Statistical and Computational Laboratory Project

Project Topic: City-wise Start-up investment prediction model

Group Members:

Yagnesh Salian (2017120054)

Amey Singh (2017120057)

Nitish Wadhavkar (2017120062)

Introduction

Problem Statement

In our project, we have created a Machine learning model that would be able to predict the investment that a given start-up from a specific domain would receive by analyzing a huge dataset containing the start-up funding data from major cities across India

Motive behind this topic

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.

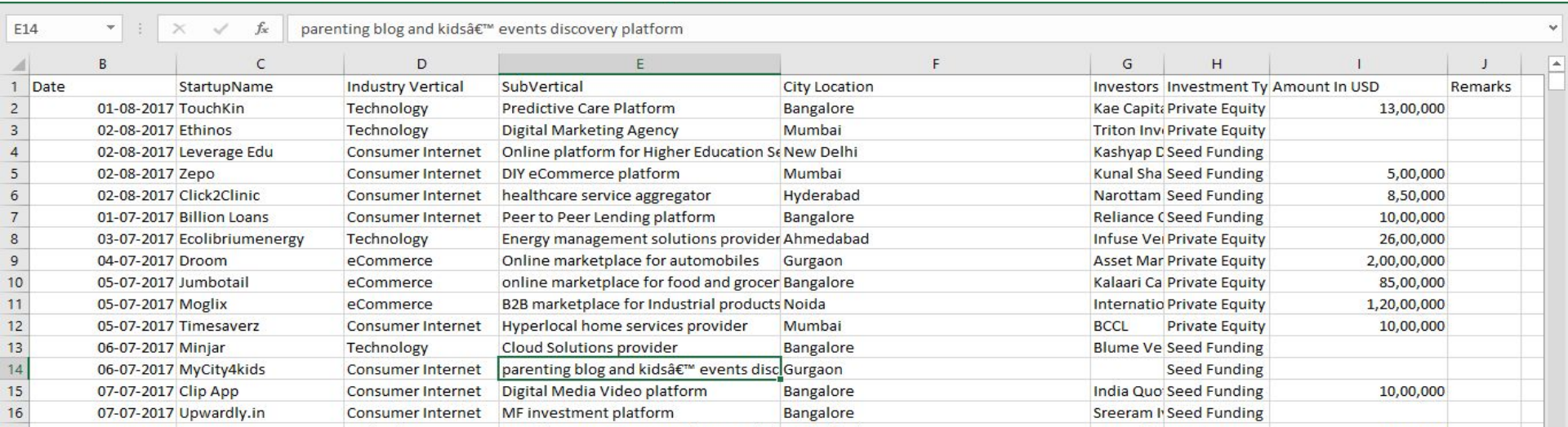
To make this process of finding proper platform easier for such entrepreneurs we've come up with this idea of predicting the expected funding offered to the start-up with certain features like the type of industry, in which city it is located, etc. using machine learning algorithm.

LET'S HAVE A LOOK AT OUR DATASET……

DATASET

Our dataset is based on startup funding given to various startups across India for last three years.

Using the dataset, we are making a model using linear regression to predict funding amount for startups depending on cities across India and Industry Vertical.



E14 | parenting blog and kidsâ€™ events discovery platform

	B	C	D	E	F	G	H	I	J
1	Date	StartupName	Industry Vertical	SubVertical	City Location	Investors	Investment Ty	Amount In USD	Remarks
2	01-08-2017	TouchKin	Technology	Predictive Care Platform	Bangalore	Kae Capita	Private Equity	13,00,000	
3	02-08-2017	Ethinos	Technology	Digital Marketing Agency	Mumbai	Triton Inv	Private Equity		
4	02-08-2017	Leverage Edu	Consumer Internet	Online platform for Higher Education Se	New Delhi	Kashyap D	Seed Funding		
5	02-08-2017	Zepo	Consumer Internet	DIY eCommerce platform	Mumbai	Kunal Sha	Seed Funding	5,00,000	
6	02-08-2017	Click2Clinic	Consumer Internet	healthcare service aggregator	Hyderabad	Narottam	Seed Funding	8,50,000	
7	01-07-2017	Billion Loans	Consumer Internet	Peer to Peer Lending platform	Bangalore	Reliance C	Seed Funding	10,00,000	
8	03-07-2017	Ecolibriumenergy	Technology	Energy management solutions provider	Ahmedabad	Infuse Ve	Private Equity	26,00,000	
9	04-07-2017	Droom	eCommerce	Online marketplace for automobiles	Gurgaon	Asset Mar	Private Equity	2,00,00,000	
10	05-07-2017	Jumbotail	eCommerce	online marketplace for food and grocer	Bangalore	Kalaari Ca	Private Equity	85,00,000	
11	05-07-2017	Moglix	eCommerce	B2B marketplace for Industrial products	Noida	Internatio	Private Equity	1,20,00,000	
12	05-07-2017	Timesaverz	Consumer Internet	Hyperlocal home services provider	Mumbai	BCCL	Private Equity	10,00,000	
13	06-07-2017	Minjar	Technology	Cloud Solutions provider	Bangalore	Blume Ve	Seed Funding		
14	06-07-2017	MyCity4kids	Consumer Internet	parenting blog and kidsâ€™ events disc	Gurgaon		Seed Funding		
15	07-07-2017	Clip App	Consumer Internet	Digital Media Video platform	Bangalore	India Quo	Seed Funding	10,00,000	
16	07-07-2017	Upwardly.in	Consumer Internet	MF investment platform	Bangalore	Sreeram I	Seed Funding		

Analysis steps

We used R-Studio for the analysis of the given dataset.

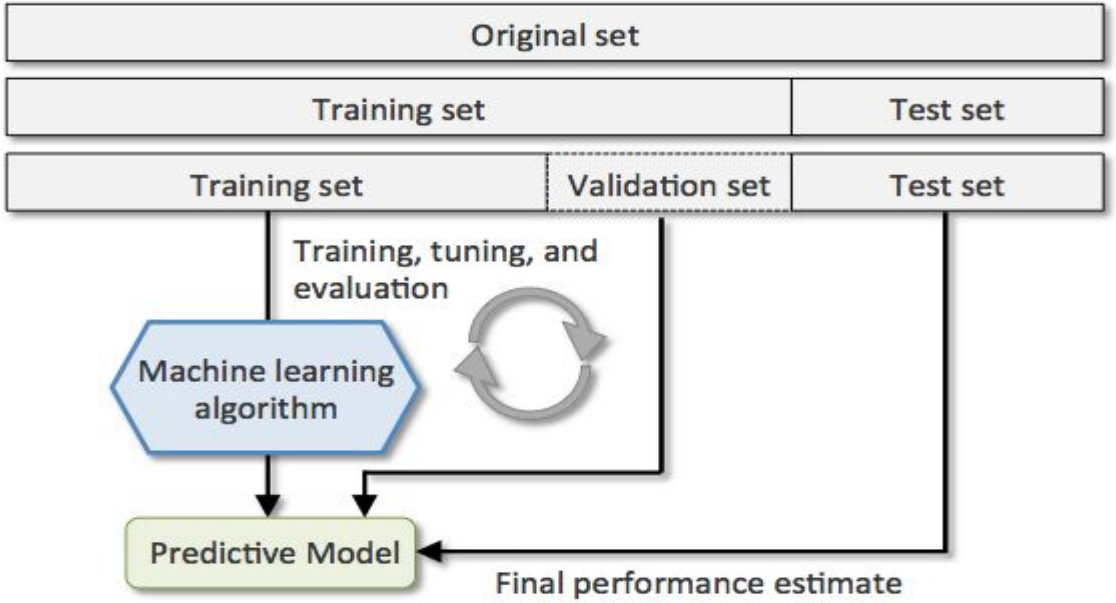
We installed packages such as xlsx,caret,tidyr,etc and loaded relevant libraries which are to be used in the analysis.

We divided the given data set into train and test data set

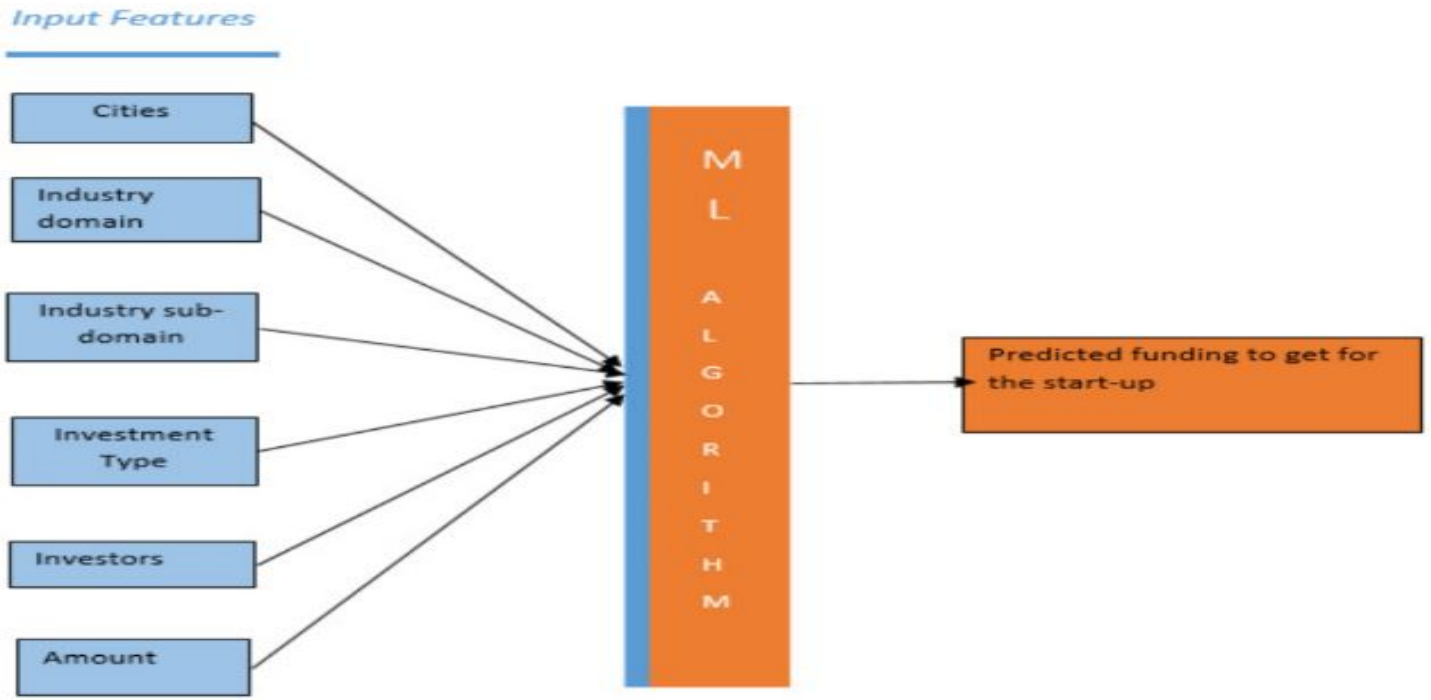
Using linear regression,we trained the dataset to get a model using train data set and tested the model by predicting the funding for the test data set.

We achieved an accuracy of 91.47% on this model.

ALGORITHM OF OUR CODE



ALGORITHM OF OUR CODE



Code Results

- 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

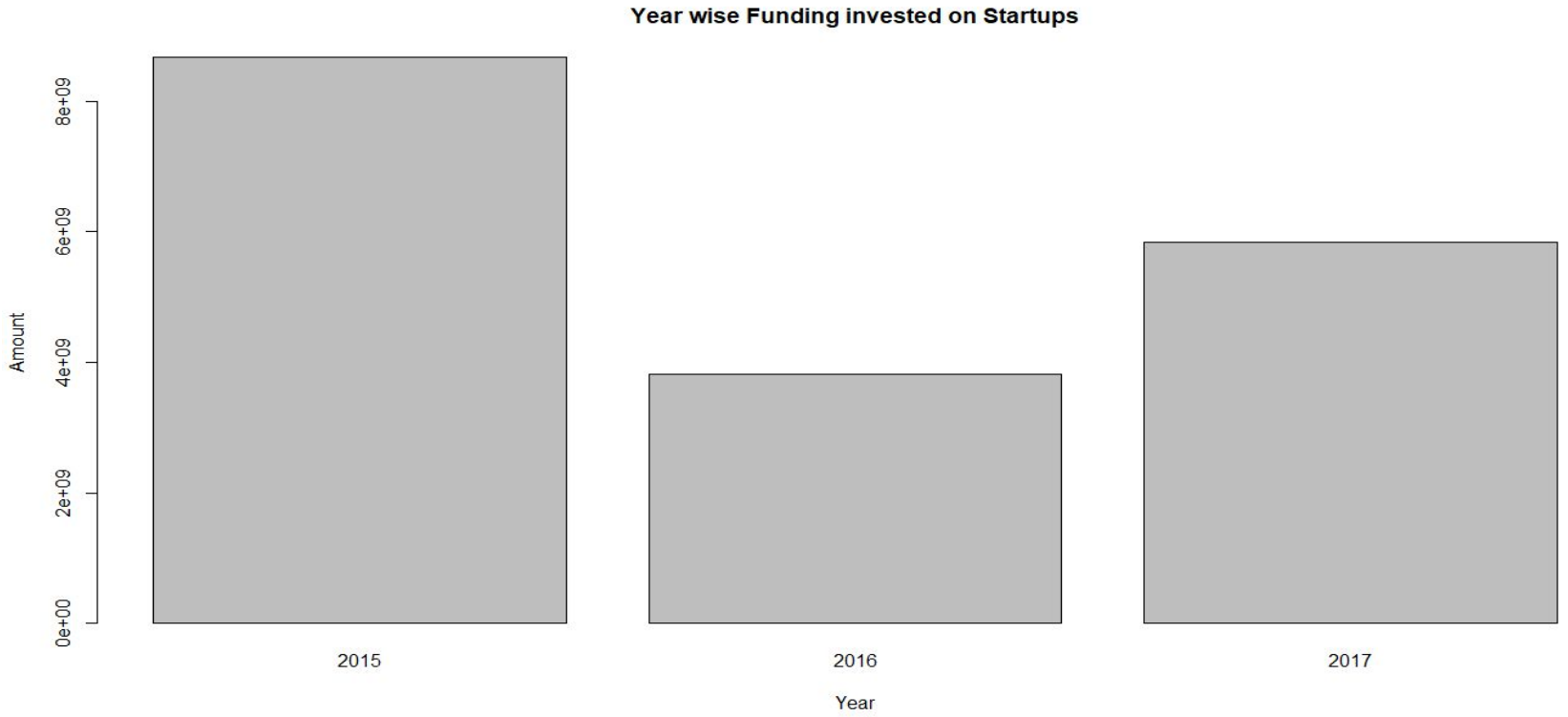
  Multiple R-squared: 0.6483,

  Adjusted R-squared: -9.626

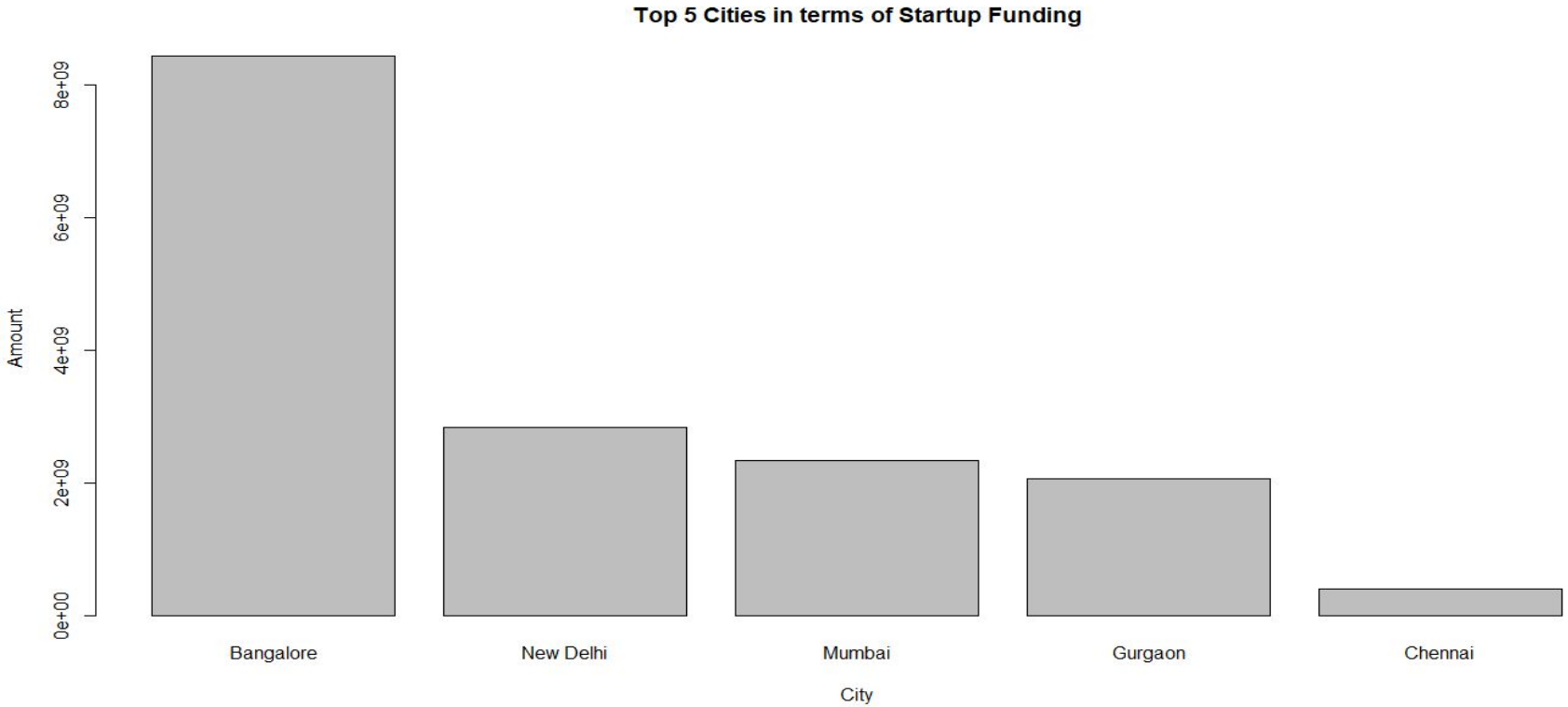
  F-statistic: 0.0631,

  p-value: 1
- Accuracy of the Model =91.47%

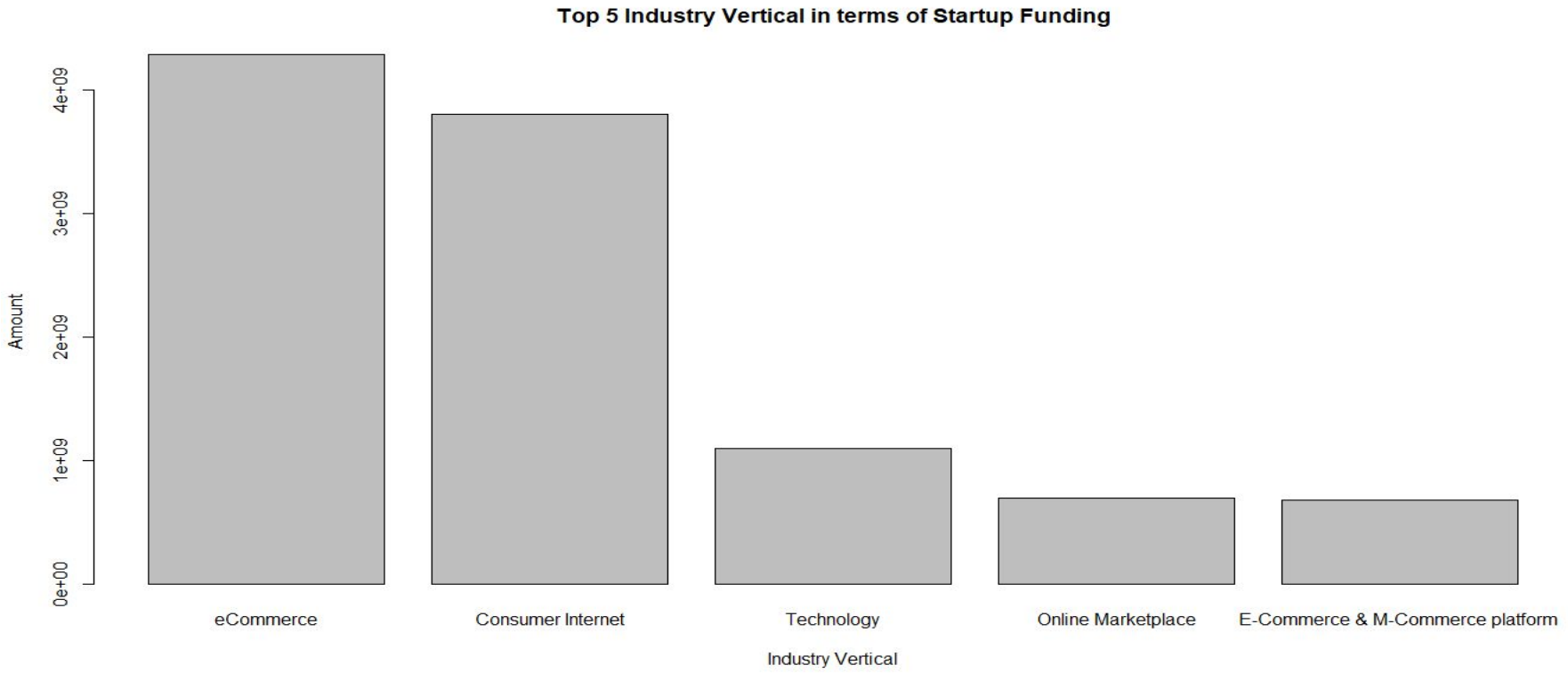
Code Results



Code Results



Code Results



Impact and Use

- ***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.***