STARTUP PROPHET

An Internship Project Report

Submitted by:

SALLA SAHAS(21UK1A6704)

SUDHATI KIRAN(21UK1A6753)

GATTIKOPPULA MANI(21UK1A6758)

AKARAPU JYOTHI(21UK1A6743)

**INDEX**

**TITLE PAGE NO**

**CHAPTER 1: INTRODUCTION**

* 1. OVERVIEW **3**
  2. PURPOSE **3**

**CHAPTER 2: LITERATURE SURVEY**

2.1 EXISTING PROBLEM **4**

2.2 PROPOSED SOLUTION **4**

**CHAPTER 3: THEORETICAL ANALYSIS**

3.1 BLOCK DIAGRAM **5**

3.2 SOFTWARE & HARDWARE DESIGNING **6-7**

**CHAPTER 4: EXPERIMENTAL INVESTIGATION 8-9**

**CHAPTER 5: FLOW CHART 10**

**CHAPTER 6: RESULTS 11-15**

**CHAPTER 7: ADVANTAGES & DISADVANTAGES 16**

**CHAPTER 8: APPLICATIONS 17**

**CHAPTER 9: CONCLUSION 17**

**CHAPTER 10: FUTURE SCOPE 17**

**CHAPTER 11: BIBILOGRAPHY 18**

**CHAPTER 1**

**INTRODUCTION**

* 1. **OVERVIEW:**

A startup or start-up is a company or project begun by an entrepreneur to seek,  develop, and validate a scalable economic model. While entrepreneurship refers to all new businesses, including self-employment and businesses that never intend to become registered, startups refer to new businesses that intend to grow large beyond the solo founder. Startups face high uncertainty and have high rates of failure, but a minority of them do go on to be successful and influential.  Startups play a major role in economic growth. They bring new ideas, spur innovation, create employment thereby moving the economy. There has been an exponential growth in startups over the past few years. Predicting the success of a startup allows investors to find companies that have the potential for rapid growth, thereby allowing them to be one step ahead of the competition.

            The objective is to predict whether a startup which is currently operating turns into a success or a failure. The success of a company is defined as the event that gives the company's founders a large sum of money through the process of M&A (Merger and Acquisition) or an IPO (Initial Public Offering). A company would be considered as failed if it had to be shut down.

* 1. **PURPOSE:**

The purpose of this project is to predict the startup prophet on machine learning model classification using the support vector machine algorithm and random forest model by this we can be capable to predict the startup prophet classification.

we will prepare the data using JUPYTER notebook and we use various models to predict the output.machine learning models are used very useful in predicting outcomes for large amount datasets. We use support vector machine and random forest model machine learning algorithm to predict the startup prophet classification.

**CHAPTER 2:**

**LITERATURE SURVEY**

**2.1 EXISTING PROBLEM**

India is seeing a growth phase under the leadership of able people. However, there still exist many challenges that need to be addressed. To solve these challenges and problems, the country as a whole must be engaged, and talent must be brought from outside the government domain, especially where domain knowledge or entrepreneurial leadership is required. People who are passionate create great things, and companies that aspire to solve bigger problems do much better than those who just look around for funding and money. A combination of talent and diverse experiences backed by strong political will are the key ingredients to coming up with out-of-the-box solutions to address the many challenges we face as a developing country. We look at some of the real issues in India that startups can aim to address. Instant access to healthcare One of the most critical needs today is access to good healthcare. Billions around the world, particularly people in the Indian subcontinent, struggle because they do not get proper access to healthcare. Even those with access have a sour experience. We have apps that let us book movie tickets and seats in a jiffy or even find that perfect restaurant! However, finding doctors is still unbelievably tough. Patient records are either maintained in fat files or if they are online, they are often not accessible or understandable. Doctors do not usually have the time to go through all the reports and this may lead to a compromise on the health front. Health-based startups can address a lot of issues plaguing instant access to healthcare in India. Healthcare is undergoing a major change and smartphones will soon replace doctors for more than 80 percent of health-related problems! Public transportation In India, the pains of a city’s chaotic public transport system.

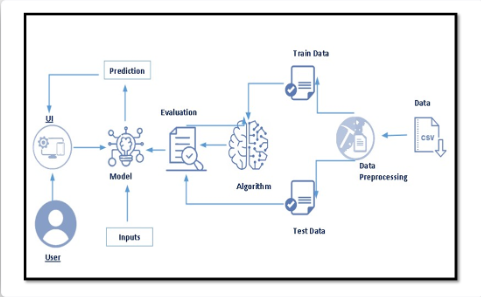
**2.2 PROPOSED SOLUTION:**

we will prepare the data using JUPYTER notebook and we use various models to predict the output. machine learning models are used very useful in predicting outcomes for large amount datasets. We use support vector machine and random forest model machine learning algorithm to predict the startup prophet classification.

**CHAPTER 3**

**THEORETICAL ANALYSIS**

**3.1 BLOCK DIAGRAM**



**3.2 SOFTWARE & HARDWARE DESIGNING:**

Software requirements

|  |  |
| --- | --- |
| **REQUIREMENTS** | **SPECIFICATIONS** |
| Anaconda Navigator | You must have anaconda installed in your device prior to begin. |
| * SPYDER ,JUPYTER Notebook, Flask * Frame work   . | * One should have SPYDER and JUPYTER notebook. * One should install flask framework through Anaconda prompt for running their web application. * We need to build the mode; using JUPYTER notebook with all the imported packages. |
| Web browser | For all Web browsers, the following must be enabled:   * Cookies * Java script |

Hardware requirements:

|  |  |
| --- | --- |
| **REQUIREMENTS** | **SPECIFICATIONS** |
| Operating system | * Microsoft windows * Unix * Linux |
| Processing | Minimum: 4 CPU cores for one user. For each deployment, a sizing exercise is highly recommended. |
| RAM | Minimum 8 GB. |
| Operating system specifications | File descriptor limit set to 8192 on UNIX and Linux |
| Disk space | A minimum of 7 GB of free space is required to install the software. |

**CHAPTER 4**

**EXPERIMENTAL ANALYSIS**

Analysis or the investigation made while working on the solution:

While working on the solution we investigated on what is Startup prophet, visualizing and analyzing the data, data pre-processing, Machine Learning service, model building. The key role on investigation is collection of data set.

**DATA PRE-PROCESSING:**

As we have understood how the data is let’s pre-process the collected data.

The download data set is not suitable for training the machine learning model as it might have so much of randomness so we need to clean the dataset properly in order to fetch good results. This activity includes the following steps.

🡪       Handling missing values

🡪     Handling categorical data

🡪     Scaling Techniques

🡪     Handling class imbalance

🡪     Splitting dataset into training and test set

Note: These are the general steps of pre-processing the data before using it for machine learning. Depending on the condition of your dataset, you may or may not have to go through all these steps.

**DATA SET COLLECTION:**

* Kaggle.com
* Machine learning repository

The data set contains thirteen classes:

1.is\_ecommerce

2.is\_otherstate

3.range

4.has\_roundA

5.has\_roundB

6.has\_roundC

7.has\_roundD

8.is\_top500

9.labels

10.has\_VC

11.funding\_rounds

12.relationships

13.milestones

**CHAPTER 5:**

**FLOW CHART**

**START**

DATA COLLECTION

VISUALZING AND

ANALYSING DATA

DATA PRE-PROCESSING

MODEL BUILDING

APPLICATION BUILDING

**CHAPTER 6**

**RESULTS**

The following code is used in our project

**STARTUP PROPHET**

Importing the libraries

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

%matplotlib inline

import seaborn as sns

sns.set(style="whitegrid")

from sklearn.preprocessing import StandardScaler

from imblearn.over\_sampling import SMOTE

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import confusion\_matrix,accuracy\_score,classification\_report

from sklearn.metrics import log\_loss

from sklearn.svm import SVC

from sklearn.ensemble import RandomForestClassifier

import pickle

import warnings

read the dataset:

df=pd.read\_csv('/content/startup prophet.csv')

df.head()

univariate analysis:

plt.figure(figsize=(12,5))

plt.subplot(131)

sns.distplot(df['relationships'],color='y')

plt.subplot(132)

sns.distplot(df['funding\_rounds'])

plt.show()

bivariate analysis:

sns.countplot(x='funding\_rounds',hue='status',data=df,palette='nipy\_spectral',order=df.funding\_rounds.value\_counts().index)

multivariate analysis:

plt.figure(figsize=(25,18))

sns.heatmap(df.corr(),annot=True,linewidth=0.5,cmap='viridis',fmt='.1f')

sns.countplot(x=df['labels'],palette='viridis')

descriptive analysis:

df.describe

**DATA PRE-PROCESSING:**

Checking for null values and dropping the values

df.isnull().sum()

df.shape

df=df.drop(columns=['Unnamed: 0', 'state\_code', 'latitude', 'longitude', 'zip\_code', 'id','city', 'Unnamed: 6', 'name',  'founded\_at'])

df=df.drop(columns=['avg\_participants','status','is\_biotech', 'is\_consulting',

       'is\_othercategory', 'object\_id','category\_code', 'is\_software',

       'is\_web', 'is\_mobile', 'is\_enterprise', 'is\_advertising',

       'is\_gamesvideo','state\_code.1', 'is\_CA', 'is\_NY',

       'is\_MA', 'is\_TX','funding\_total\_usd','closed\_at',

       'first\_funding\_at', 'last\_funding\_at', 'age\_first\_funding\_year',

       'age\_last\_funding\_year', 'age\_first\_milestone\_year',

       'age\_last\_milestone\_year'])

df.shape

handilng categorical data:

df.dtypes

scaling techniques:

x=df.drop(columns=['labels'],axis=1)

y=df['labels']

sc=StandardScaler()

x=sc.fit\_transform(x)

x

handling class imbalance:

smote=SMOTE(sampling\_strategy='auto',random\_state=42)

x\_bal,y\_bal=smote.fit\_resample(x,y)

splitting data into train and test:

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x\_bal,y\_bal,test\_size=0.3,random\_state=42)

**MODEL BUILDING:**

logistic regression:

lr=LogisticRegression()

lr.fit(x\_bal,y\_bal)

y\_pred=lr.predict(x\_test)

print(confusion\_matrix(y\_test,y\_pred))

print(classification\_report(y\_test,y\_pred))

support vector machine:

svm=SVC(kernel='rbf',C=2.0,random\_state=42)

svm.fit(x\_bal,y\_bal)

y\_predict=svm.predict(x\_test)

print(confusion\_matrix(y\_test,y\_predict))

print(confusion\_matrix(y\_test,y\_predict))

random forest classifier:

rf=RandomForestClassifier()

rf.fit(x\_bal,y\_bal)

rftest=rf.predict(x\_test)

rftrain=rf.predict(x\_train)

print(confusion\_matrix(rftest,y\_test))

print(confusion\_matrix(rftrain,y\_train))

print(classification\_report(rftest,y\_test))

print(classification\_report(rftrain,y\_train))

testing the model:

rf.predict([[3,3,3,0,0,0,1,0,0,0,0,0]])

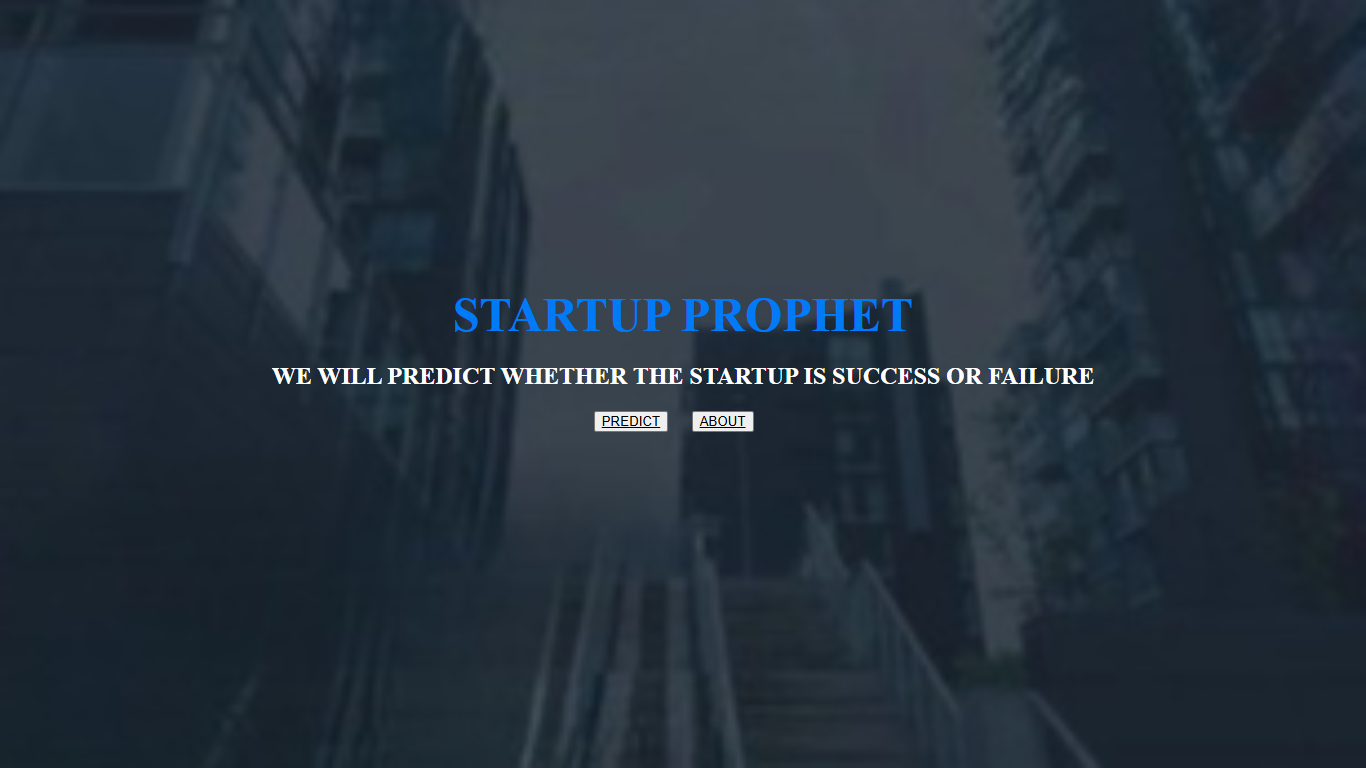
rf.predict([[9, 4,  1 ,0, 0,  1,  0,  0,  1,  1,  1,  1]])

pickle model save:

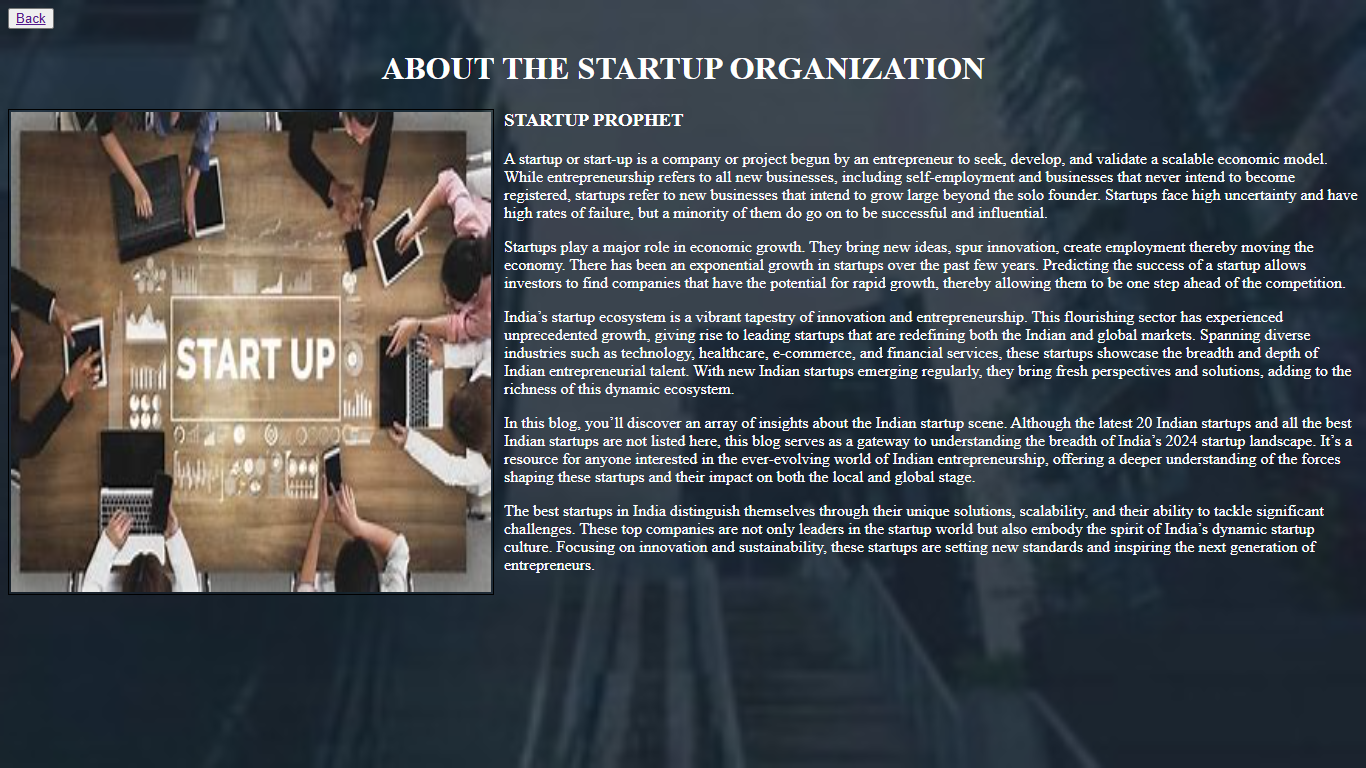
pickle.dump(rf,open("Randf.pkl","wb"))

**FINAL OUTPUT OF THE PROJECT**

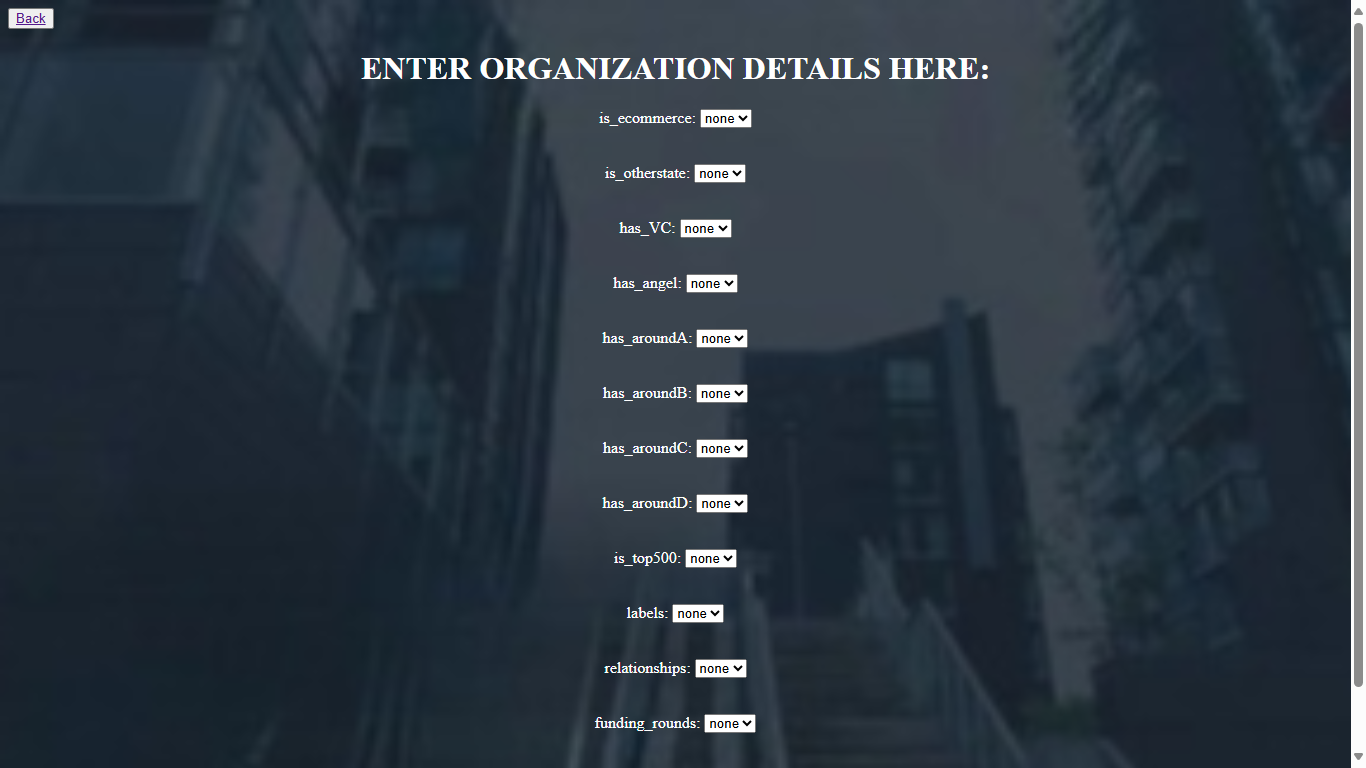
The home page will be shown as:



The info page about startup prophet



The prediction page for startup



The result page is shown as

**CHAPTER 7**

**ADVANTAGES & DISADVANTAGES**

**ADVANTAGES:**

* Being your own boss
* Flexibility
* Financial rewards
* Opportunity to innovate
* Chance to impact your community

**DISADVANTAGES:**

* High costs and limited revenue
* Lack of a developed business model
* Inadequate capital to move to the next phase
* Risk of failure is high
* Long working hours are the norm

**CHAPTER 8**

**APPLICATIONS**

**The areas where this solution can be used:**

* By stakeholders to predict the growth of the company.
* By businessman to know whether the company will

Is Success or failure.

**CHAPTER 9**

**CONCLUSION**

**FROM THIS PROJECT WE HAVE CONCLUDED THAT:**

🡪     We have Known fundamental concepts and techniques used for machine learning.

🡪     Gain a broad understanding about data.

🡪     Have knowledge on pre-processing the data/transformation techniques and some

Visualization concepts.

**CHAPTER 10**

**FUTURE SCOPE**

**Enhancements that can be made in the future**

 🡪       Know fundamental concepts and techniques used for machine learning.

🡪       Gain a broad understanding about data.

🡪       Have knowledge on pre-processing the data/transformation techniques and some visualization concepts.

**CHAPTER 11**

**BIBILOGRAPHY**

References of previous works or websites visited/books referred for analysis about the project, previous solution findings and previous

STARTUP PROPHET documents.