**Tech Funding Solutions**

A picture containing logo

Description automatically generated

**Image Credit:** [**https://www.vectorstock.com/royalty-free-vector/artificial-intelligence-ai-icon-vector-23097736**](https://www.vectorstock.com/royalty-free-vector/artificial-intelligence-ai-icon-vector-23097736)

Jason Fillbach

Jfillb@uw.edu

6/01/2022

University of Washington – Tacoma

**Abstract**

As a private equity firm, Mercury Investments are interested in how current funding rounds for technology companies are influenced by month of the year, round of funding, technology vertical and country of origin. They aim to understand the biggest factors in funding amount, understand capital markets recently, and discover if they can gain an edge through investing in niche verticals such as Artificial Intelligence.

**Data Dictionary**

**Index –** Unique ids for each company who received funding in the dataset

**Company –** The name of the enterprise who obtained funding

**Website –** External links to the company’s website

**Region –** The country of origin for the company within the dataset

**Vertical –** The vertical or category that the tech company specializes in

**Funding amount (USD) –** The amount of us dollars that the technology company received

**Funding Stage –** The stage of equity offering or funding that the money was recorded for

**Funding Date –** The month that funding was secured

**The Original Dataset can be found here:** *Shivam Bansal – Tech Company Fundings*

<https://www.kaggle.com/datasets/shivamb/tech-company-fundings-2020-onwards>

**Data Cleaning**

From the original dataset I chose to exclude a few variables for irrelevance in analysis or the lack of common structure. The usage of the variables website and index may be irrelevant in analysis of the current data sets as it presents a problem in generating any sentiment-based analysis. Beyond this, I changed many of the funding stage metrics used such as “Ukown” “Growth” and “Funding” to the “Undisclosed” Category of the variable due there being such a small sample size or was a result of a typo in data entry.

I chose to transform the “Funding Amount” Label due to its skewness of 48 when I ran my Azure ML Experiment.

**Univariate Summary**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Feature** | **Var Type** | **Data Type** | **Count** | **Missing** | **% Complete** | **Unique** | **Min** | **Q1** | **Med** | **Q3** | **Max** | **Mean** | **SD** | **Skew** | **Kurt** |
| Funding Amount | Label | Numeric | 3566 | 0 | 100% | 3566 | 40000 | 5,00,000 | 15,496,301.50 | 50,000,000 | 16,600,000,000 | 57,560,141 | 298,197,614 | 48.19 | 2,655.46 |
| Company | Feature | Categorical | 3566 | 0 | 100% | 3566 |  |  |  |  |  |  |  |  |  |
| Funding Stage | Feature | Categorical | 3566 | 0 | 100% | 19 |  |  |  |  |  |  |  |  |  |
| Vertical | Feature | Categorical | 3566 | 0 | 100% | 141 |  |  |  |  |  |  |  |  |  |
| Website | Feature | Categorical | 3336 | 230 | 93% | 3336 |  |  |  |  |  |  |  |  |  |
| Region | Feature | Categorical | 73 | 0 | 100% | 73 |  |  |  |  |  |  |  |  |  |
| Funding Date | Feature | Categorical | 3566 | 0 | 91% | 11 |  |  |  |  |  |  |  |  |  |

**Funding Amount (Numerical Label)**

Graphical user interface

Description automatically generated with medium confidenceGraphical user interface, text, application

Description automatically generated

**Bivariate Analysis**

**Funding Amount vs Stage of Funding**

**H#0***:* Due to Series funding progressing through the alphabet, and the typical amount received at later dates of funding as a company, we suspect that there will be an impact on funding amount by stage of funding

Chart, bar chart

Description automatically generatedGraphical user interface, application, table, Excel

Description automatically generated

**Summary:** There is a significant impact of Funding Stage on Funding Amount (P < 0.005)

**Funding Amount vs Region**

**H#1:** Due to certain countries often providing more venture capital, private equity, and the ability to raise capital in certain countries, there will likely be an impact on funding amount by region.

Application

Description automatically generated with medium confidence

Table

Description automatically generated

**Summary:** There is a significant impact of Region on Funding Amount P < 0.005

**Funding Amount vs Date**

**H#2:** We suspect that there will not be impact on funding amount by date due because often deals take variable times to settle.

Chart, timeline

Description automatically generated

Chart, histogram

Description automatically generated

**Regression Equation:** Funding Amount (USD) = -1.76482e+08\*Month of Funding Date^3 + 2.44389e+09\*Month of Funding Date^2 + -7.03835e+09\*Month of Funding Date + 1.8358e+10

**Accuracy:** R^2 = 0.512

**P Value** 0.147572

**Summary 1:** There is not an effect of Funding Date to Funding amount. (P > 0.005)

Graphical user interface, table, Excel

Description automatically generated

**Summary 2:** There is not a significant impact of Month over the two-year span on Funding Amount (P > 0.005)

**Funding Vs Vertical**

**H#4**: We suspect that due to certain industries within technology having higher capital costs, there will be an impact on funding vs vertical.

Chart, timeline, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Bar chart

Description automatically generated with medium confidence

Chart

Description automatically generated

Application, table, Excel

Description automatically generated

**Summary:** There is not a significant impact on the vertical being Artificial Intelligence on Funding amount (P > 0.005)

**Data Exploration Summary**

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Analysis | Effect Size | P-Value |
| Funding Stage | F-Stat | 24.9482145861247 | P < 0.005 |
| Funding Vertical | F-Stat | 0.803769321767774 | P > 0.005 |
| Funding Date | F-Stat | 1.08497248067498 | P > 0.005 |
| Funding Region | F-Stat | 24.9482145861247 | P < 0.005 |

**Correlation Matrix**

Table

Description automatically generated Chart

Description automatically generated with medium confidence

* Color Coded from negative to positive correlation (Red -> Green)
* Attributes were dummy coded to further break down all variables due to lack of variable quantity.
* The Private Equity Funding Stage and Australia Region showcased the highest relationship to funding amount which may be likely skewed to outliers in the Funding amount data.
* The relationship between Seed Funding Stage and Funding Amount has the highest negative correlation -0.10 however is not in a significant range.

**Dashboard Visualization**

Chart, treemap chart

Description automatically generated

Link to Dashboard:

[TechFunding Final - Jason Fillbach | Tableau Public](https://public.tableau.com/app/profile/jason.fillbach/viz/TechFundingFinal-JasonFillbach/Dashboard1?publish=yes)

**Different Model Testing**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature List** | **Algorithm** | **Fit** |  |
|  | Multi-Class Logistic Regression | **Overall Accuracy:** 0.055347  **MAP:** 0.055347 |  |
|  | **Multi-class Neural Network** | **Overall Accuracy:** 0.023452  **MAP:** 0.023452 |  |
|  | **Multi-class Decision Forest** | **Overall Accuracy:**  0.02391  **MAP:** 0.02391 |  |
|  | **Decision Forest Regression** | **R^2:**  0.359342  **RMSE:** 77584070.911594 |  |
|  | **Linear Regression** | **R^2: 0.71048**  **RMSE: 0.904358** |  |
|  | **Decision Forest Regression** | **R^2:** 0.694235  **RMSE:** 0.943862 |  |
|  | **Neural Network Regression** | **R^2: 0.687341**  **RMSE: 0.939803** |  |
|  | **Boosted Decision Tree Regression** | **R^2: 0.698222**  **RMSE: 0.923304** |  |
|  |  |  |  |

* Temporarily in the initial model creation process, I started with classification models that performed poorly, and were not fit for prediction, however, they are still recorded within this table to document choices made.

**Final Model Selection**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature List** | **Algorithm** | **R^2** | **RMSE** |
|  | **Linear Regression** | 0.688493 | 0.311507 |
|  | **Boosted Decision Forest** | 0.706519 | 0.293481 |

Link to Experiment: [**https://gallery.cortanaintelligence.com/Experiment/Business-Intelligence-Final-Jason-Fillbach**](https://gallery.cortanaintelligence.com/Experiment/Business-Intelligence-Final-Jason-Fillbach)

Diagram

Description automatically generated

**Findings**

* The first model was a linear regression that had an R^2 of 0.688493 and RMSE of 0.311507.
* Attributes were weighted based upon the PFI pill in ML Studio
* Attributes Chosen were Funding Stage, Vertical, Region, and Funding Date.
* The variable Vertical was split into the category of Artificial Intelligence and Other (Which included all other technology verticals).
* The variable Region was broken into the five countries with the highest funding. This included the United States, United Kingdom, India, Germany and Australia.
* Data was split 0.8 for training
* From highest to lowest, the PFI score was in the order of Funding Stage, Vertical, Vertical (2), Region(2), Region, and Funding Date. We can assume that Funding Stage is likely one of the largest contributors to Funding amount as its score is the highest at 1.201562, with the next closest being vertical at 0.472366.
* The second model was a Decision-Forest Regression that had an R^2 of 0.706519 and RMSE of 0.293481.
* Attributes were weighted based upon the PFI pill in ML Studio
* All variables chosen were separated the same as the linear regression model.
* PFI order was Funding Stage (1.394423), Vertical (0.155787), Region (0.153973), Funding\_Date (0.135465), Region(2)[ 0.028443], Vertical(2)[ 0.006223].

Based upon both models, we can find that Funding Stage is the variable that most affects funding amount. This can likely be explained by how Series funding is often chronological (Series A, Series B, etc), often leading to varying capital requirements by more mature businesses and due to the varying levels of capital flowing through funding options. We also see within this experiment that vertical does influence Funding Amount and can be likely explained by capital and investment requirements by certain verticals in tech. By categorically separating artificial intelligence from other verticals in both models, we find that it does play a role in funding amount, however its feature label is not weighted as highly as other variables. This could likely be improved by separating the verticals into more than just two groups (Artificial intelligence and other) as done within this experiment.

The dataset presented could often be improved if funding data was combined with a more historical set to take macroeconomic conditions into consideration for investment choice, and ultimately to predict the label Funding Amount. Interesting factors to consider would be the cost to borrow (interest rates), willingness to lend by investment institutions, or expected growth rates of industries within the technology sector.