



STARTUP SUCCESS PREDICTION

MACHINE LEARNING



SUBMITTED BY:
MUHAMMED SAVADK
DATA SCIENCE



Indroduction

- This is a machine learning project to predict startup success or not. Startup success prediction refers to the process of utilizing machine learning algorithms to analyze various data points and make predictions about the likelihood of a startup's success. By leveraging historical and real-time data, we can develop models that assess the potential of a startup, evaluate risks, and guide strategic decision-making. These predictive models empower entrepreneurs, investors, and stakeholders to allocate resources effectively, optimize growth strategies, and increase the probability of success.

Problem statement

Startup investment can be very risky due to the high failure rate of startups. People like angel investors and venture capitalists have a very high risk while they are investing in startups. To assist startup investors with their decisions, in this project we aim to find the important features that lead to startup success and forecast a company's success with supervised machine learning methods.



Goals

- The goal to be achieved is to determine whether a StartUp will be successful or not.



Objective

- The objective is to analyze startup behavior based on several variables, determine what variables affect startup success the most, then build a model that can predict the success of a StartUp.
- 
- 

Methodology

- Before we could use the data to train the different models, we had to clean the data and select the most important columns to be included into the model. One of the biggest problems we had with the dataset was that it had a lot of zeros and a lot of columns to choose from.
- We also realized later that the status column had around 80% of the companies as operating status and the rest as closed and acquired companies.

About data

* To train the machine learning model, we used investment data about startup companies available on Kaggle. The data has been collected from Crunchbase which is a leading website for company insights from early stage startups

*The data had around 54k rows and 29 columns. The dataset had company information such as name of the company, url, market, country, state, region, city, founded date, first funding date, last funding date. It also had data on different investment types such as seed, venture equity crowdfunding, undisclosed funding, convertible note, debt financing, angel, grant, private equity, post ipo equity, post ipo debt, secondary market, product crowdfunding, round A-H series funding

Feature explanations:

- **age-first funding year**-age first funding year is the age of the company in years since it got first funding. similar for age last funding year.
- **milestone**-milestone for any startup is a tracking mark for startups. Just like a milestone on the side of a road marks how far you've gone, a milestone in startups tracks progress as an startup grow and implement their plan.
- **relationships**- it says how many relationship does a startup have. For example a start up can have relationships with accountants, investors, vendors, mentors, etc.



con->

venture capitalists(seed)-->its a form of small financing companies that provides funds for upcoming startups,its help to get pmf

angel-->angel investing means individuals who provide financial support or fund to upcoming startups or for developing a startup in exchange a share of owner ship is given.

Avg-participants

is Top 500

category or industries

total funding amount





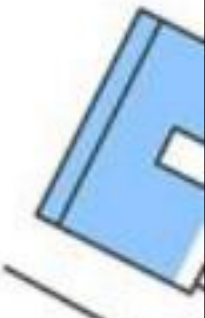

con->

A- Round A is the first stage of financing that a startup raises after its initial step. It is used for further development of product, service, and expand the team and scale of business operations.

B- Round B comes from after a successful round A. By this time, it might have achieved targets and market traction. This fund is used for increasing marketing effort, advancement of infrastructure and hire additional talent.

C- Round C occurs when a company achieves growth and success. This fund is used for expanding new markets and acquiring complementary business.

D- Round D means late stage. That is, a company requires additional capital to fuel its expansion. At this stage, startups are well established and preparing for initial public offering. This fund is used for further growth, research and development also to increase market share.



```
graph TD; A[Data collection] --> B[Data preprocessing]; B --> C[Data visualization]; C --> D[Model selection]; D --> E[Model evaluation]; E --> F[Model deployment];
```

Data collection

Data preprocessing

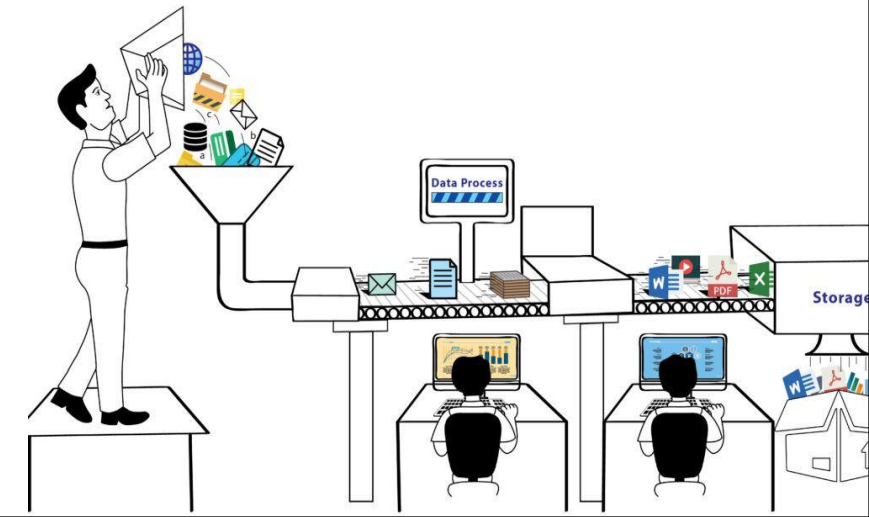
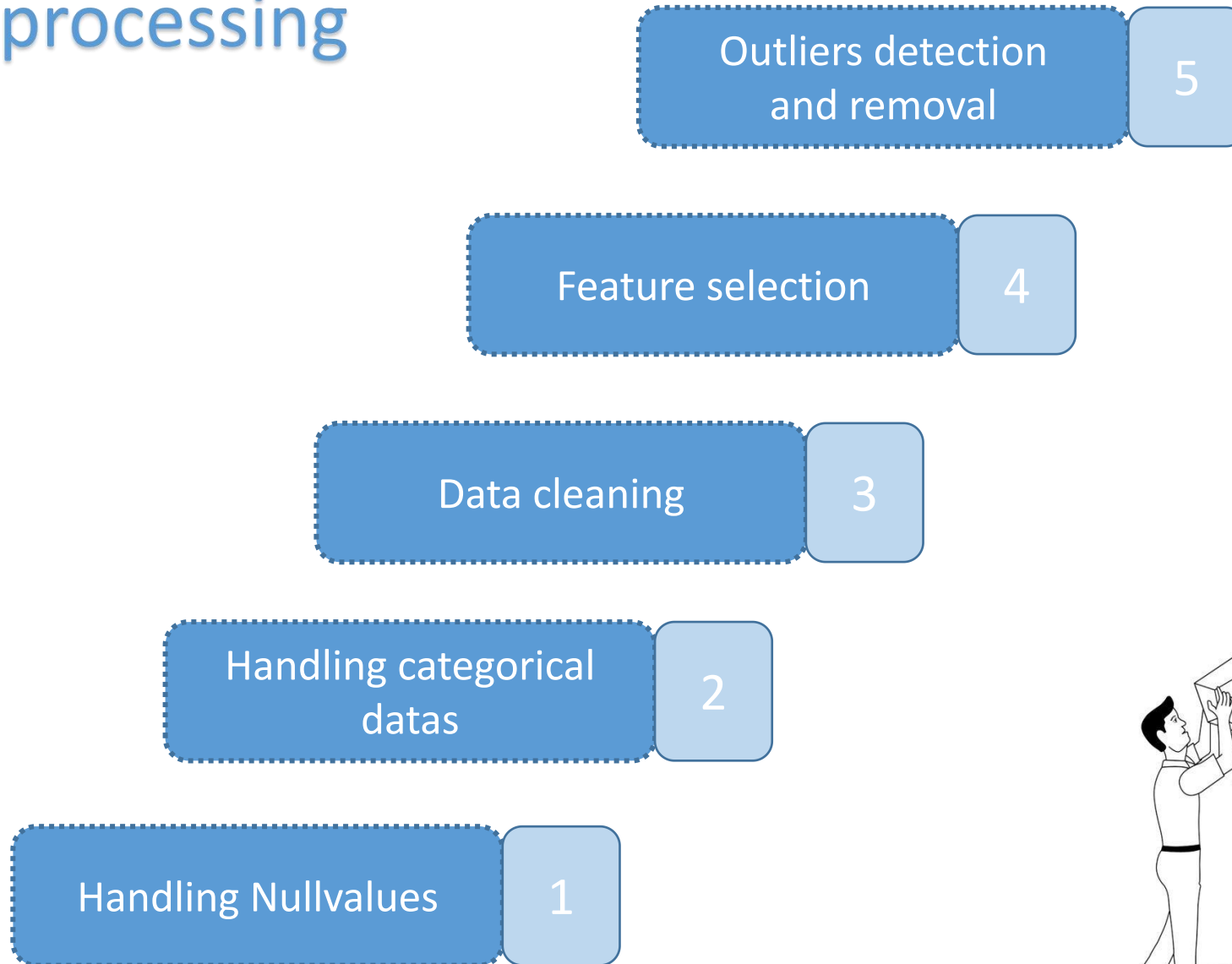
Data visualization

Model selection

Model evaluation

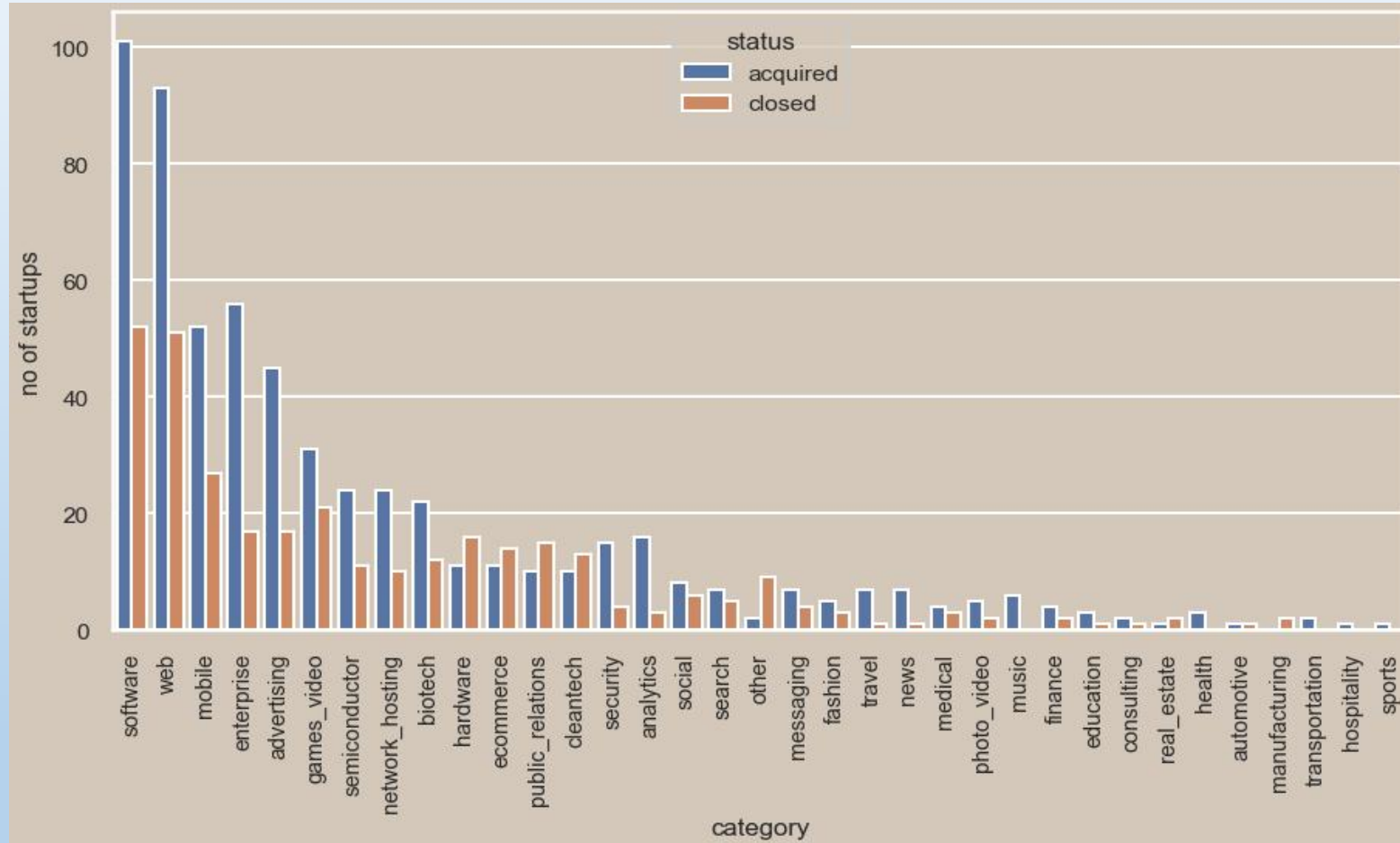
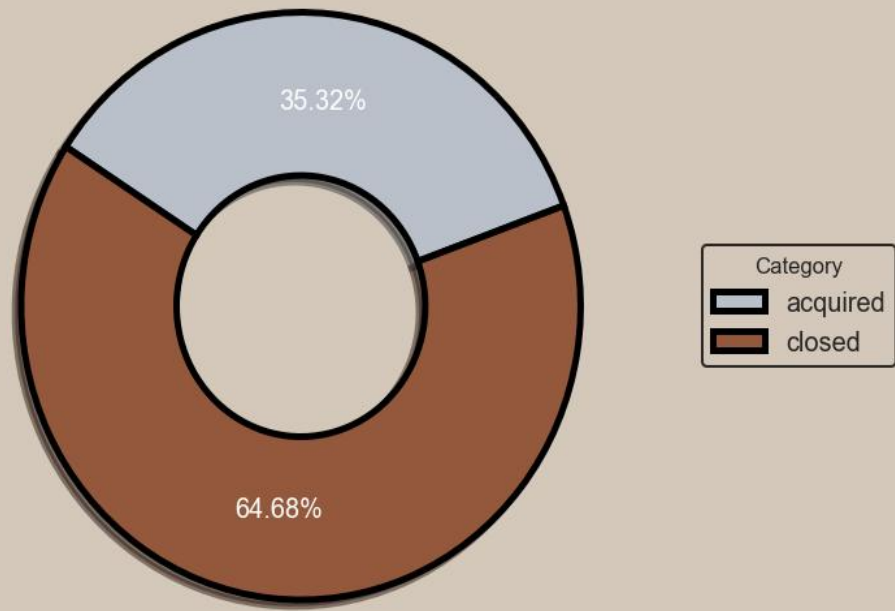
Model deployment

Data preprocessing



Data visualization

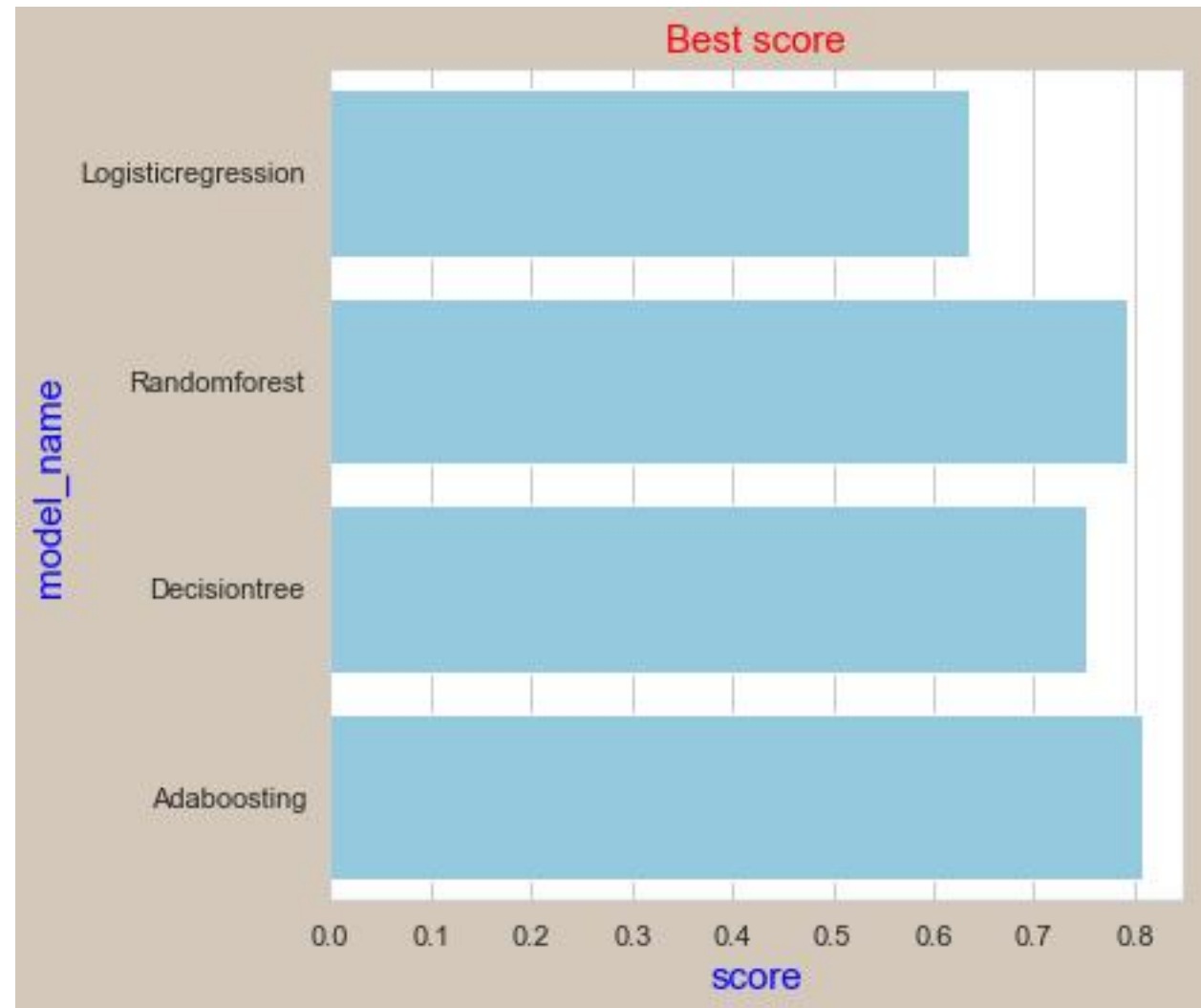
statup status



Model selection&training

- Here iam Using hypertuning for selecting best model
- Here RandomforestRegressor and Adaboost both are giving good scores

Model comparison:



AdaBoosting

- its a supervised machine learning algorithm
- ensemble->boosting
- sequential
- weaklearner combine to build strong learners.
- all the observation gives same weights
- if the prediction is incorrect, give higher weight to observation which have been predicted incorrectly
- the process is continue til a limit is reached in the model performance

Model Deployment

- Here model Deployed using Django Frame work
- Django is a popular python frame work

Conclusion

- Predicting startup success can help investors, entrepreneurs, and decision-makers allocate resources more effectively and make informed strategic decisions
- As the field of machine learning continues to advance, it will undoubtedly bring new opportunities and challenges for predicting startup success, leading to a more dynamic and informed startup ecosystem.
- The limitations -Data quality and availability: The quality and availability of data can significantly impact the accuracy of predictions.
- Limited feature set: Machine learning models rely on features or variables to make predictions. Identifying the most relevant features for startup success prediction can be challenging, and overlooking critical factors can lead to inaccurate predictions.
- Unique and unpredictable factors: Startups operate in dynamic and rapidly changing environments. Factors such as market trends, competition

The background features a collection of blue line-art icons representing various office and school supplies. These include a laptop, a calculator, a smartphone, a pen, a notepad, a coffee cup, a book, and a paper airplane, scattered around the central text.

THANK YOU FOR YOUR ATTENTION