Your first machine learning project

Sayak Paul | Deep Learning Associate at PyImageSearch

January 04, 2020

Mehsana, Gujarat, India

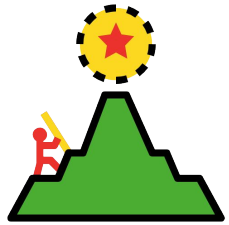
Agenda

- Your first ML project
  - Thinking about a problem statement
  - Things to consider for the project
  - Executing the project
  - Presenting the project and having feedback
  - ...
- Next steps

Motivation

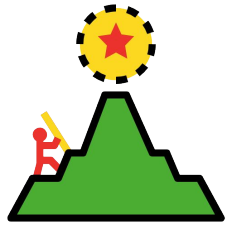
Source

Your first ML project - Challenges



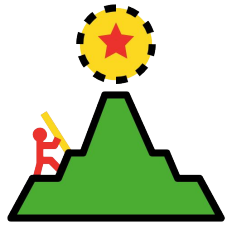
- How much should I know before I start my ML project?

Your first ML project - Challenges



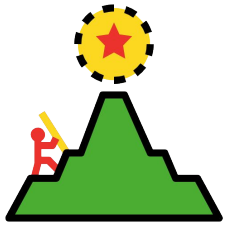
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?

Your first ML project - Challenges



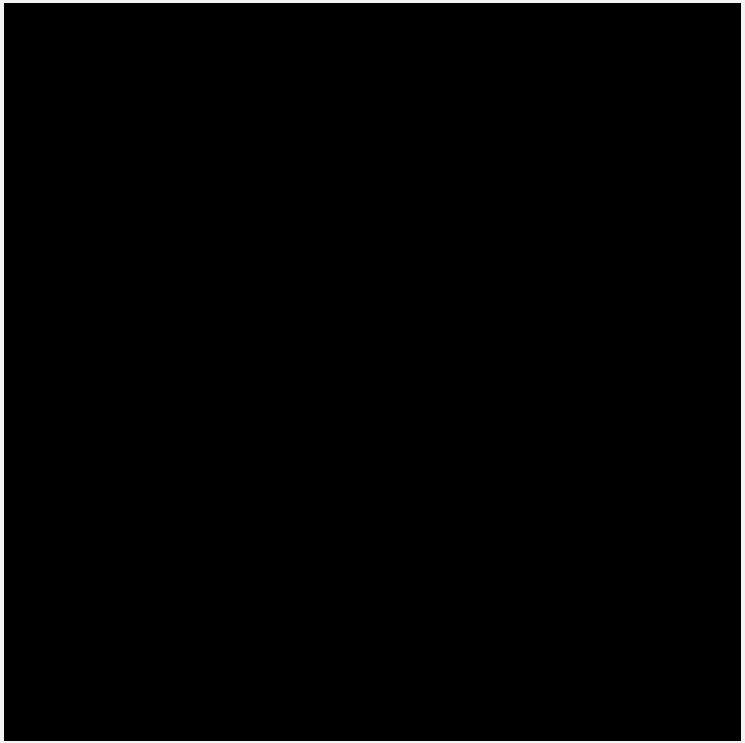
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?

Your first ML project - Challenges

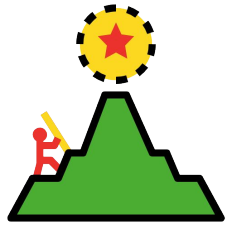


- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
- Lots of moving parts in the project. Where do I start?

Let’s talk about solutions now!

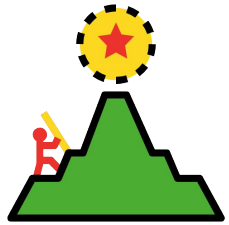


Your first ML project - Solutions



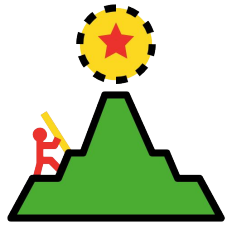
- How much should I know before I start my ML project?
  - You ***don't need to know everything*** before you start the project. Just basic concepts should be enough to start off.

Your first ML project - Solutions



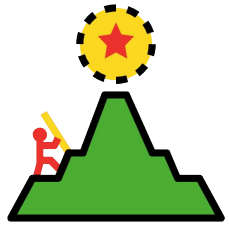
- How much should I know before I start my ML project?
  - You ***don't need to know everything*** before you start the project. Just basic concepts should be enough to start off.
  - I typically follow a **60:40** ratio (things I know : learning opportunities). Make your own ratios and adjust them over time.

Your first ML project - Solutions



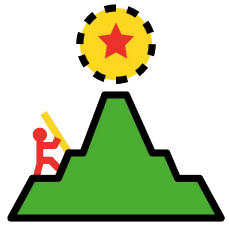
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
  - You cannot be a master of everything at any given time.

Your first ML project - Solutions



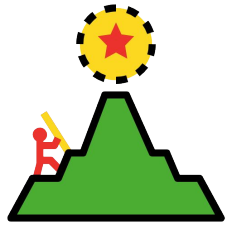
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
  - You cannot be a master of everything at any given time.
  - Focus on the basics that are still missing. Focus on the things relevant for executing the project.

Your first ML project - Solutions



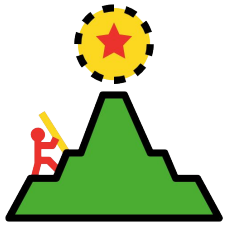
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
  - You cannot be a master of everything at any given time.
  - Focus on the basics that are still missing. Focus on the things relevant for executing the project.
  - Don’t rush it out!

Your first ML project - Solutions



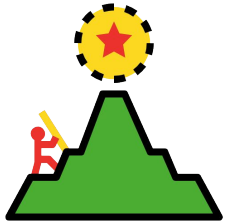
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
  - You cannot be a master of everything at any given time.
  - Focus on the basics that are still missing. Focus on the things relevant for executing the project.
  - Don’t rush it out!
  - **So, how much understanding is needed?**

Your first ML project - Solutions

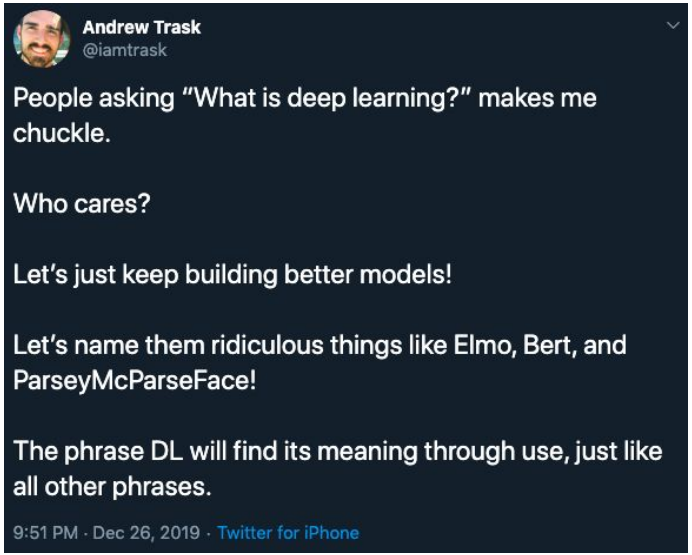


- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
  - [...]
  - Don't rush it out!
  - **So, how much understanding is needed?**
    - **Just enough to convince yourself!** Understanding can always be iterated!

Your first ML project - Solutions



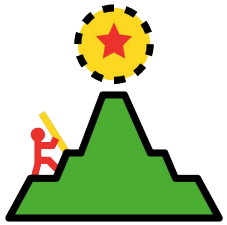
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?



Source

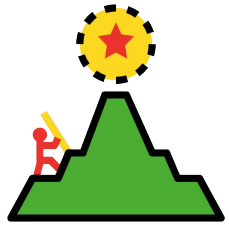
sayak.dev

Your first ML project - Solutions



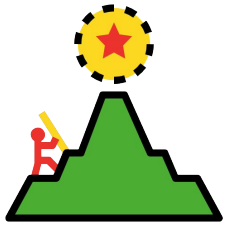
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
    - Try identifying the problems in your surroundings - places you visit, platforms you use and so on.

Your first ML project - Solutions



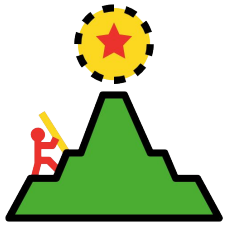
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
    - Try identifying the problems in your surroundings - places you visit, platforms you use and so on.
    - Decide if the problem is machine learning friendly (problem framing).

Your first ML project - Solutions



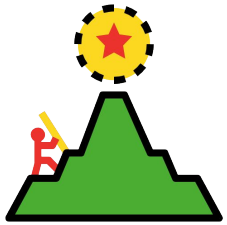
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
  - Try identifying the problems in your surroundings - places you visit, platforms you use and so on.
  - Decide if the problem is machine learning friendly (problem framing).
  - Participate in **Kaggle competitions**, but ...

Your first ML project - Solutions



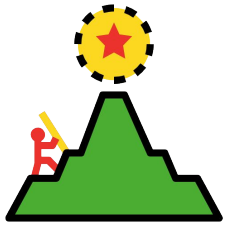
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
  - [...]
  - Participate in **Kaggle competitions**, but ...
    - Don't chase the leaderboard!

Your first ML project - Solutions



- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
  - [...]
  - Participate in **Kaggle competitions**, but ...
    - Don't chase the leaderboard!
    - Always think of the bigger picture.

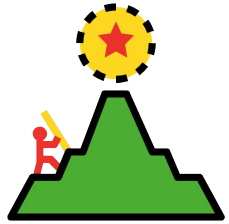
Your first ML project - Solutions



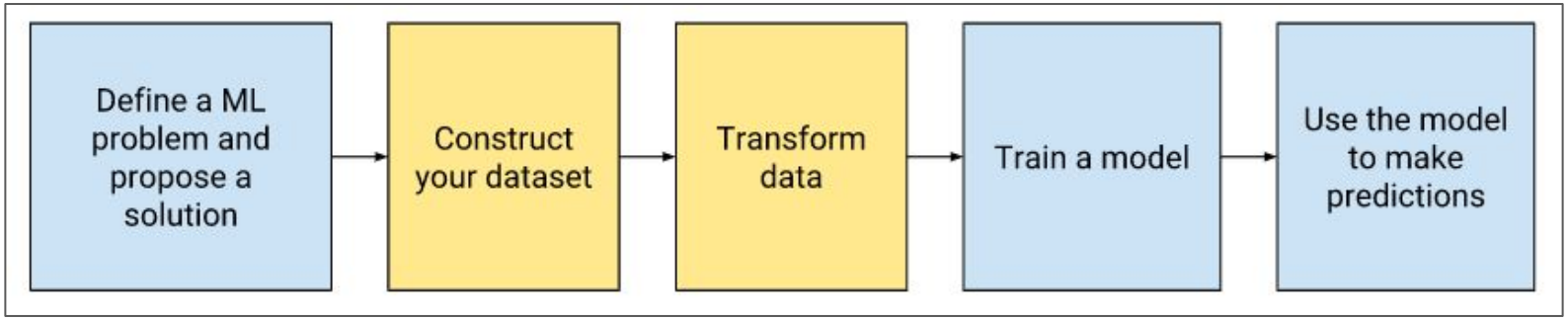
- How much should I know before I start my ML project?
- ML is **interdisciplinary**. How to not get overwhelmed?
- What problem statement do I choose for the project?
  - [...]
  - Participate in **Kaggle competitions**, but ...
    - [...]
    - Here's an amazing resource to learn creative ML: https://mlwave.com/

Executing an ML project: Some thoughts

Some things to consider

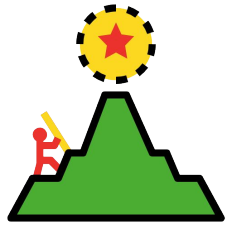


- I find it consistent to start with a **sequential** flow:



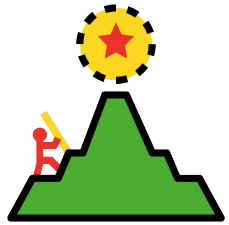
Source

Some things to consider



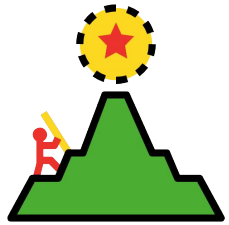
- I find it consistent to start with a **sequential** flow.
- **Data!**

Some things to consider



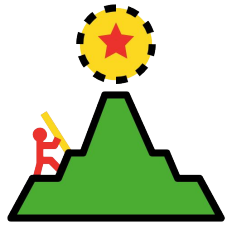
- I find it consistent to start with a **sequential** flow.
- **Data!**
  - How do I collect data?

Some things to consider



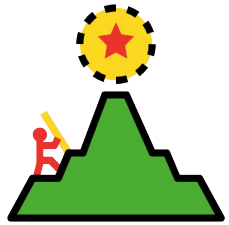
- I find it consistent to start with a **sequential** flow.
- **Data!**
  - How do I collect data? (Kaggle and other online resources, web scraping, manually collect data)

Some things to consider



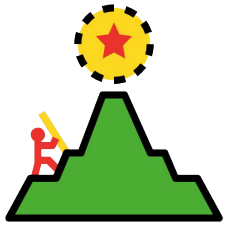
- I find it consistent to start with a **sequential** flow.
- **Data!**
  - How do I collect data?
  - Is there a similar dataset available?

Some things to consider

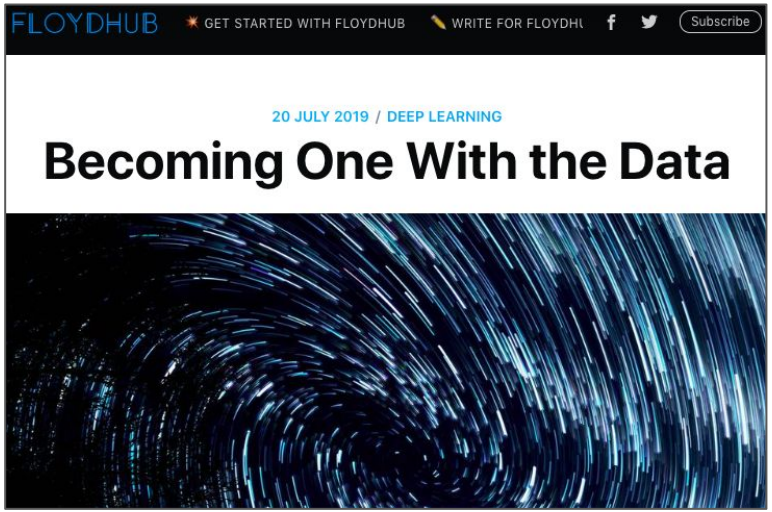


- I find it consistent to start with a **sequential** flow.
- **Data!**
  - How do I collect data?
  - Is there a similar dataset available?
  - How do I become one with the data?

Some things to consider

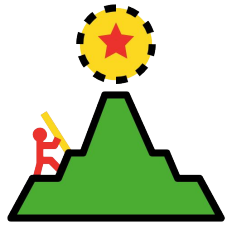


- I find it consistent to start with a **sequential** flow.
- **Data!**
  - [...]



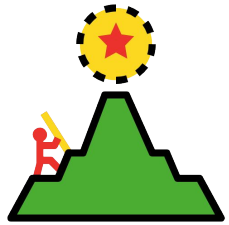
Article link

Some things to consider



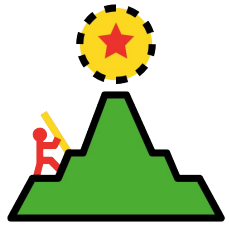
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**

Some things to consider



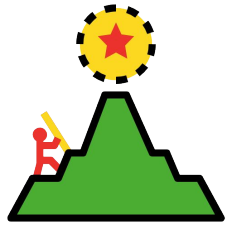
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
    - Which model should I consider?

Some things to consider



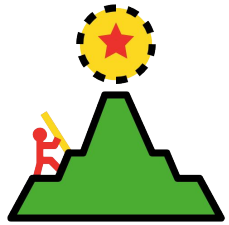
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
  - Which model should I consider?
  - How do I train a model?

Some things to consider



- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
  - Which model should I consider?
  - How do I train a model?
    - How can I train my model faster?

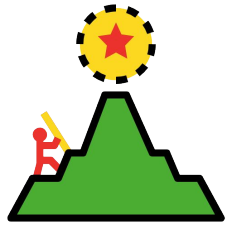
Some things to consider



- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
  - Which model should I consider?
  - How do I train a model?
    - How can I train my model faster?
    - How can I train it better?

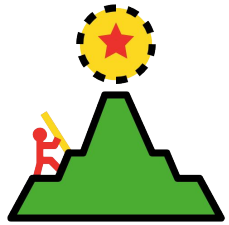
sayak.dev

Some things to consider



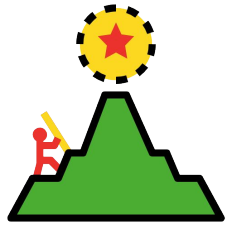
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
  - Which model should I consider?
  - How do I train a model?
  - How do I validate a model?

Some things to consider



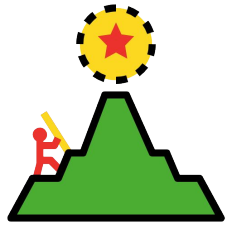
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
  - Which model should I consider?
  - How do I train a model?
  - How do I validate a model?
  - How do I debug a model?

Some things to consider



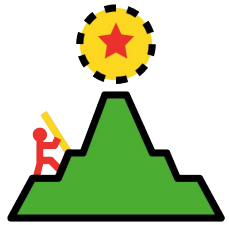
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
  - [...]
  - How do I debug a model?
    - Check out this course: https://developers.google.com/machine-learning/testing-debugging

Some things to consider



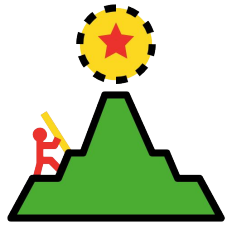
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
- Does my ML model integrate well with other systems?

Some things to consider



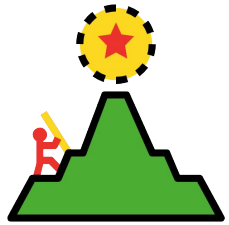
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
- Does my ML model integrate well with other systems?
  - A web application

Some things to consider



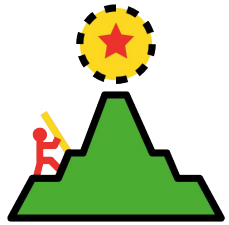
- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
- Does my ML model integrate well with other systems?
  - A web application
  - A mobile application

Some things to consider



- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
- Does my ML model integrate well with other systems?
  - A web application
  - A mobile application
  - A Raspberry Pi

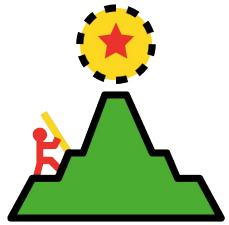
Some things to consider



- I find it consistent to start with a **sequential** flow.
- **Data!**
- **Modeling!**
- Does my ML model integrate well with other systems?

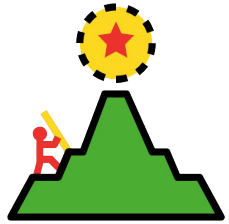
Possibilities are endless here!

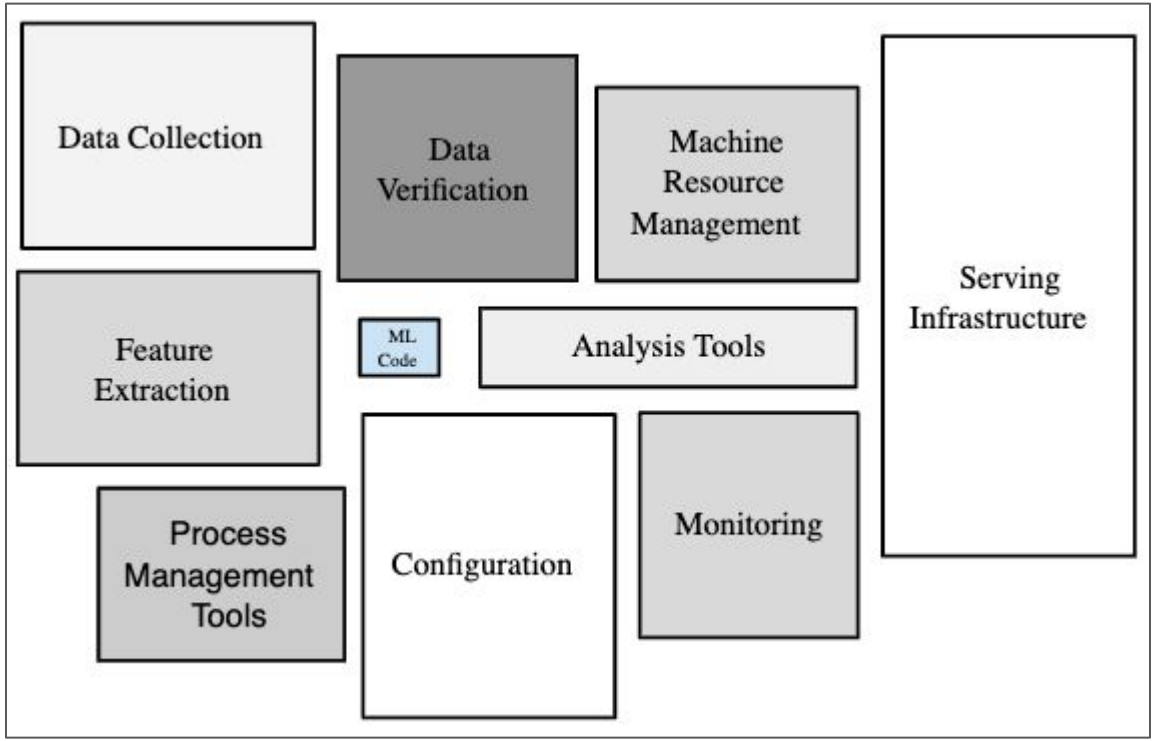
Some things to consider



- [...]
- Does my ML model integrate well with other systems?
- Think about the other components of your project too, ***not just ML***!

Some things to consider



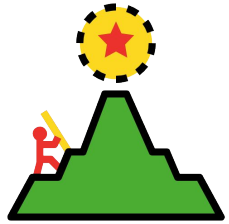


Source

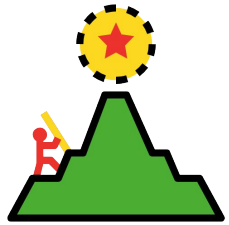
Presenting your project

Project out in the wild

- Nothing like a structured GitHub repository

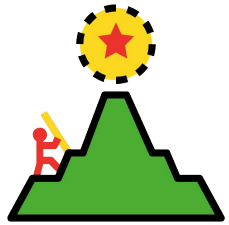


Project out in the wild



- Nothing like a structured GitHub repository
  - Be sure to add a proper **README** including a demo of your project

Project out in the wild



- Nothing like a structured GitHub repository
  - Be sure to add a proper **README** including a demo of your project
  - A polished directory structure

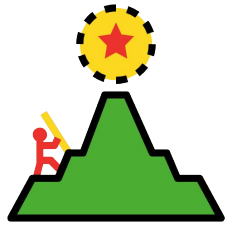
Project out in the wild

```
apparel_classifier/             # Package that can be deployed as a self-contained prediction system
  __init__.py

  apparel_predictor.py      # Takes a raw image and obtains a prediction

  datasets/                   # Code for loading datasets
      __init__.py
      dataset.py              # Base class for datasets - logic for downloading data (if not available)
      fmnist_dataset.py
      fmnist_essentials.json
      dataset_sequence.py

  models/                     # Code for instantiating models, including data preprocessing and loss functions
      __init__.py
      base.py                 # Base class for models
      image_model.py

  networks/          # Code for building neural networks (i.e., 'dumb' input->output mappings) used by models
      __init__.py
      mlp.py

  tests/
      support/                        # Raw data used by tests
      test_apparel_predictor.py       # Test model on a few key examples

  weights/                            # Weights for production model
      Image_Model_FMNIST_Dataset_weights.h5

  util.py

training/                       # Code for running training experiments and selecting the best model.
    run_experiment.py           # Parse experiment config and launch training.
    util.py                     # Logic for training a model with a given config
```

Click to enlarge.

sayak.dev

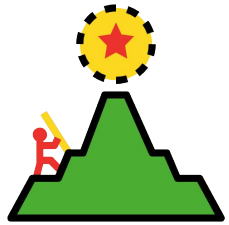
Project out in the wild



- Nothing like a structured GitHub repository
- Write out a blog on the project and be as detailed as you can be!

sayak.dev

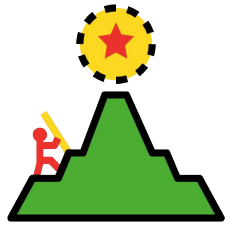
Project out in the wild



- Nothing like a structured GitHub repository
- Write out a blog on the project and be as detailed as you can be!
  - FloydHub AI Writer Program
  - Weights and Biases Content Developers
  - Nanonets Writers

Paid writing opportunities.

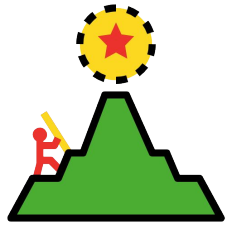
Project out in the wild



- Nothing like a structured GitHub repository
- Write out a blog on the project and be as detailed as you can be!
- Share your project with the communities

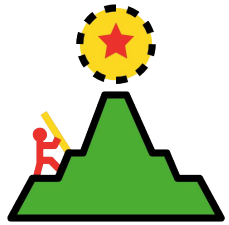
sayak.dev

Project out in the wild



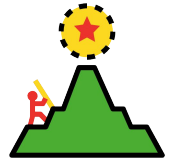
- Nothing like a structured GitHub repository
- Write out a blog on the project and be as detailed as you can be!
- Share your project with the communities
  - Kaggle
  - AIDL Facebook Group
  - Twitter
  - FastAI Forums

Project out in the wild



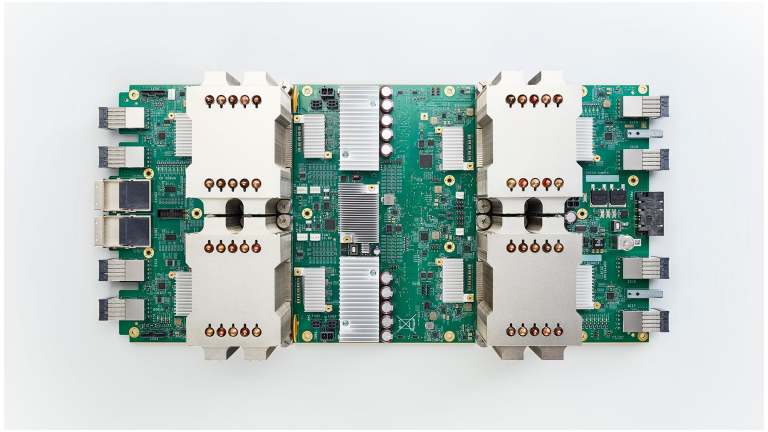
- Nothing like a structured GitHub repository
- Write out a blog on the project and be as detailed as you can be!
- Share your project with the communities
- Be open to constructive feedback.

Departing thoughts



- Figure out what interests you. Machine learning is a huuuuuge field!
  - Some good directions here: https://www.sayak.dev/interviews
- Discuss your work with like-minded people.
- Finally, apply, learn, make mistakes and repeat!

Free Cloud TPUs to support your ML research



7 on-demand + 20 preemptible Cloud TPU devices per person

Free access available for several months

Visit g.co/tputalk and enter event code DL20 to get started!

If you are an entrepreneur

Source

Just to finish off in style

Slides available here: http://bit.ly/dloop20

See you next time

@RisingSayak

Thank you very much :)

Experts

pyimagesearch