

Start-up Investment Prediction and Analysis

Yagnesh Salian
2017120054
Batch E
TE EXTC
SPIT

Amey Singh
2017120057
Batch E
TE EXTC
SPIT

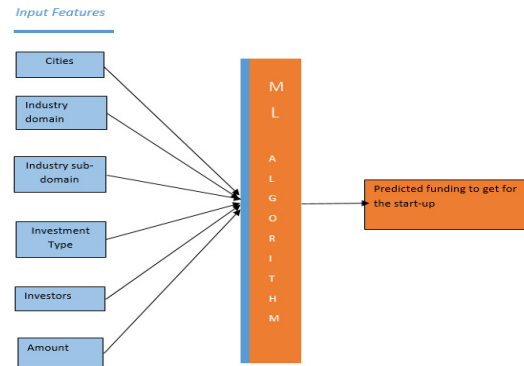
Nitish Wadhavkar
2017120062
Batch E
TE EXTC
SPIT

Abstract—When few people come together propose an innovative solution to a certain problem, they create a prototype test it to decide to turn this idea into a start-up company. This is where start-up investors come into play. Investments and funding are major factors deciding the fate of the start-up, thus before starting any company it is very important to choose the right environment which will help it to flourish. To make this process of finding proper platform easier for such entrepreneurs we've come up with this idea of predicting the expected funding offered to the start-up with certain features like the type of industry, in which city it is located, etc. using machine learning algorithm. This will reduce the struggle of innovators keen to start a company who have less experience in business and help to a great extent for planning the investment strategy.

I. INTRODUCTION

The funding or investment provided to a start-up available or provided by an investor change with changes in geographical position, type of industry and its field of specialization. I also depends on the type of investment needed e.g. private equity, seed funding, etc. Before starting any company rough prediction of the funding that one get is necessary for the creation of its business model. Only a successful solid model will be able to move the company in the right direction. Thus, the right prediction of the investment or funding depending upon the specifications of the company will make the business model more rugged. India is a breeding ground for cultural, grassroots and frugal innovation. Add a population of over one billion to the mix, it becomes an exciting geography for startups to create scalable and repeatable business models. India is home to the world's third largest startup ecosystem, having added over 1,300 tech startups. Also, Startup India, an initiative by Government of India has boosted the startup culture in India. Investments and funding are major factors deciding the fate of the start-up, thus before starting any company it is very important to choose the right environment which will help it to flourish. To make this process of finding proper platform easier for such entrepreneurs we've come up with this idea of predicting the expected funding offered to the start-up with certain features like the type of industry, in which city it is located, etc. using machine learning algorithm.

II. IMPLEMENTATION



A. Dataset

The dataset is based on start-up funding given to various start-ups across India for last three years. Using the dataset, we are making a model using linear regression to predict funding amount for start-ups depending on cities across India and Industry Vertical. To create the model of the algorithm to work with good accuracy a large data set was collected using different resources. The data set consisted on various input features which were type of industry, its subdomain, city where it's located and the investment type with each set having a particular value of funding provided.

B. Creating a model

This data set was divided in proportion into data used for training and testing. Using R script and using linear regression algorithm of machine learning this data set was trained to get a working model which was further tested on the test data set.

C. Code

```
installing packages and loading relevant libraries
install.packages("xlsx"); install.packages("heuristic")
install.packages("caret") install.packages("tidyr")
library("caret") library("heuristic") library("tidyr")
library("xlsx") library("rpart")
loading the dataset
data<- read.xlsx("startup_funding.xlsx", sheetIndex =
1); model <- head(data, 1000) print(head(model));
data wrangling ,cleaning and processing
model[is.na(model)]~0; Oj~ -15; data[is.na(data)]~0;
```

```

modelCity.Location[is.na(modelCity.Location)] =
  "BANGALORE";
modelRemarks <- NULL; modelCity.Location;
toupper(modelCity.Location); modelCity.Location;
trimws(modelCity.Location, which =
  "right"); modelCity.Location;
as.factor(modelCity.Location); head(model)

```

Dividing the dataset into train and test dataset

```

set.seed(1234) ind <- sample(2,nrow(model),replace =
TRUE,prob = c(0.8,0.2)) train_data <- model[ind ==
1,] test_data <- model[ind == 2,]

```

training the dataset to get a model using linear regression

```

mod <- lm(formula =
Amount.In.USD ~ City.Location+Industry.Vertical+
SubVertical+Date+Investment.Type,data = model)
summary(mod)

```

using the test data set ,predicting the funding amount for the test data set

```

pred <- predict(object = mod,newdata = test_data,type =
"response") print(head(pred)) pred[is.na(pred)] <- 0

```

calculating the accuracy

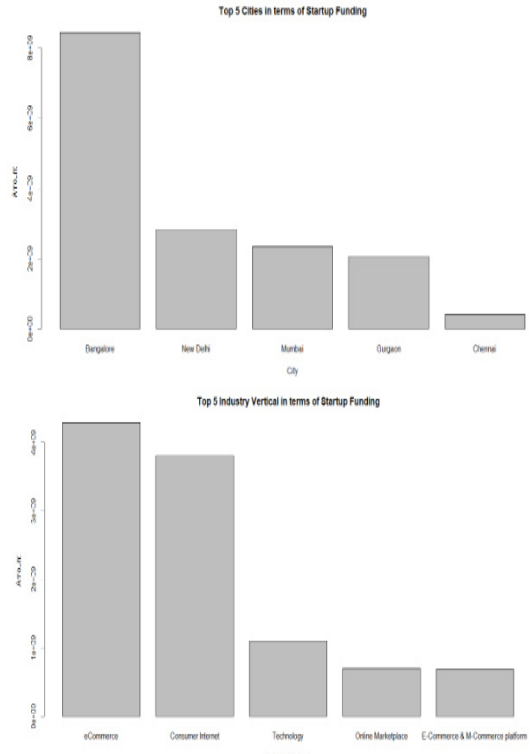
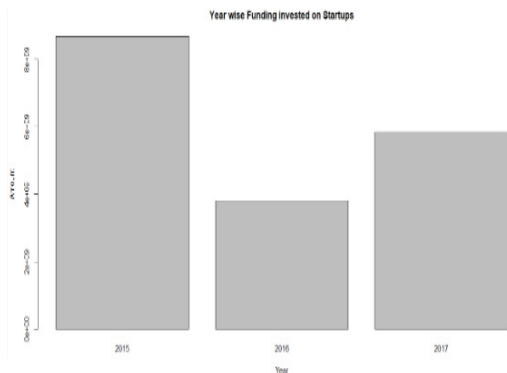
```

acc <- pred/test_data$Amount.In.USD acc[is.infinite(acc)] <- 0
acc <- ifelse(acc < 0,0,acc) accuracy <- sum(acc)/length(acc)
accuracy

```

D. Analysis

We used R-Studio for the analysis of the given dataset. We installed packages such as `xlsx`, `caret`, `tidyr`, etc and loaded relevant libraries which are to be used in the analysis. We divided the given data set into train and test data set Using linear regression, we trained the dataset to get a model using train data set and tested the model by predicting the funding for the test data set. We achieved an accuracy of 91.47 on this model.



E. Output

Summary of the Model Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1 Residual standard error: 214100000 on 33 degrees of freedom Multiple R-squared: 0.6483, Adjusted R-squared: -9.626 F-statistic: 0.0631, p-value: 1 Accuracy of the Model =91.47

III. CONCLUSION

Such a model can be vast help to budding innovators and entrepreneurs for targeting markets as per cities and would have much higher chances of succeeding. More data being added to the given dataset will increase the accuracy of the model.

ACKNOWLEDGMENT

We would thank our professor Dayanand Ambawade for mentoring us on this project idea and inspiring us to think with our full potential. We would also acknowledge our college SPIT for giving us a platform to work on our idea.

REFERENCES

- [1] H. Kopka and P. W. Daly, *A Guide to L^AT_EX*, 3rd ed. Harlow, England: Addison-Wesley, 1999.