Predicting the Funding Among for Academic Research Projects

Alexander Marine   **Ali Oku**
Northern Illinois University

ABSTRACT

The funding amount is an important, sometimes unseen part of any research project of a substantial size. This is because it gives researchers useful information on how much money to expect when designing projects and writing grant applications. Although predicting grant funding is essentially a regression task, we utilize an intuitive method of binning to group funding amount hence, making it a classification task. However, our work does include both a classification and regression approach in predicting research funding. Our models are built on features ranging from the length of the title and abstract of the project to the discipline and scope of the research project. We build several classification and regression models to predict the funding amount a research gets. Our classification models initially required grouping the funding into four bins and predicting which bins best represented a research project. Our regression models involve using both multiple linear regression and artificial neural networks to predict research funding. We find that the most important feature in predicting funding is the duration of the research project. We also find that research topics covering computational sciences and/or engineering tended to receive higher funding amount in comparison to other research disciplines. Although our project shows the difficulty in predicting grant funding, it gives insight on binning techniques that can be used in this field.

INTRODUCTION

The funding amount is an important, sometimes unseen part of any research project of a substantial size. It allows the researchers the capital they need to go about their studies in the fashion that they believe will result in the most accurate outcome possible. We believe that having the ability to predict the funding amount of a research project given various details about it such as the research topic, the research organization, etc., can allow those seeking a grant to have an idea of how much money they could potentially receive based on how much research projects under similar circumstances got. Possessing this ability has the advantages of allowing researchers to feel more comfortable asking for a larger grant if similar projects received more or simply gives a baseline idea of what is considered a lot or a little when it comes to the funding for certain projects.

Our hypothesis is that research projects that come from universities ranked within the top twenty in the United States and fall under the category of engineering will on average have a larger funding amount than those coming from smaller universities or independent research institutions. Also, we predict that projects that last over five years will have a noticeably larger funding amount than those that last less than five years.

In order to complete the task of finding the features within the Grants Dataset, we will use several classification and regression models including Random Forests, Naïve Bayes, Logistic Regression, K Nearest Neighbors, Support Vector Machines, Neural Networks, and Feature Selection with Decision Trees. The final evaluations will be made after a five-fold cross validation.

Our studies differ from earlier work on predicting grants because there have not been any studies done on if the research institution or timeline of the study correlates to a different funding amount. While there have been studies that tread the same general path as ours, we hope that information available in the Grants Dataset allows us to draw conclusions unique from those of other research projects. Our proposal also differs from previous work since we would be integrating neural networks and autoencoders for predictions.

DATASET DESCRIPTION

The dataset that was used for our experiments was obtained from Altmetrics and contains over 50,000 rows of data about grants that were received from various organizations for the funding of research papers. There are sixty-three columns present containing information for each project, but many of these columns contain sparse, or unhelpful data that will be dropped during data cleaning. Before any feature extraction of the data, the columns that were used in our research models were *start_year* and *end_year*, as we thought there was potential for these being useful to predict the funding amount in their base state.

DATA CLEANING AND FEATURE EXTRACTION

There were many columns present within the data that we found to not serve any use in predicting the funding amount of the research projects. These columns were dropped from our dataset and not considered in our models. Also, all rows that contained any null values were dropped from consideration due to them being such a small percentage of our total data. One hurdle that was encountered early on in our data cleaning process was that many of the research projects were assigned multiple subjects. In order to reduce model complexity, we dropped all of the subjects except the first one, as we believe that it is sufficient to aid in training our models.

The next step in feature extraction that we took involved making new columns in our dataset based on the information that is available in the other columns. We started this process with getting the total number of words used in both the title of the paper, and the abstract. By doing this, we can have integer values in our dataset for each project's title and abstract respectively, which makes it much easier to feed into our models. Although we lose some data by doing this, we believe that the simplification of the data works best for our circumstances. Next, we took the start and end date and found the duration of how long the projects took, as it seemed possible that projects with a longer duration could have received more money. Using the *to_datetime* feature within Pandas, we were able to convert the *end_date* and *start_date* columns to a format with which subtraction is much simpler. By having the duration, it allows us to compare the different lengths of time that projects were active in comparison to others.

Regarding the research organizations, we categorized them as either an academic institution, or not one. Then within the ones that were academic institutions, we listed if that school was included as a top twenty school in the United States. Since this dataset included a large number of universities, we wanted to compare how the research organizations that were schools compared to ones that were not. After having a list of all the organizations that were schools, it was natural to see if the ones that were highly ranked would earn more funding on average. Another feature that was extracted from the data was the number of researchers that were present for each project. Without having a reference to who all these researchers are or what they have accomplished in the

past, their names are of little use to us. Therefore, by counting the number of people listed for each project, it gives us an integer value that we can compare between projects.

Due to the amount of different subject matters for all of the projects, we decided to group these subjects into broader categories for easier model predictions. The fields that we ended up with are engineering, computational, health and medicine, and other. We decided upon these groups since a large amount of the categories available in the dataset seemed to fall somewhat evenly between them, and only a few having to go into the other category. These groups are important since some research topics are more likely to receive higher funding amount in comparison other research disciplines.

The final piece of feature extraction we performed was creating the bins for the funding amounts that will be used as what we are trying to predict. We underwent several iterations of our funding bins in terms of the number of bins, as well as the range in which each bin covers monetarily. We ultimately ended up with a total of four bins, where *bin_1* covers the projects that received zero funding, *bin_2* covers the range of zero to $99,999, *bin_3* covers the range of $100,000 to $299,999, and *bin_4* covers everything about $300,000. Because there were a number of projects in the dataset that were listed as having no funding, we thought it was necessary to have a bin dedicated just to that, since it is a unique circumstance among the others. As for the other three groups, we found that the bins we decided on offered a somewhat even distribution between them, while giving us a low number of bins to allow for easier predictions.

S/N	Feature	Data Type	Description
1	title_len	integer	The number of words in the project title
2	abstract_len	integer	The number of words in the abstract
3	funding_amount	float	The total funding awarded to the project
4	start_year	integer	The starting year of the project
5	end_year	integer	The ending year of the project
6	duration	integer	The duration of the project in days
7	no_researcher	integer	The total number of researchers involved in the project
8	research_org	string	The research organization performing the research project
9	funder	string	The organization funding the research project
10	is_top_twenty	boolean	True, if the research organization belongs to top twenty research colleges in the United States.
11	is_academic	boolean	True, if the research organization is a college or university
12	funding_bin_1	string	"bin_1": project received no funding "bin_2": project received between $0 - $100000 "bin_3": project received between $100000 - $300000 "bin_4": project received more than $300000
13	field_bins	string	"computational": projects involving computer sciences "engineering": projects involving all fields of engineering "health and medicine": projects involving health sciences, public health and medicine "natural science": projects involving biology and chemistry

14	start_date	date object	"other": projects involving other fields not listed above
14	start_date	date object	The proposed start date of the project
15	end_date	date object	The proposed end date of the project
16	research_city	string	The city of the organization performing the research
17	research_state	string	The state (location) of the research organization
18	title	string	The title of the research as detailed on the grant application
19	abstract	string	The research abstract as detailed on the grant application

Table 1: The features/variables used obtained from feature generation.

CLASSIFICATION

Due to the big variation in the funding amount received, we grouped all funding amounts received into four different bins (*funding_bin_1*). The goal of classification was to predict the funding bin a project belongs to using different features in the dataset. We originally used six bins ensuring that the different bins contained almost equal number of objects. However, after running several models, we observed that using four bins better represented the data. The final bins used in the models are described in Table 1 (*funding_bin_1*). Classification was performed using K nearest neighbors (KNN), Naïve Bayes Classification, and Logistic Regression. The models were evaluated over a five-cross validation and the mean accuracy was also recorded.

Logistic Regression:

For Logistic Regression, we used the *LogisticRegression* portion of the *sklearn.linear-model* library. When calling *LogisticRegression*, we used the default solver of *lbfgs*, and *multi_class* was set to *multinomial*. By setting it to multinomial, the loss is minimized over the multinomial loss fit across the probability distribution, even when the data is binary. Finally, we set the *max_iter* value to 10,000 to properly iterate over the entire dataset. We then called fit on our variable created from calling *LogisticRegression*, passing in our *X_train* and *y_train* variables achieved by *train_test_split*. Next, we created *y_pred* by calling *predict* and passing *X_test*. To visualize the results of our model, we created a confusion matrix by using the *sklearn* library, as well as printing out the *accuracy_score* with the parameters *y_test* and *y_pred*.

Naïve Byes:

We followed a similar order of operations for the Naïve Bayes Theorem. We first imported *GaussianNB* from the *sklearn.naive_bayes* library. When calling *GaussianNB,* we left both parameters to their default, with *prior* being set to *None*, and *var_smoothing* being set to *1e-9*. We then followed the same steps as listed above for *LogisticRegression* by performing *fit* and *predict* on our variables created by *train_test_split*.

K-Nearest Neighbor (KNN):

KNN was implemented using the *sklearn* library available on python. This model was used in predicting the likelihood an object fell within one of the four funding bins described in Table 1 (*funding_bin_1*). In selecting the number of neighbors when building the algorithm, we evaluated the best possible option range from k = 1 to k = 13. We noted that the training and testing accuracy converged around 11 neighbors as shown in Figure 1. The average accuracy using k = 11 was about 64%.
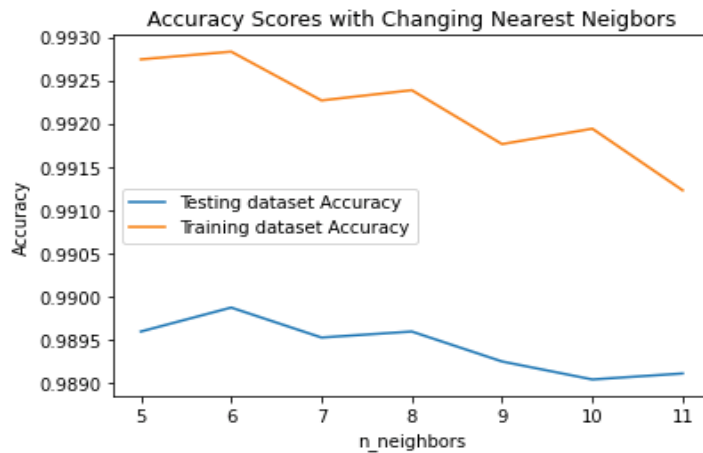
Figure 1: Training and testing accuracy over different number of nearest neighbors. K=11 was chose since training and test accuracy covered around this value.

REGRESSION

The goal of regression was to predict the funding amount received. We carried out Multiple Linear Regression using *scikit-learn* library. Due to the high variance in the funding received, we also used neural networks with two hidden layers to capture the non-linear complexities that might exist in the data.

Multiple Linear Regression:

Multiple Linear Regression was achieved using *linear_model* package available on *sckit-learn*. The funding amounts were also normalized using min-max normalization bringing all the value between 0 and 1.

Neural Network (Regression):

To capture possible non-linear and complex relationships that might exist in the data, we implemented a neural network for regression using *tensorflow* and *keras* packages on python. The neural network consisted of an input layer, two dense hidden layers of sizes fifty and twenty-five respectively and a final output layer of size one (see Figure 2). The output layer of size one consisted of a linear activation to match the funding amount. The data was scaled before running the model to accommodate the huge variation that existed in the different columns. The funding amounts were also normalized using min-max normalization bringing all the value between 0 and 1. The neural network was trained over ten epochs with a batch size of fifty to prevent overfitting as shown in Figure 3.

```
Model: "sequential_1"
_____
Layer (type)                 Output Shape              Param #
=================================================================
dense_3 (Dense)              (None, 50)                550
_____
dense_4 (Dense)              (None, 25)                1275
_____
dense_5 (Dense)              (None, 1)                 26
=================================================================
Total params: 1,851
Trainable params: 1,851
Non-trainable params: 0
_____
```

Figure 2: Neural network for predicting the funding amount. The size of the last layer is set to 1 to match the funding amount.
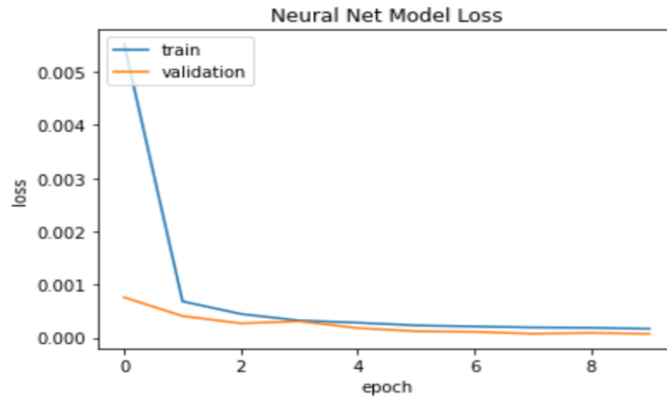

Figure 3: Training and validation loss for regression neural network used in predicting the total funding received

FEATURE SELECTION

All models used in both the classification and regression tasks were initially evaluated using most of the numerical features in the dataset. The categorical features were also encoded using one-hot encoding and used in building and evaluating the models. The features included *start_year*, *duration*, *end_year*, *title_len*, *abstract_len*, *no_researchers*, *is_academic*, *funder*, and *field_bins*. These features are explained in Table 1. Random Forest classifier was also used to rank the importance of the features in predicting the total funding received. The Random Forest classifier utilized GINI index to select the features and nodes that led to the greatest decrease in impurity (nodes at the top of the tree). According to this method, the most important features in classifying the objects into the four funding bins was the duration of the research study. The names of the top ten features and their corresponding GINI importance values are detailed in Table 2.

For the regression task, a similar technique called Random Forest Regressor was also used to pick the top features that were predictive of the funding received in the dataset. The results of this model also indicated that the duration was the most important feature at predicting the funding amount. The names of the top ten features and their corresponding importance scores are detailed in Table 3 and Figure 2.

The duration of the research project was the most important feature for both the classification and regression tasks. This was quite intuitive since research projects that span a longer time would require more funds and support. For the abstract length, it appeared that abstracts that were below or above a threshold were more likely to receive lower funding even though most abstract length fell between 200 and 400 words. For the title length, visualization showed that research projects with a title length over 20 words received lower funds.

Rank	Feature Name	Gini Importance Value
1	duration	0.290
2	abstract_len	0.203
3	title_len	0.118
4	end_year	0.106
5	start_year	0.092
6	no_researchers	0.061
7	is_academic	0.018
8	'funder_Directorate for Computer & Information Science & Engineering'	0.016
9	'funder_Directorate for Mathematical & Physical Sciences'	0.010
10	'field_bins_computational'	0.007

Table 2: Features with corresponding GINI values at predicting the funding bins an object falls into using random forests classifier.

Rank	Feature Name	Importance Score
1	duration	0.295
2	start_year	0.264
3	abstract_year	0.160
4	end_year	0.084
5	funder_Directorate for Geosciences	0.070
6	no_researchers	0.037
7	title_len	0.034
8	is_academic	0.015
9	funder_Directorate for Mathematical & Physical Sciences	0.007
10	is_top_twenty	0.007

Table 3: Features with corresponding importance scores at predicting the funding amount using random forest regressor.
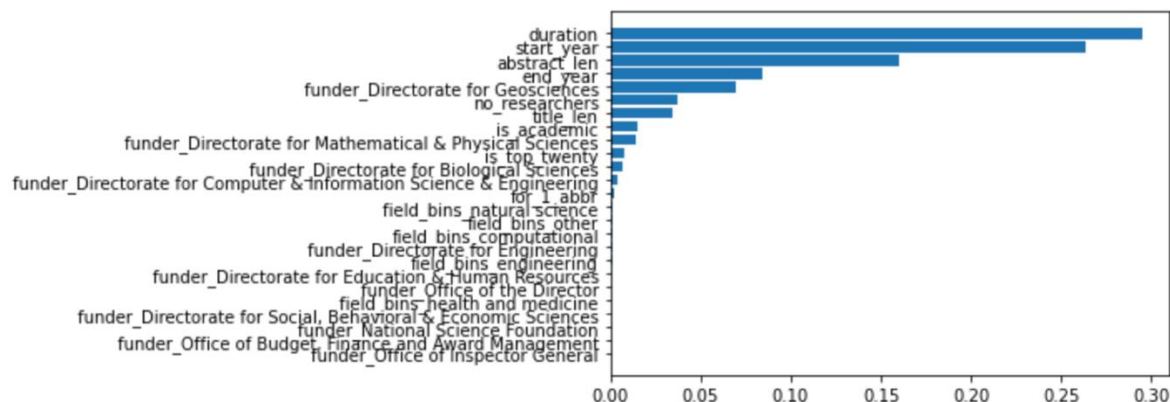
Figure 4: Figure ranking the importance of features using random forest regressor.

RESULTS

Without using the top ten features identified by our feature selection unique, classification with Logistic Regression achieved an average of 64% accuracy evaluated over a five-fold cross validation. Naïve Bayes achieved an average classification accuracy of 30% evaluated over a five-fold cross validation while K-Nearest Neighbors achieved an average accuracy of 98% over a five-fold cross validation. However, using the ten features identified using Random Forests produced a significantly higher accuracy in comparison to using all applicable features extracted from the dataset as indicated within Table 4. Analyzing the confusion matrix produced from the classification results revealed a class imbalance especially in the funding bin capturing grant applications that did not receive any funding (*bin_1*). To accommodate this class imbalance, oversampling was used to increase the number of objects in the lacking class so that it had equal weight and frequency with the other classes. After oversampling, the classification models re-ran, and the results are detailed within Table 5. Overall, there was a slight decrease in prediction accuracy after oversampling.

Regression was more difficult than classification due to the high variance in the funding amounts received. Regression was initially carried out using all of the application features, but this generally led to very low performance. A set of ten features was then identified using Random Forest regressors which led to slight increase in accuracy. The Neural Networks performed relatively poorly at predicting the funding received with some predicting values falling below zero. Although using just the top ten features extracted from feature extraction slightly improve prediction, subsequent reduction in the number of features did not seem to affect the overall accuracy of the model. A slice of the first thirty actual funding amounts and their corresponding features are shown in Figure 5.

Multiple Linear Regression using the selected features also performed poorly due to the large mean square error values. However, the mean squared error values produced from Multiple Linear Regression were lower than those produced using Neural Networks. A lower mean squared value indicates that the Multiple Linear Regression model performed better than the Neural Network at predicting the total funding received.

Model	Accuracy (using all applicable features)	Accuracy (using only top ten identified features)
Logistic Regression	44%	64%
Naïve Bayes	34%	62%
KNN	61%	60%

Table 4: Classification algorithms and their corresponding accuracy scores using all applicable features and selected features using random forests without oversampling (class imbalance is present).

Model	Accuracy (using all applicable features)	Accuracy (using only top ten identified features)
Logistic Regression	56%	54%
Naïve Bayes	38%	56%
KNN	72%	72%

Table 5: Classification algorithms and their corresponding accuracy scores using all applicable features and selected features using random forests and oversampling (no class imbalance).
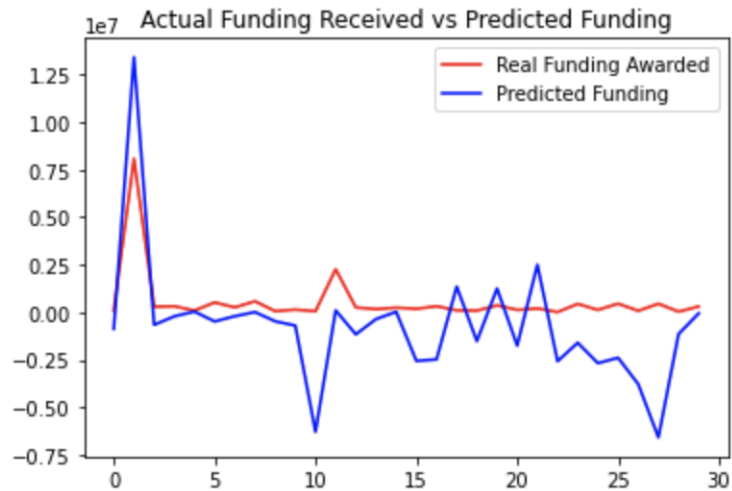

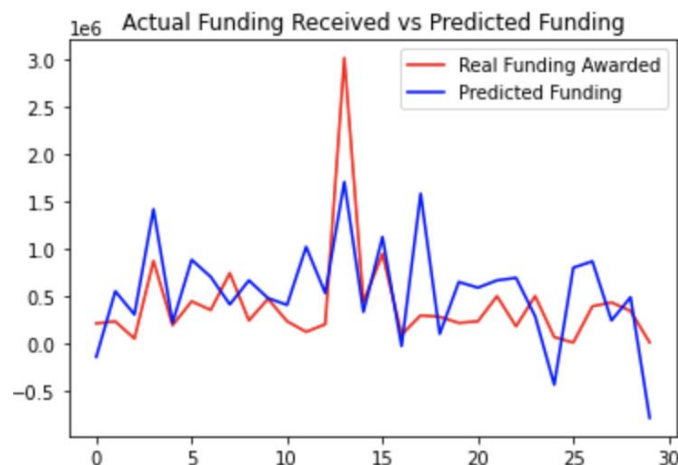**Figure 5:** Regression to predict funding amount using neural network. Some predicted funding fell below zero.

Figure 6: Multiple linear regression to predict the funding amount.

Model	MSE
Multiple linear regression	4.8×10^{12}
Artificial neural network	2.4×10^{13}

Table 6: The Mean Squared Error (MSE) of the multiple linear regression model and the neural net-based model at predicting the total funding received. On average, multiple linear regression had a lower MSE than the neural network model.

DISCUSSION

The funding amount received for a grant is based on very complex variables that can differ from one application to another. Since grant applications are mainly reviewed by other scholars, there might be personal decisions that can vary from one person to another and can lead to not being able to be captured by the models. To account for the high variability in funding received, we created four bins with the goal of predicting what bin a grant application falls into. This is equally an important task since it can inform researchers of how much grant funding, they expect to receive for their research work. To address this classification problem, we used Logistic Regression, K-Nearest Neighbor, and Naïve Bayes classifier while accommodating for class imbalance. From our data, KNN tended to outperform the other two classification models used on both the class imbalanced and class balanced data (see Table 4 and Table 5). Generally, there was a decrease in accuracy after oversampling indicating that the class imbalanced introduced a small classification bias when training the models. While the results of classification can provide useful information, predicting the actual funding amount is essential. To do this, we implemented Multiple Linear Regression and artificial Neural Networks (see Figure 5 and Figure 6). We hoped that the Neural Network model with multiple hidden layers would capture the non-linear complexities that might exist in the data. However, Multiple Linear Regression outperformed Neural Networks at predicting the funding amount as observed by comparing the mean square error of both models. The mean square error produced by both regression models were quite high

as detailed on Table 6. Since mean square error involves squaring the difference between the actual value and the predicted value, large prediction errors would produce very large mean square error values. However, when using scaled funding values (min-max normalization), the MSE reported was quite low but not zero. In both cases, the MSE from the multiple linear regression was lower than the MSE from the neural network model (Table 6). This indicates that the multiple linear regression model performed better than the neural network at predicting grant funding.

The limitation of our classification models is that they may lack generalization overtime since the amount of funding grants awarded may change overtime. A possible improvement to our neural network model is to preprocess the data by using an autoencoder-based model for dimensionality reduction which can better capture the non-linear relationships that exist in the data. This process data can then be fed into the neural network to be used in predicting the funding amount.

There were many variables that limited the ability to accurately predict the funding amount. The dataset did not include important information and content of the research that might be considered in awarding a grant. The funding amount a research receives is highly dependent on the possible impact of the research and its role in solving current problems. However, our dataset did capture this information and others that might impact whether funding is awarded and how much funding is awarded. The different funding awards were gifted by different organization which might greatly impact whether funding is allocated since each of the departments/organizations might have different funding capacities. For example, the Directorate for Computer, Information Science and Engineering might have a greater funding budget in comparison to the Department of Geosciences. Furthermore, awarding departments might use different criteria in evaluating grant applications making it more difficult to find patterns that can be generalized to the entire dataset. This would also mean that the funding organization might play a stronger role in comparison to other features in the data.

For future research, more features in the title and abstract can be utilized in building both classification and regression models. The title and abstract can hold essential information about what fields are covered by the research. Along with this, there is an abundance of information that is not taken into consideration when removing the names of the researchers who worked on the project. Since the name value of some people can greatly alter the amount of funding that a research project could amass, it would be beneficial if it could be added to the predictions. The topics that the projects cover is another area where some improvements on our results could be made. As stated earlier, we removed all the sub-topics from our model for simplification purposes and only left the one that was listed first. By creating a scenario where all of the topics are included when predicting the funding amounts, it can potentially show a correlation that was not visible within our models.

One aspect that we ignored in our computations was inflation. The range of start years for the research projects in the dataset are from 1992 to 2020. According to *in2013dollars.com*, $100 in 1992 would be $184.47 in the year 2020 for an 84.47% increase. We believe that although this amount has a smaller impact on our models to the large number of samples that cover all the different years within the time frame, it is something that can be added upon in future experiments to potentially increase accuracy.

CONCLUSION

After evaluating all models used for both classification and regression, we find that binning and classification is more useful at predicting the funding amount than our regression-based models. We also find that the K-nearest model with a nearest value set of 11 had the best performance at predicting the funding bins in comparison to logistic regression and naïve bayes classifier. Multiple linear regression also outperformed our neural network model as shown by the lower MSE score on table 6.

RELATED WORKS

There have been many similar research projects that share the goal of finding the best predictors for the funding amount of a research project. Prediction by the topic of a research project [7, 13, 18] is one of the features that has been shown to have the highest correlation to the funding amount. Topics that are deemed to be highly prominent or hot tend to receive more funding than projects that are not. The genders of the people on the research team have had conflicting results in the past, with some projects finding that there was not a noticeable difference in the amount of funding that women lead projects obtained when compared to men lead projects, and others the suggested otherwise [2, 16]. There have been many studies done on how the authors' social circle and their previous work can impact the funding amount of their project [5, 8]. It was even found in one study that the authors' network was more important in securing funding than the publication provider or whether they have been cited. Some other features that have been studied include how papers with abstracts that contain few common words obtain more funding than ones that do [4], how the number of funding acknowledgements does not increase the funding amount of the project [11], and how the number of doctorate degrees awarded by the school per year can help predict the amount [1]. There have also been many research projects down on adjacent topics to predicting funding amounts such as how funding organizations who distribute their money to several smaller research groups rather than giving larger amounts to more well-known researchers has been shown to increase the performance of those research projects [17, 19] and the fairness that is involved when picking the projects to fund [3].

Given its similar circumstances to the funding of research projects, we also investigated studies done on the predictors in the funding amounts of crowd-funded products [10, 12]. There are many similar indicators that overlap between the two such as the project description being a good predictor of the funding amount [6, 20], but also some unique findings like how the market environment can impact the funding [9] and the social media engagement impact as well [14].

REFERENCES

[1]   Ellyson, E.J., Krueger, J.P. Predicting federal research funding at colleges and universities. Res High Educ 13, 131–136. 1980

[2]   Hechtman L.A., Moore N.P., Schulkey C.E. ,et al.NIH funding longevity by gender. Proc Natl Acad Sci USA. 2018

[3]   Mutz, R., Bornmann, L. Daniel, H.-D. Testing for the fairness and predictive validity of research funding decisions: A multilevel multiple imputation for missing data approach using ex-ante and ex-post peer evaluation data from the Austrian science fund. J Assn Inf Sci Tec, 66: 2321-2339. 2014

[4]   Markowitz, David M. What Words Are Worth: National Science Foundation Grant Abstracts Indicate Award Funding. Journal of Language and Social Psychology 38, no. 3. 264–82. 2019

[5]   Martinez, R.F., Lorza, R.L., Santos Delgado, A. A., Piedra, N. Use of classification trees and rule-based models to optimize the funding assignment to research projects: A case study of UTPL, Journal of Informetrics, Volume 15, Issue 1. 2021

[6]   Zhou, M., Lu, B., Fan, W., et al. Project description and crowdfunding success: an exploratory study. Inf Syst Front 20, 259–274. 2018)

[7]   Neumann, Y. "Predictors of Funding in Academic Fields." Research in Higher Education, vol. 9, no. 2, Springer, pp. 115–22, 1978

[8]   Vaschenko, K. Predicting Probability of Feasibility Funding Based on Academic Characteristics and Entrepreneurial Orientation. 2017.

[9]   Wu, L., Li, Z., Zhao, H., Liu, Q., Chen, E., Estimating fund-raising performance for start-up projects from a market graph perspective. Pattern Recognition, Volume 121. 2022

[10]   Chung, J., Lee, J., A Long-Term Study of a Crowdfunding Platform: Predicting Project Success and Fundraising Amount. In Proceedings of the 26th ACM Conference on Hypertext & Social Media (HT '15). Association for Computing Machinery, 211–220. 2015

[11]   Rigby, J., Looking for the Impact of Peer Review: Does Count of Funding Acknowledgements Really Predict Research Impact? Scientometrics, vol. 94, no. 1, pp. 57–73. 2013

[12]   Zhang, X., Lyu., H., Luo, J. What Contributes to a Crowdfunding Campaign's Success? Evidence and Analyses from GoFundMe Data, in Journal of Social Computing, vol. 2, no. 2, pp. 183-192, 2021

[13]   Klavans, R., Boyack, K. W. Research portfolio analysis and topic prominence, Journal of Informetrics, Volume 11, Issue 4, 1158-1174, 2017

[14]   Yeh, J.-Y., Chen, C.-H., A machine learning approach to predict the success of Crowdfunding fintech project, Journal of Enterprise Information Management, Vol. ahead-of-print No. ahead-of-print. 2020

[15]   Ebadi A, Schiffauerova A, How to Receive More Funding for Your Research? Get Connected to the Right People! 2015

[16]   Leberman S., Eames B., Barnett S., Unless you are collaborating with a big name successful professor, you are unlikely to receive funding. Page 644-661. 2015

[17]   Liefner, I. Funding, resource allocation, and performance in higher education systems. Higher Education 46, 469–489. 2003

[18]   He, Dan, and D. S. Parker. Learning the Funding Momentum of Research Projects. PAKDD'11 Proceedings of the 15th Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining - Volume Part II, pp. 532–543. 2011

[19]   Mongeon P., Brodeur C., Beaudry C., Larivière V., Concentration of research funding leads to decreasing marginal returns, Research Evaluation, Volume 25, Issue 4, Pages 396–404, 2016

[20]   Westerlund M., Singh I., Rajahonka M., Leminen S., Can show-text project summaries predict funding success on crowdfunding platforms? ISPIM Conference Proceeding, 2019