Grant Program Result Prediction: a Real-Time Management System to Estimate Results for a Given Budget

Luna Maltseva
Dept. of Software Engineering of
American University of Central Asia
Bishkek, Kyrgyzstan
md12366@auca.kg

Abstract*—The integration of ARTeMiS by AUCA's Center for Civic Engagement (CCE) as a management system for their Student Initiative Development Program (SIDP) grant program landmarks a steep increase in the SIDP Committee Members' ability to make data-driven judgement calls. However, the system can be further improved by providing forecasts of the potential impact each project will have, wherefrom arises the need for a grant result program prediction. This paper will use regression, data mining, and dynamic programming to estimate the optimal allocation of budget to yield maximal impact.

***Keywords*—SIDP, civic engagement, grant management, grant selection**

I. Introduction

A. Background

The American University of Central Asia, in collaboration with Bard College, had introduced the Student Initiative Development Program in 2015 as a part-time grant program. SIDP went on to become a full-time program by 2017. Since its establishment, its purpose has been to serve as a platform for AUCA students to advance their leadership potential and professional skills and qualities. Within the framework of the program, Student Leaders implement their civic engagement community-based initiatives, receiving mentorship, individual consultations, workshops, and trainings conducted by AUCA's Center of Civic Engagement. The primary goal of the program is to promote and develop Civic Engagement culture and increase the sense of agency among AUCA students who, as future leaders of the region, will encounter complex problems[1].

B. Center for Civic Engagement

The Center for Civic Engagement is an umbrella entity which oversees several subdivisions; specifically: AUCA Sustainable, Legal Clinic AUCA, Institute for Behavioral Health AUCA, and a suite of grant programs—SIDP, ACEP, and Capstone. CCE performs calls for applications for SIDP within the two first two months of both the Fall and Spring semesters. The call for applications is closed within two weeks of its announcement. Following the deadline for submission of applications for the grant, an assembly of Committee Members, specializing in fields related to civic engagement, evaluate and select submitted applications, providing suggestions for the implementation of the project. The results of this process are announced within two to three weeks after the deadline, after which Project Leaders have upwards of five months to implement their community-oriented project.

C. Application Documents

The project applications consist of three items: the proposal, the project leader information, and the budget form.

1) Proposal: The project proposal includes several sections, in which Project Leaders must describe the problem statement, indicate their target group, formulate their goal statement, propose activities,

outline a timeline, list requested resources and staff responsibilities, enumerate partners of the project, brief on the budget, and design the assessment process. The program permits a per-project budget upwards of one thousand dollars at the time of filing.

2) Project leader information: The project leader information similarly includes several sections, in which the Project Leaders must provide basic personal identification, share their aspirations, describe the mission of their project, perform self assessment, and share their vision for the project post its realization.

3) Budget Form: In the budget form, Project Leaders must precisely calculate the expenses the implementation of their project will invoke, such as travel, accommodation, provision, payments, meetings, and other expenses. It is important to note that Project Leaders are not allowed to allocate any of the funds for personal expenses, and will have to uphold every transaction with receipts following the implementation of their project.

D. Post-Application Documents

As an intermediate step to monitor project activity, the Project leader maintains a progress report, which updates the center with key information regarding the ongoing implementation of the project, specifically: project staff, the importance of the project, its outline, the strategic plan, a success story captured within the framework of the project, the current location, and an actions timetable.

Post project implementation, the project leader submits an Advance Report alongside transaction receipts in addition to a Narrative Report.

1) Narrative Report: Within the Narrative Report, the Project Leader details the implementation of the project, featuring an executive summary, a detailed recount and reflection upon project planning, evaluation of the management of the project, a critical assessment of the project execution, feedback on the the project's results and the personal experience of implementing to be used as advice for similar future projects, the financial analysis of the project, and recommendations for oneself were they to attempt the project again.

2) Advance Report: The Project Leader must fill out an official accounting form that outlines financial transactions between all parties involved in the implementation of the project, such as budget allocation, budget spending with details on every transaction executed within the implementation of the project—verified and crosschecked with scanned receipts, and the remainder of the budget. The form is then processed by CCE and forwarded to the Financial Office.

All the documents outlined, circa 2019, were submitted and packaged in an electronic format—specifically, in a .docx format for narrative reports and an .xlsx format for financial reports—via email to sidp@auca.kg.

II. Real Time Management System

A. Construction of the System

After the CCE requested to perform rudimentary data analysis in the Fall of 2024 to provide an overview of the results and impact their suite of grant programs had, which required manual data retrieval to yield accurate results, by November of 2024 it became apparent that there was a need for an automatic data-collection and tracking system. In early February of 2025, the development of such a system was greenlit. For the next eight months, the following development cycle was taking place.

1) Declaration: The CCE declared the need for a set of features within the framework of the system to be implemented, with detailed inquiries into the specifics and vision for the implementation of each individual feature following suit.

2) Implementation: After the technical specifications were compiled, deliverables were implemented following the requirements as closely as possible, relying on internal and external resources to make judgement calls when specifics were missing.

3) Feedback: Upon a MVP implementation of the requested features, the work was presented to the CCE, and with further exploration of implementation details, a new set of requirements was compiled.

4) Revision: Finally, based on the feedback, the product was polished, finalized, and integrated into the system, starting the cycle anew.

B. Digitalized Application Process

The mission critical aspect of the implementation was digitalizing the application process. As a result of consulting internal documents, overviewing past applications to recognize consistent patterns, evaluating other grants' application processes—ADB, EBRD, USAID, C5+1—three easily parsable forms were compiled: the Application form, the Report form, and

the Sign-off form. Upon submission, the forms automatically synchronize with an internal database, where they have metadata attached and are automatically linked to one another.

1) Application form: The Application form was compiled by reviewing past applications and internal and external documents. It merged the Project Proposal, Project Leader Information, and Project Budget into a single form which is designed as a draft and project modeling tool for the framework of the students' civic engagement community-oriented projects. It features twelve sections: Leader's Information, Leader's Motivation, Project Description, Target Group, Goal Statement, Activities, Timeline, Resources, Partnerships, Budget, and Monitoring.

2) Report form: The Report form is an intermediate form that is filled out by the Project Leader throughout the implementation of the project. By design, it features a single section, gathering five metrics: project engagement, updates, and budget allocation. Project Leaders are incentivized to frequently fill out the form as a medium of communication with the CCE, which, with ARTEMIS, allows to track the progress of any accepted project in the system.

3) Sign-off form: The Sign-off form is filled out upon the completion of the project, largely expanding on the information outlined within the Application form, asserting the veracity of the initial claims, as well as gathering specific metrics not detailed in the Report form, and as such capturing the project's results. It combines the Narrative Report and the Advance Report of the former system, resulting in seven total sections: executive summary; reflection, evaluation, and experience; target group report; finalized timeline; staff enumeration, their responsibilities, and resources utilized; partnerships, cooperation, and affiliates; and the budget form.

The system managing the database of documents in real-time received the codename "A Real-Time Management System," or ARTeMiS.

C. Problem Statement

With Applicant Tracking Systems becoming prevalent in the realities of the modern world, there is yet further opportunity for improvement within the scope of ARTeMiS. While many aspects of modern ATS, such as screening or multi-step selection processes[2] could not be considered simply out of the spirit of each application being reviewed with human insight and experience, some of its features: flagging unwanted and unrealistic fields, tracking recurrent projects and applications by Project Leaders, and, specifically, best possible resource allocation to yield maximal results[2] will be explored within this paper to reach conclusions on the future trajectory of the development of ARTeMiS.

Current ATS techniques such as Resume2Vec[3], which parses keywords within applications and feeds them into LLMs (Ollama, Gemini, etc.) to screen applications (and in some cases, assign scores) resulting in high coefficients of determination [$R^2 \geq 0.85$f]). That being outside of the scope of this course, this paper will impose an artificial limit of using simple data science and analysis techniques, alongside single-threaded algorithms to evaluate applications.

This paper will implement grant program result prediction by: (1) attempting to use ridge regression[5] to predict ratings for applications; (2) use datamining[4,6] to extract quantitative confidence intervals[7] of previous rounds of applications; and (3) select applications via multidimensional dynamic programming[8] to saturate a given budget such that the subset yields maximal immediate results.

D. Incoming Limitations

It is safe to assume that the first goal will not be met via the proposed methodology due to the dataset being extremely limited due to the recent debut of ARTeMiS. What furthermore complicates the task are the multiple, inconsistent, and uninferable individual experience-based selections that are made by Committee Members. The aim being not to classify the resulting status of the project—whether it is accepted or rejected—but to predict quantitative ratings of the project limits the selection of algorithms to regularized regression and random forest. The limitations will be further explored and explained in the paper.

What is more, the previous model of applying for SIDP, which consists of a proposal, personal information, a budget form, an intermediate report, a narrative report, and an advance report, being filled out over several .docx and .xlsx files, lack structural consistency, likely making data mining quite difficult. Shall that prove to be the case, the paper will fall back onto a manually collected database of accepted projects in the timeframe between 2019 and 2023.

III. Mathematical Model

A. Briefing On Models Used

This paper will utilize a variety of algorithms to perform the task of grant result prediction. This section will explore the theoretical and mathematical frameworks of Polynomial Regression with Ridge Regularization[5] paired with Term Frequency-Inverse Document Frequency[4], Sequence Labeling alongside Constraint Satisfaction for Result Extraction[6], Confidence Interval[7] creation for estimating the likely range of projects, and singlethreaded Multidimensional Dynamic Programming[8] for project selection.

B. Regression

Application rating prediction is a polynomial multiple regression algorithm which, for a set of objects X (1), finds a set of weights ω (2)

$$X = \left\{1, x_1, x_2, \ldots, x_p\right\},\ X \in \mathbb{R}^{n \times p} \tag{1}$$

$$\omega = \left\{\omega_0, \omega_1, \omega_2, \ldots, \omega_p\right\},\ \omega \in \mathbb{R}^{n \times p} \tag{2}$$

to approximate the value of interest y (3) such that the error ε (4) is minimized.

$$y \sim \hat{y} = X\omega + \varepsilon,\ y \in \mathbb{R}^n \tag{3}$$

$$\varepsilon \sim N\left(0,\ \sigma^2 I\right) \tag{4}$$

To aid regression analysis, categorical columns are processed by term frequency-inverse document frequency algorithm, which, for a set of documents D and terms V calculates the frequency of every term (5)

$$tf_{ij} = \frac{f_{ij}}{\sum_{k=1}^{M} f_{ik}} \tag{5}$$

by boosting the values of scarce terms (6) to construct a matrix X (7).

$$idf_j = \log \frac{N}{1 + n_j} \tag{6}$$

$$X_{ij} = tf_{ij} \times idf_j \tag{7}$$

The regression algorithm is then run through ridge regularization to tune down volumes with excessive noise (8).

$$\hat{\omega}_\lambda = \underset{\omega \in \mathbb{R}^p}{arg\ min} \left\|X\omega - y\right\|_2^2 + \lambda \|\omega\|_2^2 \tag{8}$$

C. Datamining

For the purpose of datamining, sequence labeling, constraint satisfaction, and result extraction are used. The first is a conditional probability model which, for an input x, computes likelihood of a label sequence y using the normalized exponential (11), clamping the sum to [0;1] (12).

$$p(y|x) = \frac{1}{Z(x)} e^{\theta^\top F(x,y)} \tag{9}$$

$$Z(x) = \sum_{y'} e^{\theta^\top F(x,y')} \tag{10}$$

Second, constraint satisfaction, selects optimal set B by maximizing summed penalties (13).

$$\hat{B} = \underset{B}{argmax} \sum_{(n,r) \in B} -\left|pos(n) - pos(r)\right| \tag{11}$$

Last, Result extraction, given a state s, computes the probability of label y using logistic mapping (14), transforming s into a probability through weight vector.

$$p\left(y|s_i\right) = \frac{1}{1 + e^{-\omega^\top \varphi\left(s_i\right)}} \tag{12}$$

On top of that, this paper requires parameter estimation through confidence intervals via standard normal-based expression (15).

$$\hat{\theta} \pm z_{a/2} \sqrt{V(\hat{\theta})} \tag{13}$$

D. Dynamic Programming

Finally, the paper uses multidimensional dynamic programming with a finite planning horizon to compute the minimal achievable outcome. State transitions *s'* consist of deterministic plus noise update (16). The value of any state at time t is obtained by minimizing

all feasible outcomes taking into account both immediate cost and expected future cost (17).

$$s' = f(s, a, \xi) \tag{14}$$

$$V_t(s) = \min_{a \in A(s)} \mathbb{E}\left[c_t(s, a) + V_{t+1}(s') \right] \tag{15}$$

IV. Data

A. Raw Data

This section overviews three sources of raw data. First, ARTeMiS' application data consists of structured data collected in the Fall of 2025. While the relational database's raw application data is private, with further data being absent at the time of this paper being written, a demonstration version of the dataset is provided (40). Nonetheless, the dataset itself has the following form:

```
ID Timestamp                ... R2   R3   Sum
1  2025-10-03 19:43:21.028  ... 0.95 0.95 90.00
2  2025-10-07 12:50:49.238  ... 0.20 0.80 50.00
3  2025-10-08 21:24:50.642  ... 0.75 0.75 78.75
4  2025-10-08 22:43:11.187  ... 0.80 0.90 87.50
5  2025-10-09 15:23:53.319  ... 0.10 0.10 15.00
```

with dimensions `[39 rows x 68 columns]`, consisting of four rating categories, nine numerical categories, with the rest being quasi-structured and structured columns (33).

Second is an assembly of applications and implemented projects for SIDP in the Fall of 2024 and the Spring of 2025, with sixty incoming applications and nineteen implemented projects, formatted according to the description outlined in Section I (34, 35).

Lastly, a manually analyzed collection of SIDP/ACEP project results through 2019 to 2023 (36). While the dataset itself is private, it has the following shape:

```
ID Name                 Theme      ... Duration
1  Photography Training Education  ... 10d
2  Young Debater School Education  ... 4d
3  Juvenile ████        Healthcare ... 5d
4  Basketball for All   Culture    ... 60d
```

with dimensions `[106 rows x 13 columns]`.

B. Data Cleaning and Extraction

For the first item on the list, the ARTeMiS dataset, the dataset was cut down and written into three tables for specific further use. For regression, only informative numerical columns had been left: duration, participants, and budget – to be compared with rating, and stored in `artemis_data_for_regression.xlsx` (19, 38). Needless to say, the presented approach is oversimplified and is unlikely to yield informative results; however, going off previous attempts of harvesting data columns with TF-IDF (5), the results were appalling, yielding R^2 and CV in the range of [≈–21000; –8.31]. This indicates that more sophisticated data extraction methods should be used for informative metrics, such as previously mentioned Resume2Vec[3]. For numeric visualization and testing data, everything from regression and the project's theme, country, and ID have been stored in `artemis_data_numeric.xlsx` (20, 39). Lastly, for dynamic programming data, the ARTeMiS dataset had been taken and had its numeric parameters tuned such that projects with greater feasibility, based on an interval constructed when considering past projects, would become more favorable for selection and stored in `artemis_data_for_DP.xlsx` (21, 37); the specifics of the approach to retrieve data for dynamic programming will be discussed below.

As for SIDP projects in the 2024-2025 timeframe, text-mining has been utilized (9, 11, 12) to harvest raw submitted data of `artemis_data.xlsx` and narratives/* (31, 34). However, the results yielded by the text-mining were unsatisfactory, extracting useful information out of four narrative reports and unable to do so with applications (35), thus establishing a causal chain between the "promised results" and "actual results" impossible to establish via this method. This is mostly due to the inconsistency of data fields in the documents, with some documents not outlining quantitative data, which highlights the need for a structured system such as ARTeMiS. Not being able to utilize a direct causal chain, this paper falls back to the collection of SIDP/ACEP projects and attempts to establish a disconnected causal chain. That will be done by using present projects as precursors for the results of past projects.

Thus, the SIDP/ACEP collection from 2019 through 2024 undergoes a multitude of transformations (36). First, normalized by removing fields and capping them with IQR such that the data fields match the data of current ARTeMiS applications (32, 39, 41). Then, the basis of the past selected projects is masked onto current ARTeMiS' applications by searching for

projects under the same thematic category and possessing similar incoming values such that there is a minimal and maximal randomized overlap. All of that is done to derive the probable min, max, mean, and confidence intervals (13), which are further normalized with expected results (i.e. rating) to prepare the data for Dynamic Programming and ultimately stored in `artemis_data_for_DP.xlsx` (27, 28, 29, 30, 41, 42, 43, 44, 45).

With the data collected and successfully cleaned, the project is ready to proceed with selecting grants via dynamic programming.

C. Visualization of Data

To get a better understanding of the data, the reader is implored to examine the following five graphs in appendices: first of all, ARTeMiS distribution of key variables, generated by code in appendix item 23 (47); second, ARTeMiS project diversity, generated by code in appendix item 24 (48); next, ARTeMiS polynomial regression and ARTeMiS correlation generated by code in appendix item 26 (49); and lastly, ARTeMis range based on past projects generated by code in appendix 30 (50).

Examining the graphs, the following observations can be made: (1) Based on the distributions in appendix items 47 and 48, project applications vary significantly; (2) The measly correlation between numeric values in appendix item 49 indicates that regression, or any other model, will not be able to achieve significant results; (3) The peaks and valleys in the confidence intervals for quantitative variables in appendix item 50 is likely explained by the confidence interval value shrooming as a result of an unlikely proportion of participants over a budget, &c.

V. Result Preiction

A. Choice of Algorithms

For the task of predicting the scores based on the fields of the projects, polynomial regression with ridge regularization has been chosen in favor of random forests, linear support vector machines, and machine learning algorithms as regression would require magnitudes less data to construct a model (21). Given the fact that the incoming dataset only contains 39 entries (38), it will be the best choice. Multidimensional dynamic programming has been chosen over block cipher mode of operation or integer programming as it allows for conditional statements within its execution, which is critical as the task demands project diversity (25, 37).

B. Selection Score Prediction

Ridge regression (α=10) (10) is constructed on a 67/33 split on the `artemis_data_for_regression.xlsx` (38) dataset, fitted, and tested via (21):

```
15 model = Pipeline([
16     ("poly", PolynomialFeatures(degree=2,
   include_bias=False)),
17     ("scaler", StandardScaler()),
18     ("ridge", Ridge(alpha=10))
19 ])
...
23 model.fit(X_train, y_train)
24 test_r2 = r2_score(y_test,model.predict(X_test))
```

Resulting in a cross validated coefficient of determination CV of `-0.145` and a test coefficient of determination R^2 of `-0.083`. The results are to be expected based on the visualizations of the values' correlation (49). The result of regression being effectively useless, the model will not be applied to predict ratings, instead, if at least for the time being, relying only on the Committee Members' personal judgements to select applications.

C. Application Selection for Max Impact

With ARTeMiS' data prepared for dynamic programming (15), a multidimensional dynamic programming recursion loop which is to run with the `artemis_data_for_DP.xlsx` (37) passed as path, delimited by budget of `9700`, and diversity factors' threshold set to `0.8`, is constructed.

```
15  def select_projects_dp(
16      filepath: str,
17      max_budget: int,
18      theme_diversity_factor: float,
19      country_diversity_factor: float,
20      max_states: int = 200_000,
21      verbose: bool = False
22  ) -> List[int]:
```

The reader is invited to examine the function `select_projects_dp.py` located in the code of appendix item 25. The code begins by initializing items in lines 37 to 45. Then, in the for-loop beginning on line 54, IDs are selected such that the rating (feasibility) and participation (impact) is maximized within the constraints of the budget. In the same for-loop, if the

number of lists exceeds 200k, less optimal results are cleaned. Lastly, in the for-loop starting on line 85, the best list of IDs is selected that does not violate the diversity threshold.

D. Limitations

With the artificial and chronological constraints of this project, there is one critical nuance to every algorithm applied.

Ridge regression, or any other classification method, cannot estimate the rating of a given project in a dataset containing 39 entries, mandating the number of columns to be minimized, leaving only the ones containing the most informative details. According to existing studies of ATS systems[3], even with a large dataset, R^2 is not likely to exceed 0.7. What complicates matters in the case of this project is the volatility of both the applications and the ratings of the committee members, which are made on a case-by-case basis by taking into account numerous factors and an individualized background of experience, making the previous limit to be desirable.

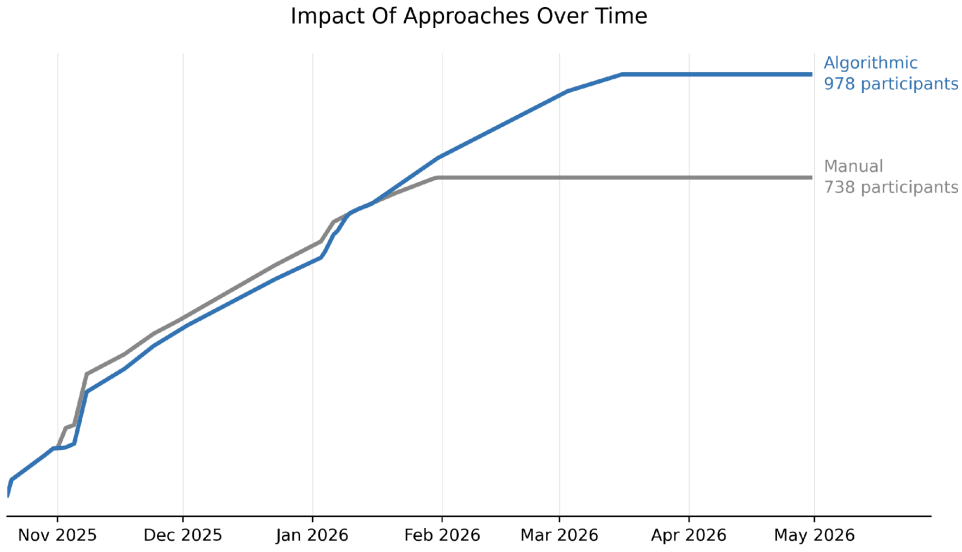
When it comes to estimating the boundary of impact and constructing confidence intervals for current ARTeMiS projects based on previously selected projects, the most apparent issue is the lack of the link between the application and the implementation of the project. While the algorithms attempt to manage that with clamping data, theme-deliniation, and limited proximity-based reference points, the predictions are innately imperfect.

Lastly, having dynamic programming as a project selection tool provides only a rough approximation of the optimal results, as its core idea is to maximally optimize for certain parameters while ignoring unspecified ones.

E. Predictions

Having obtained all the data by running every script in order to obtain all the necessary data, the project selection script (25) is run with the tuned for it ARTeMiS dataset (37), delimited by a total budget of 9700$ and a 0.8 diversity threshold. Resulting, the algorithm selected thirteen projects (compared to the Committee Members' eleven), with the overlap of the list produced by the dynamic programming approach with the list of projects selected by committee members being 72.7%, which both suggests that the committee members are effective in selecting optimal projects and the algorithm is able to find gains. Specifically, it gains an immediate affect of 240 on the same dataset (37) (978 subjects compared to the selected 738 subjects). The progression of the immediate direct impact is demonstrated in the following graph. The two line graphs demonstrate that an overall greater result is achieved, though over the span of a greater period of time.

(51) Comparison between the cumulative results of manual project selection against algorithmic project selection



VI. Conclusion

A. Conclusion

This paper examined the possibility of grant selection for a given budget given the constraints of utilizing classic data science and algorithmic methods, alongside roadblocks in the form of data missing due to the recent debut of ARTeMiS with the difficulty of compensating for it through a clear cause-effect chain of previous projects, along with a shortage of data to train models on. The paper presented solutions to those challenges in the form of leaving, for the time being, the status quo selection process unintruded; estimating the likely resulting range of the projects and shifting it according to the feasibility of the project. In the end, the paper used a dynamic programming approach to select projects, resulting in an overlap of 73% with the manually selected projects and distributing the budget across more projects, resulting in an overall increase in the number of directly affected subjects.

B. Recommendations

Given the transient nature of the exposition of this project, it is impossible to give definitive recommendations for projects of similar constraints. In the case of this project, however, two recommendations can be made to be enacted down the line: first, more data needs to be collected, and that will be done as time passes, hence it would be best to resume work on the project after a year has elapsed; second, more advanced machine learning algorithms need to be used, potentially resorting to existing transformer models on the market that will be hosted locally. Shall these recommendations be followed, the future prospects of the development of ARTeMiS looks bright.

C. Credits

With the paper concluded, the author finds it worthwhile to mention the people thanks to whom the ARTeMiS project is made possible. It is first and foremost thanks to the supervision of the CCE Director Nurzhamal Karamoldoeva and the CCE Coordinator Aliia Iusupova that the project became a reality. The ongoing status of the project is thanks to the SIDP Committee Members, Daniyar Karabaev, Sagynbek Orunbaev, and Phillipe Boizeau, who were willing to be early adopters of the project. Lastly, the author recognizes that the initial steps of the projects have been shaped in collaboration with the Bard's Civic Engagement Office's Director, Jonathan Becker, and Coordinator, Erin Cannan.

Appendix

A. Source Code

(18) ARTeMiS Data Cleaner for Dynamic Programming. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_cleaner_for_DP.py
(19) ARTeMiS Data Cleaner for Regression. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_cleaner_for_LR.py
(20) ARTeMiS Numeric Data Cleaner. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_cleaner_numeric.py
(21) ARTeMiS Regression Model. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_regression.py
(22) ARTeMiS Structure Overview. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_review.py
(23) ARTeMiS Key Variable Visualization. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_review_bar_visualizer.py
(24) ARTeMiS Diversity Visualization. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_review_pie_visualizer.py
(25) ARTeMiS Dynamic Programming Project Selection and Visualization. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_selector.py
(26) ARTeMiS Regression Visualization: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/artemis_visualizer.py
(27) ARTeMiS Result Prediction by Past SIDP & ACEP Projects. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/merger.py
(28) Simplified ARTeMiS Result Prediction by Past SIDP & ACEP Projects. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/merger_simple.py
(29) ARTeMiS Range Prediction Normalized. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/merger_output_normalizer.py
(30) ARTeMiS Result Prediction by Past SIDP & ACEP Projects Normalized Visualized. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/merger_visualizer.py
(31) Recent SIDP Text Mining. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/narrative_parser.py
(32) Past SIDP & ACEP Projects Cleaner. Available: raw.githubusercontent.com/lunamaltseva/ARTeMiS-Data-Science/refs/heads/main/code/past_project_cleaner.py

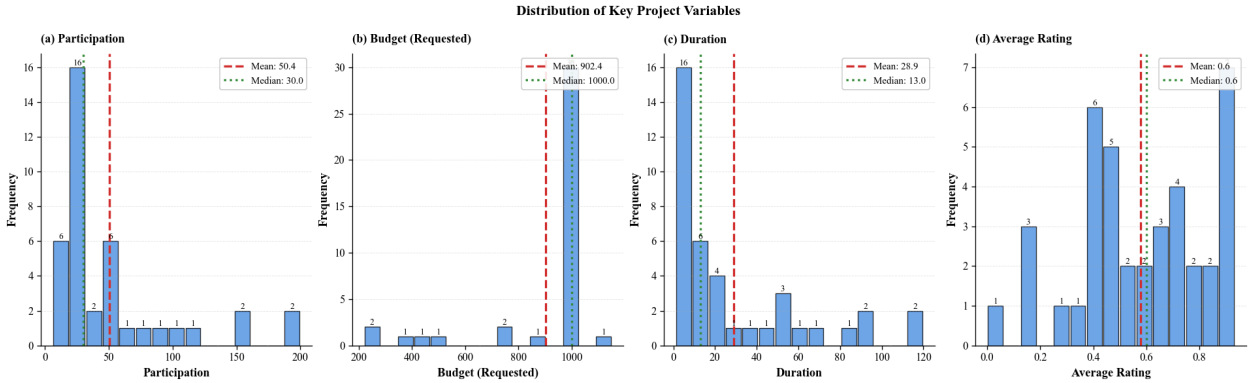
B. Raw and Cleaned Data

(33) ARTeMiS Data. UNAVAILABLE.
(34) SIDP Project Proposals. UNAVAILABLE.
(35) SIDP Implemented Projects. UNAVAILABLE.

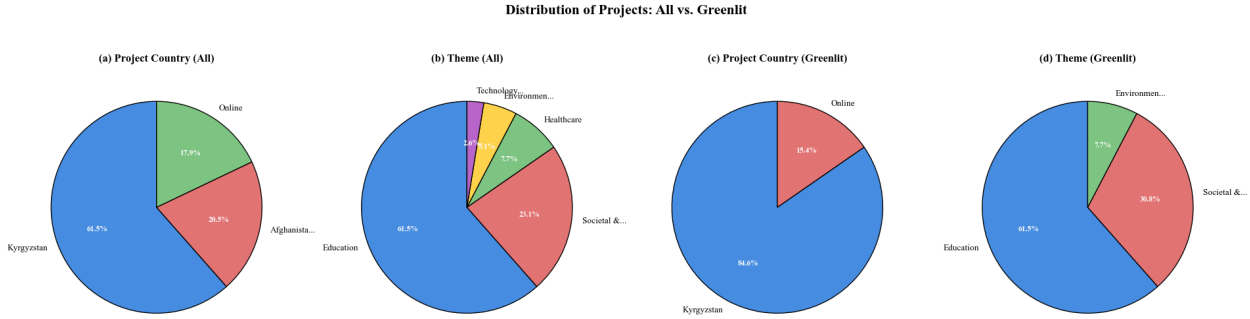
(36) Past SIDP & ACEP Projects. UNAVAILABLE.
(37) ARTeMiS Data Prepared for Dynamic Programming. Available: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/artemis/artemis_data_for_DP.xlsx
(38) ARTeMiS Data Prepared for Regression. Available: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/artemis/artemis_data_for_regression.xlsx
(39) ARTeMiS Data Prepared for Visualization. Available: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/artemis/artemis_data_numeric.xlsx
(40) ARTeMiS Demonstration Data. Available: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/artemis_demo/artemis_demo_data.xlsx
(41) Past SIDP & ACEP Projects Cut-down Data. Available: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/comparison_data/previous_projects_data.xlsx
(42) Past SIDP & Acep Projects Cut-down Cleaned Data. Available: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/comparison_data/previous_projects_data_cleaned.xlsx
(43) ARTeMiS Raw Interval Data: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/interval_data/interval_analysis.xlsx
(44) ARTeMiS Simplified Raw Interval Data: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/interval_data/interval_analysis_applied.xlsx
(45) ARTeMiS Interval Data Normalized: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/interval_data/interval_analysis_applied.xlsx
(46) Past SIDP Projects Narratives Parsed Merged: https://github.com/lunamaltseva/ARTeMiS-Data-Science/blob/main/data/narratives_parsed/merged.xlsx

C. Graphs

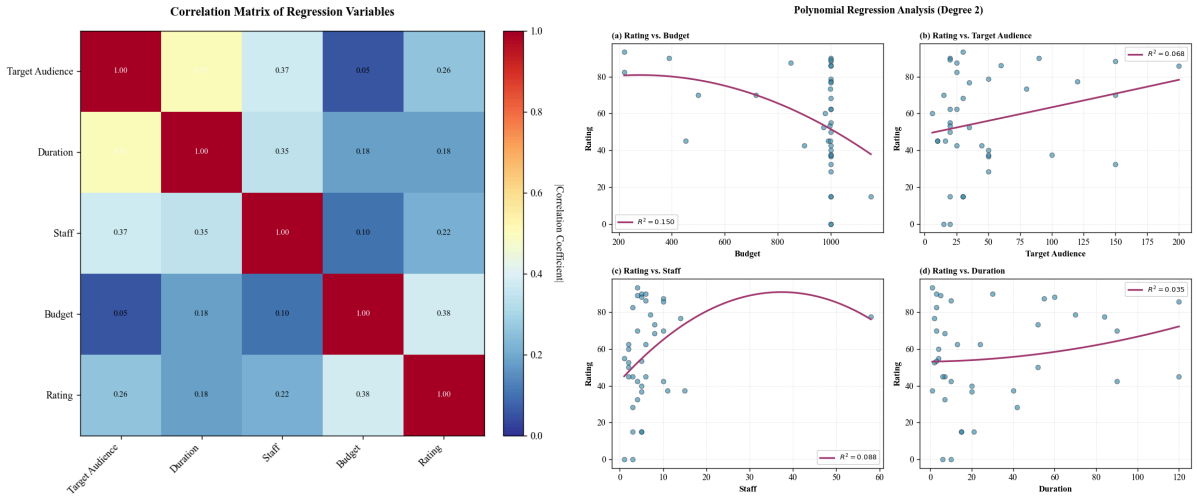
(47) ARTeMiS distribution of key variables, graphed.



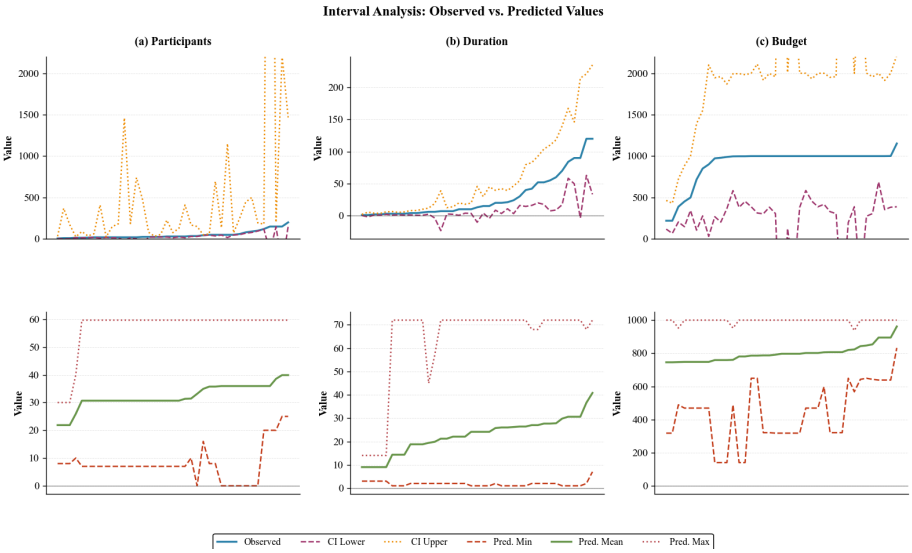
(48) ARTeMiS project diversity, graphed.



(49) ARTeMiS polynomial regression and ARTeMiS correlation



(50) ARTeMis range based on past projectss, graphed.



References

[1] American University of Central Asia - Student Initiative Development Program. auca.kg/en/sts_sidp.

[2] L. Weber, “Your Résumé vs. Oblivion,” Wall Street Journal, Jan. 2012, [Online]. Available: https://www.wsj.com/articles/SB10001424052970204624204577178941034941330

[3] R. V. K. Bevara et al., “Resume2Vec: Transforming Applicant Tracking Systems with Intelligent Resume Embeddings for Precise Candidate Matching,” Electronics, vol. 14, no. 4, p. 794, Feb. 2025, doi: 10.3390/electronics14040794.

[4] A. Ushio, F. Liberatore, and J. Camacho-Collados, “Back to the basics: a quantitative analysis of statistical and Graph-Based term weighting schemes for keyword extraction,” arXiv (Cornell University), pp. 8089–8103, Apr. 2021, doi: 10.48550/arxiv.2104.08028.

[5] G. P. Adhikari, “Interpreting the basic results of multiple linear regression,” Scholars Journal, pp. 22–37, Dec. 2022, doi: 10.3126/scholars.v5i1.55775.

[6] M. Sajan., J. T. Abraham, and S. J. Kalayathankal, “Data mining techniques and methodologies,” International Journal of Civil Engineering and Technology (IJCIET), Jul. 10, 2018. https://iaeme.com/Home/article_id/IJCIET_09_07_025

[7] K. Kang et al., “Accurate confidence and Bayesian interval estimation for non-centrality parameters and effect size indices,” Psychometrika, vol. 88, no. 1, pp. 253–273, Feb. 2023, doi: 10.1007/s11336-022-09899-x.

[8] Y. Zhang, “A survey of dynamic programming algorithms,” Applied and Computational Engineering, vol. 35, no. 1, pp. 183–189, Feb. 2024, doi: 10.54254/2755-2721/35/20230392.