gofundme™

Karen Ouyang

METIS

How much of your goal can you expect to receive?

- Category
- Goal Amount
- Story

METIS

Methodology

Web Scraping → Data Preprocessing → Data Analysis → Feature Engineer → Model

Web Scraping
- Beautiful Soup
- Selenium
- Pickle

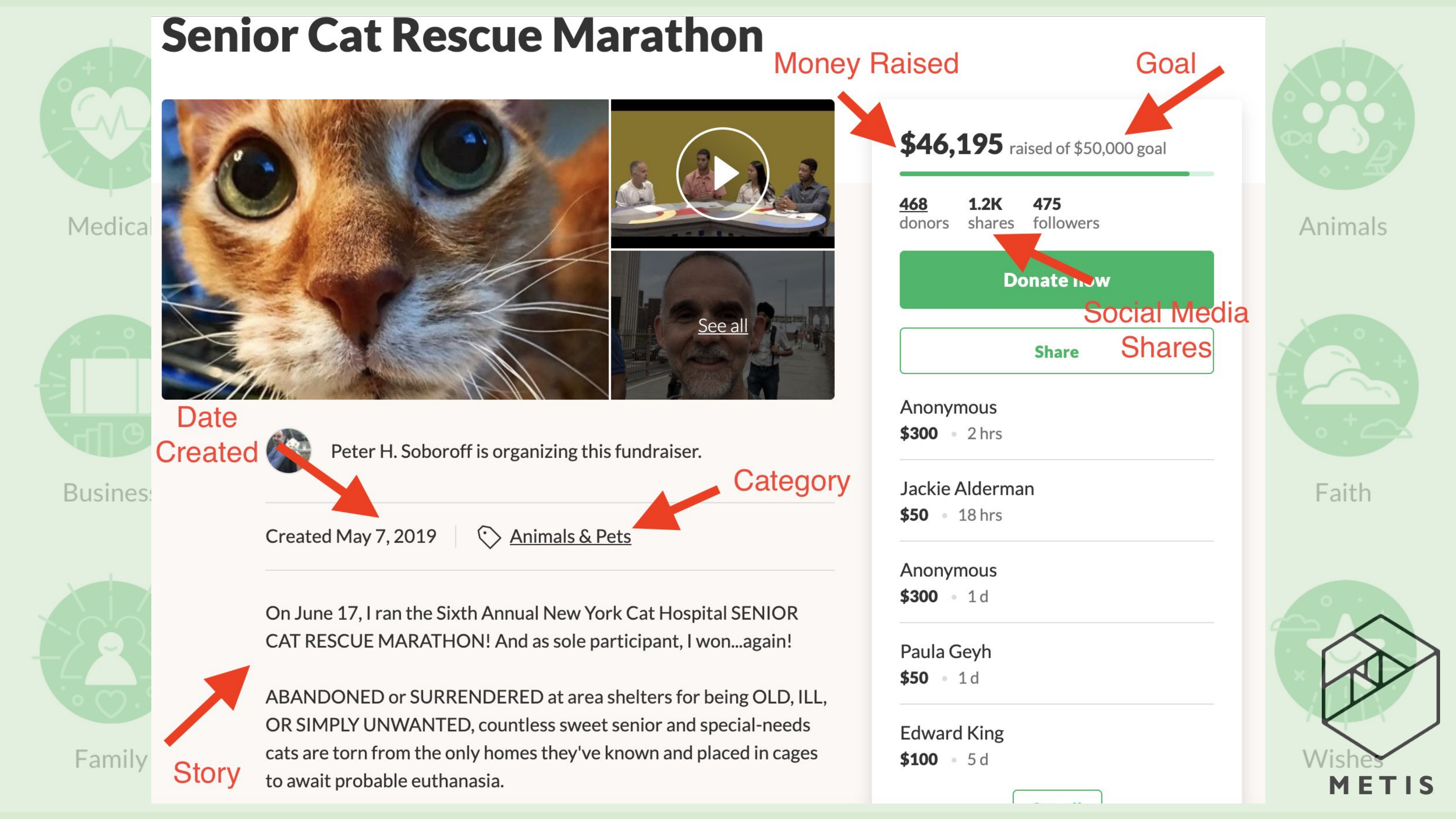
Data Preprocessing
- Formatting strings to numeric inputs

Data Analysis
- Locate and remove outliers
- Residuals

Feature Engineer
- Polynomial Features
- One hot encoding categories

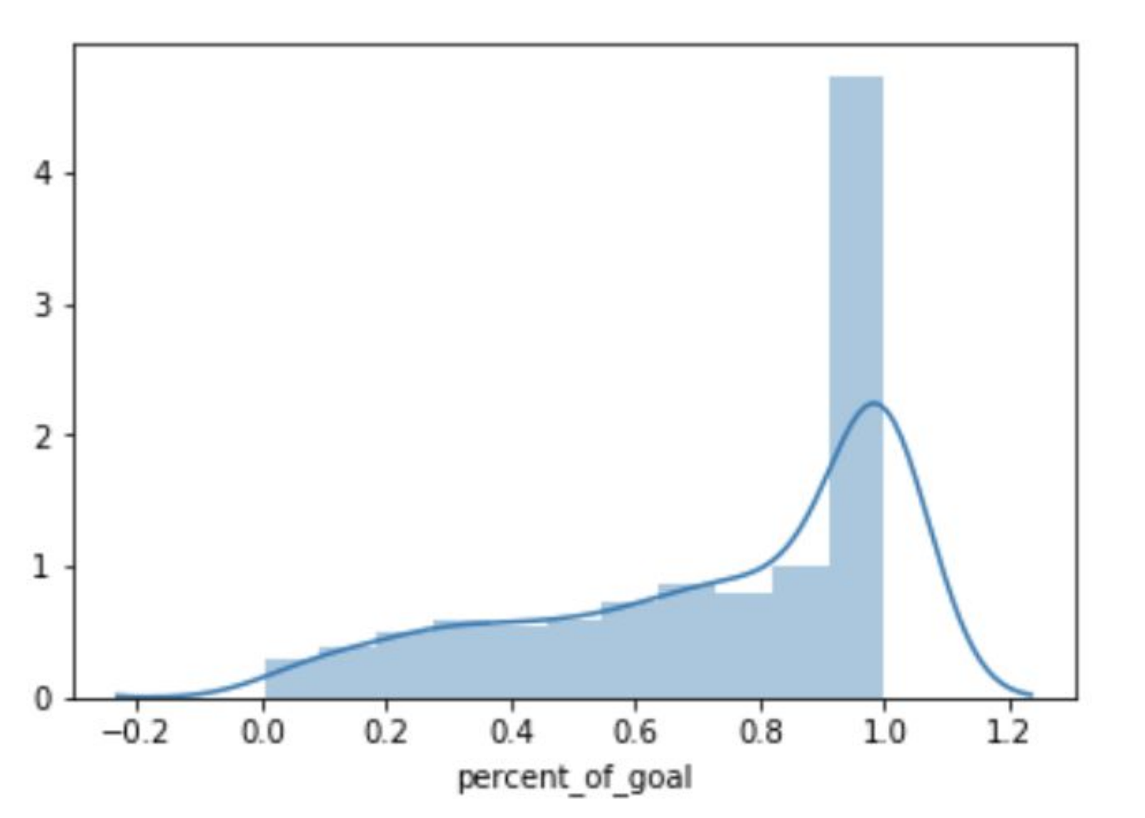
Model
- Cross Validation
- Lasso

METIS



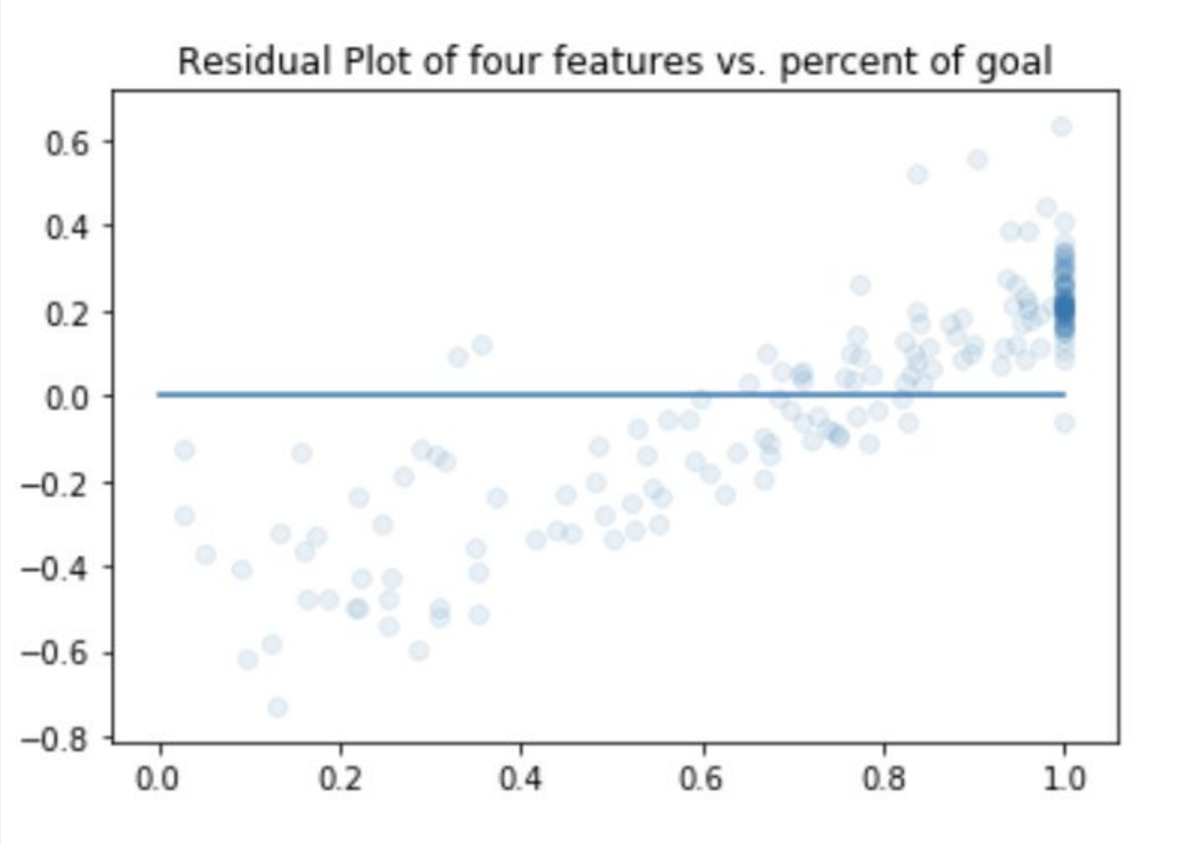
Senior Cat Rescue Marathon
Money Raised
Goal
$46,195 raised of $50,000 goal
468 donors
1.2K shares
475 followers
Donate now
Social Media Shares
Share
See all
Date Created
Peter H. Soboroff is organizing this fundraiser.
Category
Created May 7, 2019
Animals & Pets
On June 17, I ran the Sixth Annual New York Cat Hospital SENIOR CAT RESCUE MARATHON! And as sole participant, I won...again!
ABANDONED or SURRENDERED at area shelters for being OLD, ILL, OR SIMPLY UNWANTED, countless sweet senior and special-needs cats are torn from the only homes they've known and placed in cages to await probable euthanasia.
Story
Anonymous
$300 • 2 hrs
Jackie Alderman
$50 • 18 hrs
Anonymous
$300 • 1 d
Paula Geyh
$50 • 1 d
Edward King
$100 • 5 d
Medical
Business
Family
Animals
Faith
Wishes
METIS

Target Distribution



METIS

Polynomial Feature Engineering might help



METIS

Multiple Linear Regression Models (941 data points)

- Polynomial transformations of goal amount, word count, polarity and subjectivity scores (Lasso) : R^2 = .20

METIS

Multiple Linear Regression Models (941 data points)

- Polynomial transformations of goal amount, word count, polarity and subjectivity scores (Lasso) : $R^2 = .20$

- Just 18 categories dummified (Lasso): $R^2 = 0.01$

METIS

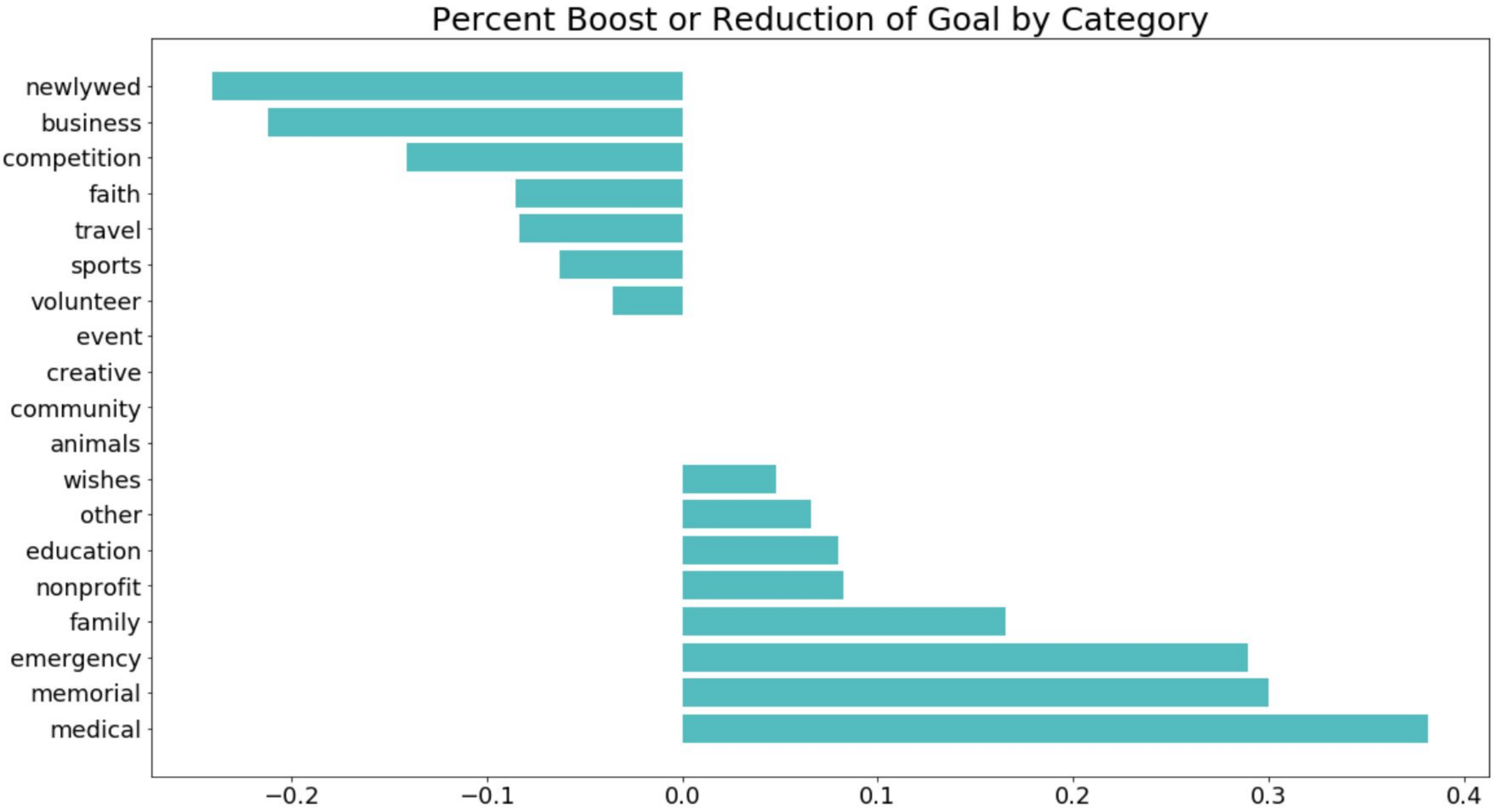
Multiple Linear Regression Models (941 data points)

- Polynomial transformations of goal amount, word count, polarity and subjectivity scores (Lasso) : R^2 = .20

- Just 18 categories dummified (Lasso): R^2 = 0.01

- Combined (Lasso): R^2 = .43

METIS


Percent Boost or Reduction of Goal by Category
newlywed
business
competition
faith
travel
sports
volunteer
event
creative
community
animals
wishes
other
education
nonprofit
family
emergency
memorial
medical
−0.2
−0.1
0.0
0.1
0.2
0.3
0.4

Conclusions

Top 3 Unpopular Categories (Reduce Predicted Percent of Goal Raised):

Newlywed, Business, Competition

Top 3 Popular Categories (Boost Predicted Percent of Goal Raised):

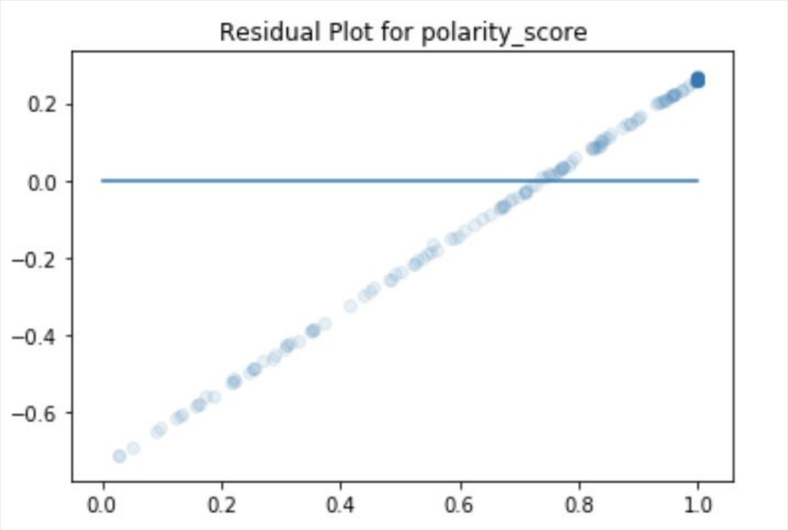
Medical, Memorial, Emergency

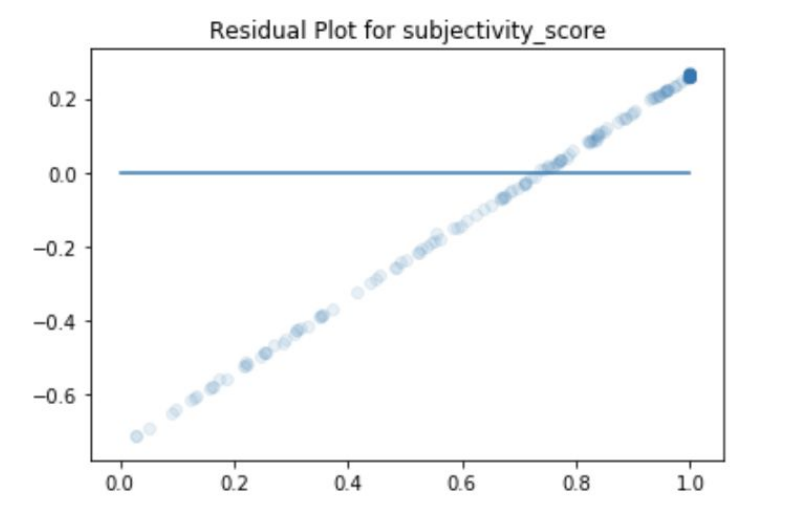
R^2 score (Lasso) = .43

(Model has some predictive power.)

METIS

Appendix A





METIS

Appendix B

OLS Regression Results

Dep. Variable:	percent_of_goal	**R-squared:**	0.158
Model:	OLS	**Adj. R-squared:**	0.157
Method:	Least Squares	**F-statistic:**	178.1
Date:	Thu, 18 Jul 2019	**Prob (F-statistic):**	2.33e-37
Time:	14:04:42	**Log-Likelihood:**	-86.063
No. Observations:	952	**AIC:**	176.1
Df Residuals:	950	**BIC:**	185.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.8188	0.011	77.457	0.000	0.798	0.840
goal_clean	-1.518e-06	1.14e-07	-13.345	0.000	-1.74e-06	-1.29e-06

Omnibus:	73.693	**Durbin-Watson:**	1.474
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	72.122
Skew:	-0.619	**Prob(JB):**	2.18e-16
Kurtosis:	2.465	**Cond. No.**	1.14e+05

OLS Regression Results

Dep. Variable:	percent_of_goal	**R-squared:**	0.022
Model:	OLS	**Adj. R-squared:**	0.021
Method:	Least Squares	**F-statistic:**	20.72
Date:	Thu, 18 Jul 2019	**Prob (F-statistic):**	6.03e-06
Time:	13:25:22	**Log-Likelihood:**	-152.00
No. Observations:	941	**AIC:**	308.0
Df Residuals:	939	**BIC:**	317.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.7905	0.015	53.849	0.000	0.762	0.819
word_count	-0.0002	3.57e-05	-4.551	0.000	-0.000	-9.25e-05

Omnibus:	101.653	**Durbin-Watson:**	1.796
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	109.934
Skew:	-0.790	**Prob(JB):**	1.34e-24
Kurtosis:	2.445	**Cond. No.**	650.

METIS

Appendix C

OLS Regression Results

Dep. Variable:	percent_of_goal	**R-squared:**	0.000
Model:	OLS	**Adj. R-squared:**	-0.001
Method:	Least Squares	**F-statistic:**	0.02672
Date:	Thu, 18 Jul 2019	**Prob (F-statistic):**	0.870
Time:	13:26:05	**Log-Likelihood:**	-162.25
No. Observations:	941	**AIC:**	328.5
Df Residuals:	939	**BIC:**	338.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.7406	0.015	50.652	0.000	0.712	0.769
polarity_score	-0.0115	0.070	-0.163	0.870	-0.149	0.126

Omnibus:	111.029	**Durbin-Watson:**	1.796
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	119.731
Skew:	-0.823	**Prob(JB):**	1.00e-26
Kurtosis:	2.414	**Cond. No.**	7.66

OLS Regression Results

Dep. Variable:	percent_of_goal	**R-squared:**	0.009
Model:	OLS	**Adj. R-squared:**	0.008
Method:	Least Squares	**F-statistic:**	8.259
Date:	Thu, 18 Jul 2019	**Prob (F-statistic):**	0.00415
Time:	13:29:15	**Log-Likelihood:**	-158.14
No. Observations:	941	**AIC:**	320.3
Df Residuals:	939	**BIC:**	330.0
Df Model:	1		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
const	0.6432	0.035	18.615	0.000	0.575	0.711
subjectivity_score	0.2055	0.072	2.874	0.004	0.065	0.346

Omnibus:	105.242	**Durbin-Watson:**	1.806
Prob(Omnibus):	0.000	**Jarque-Bera (JB):**	119.148
Skew:	-0.831	**Prob(JB):**	1.34e-26
Kurtosis:	2.471	**Cond. No.**	9.33

METIS

Appendix D

```
a = sorted(list(zip(X_train.columns, lasso_model.coef_ )), key=lambda x:(x[1]), reverse=True)
a
```

executed in 12ms, finished 17:04:02 2019-07-18

```
[('medical', 0.4053419307620743),
 ('memorial', 0.3220248395811013),
 ('emergency', 0.3103948793339755),
 ('family', 0.18115671509736306),
 ('subjectivity_score_sq', 0.121712777245019),
 ('nonprofit', 0.10337175372043461),
 ('education', 0.09372917643687535),
 ('polarity_score_sq', 0.08766756195621231),
 ('other', 0.08515516741341855),
 ('wishes', 0.05426811376053922),
 ('polarity_score', 0.01728347245747426),
 ('polarity_score*subjectivity_score', 0.002876832390037462),
 ('word_count', 1.9494874410936483e-05),
 ('goal_clean*subjectivity_score', 2.783554809979415e-07),
 ('goal_clean_sq', 7.241047849834209e-12),
 ('subjectivity_score', 0.0),
 ('animals', 0.0),
 ('community', 0.0),
 ('goal_clean*word_count', -2.8885300286703573e-10),
 ('word_count_sq', -1.4712091437890873e-08),
 ('goal_clean*polarity_score', -2.163443742647784e-06),
 ('goal_clean', -4.6806268435334145e-06),
 ('word_count*subjectivity_score', -0.00010646638552996634),
 ('word_count*polarity_score', -0.00016905022522632486),
 ('event', -0.008614321883326824),
 ('creative', -0.011097453692479492),
 ('volunteer', -0.05111899491494909),
 ('sports', -0.09201225832626436),
 ('faith', -0.09618460495902338),
 ('travel', -0.09987170235078825),
 ('competition', -0.1640158246042643),
 ('business', -0.22321901440607678),
 ('newlywed', -0.25958480494288333)]
```

METIS

Appendix E

5 Fold Cross Validation scores for combined features using simple linear regression, ridge regularization, and lasso regularization

Simple regression scores: [0.3574515650848509, 0.39187393007051596, 0.3796770254075955, 0.34630392817934286, 0.5067102950987303]
Ridge scores: [0.3695460732306116, 0.39838596338521015, 0.3711886254455937, 0.3492615529619597, 0.5024877225398371]
Lasso scores: [0.3807339899745644, 0.4066521482353215, 0.37657345481029625, 0.3593620064465233, 0.5101509640509447]

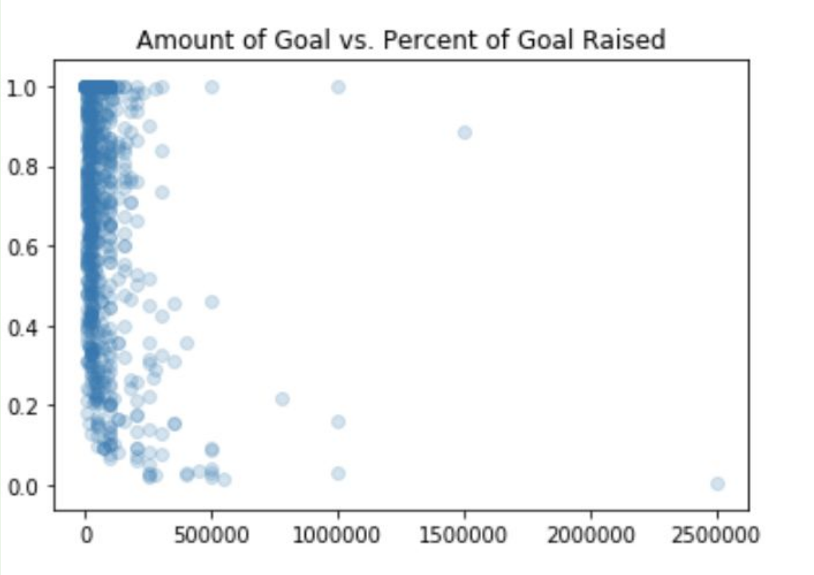
Simple mean cv r^2: 0.396 +- 0.057
Ridge mean cv r^2: 0.398 +- 0.054
Lasso mean cv r^2: 0.407 +- 0.054

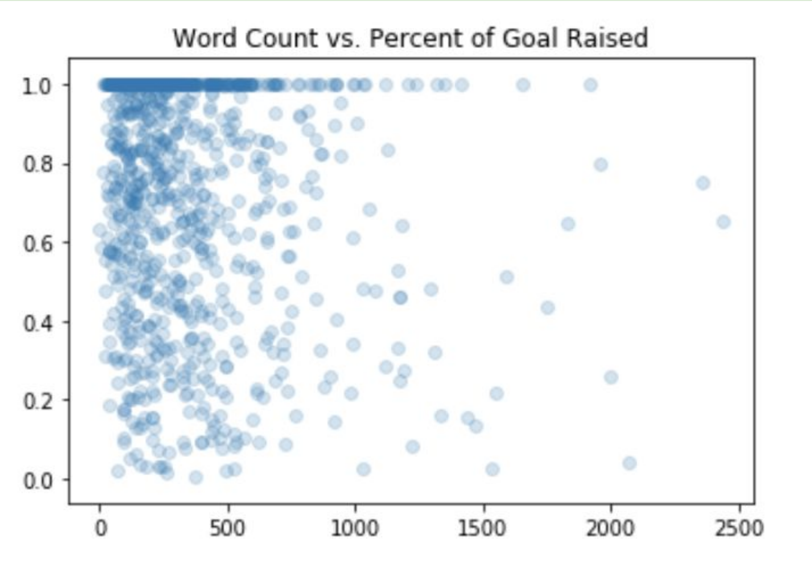
METIS

Appendix F

	goal_clean	percent_of_goal	word_count	days_active	polarity_score	subjectivity_score
count	941.000000	941.000000	941.000000	941.000000	941.000000	941.000000
mean	53583.940489	0.738747	318.285866	108.533475	0.160025	0.465167
std	74213.957804	0.287662	259.834338	234.339861	0.133964	0.130704
min	1.000000	0.019842	1.000000	1.000000	-0.500000	0.000000
25%	10000.000000	0.533200	133.000000	62.000000	0.088235	0.407762
50%	25000.000000	0.837592	244.000000	105.000000	0.151662	0.473718
75%	75000.000000	1.000000	428.000000	143.000000	0.226376	0.529816
max	500000.000000	1.000000	1468.000000	7132.000000	1.000000	1.000000

Appendix G





METIS

Karen Ouyang

karenouyang93@gmail.com

morningkaren

morningkaren

METIS

Questions & Comments