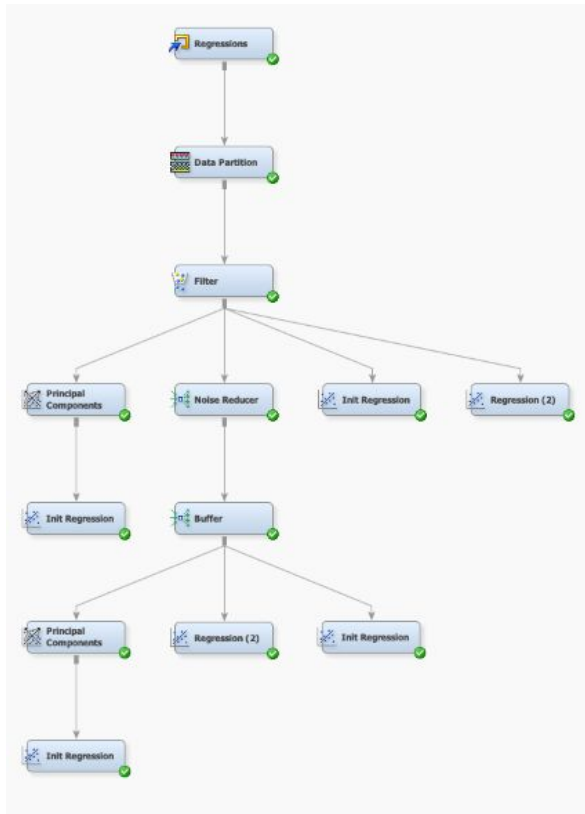
Decision Trees

- The variable success can hold one of two values:
 - “True” - denoting that it raised more money than it requested
 - “False” - the project failed to reach its goals.
 - Because we were aiming to predict a binary variable, we decided to use a Decision Tree.
- Could use any variable (except for state), and SAS was given the ability to generate the tree as it seemed fit
 - Because this initial Decision Tree proved to be very useful, we then decided to take the next step and create a secondary Decision Tree aimed this time at the “state” variable.
- In these first two decision trees, two sets of variables contained similar information
 - “pledged”, “usd_pledged”, and “usd_pledged_real”
 - “goal”, “usd_goal_real”
- The state Decision Trees are slightly less accurate due to some overlap in the senary values
 - “CANCELLED”, “FAILED”, and “SUSPENDED” all denoting that the project had failed.
 - We should note that due to this overlap we believe that it is acceptable that there are multiple values in the Tree’s terminal nodes and for the initial question that was asked does not lessen the model to any degree.

Project Conclusions

Final Diagram Used

- Separate diagram for each analysis technique (Regressions, Auto Neural Networks, and Decision Trees)



Project Conclusions (Cont.)

Conclusions

- Decision Trees gave us the best results
 - Especially when repeated data was filtered out
- Most important factors:
 - Project's financial goal ("goal")
 - Pledged amount ("pledged")
 - Number of backers ("backers")
- There is no mixture of the given variables that accurately determines a project's "success_factor"

Project Conclusions (Cont.)

Further Conclusions

- Discovered our hypothesis was false
 - Data showed that the number of “backers” superseded all other variables
 - If the project maintained 17 or more backers, it had a significantly higher chance of succeeding
 - 71.50% “True”
 - 28.50% “False”
- All of the terminal nodes managed to predict the correct values with less than 5% error
 - The exception was a singular terminal node containing 12 values with a near even split between “True” and “False”

Project Findings

- What creators do with the money is not important to the project's financial success
- Picking an appropriate goal is critical

Best predictors:



Project's Goal



Pledged \$



of Backers

Thank you. Questions?

Presenters:

Nolan Beck

Matthew Costello

Spencer Daugherty

Brian Freiss

Jake Gadaleta

