taplingua

TALENT COCO

AI/ML Projects

Predicting heart disease through classification

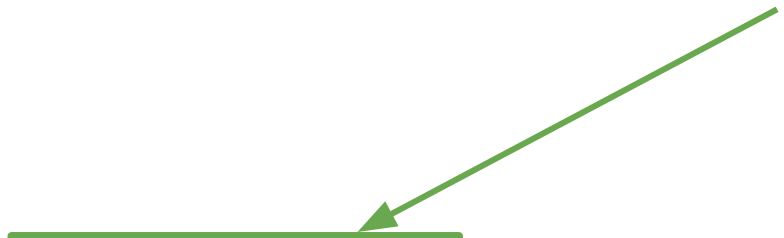
End to End Machine Learning Project

Heart Disease Classification based on clinical data

The client is a major US hospital that has hired us to see if we can help their doctors more easily predict heart disease in patients. They have shared with a large dataset of clinical data of past patients. They want us to see if we can build an ML that can model the dataset and make predictions.

The project is an end to end machine learning project that goes through the entire machine learning workflow including deployment. It will allow you to gain hands on experience of the various tools and techniques in implementing a number of models like Logistics Regression, RandomForestClassifier, KNeighborsClassifier. Full handson.

Let’s introduce the project and what we have been tasked



What’s the project all about?

Why the project and what you will get out of it

How are we going to do the project?

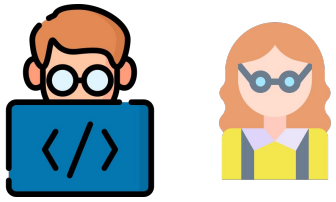
A very brief introduction to AI/ML

Problem definition

Big Hospital wants to help doctors do their job better shared data of past patients (a CSV file). We will deliver a model that predicts given the data

The dataset is a **labeled** and **structured** and has various clinical data about the patient like cholesterol, ECG, heart rate, chest pain etc.

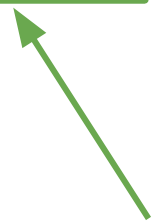
Our task is to build the model End to End and then provide a way for the doctors to access our prediction model to assess new patients.



What is the project about?

How are we going to do the project?

A very brief introduction to AI/ML



Let's discuss how the next 10 days look like

It’s a real world project that will deliver from start to finish.
The goal is that you can put this experience in your resume and take a recruiter through the whole process if asked.

A real world project is critical for

1. Fastest way to get the skills to get hired
2. Recruiters want to talk to you ONLY if have project experience
3. Discuss the project in the interview


CERTIFICATE OF COMPLETION

Full stack Project/Internship

THIS IS TO CERTIFY THAT

Ashwin

has completed the

Airport Parking Search Portal Project/Internship

on 21-03-2023

with grade A

Abinaya Poojary
Trainee
tce.

Santanu Dasgupta
Trainer
taplingua

You will get a shiny **PROJECT CERTIFICATE** for all your hard work.

Certificate does have grades:
Primarily dependent on how well you do the project and participation in the WhatsApp group.
Grade: **Points for asking questions as well as for answering questions. No question is silly or basic. Timely submission!!!**

Project structure

- Fast paced - please ensure you can have extra couple of hours per day
- Take home assignments for Foundation (python, pandas, numpy + Project code)
- WhatsApp group for the course - please participate actively
- 10 days approx - 1 hour class and 2 hours work (approx) for hands on coding

Prerequisites

- Some python or another programming language is helpful
- Please do assignments along with the project and submit
- Share assignment for Foundations at end of class

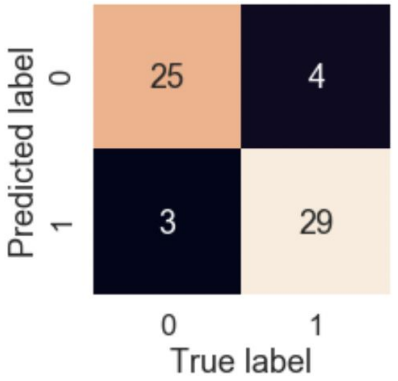
Daily submissions and evaluation in WhatsApp

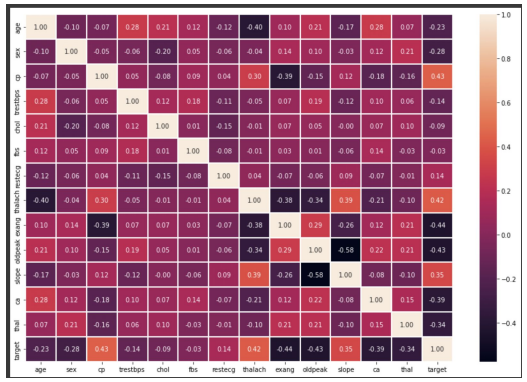
- Colab and Github
  - Work in Google colab -> export to github link -> share in Whatsapp
  - Share 2 files foundation.ipynb and heart-classification.ipynb
  - To get certificate code must be shared and also pass automated tests in system

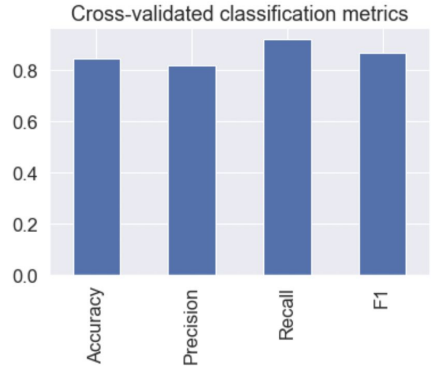
Please say hello in the WhatsApp Group today so we know that you are engaged. Say “ Hello from Shreyas” etc. as an example!!!

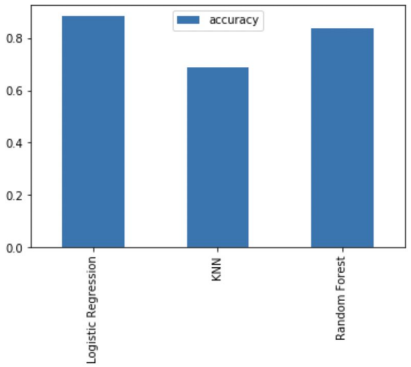
Outcome of the project

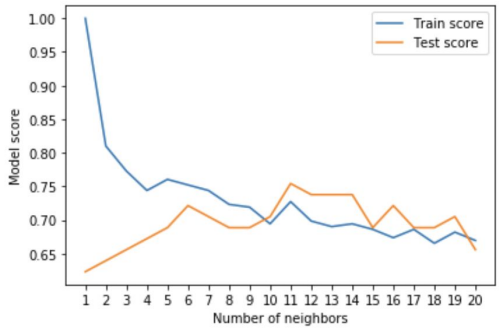
Notebook with a lot of analysis, graphs, and observations which you will share in group.

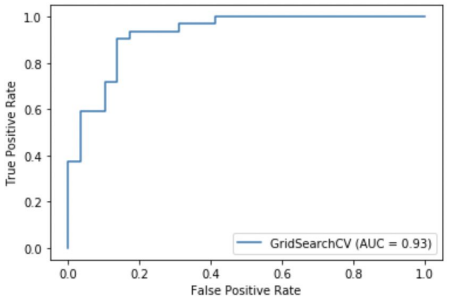












Project outline for Predicting Heart Disease Project

Day 1
Github setup, Problem Definition, and introduction
Brief recap of concepts in workflow (types of data, splitting data)

Day 2
Introduce project and the workflow. Overview of pandas, numpy. Sklearn and how these map to the workflow.
Types of data sources available (web scraping, own data, IoT, kafka streams etc.)

Day 3
Exploring data and finding patterns - (missing values, imbalanced datasets)

Day 4
Exploring data and finding patterns (contd.)

Day 5
Transforming the data and feature engineering (handle missing values, normalization, scaling),

Day 6
Feature engineering (categorical encoding, new features from existing features)

Day 7
Feature selection (dropping features, correlation, univariate selection, forward elimination, backward elimination)

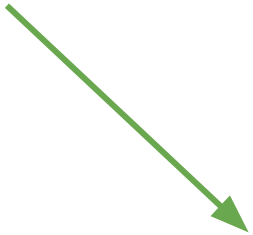
Day 8
Split the data, Model selection (including ensemble techniques). Kfold Cross validation and random seed.

Day 9
More models, Hyper parameter tuning (Randomized Search)

Day 10
Confusion matrix, Classification recall - Precision recall, F1 scores, Evaluating model (ROC, AUC)

Day 11
Pipeliness, deployment and project review. Dockers/Kubernetes

Let’s get a 30,000 feet view of AI/ML and what a real world machine learning project looks like



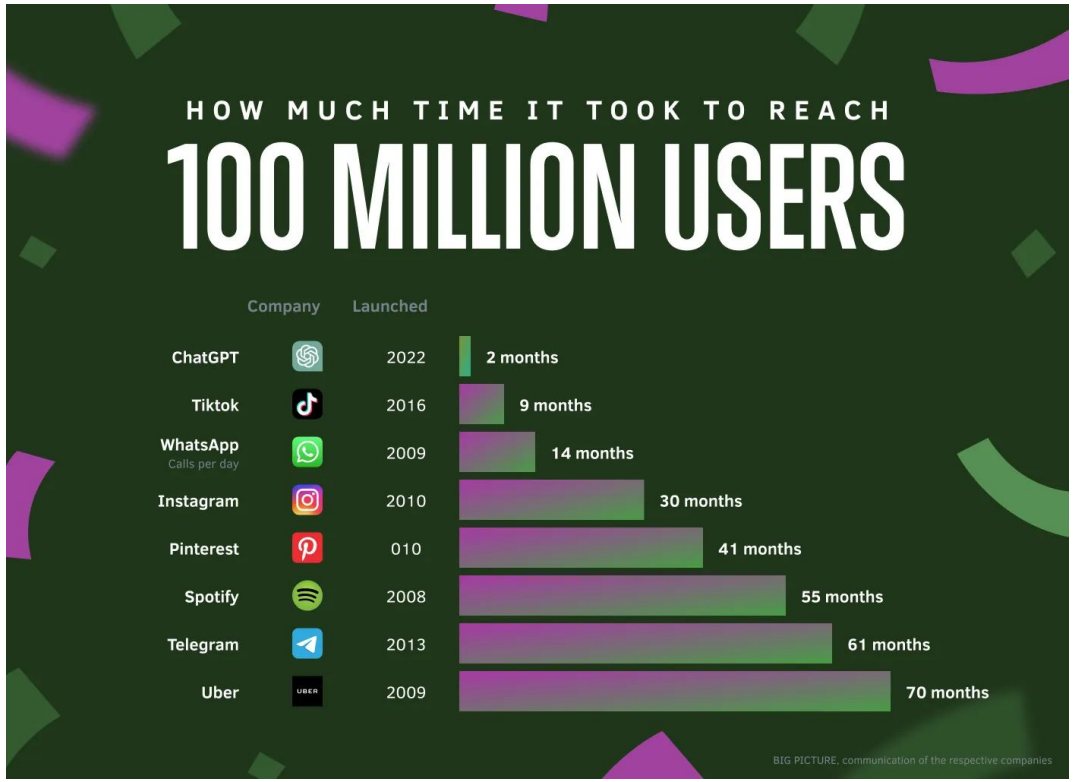
What is the project about?

How are we going to do the project?

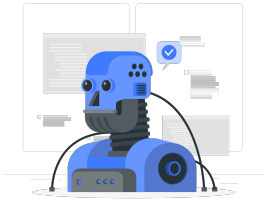
A very brief introduction to AI/ML (ML 101)

What’s all the excitement about the future?

ChatGPT-4



How to get computers to take decisions like humans?



Experience

~~**Experience**~~
Data

Storing and retrieving data

Humans take decisions from experience

Instead of experience, computers make decisions though lots of data

Computers are good at storing and retrieving data

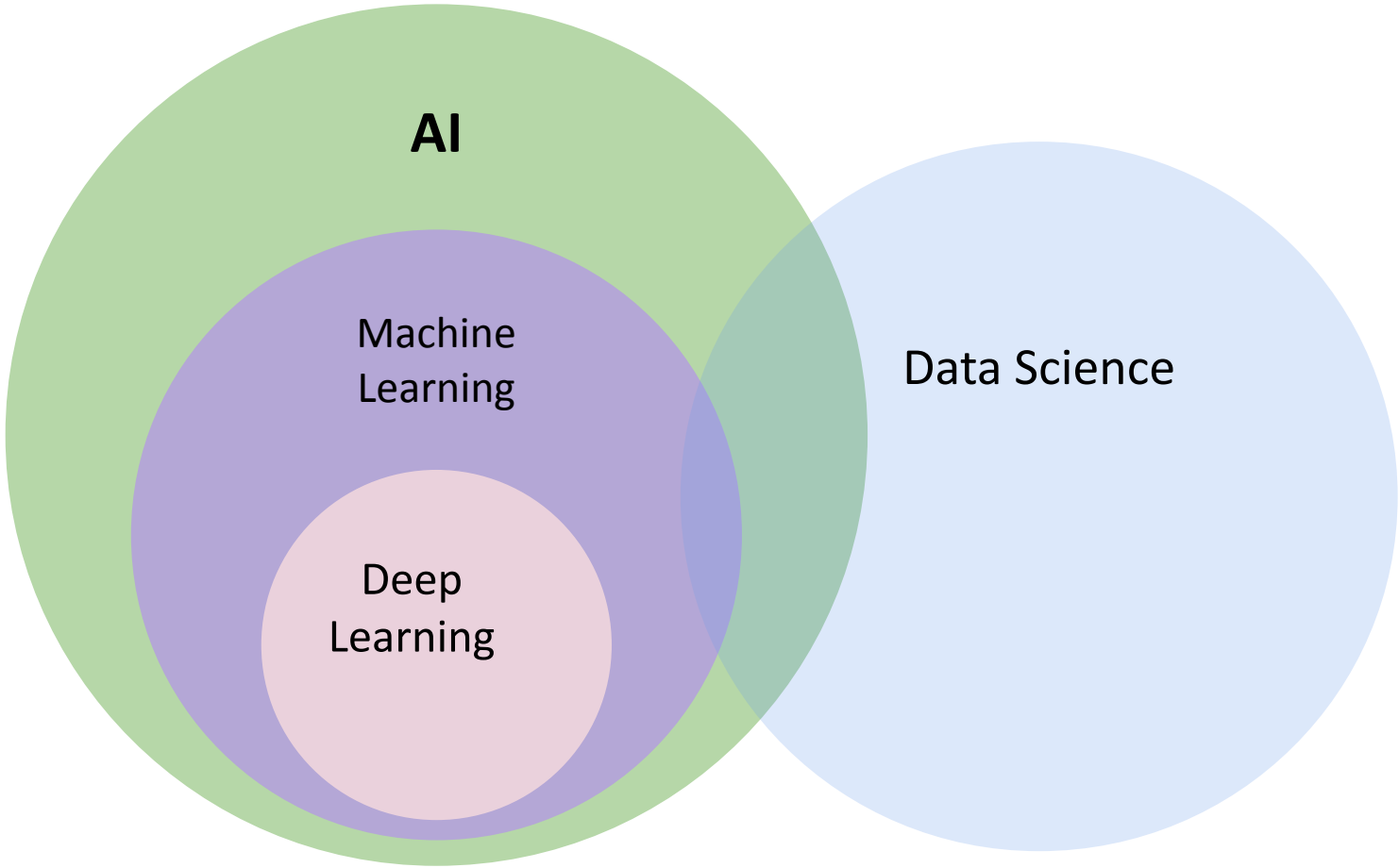
What it aims to do…

Find patterns in the data

Summarize the patterns in precise ways

... using lots of data and without manual intervention

AI/ML and Data Science



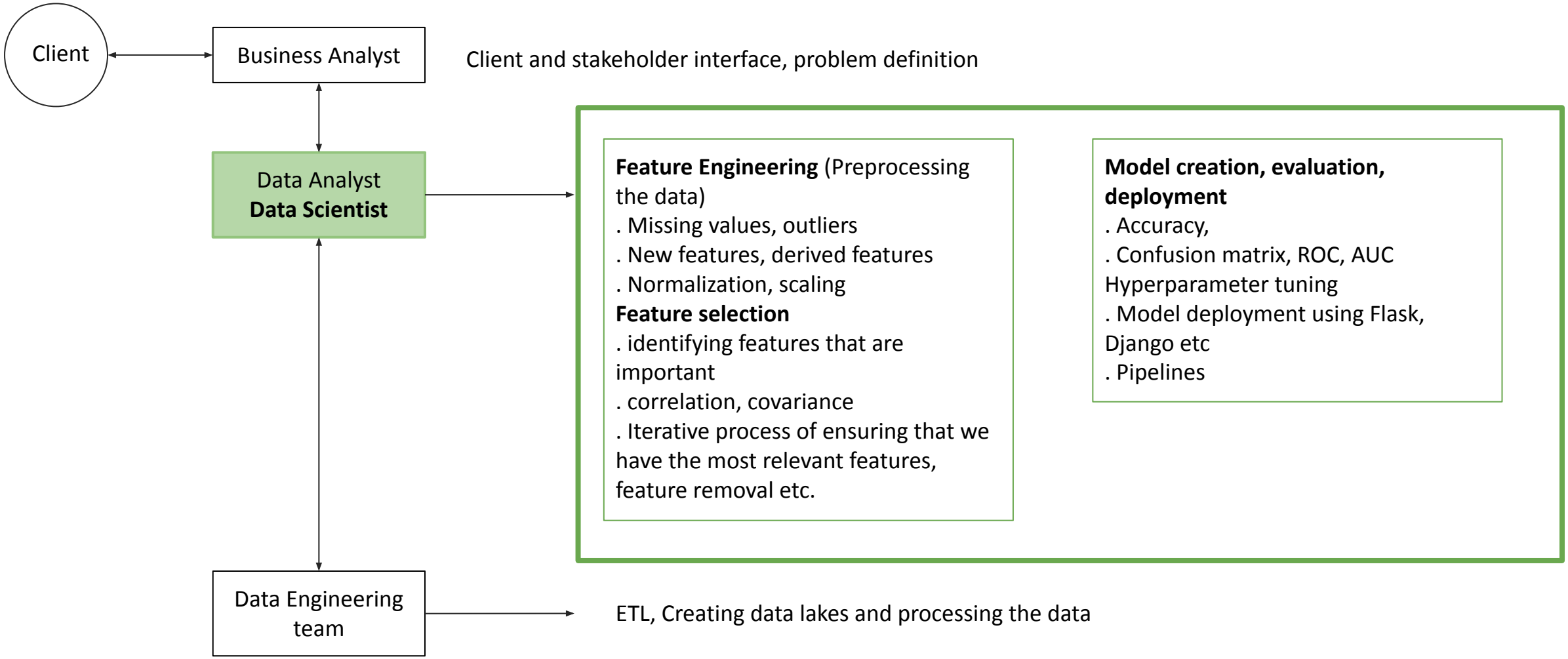
Machine Learning

Using past data to make predictions

Normal algo gets **instructions** and what to do and then does it

Machine learning algo gets the **data** and figures out what to do

Roles, responsibilities and tasks in for Data Science projects for Data Analysts (fresher roles)



End to End project for Machine Learning

It’s not just about the model, process and steps are as important

End to End project flow

Collect Data

Data Analysis (Feature engineering and selection) → Machine Learning

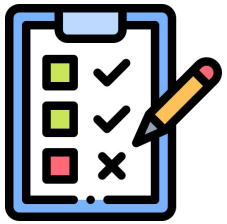
1. Define Problem

What is the problem statement

2. Get Data

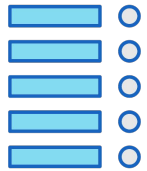
Get the data

3. Evaluate



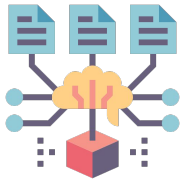
Define success

4. Features



Pick the features

5. Model



Pick models

6. Experiment

Try different models and approaches

Iterate

Deploy model

Libraries and Environments we will use

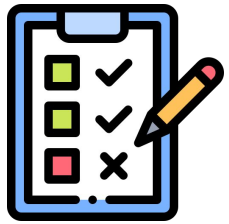
1. Define Problem

Problem statement?

2. Get Data

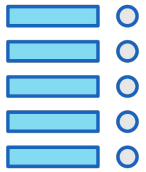
Get the data

3. Evaluate



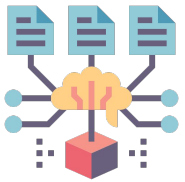
Define success

4. Features



Pick the features

4. Model



Pick models

5. Experiment

Try different models approaches

Pandas NumPy matplotlib

Data Analysis & Manipulation

scikit learn TensorFlow PyTorch dmlc XGBoost

Machine Learning/ Deep learning

Environment

jupyter colab CONDA

Assignment

1. Connect Github and Colab
2. Create your own python.ipynb
3. Follow the tasks in the file and run the code in the python.ipynb.
4. Do it in your own file and once done share your code by tomorrow
5. You will have until end of the week to submit the python.ipynb, pandas.ipynb, numpy.ipynb

Colab link for Python Notebok

https://colab.research.google.com/drive/1b1cEV7UMNoOF1c4TZllstltDWLbRI4Va#scrollTo=Zxjk0itOGGi_

Appendix
Colab and Github Setup

github.com

Sign in / Sign up

https://colab.research.google.com/

Instructions on how to push to code to github

1. Sign in to colab.research.google.com

2. Download your colab file as a notebook to your local machine
File-> Download (Choose .ipynb)

3. Login to Github and create a new repo (see Fig 3)

Download .ipynb notebook and push to github using command line.

4 Download git from https://gitforwindows.org/

github.com

5 Follow instructions from git website when you create the repo on how to push your 1st commit from your desktop

```
echo "# dummy" >> README.md
git init
git add README.md
git commit -m "first commit"
git branch -M main
git remote add origin
https://github.com/talentcoco/dummy
.git
git push -u origin main
```

Fig 3. Create a repository

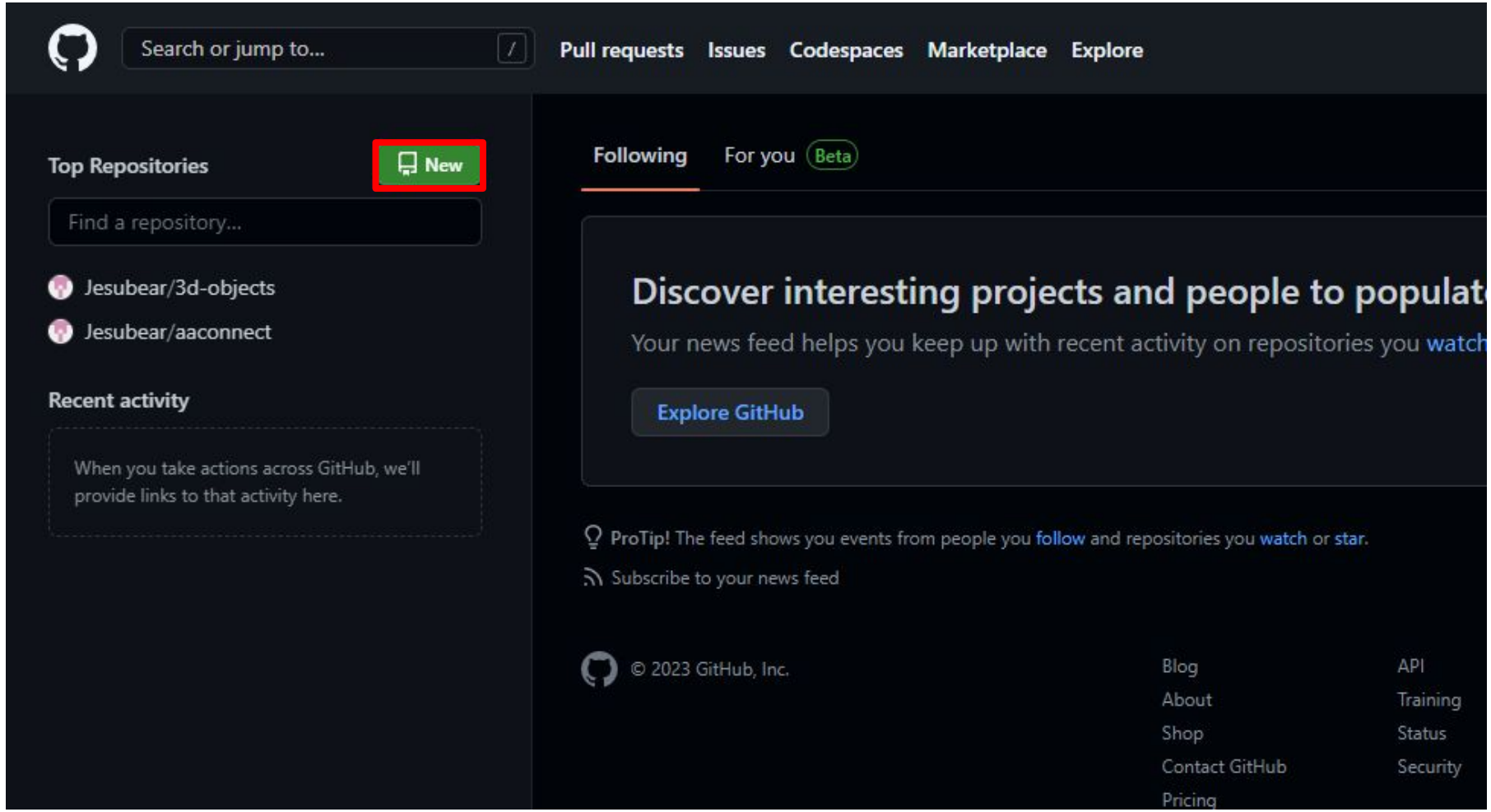
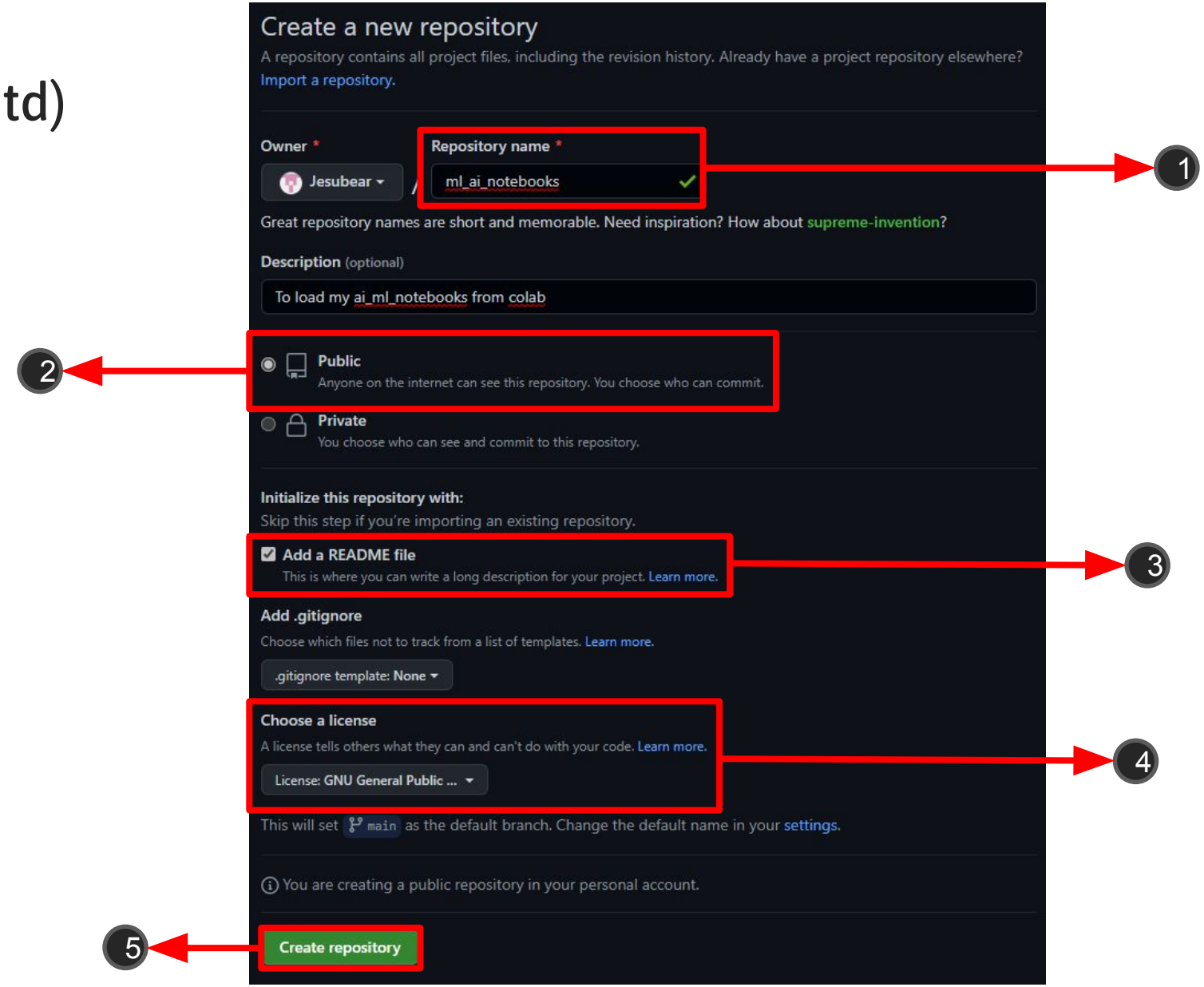
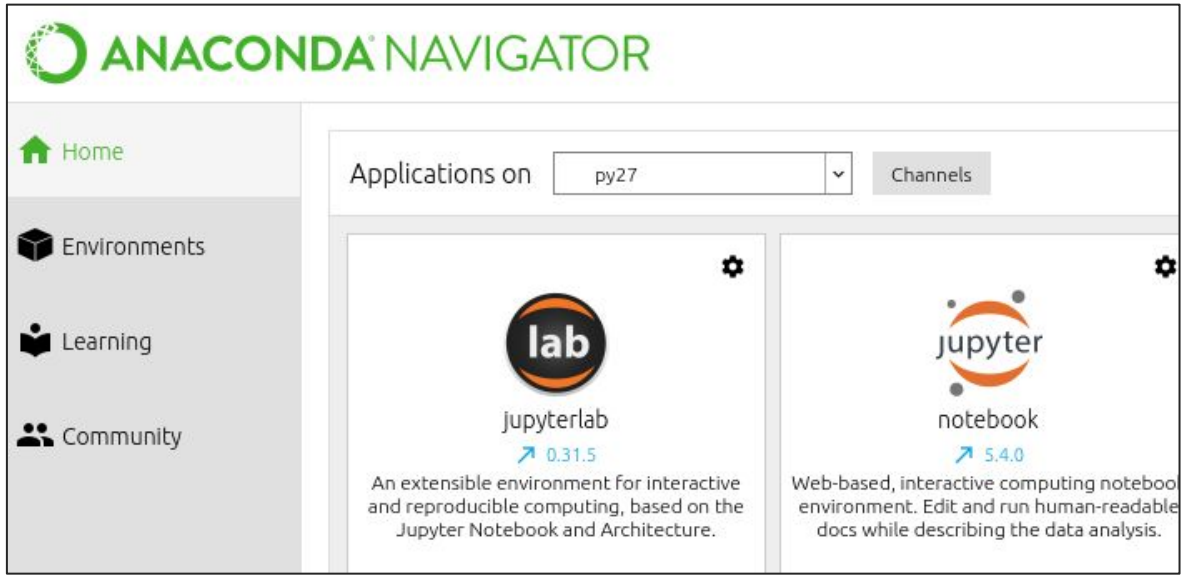


Fig 3. Create a repository (contd)



Alternative tools



jupyterlab TNG

WinPython

Amazon SageMaker