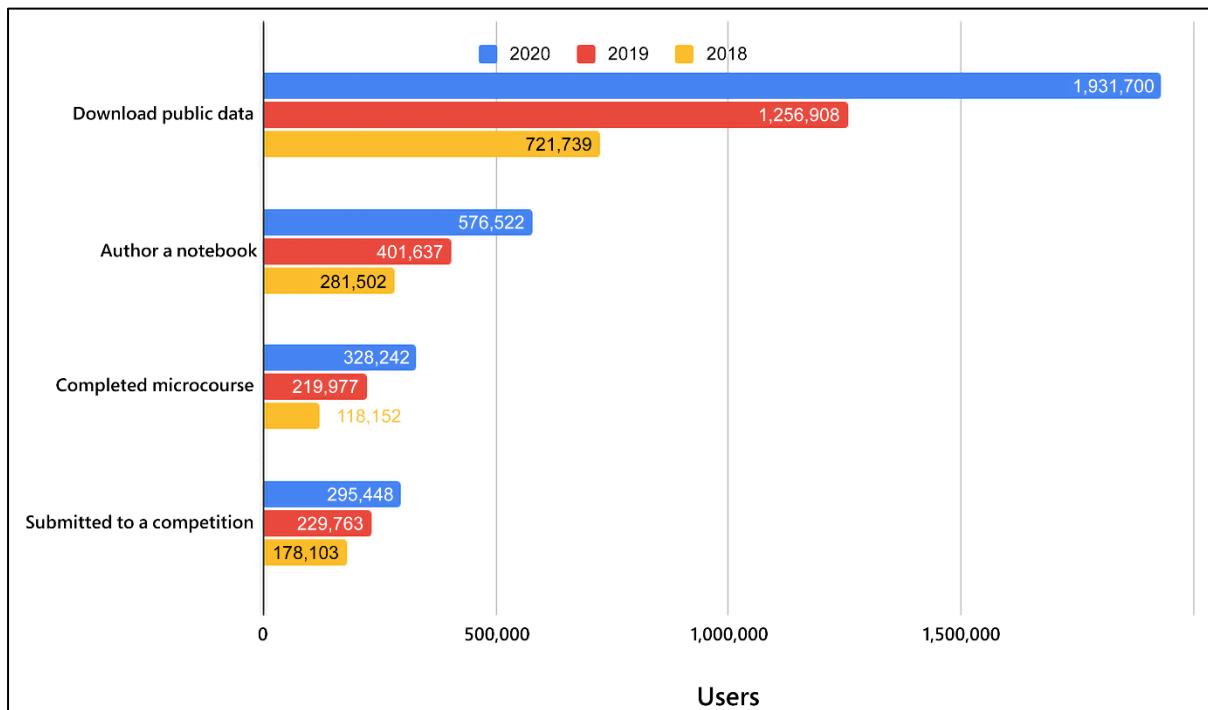


Chapter 1: Introducing Kaggle and Other Data Science Competitions



The screenshot shows the H&M Personalized Fashion Recommendations competition page on Kaggle. The top banner features the competition title, a \$50,000 prize, and a call-to-action button. Below the banner, there's an overview section with tabs for Overview, Data, Code, Discussion, Leaderboard, and Rules. The Description tab is active, providing details about the H&M Group and its products. The Evaluation, Timeline, and Prizes sections are also present. The main content area contains a detailed description of the competition's purpose and the data available for recommendation models.

H&M Personalized Fashion Recommendations
Provide product recommendations based on previous purchases

\$50,000
Prize Money

H&M Group · 1,283 teams · 2 months to go (a month to go until merger deadline)

Overview Data Code Discussion Leaderboard Rules Join Competition ...

Description
H&M Group is a family of brands and businesses with 53 online markets and approximately 4,850 stores. Our online store offers shoppers an extensive selection of products to browse through. But with too many choices, customers might not quickly find what interests them or what they are looking for, and ultimately, they might not make a purchase. To enhance the shopping experience, product recommendations are key. More importantly, helping customers make the right choices also has a positive implications for sustainability, as it reduces returns, and thereby minimizes emissions from transportation.

In this competition, H&M Group invites you to develop product recommendations based on data from previous transactions, as well as from customer and product meta data. The available meta data spans from simple data, such as garment type and customer age, to text data from product descriptions, to image data from garment images.

There are no preconceptions on what information that may be useful – that is for you to find out. If you want to investigate a categorical data type algorithm, or dive into NLP and image processing deep learning, that is up to you.

A screenshot of a Jupyter Notebook interface. The left pane shows a code cell with Python code for data processing, including imports for numpy and pandas, and a loop to list filenames in a directory. The right pane displays various notebook settings: Language (Python), Environment (Preferences), Accelerator (None), and Internet (checkbox checked). A search bar at the bottom right says "Search for examples of how to do things".

```

[1]:
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

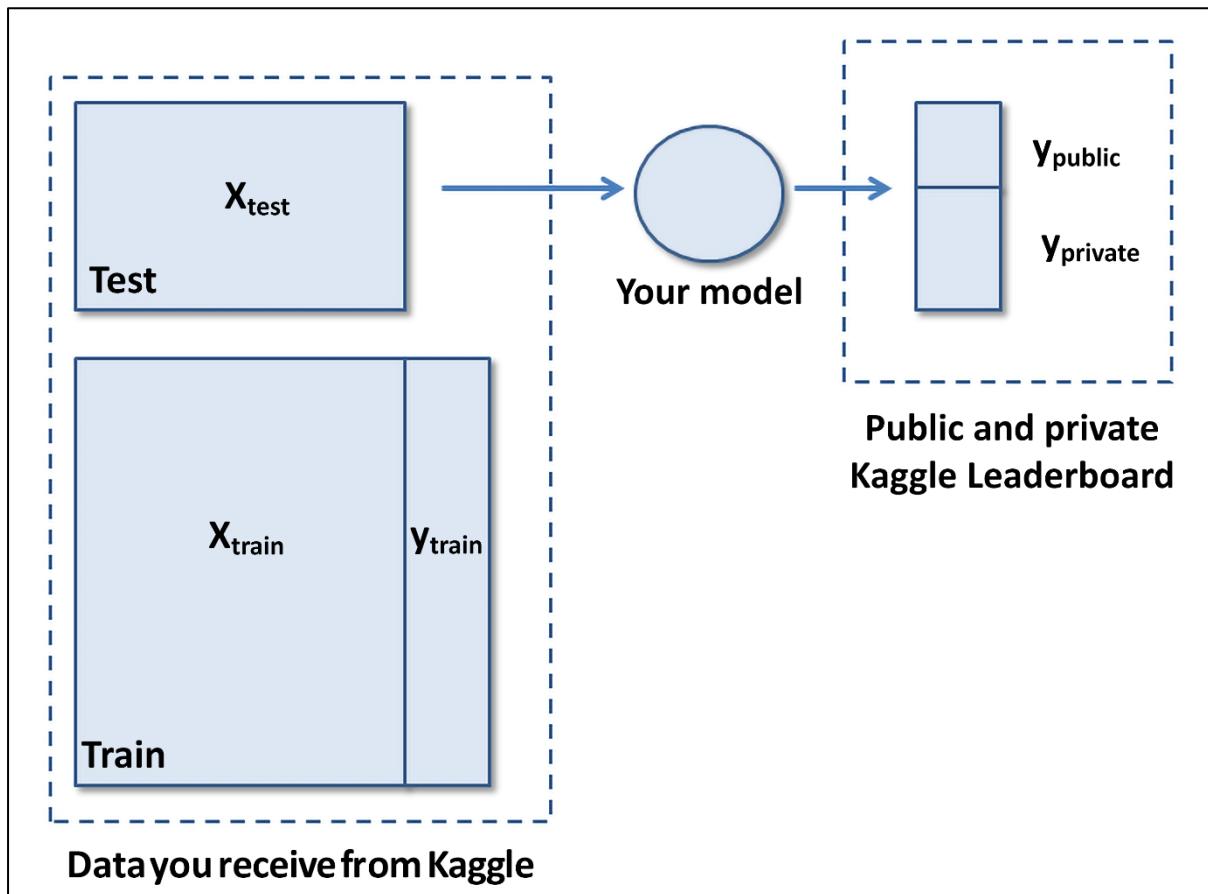
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking run or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 2GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version using "Save & Run"
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session

```



Chapter 2: Organizing Data with Datasets

Trending Datasets

See All



Python Course Discussion JSON
Bruno Ferreira · Updated 2 hours ago
Usability 10.0 · 80 KB
1 File (JSON)

0



Largest Cities in the World
Rishi Damarla · Updated 3 hours ago
Usability 9.4 · 12 MB
1 File (other)

1



Deforestation and Forest Loss
Chiticariu Cristian · Updated 7 hours ago
Usability 7.1 · 4 KB
1 File (CSV)

2

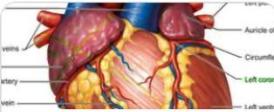


Bitcoin Price
ELOO · Updated 9 hours ago
Usability 10.0 · 271 KB
1 Task · 1 File (CSV)

7

Popular Datasets

See All



Heart Attack Analysis & Prediction Dataset
Rashik Rahman · Updated a month ago
Usability 10.0 · 4 KB
1 Task · 2 Files (CSV)



World Happiness Report 2021
Ajaypal Singh · Updated a month ago
Usability 9.7 · 55 KB
1 Task · 2 Files (CSV)



Stroke Prediction Dataset
fedesoriano · Updated 3 months ago
Usability 10.0 · 67 KB
1 Task · 1 File (CSV)



Reddit Vaccine Myths
Gabriel Preda · Updated a day ago
Usability 10.0 · 227 KB
2 Tasks · 1 File (CSV)

Datasets

Explore, analyze, and share quality data. [Learn more](#) about data types, creating, and collaborating.

+ New Dataset Your Work

 Search datasets

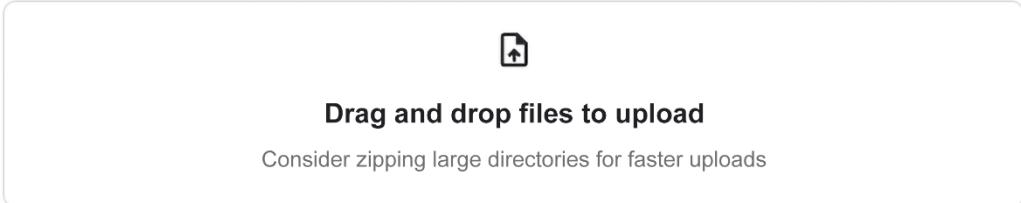
 Filters

[Datasets](#) [Tasks](#) [Computer Science](#) [Education](#) [Classification](#) [Computer Vision](#) [NLP](#) [Data Visualization](#)



 Enter Dataset Title 

www.kaggle.com/konradb/

 
 **Drag and drop files to upload**
Consider zipping large directories for faster uploads

 or

 **Browse Files**

  **Private**  **Create**

 **efnet_pytorch** 

www.kaggle.com/konradb/efnet-pytorch

 **Import GitHub repository**

Create a dataset from a GitHub repository archive. Use the repo URL or any deep link.

 GitHub URL  <https://github.com/lukemelas/EfficientNet-PyTorch>

 **Preview**

-  .github
-  workflows
-  main.yml (463 B)
-  .gitignore (1.44 KB)
-  LICENSE (11.09 KB)

 **Private**  **Create**

Easy to understand and includes essential metadata	Rich, machine readable file formats and metadata	Assurances the dataset is maintained
✓ Add a subtitle	○ Add file information Help others navigate your dataset with a description of each file	✓ Specify provenance
✓ Add tags	○ Include column descriptors Empowers others to understand your data by describing its features	✓ Specify update frequency
✓ Add a description	✓ Specify a license	○ Publish a notebook Provide an example of the data in use so other users can get started quickly
✓ Upload an image	✓ Use preferred file formats	○ Add a task Suggest an analysis users can do with this dataset

Filters

[Datasets](#) [Tasks](#) [Computer Science](#) [Education](#) [Classification](#) [Computer Vision](#) [NLP](#) [Data Visualization](#)

71 Tasks See All

Are the tweets are really effective?
[Mustafa Dogan](#) • 5 Submissions



Tweets about theTop...
[Omer Metin](#) • Usability 10.0

Detection of depressing tweets
[Baba Bulls Eye](#) • 0 Submissions



Depression Analysis
[Baba Bulls Eye](#) • Usability 3.5

Analyzing the influence of the twitter chat over Bitcoin price ...
[laetitia hoquetis](#) • 4 Submissions



Bitcoin tweets - 16M t...
[Alex](#) • Usability 9.7

Perform Sentiment Analysis on the US Election Candidates
[Manch Hui](#) • 2 Submissions



US Election 2020 Twe...
[Manch Hui](#) • Usability 10.0

695 Datasets Most Votes ▾ grid

 **Sentiment140 dataset with 1.6 million tweets**
[Maptoç Mıxənəliç KazAnova](#) • Updated 4 years ago
Usability 8.8 - 1 File (CSV) • 81 MB - 1 Task

1085 Gold ...

 **Twitter US Airline Sentiment**
[Figure Eight](#) Updated 2 years ago
Usability 8.2 - 2 Files (SQLITE, CSV) • 3 MB

747 Gold ...

 **News Headlines Dataset For Sarcasm Detection**
[Rishabh Misra](#) • Updated 2 years ago
Usability 10.0 - 2 Files (JSON) • 3 MB - 1 Task

667 Gold ...

 **Dataset**

Explosive Violence Monitoring Project

Recorded casualty data from explosive violence from English-language media

 Konrad Banachewicz • updated 18 days ago (Version 1)

[Data](#) [Tasks](#) [Code](#) [Discussion](#) [Activity](#) [Metadata](#) [Settings](#) [Download \(5 MB\)](#) [New Notebook](#) ⋮

▲ Make your dataset easy to use Usability 4.1

notebook27227176cf Draft saved

File Edit View Run Add-ons Help

Code ▾ Draft Session off (run a cell to start) ⌂ Save Version 0 >|

```
# This Python 3 environment comes with many helpful analytics libraries installed
# It is defined by the kaggle/python Docker image: https://github.com/kaggle/docker-python
# For example, here's several helpful packages to load

import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)

# Input data files are available in the read-only "../input/" directory
# For example, running this (by clicking "Run" or pressing Shift+Enter) will list all files under the input directory

import os
for dirname, _, filenames in os.walk('/kaggle/input'):
    for filename in filenames:
        print(os.path.join(dirname, filename))

# You can write up to 2GB to the current directory (/kaggle/working/) that gets preserved as output when you create a version
# You can also write temporary files to /kaggle/temp/, but they won't be saved outside of the current session
```

+ Code + Markdown

Data + Add data ^

Input

Output ↗ /kaggle/working ⌂

Settings

Language Python ▾

Environment Preferences

Accelerator None ▾

Internet

Schedule a notebook run ^

Schedule this notebook to run and save a new version on a future date. [View all your scheduled notebooks.](#)

API

Using Kaggle's beta API, you can interact with Competitions and Datasets to download data, make submissions, and more via the command line. [Read the docs](#)

[Create New API Token](#)

[Expire API Token](#)

Drive

Search in Drive

My Drive > Kaggle ▾

On September 13, 2021, a security update will be applied to some of your files. [Learn more](#) See files X

Name	Owner	Last modified	File size
kaggle.json	me	May 9, 2021 me	63 bytes

New

My Drive

Computers

Shared with me

Recent

Starred

Trash

Storage

28.81 GB of 201 GB used

[Buy storage](#)



A scenic view of a coastal town built into a steep hillside overlooking the sea. The town features colorful buildings with terracotta roofs, nestled among greenery and rocks. The sea is visible in the background under a clear sky.

Download (155 KB)

New Notebook

⋮

Tags health, religion and belief sys
beginner, healthcare

- Copy API command**
- Social share...**
- Report issue...**
- Bookmark**

Chapter 3: Working and Learning with Kaggle Notebooks

Code

Explore and run machine learning code with Kaggle Notebooks. Find help in the [Documentation](#).

+ New Notebook Your work

Search public notebooks Filters

All notebooks Recently Viewed Python R Beginner NLP Finance Random Forest GPU TPU Competition notebook

Trending

See all (251)

Spam Classifier with the use of Stemming

Updated 9 hours ago

Email Spam

TPS-07 Turtles on turtles...

Updated 13 hours ago

Tabular Playground Series - Jul 2021+2

Hotel_Booking_Dataset_Pract

Updated 44 minutes ago

Hotel Booking

STARTER Notebook(Simple EDA+ XGB)

Updated 2 hours ago

Tabular Playground Series - Jul 2021

Dataset

Temperature Forecast Project using ML

Predict Minimum and Maximum Temperature during a day based on various features

Ayush Yadav • updated 12 hours ago (Version 1)

Data Tasks Code (1) Discussion Activity Metadata Download (2 MB) New Notebook

Usability 7.4 Tags intermediate, environment, regression, categorical data, korea

Description

Temperature Forecast Project using ML

Problem Statement:

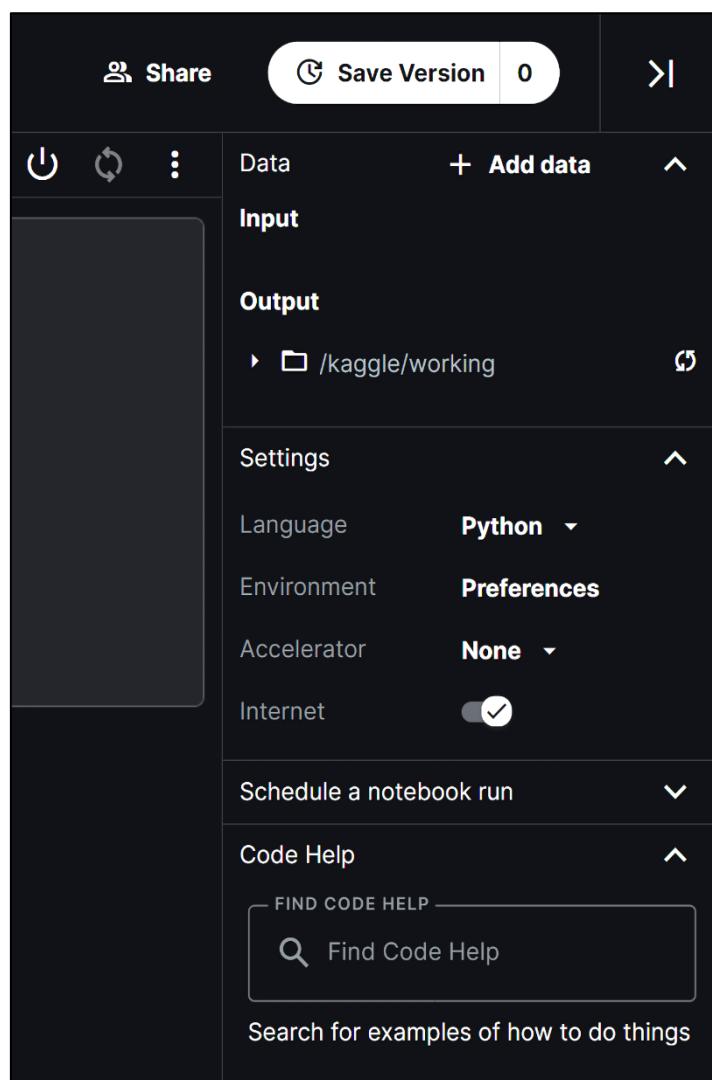
Data Set Information:

This data is for the purpose of bias correction of next-day maximum and minimum air temperatures forecast of the LDAPS model operated by the Korea Meteorological Administration over Seoul, South Korea. This data consists of summer data from 2013 to 2017. The input data is largely composed of the LDAPS model's next-day forecast data, in-situ maximum and minimum temperatures of present-day, and geographic auxiliary variables. There are two outputs (i.e. next-day maximum and minimum air temperatures) in this data. Hindcast validation was conducted for the period from 2015 to 2017.

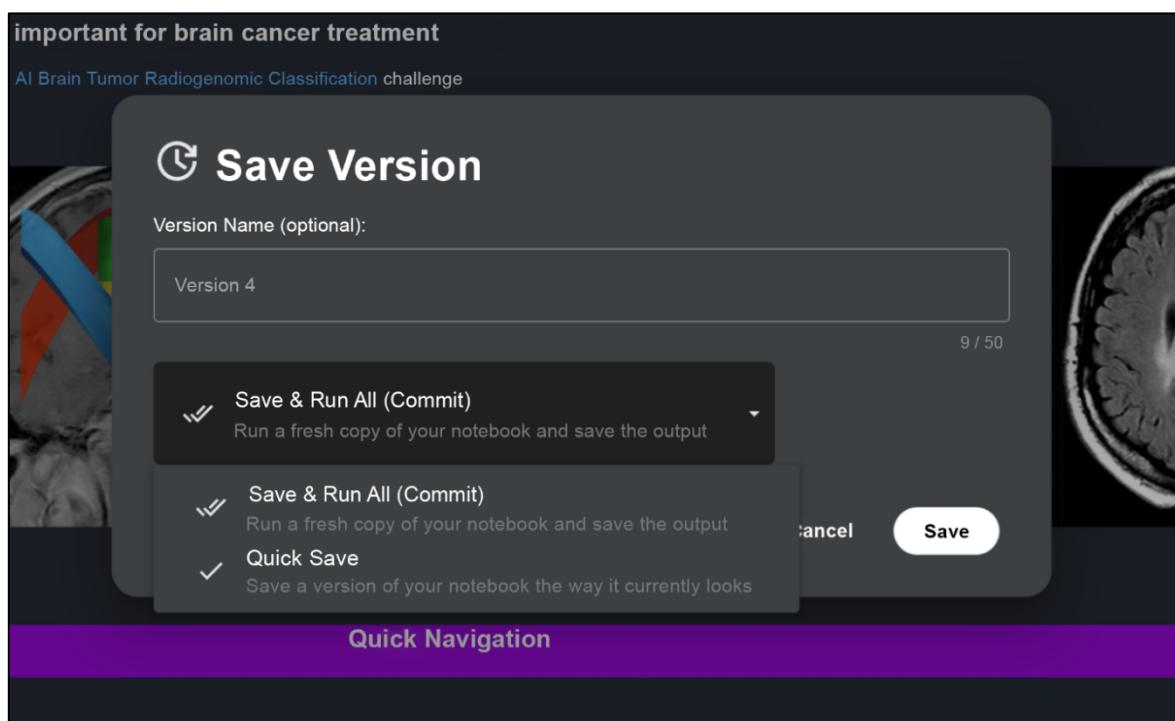
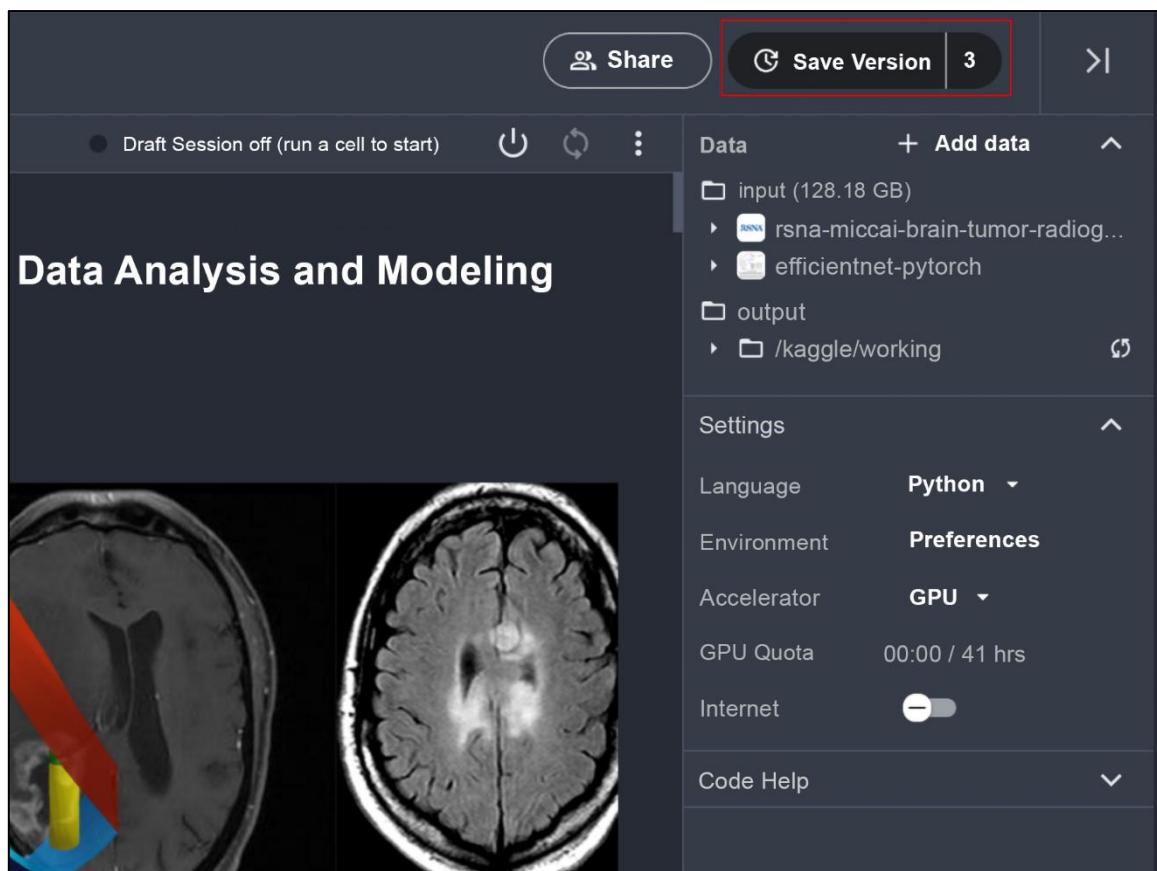
Attribute Information:

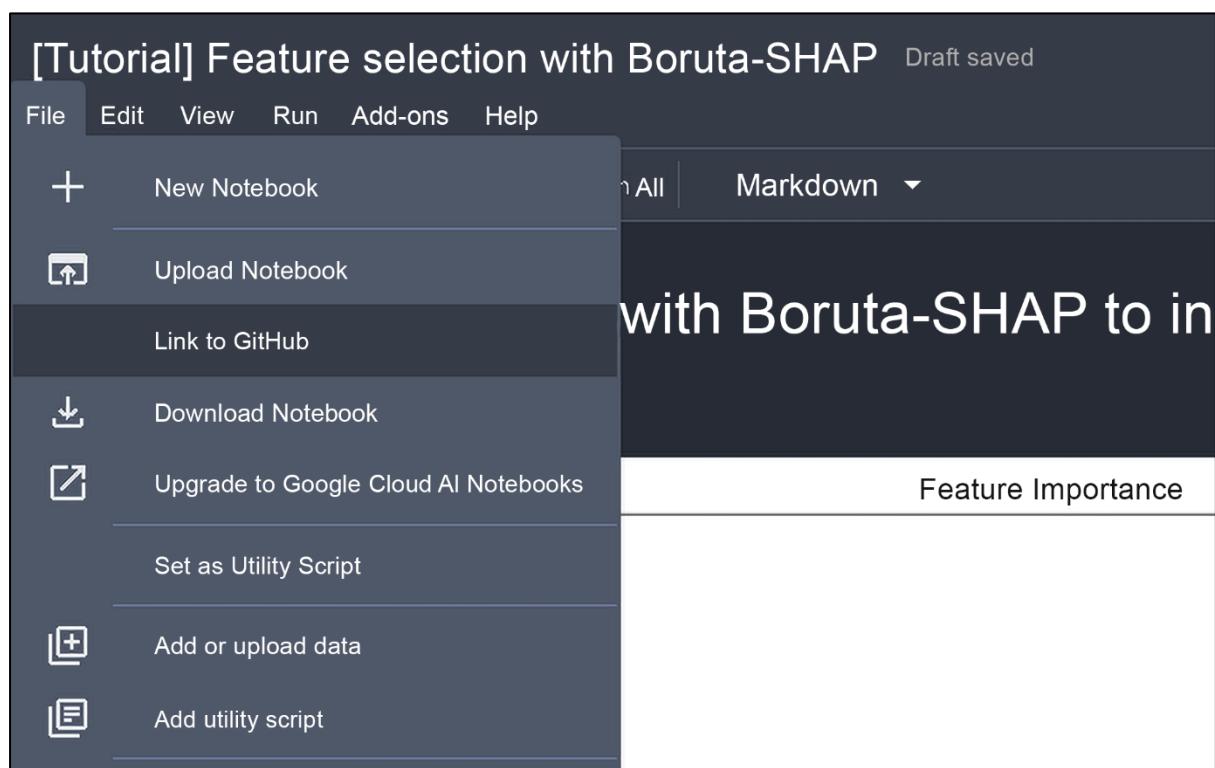
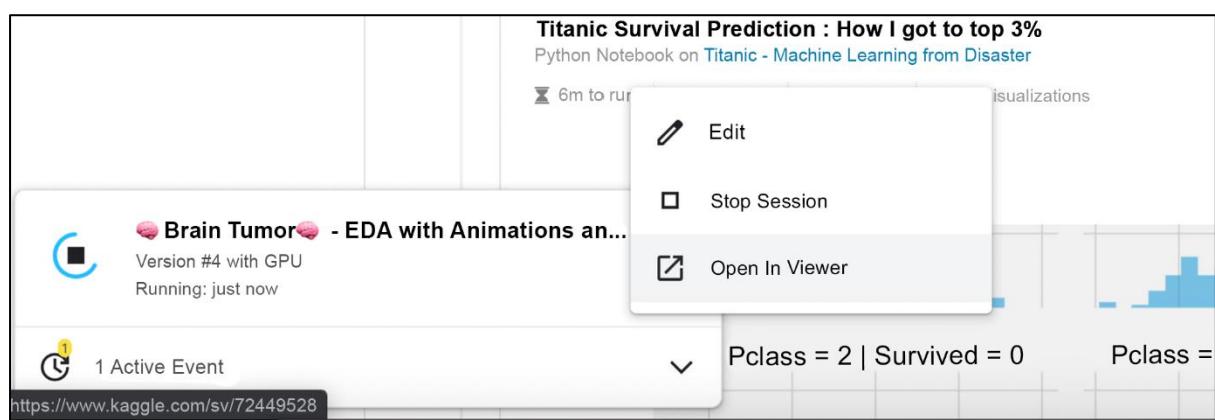
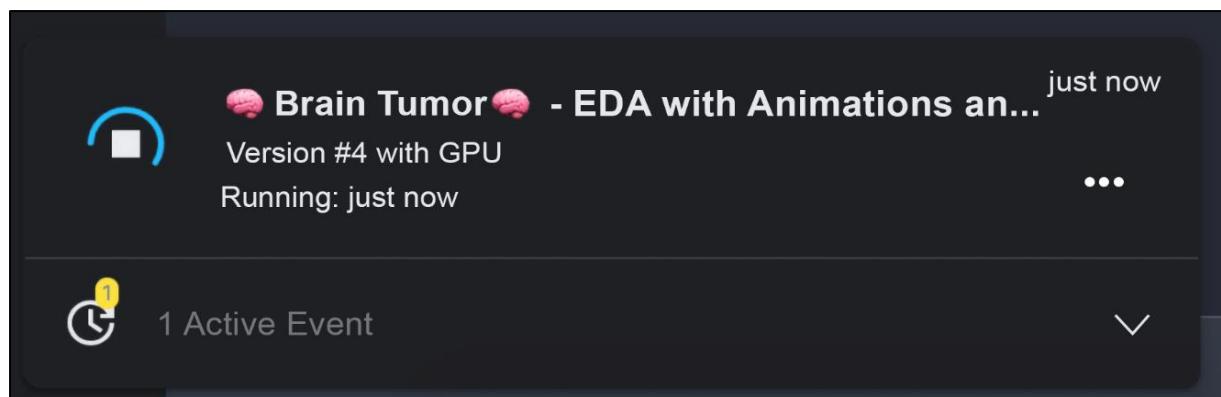
For more information, read [Cho et al, 2020].

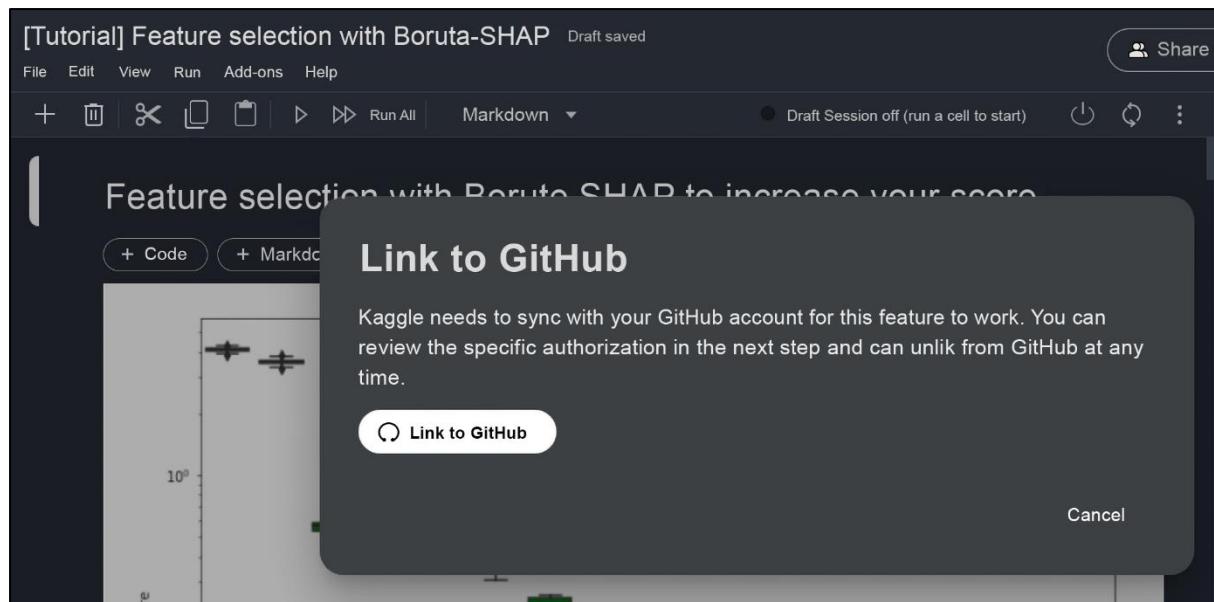
This screenshot shows a Jupyter Notebook interface. The main area displays a single code cell containing Python code related to file processing and data loading. The code includes imports for numpy, pandas, and os, along with logic to walk through a directory and print filenames. The notebook title is "notebook729b4575b" and it is marked as "Draft saved". The top bar includes standard menu options like File, Edit, View, Run, Add-ons, and Help, along with buttons for Share, Save Version (0), and Run All. On the right side, there are sections for Data, Input, Output, Settings, and Code Help. The Settings section shows the Language as Python, Environment as Preferences, Accelerator as None, and Internet as On. The Output section shows the current working directory as /kaggle/working. A search bar at the bottom right is used for finding examples of how to do things.



This screenshot shows a Kaggle notebook submission page for "Sho except + LightGBM + CatBoost + ANN". The page includes a summary card for the best submission, which failed and was submitted 22 days ago. It also shows the public score of 1.3360. The right side of the page displays the "Version 1 of 2" of the notebook, which was copied from "LightGBM + CatBoost + ANN 250512 (+1+1)". The notebook title is "MLB Player Digital Engagement Forecasting" and the description is "LightGBM + CatBoost + ANN 2505f2". A red box highlights the "Copy and Edit" button in the top right corner of the notebook preview.







Save copy to GitHub

Upload this ipynb to GitHub under your GitHub account lmassaron. This can only be undone directly on GitHub.

REPOSITORY

lmaassaron/kaggle_public_notebooks ▾

BRANCH

main ▾

FILE NAME

tutorial-feature-selection-with-boruta-shap.ipynb

COMMIT MESSAGE

Kaggle Notebook | [Tutorial] Feature selection with Boruta-SHAP | Vers

Include a link to Kaggle

 **Konrad Banachewicz** X

-  Your Profile
-  Account
-  Sign Out

 **Your accelerator quota**

GPU 41h available of 41h 

TPU 30h available of 30h 

 ...

^ 11 Share Edit ...

Version 34 of 34

Notebook

Groundwork

- Patterns
- Dependence
- Stationarity

Input (2)

Execution Info

Log

Comments (5)

-  Copy and Edit Notebook
-  View Copies (1)
-  Download code
-  Embed notebook
-  Copy API command
-  Upgrade to Google Cloud AI Notebooks
-  Unfollow Notebook
-  Unfollow Comments
-  Bookmark
-  Delete Notebook

Upgrade to Google Cloud AI Platform Notebooks



Google Cloud

Access more compute power by exporting your notebook and its dependencies to Google Cloud AI Platform Notebooks where you can customize a virtual machine without quotas or runtime limits.

This process is three steps:

1. Setup a billing-enabled Google Cloud Project
2. Setup your notebook instance and optionally customize your machine
3. Run your code without limits

[Cancel](#)

[Continue](#)

adversarial validation code

X

Microphone



All

Images

Videos

News

Maps

More

Tools

About 1.100.000 results (0,45 seconds)

<https://towardsdatascience.com/adversarial-validation-ca...>

Learning the Adversarial Validation model - Towards Data ...

Go there for better rendering of in-line **code**. Introduction. If you were to study some of the competition-winning solutions on Kaggle, you might notice ...

21 Jan 2020 · Uploaded by WelcomeAIOverlords

<https://www.kdnuggets.com/2016/10/adversarial-val...>

Adversarial Validation, Explained - KDnuggets

This post proposes and outlines **adversarial validation**, a method for selecting training examples most similar to test ... The **code** is available at GitHub.

People also ask

What is adversarial validation?



What is the difference between cross-validation and holdout validation?



What is the validation set in the 3 way cross-validation for?



What is cross-validation and validation machine learning?

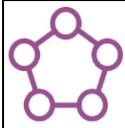


Feedback

<https://www.kaggle.com/konradb/adversarial-validat...>

Adversarial validation and other scary terms | Kaggle

Explore and run machine learning **code** with Kaggle Notebooks | Using data from Sberbank Russian Housing Market



Expert

You've completed a significant body of work on Kaggle in one or more categories of expertise. Once you've reached the expert tier for a category, you will be entered into the site wide Kaggle Ranking for that category.

Competitions

2 bronze medals

Datasets

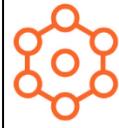
3 bronze medals

Notebooks

5 bronze medals

Discussions

50 bronze medals



Master

You've demonstrated excellence in one or more categories of expertise on Kaggle to reach this prestigious tier. Masters in the Competitions category are eligible for exclusive Master-Only competitions.

Competitions

1 gold medal

2 silver medals

Datasets

1 gold medal

4 silver medals

Notebooks

10 silver medals

Discussions

50 silver medals

200 medals in total



Konrad Banachewicz

I interrogate data for a living at TNG Quant Consultancy
Haarlem, North Holland, Netherlands

Joined 11 years ago · last seen in the past day

[Q](#) [T](#) [in](#)

Followers 642



Discussion
Grandmaster

[Home](#)

[Competitions \(107\)](#)

[Datasets \(23\)](#)

[Code \(49\)](#)

[Discussion \(1,240\)](#)

[Followers \(642\)](#)

...

[Edit Public Profile](#)

Competitions Master

Current Rank **470**
of 162,835

Highest Rank **34**

9 29 16

[Homesite Quote...](#)
 5 years ago
Top 1%

[Machinery Tube...](#)
 6 years ago
Top 1%

[AMS 2013-2014...](#)
 8 years ago
Top 4%

Datasets Expert

Current Rank **76**
of 42,360

Highest Rank **59**

0 1 14

[Augmented trai...](#)
 2 months ago

[Pytorch-Tabnet](#)
 10 months ago

[Text recognition...](#)
 2 years ago

Notebooks Master

Current Rank **152**
of 177,997

Highest Rank **54**

4 10 24

[Adversarial valid...](#)
 4 years ago

[We need to go d...](#)
 10 days ago

[Linear baseline ...](#)
 2 months ago

Discussion Grandmaster

Current Rank **79**
of 206,287

Highest Rank **7**

60 88 586

[Augmented dat...](#)
 2 months ago

[Shrink the data ...](#)
 3 years ago

[Minimalistic BER...](#)
 2 years ago

Chapter 4: Leveraging Discussion Forums

The screenshot shows the 'Discussions' section of the Kaggle website. On the left is a sidebar with links to Home, Competitions, Datasets, Code, Discussions (which is selected), Courses, and More. The main area has a search bar at the top. Below it is a section titled 'Discussions' with a sub-section 'Your Discussions'. To the right is a cartoon illustration of a hand holding a smartphone, pointing at a computer screen displaying a discussion interface. The main content area lists four forum categories: 'General', 'Getting Started', 'Product Feedback', and 'Questions & Answers'. Each category has a small icon, a title, a brief description, and a timestamp indicating when the last post was made. Below each category is a link to 'Recent topics'.

- General**
Announcements, resources, and interesting discussions
last post 2 minutes ago by Bambula
- Getting Started**
The first stop for new Kagglers
last post 7 minutes ago by Emma
- Product Feedback**
Tell us what you love, hate, and wish for
last post 2 hours ago by HuzaifaMS
- Questions & Answers**
Technical advice from other data scientists
last post 2 hours ago by Jae-Yu Yeh

The screenshot shows the 'Discussion from across Kaggle' page. At the top is a search bar and a 'Filters' button. Below are several tabs: 'Your Activity', 'Bookmarks', 'Beginner', 'Data Visualization', 'Computer Vision', 'NLP', 'Neural Networks', and 'more'. A dropdown menu shows 'Hotness' is selected. The main content area lists recent posts with user profiles, titles, and interaction counts (upvotes, comments, and more). Each post includes a link to the full discussion.

- Received suspicious team merge offer**
kaeruru - in CommonLit Readability Prize · Last comment 8h ago by Minh Tri Phan
44 upvotes, 13 comments
- DICOM to PNG dataset (128 GB → 5.2 GB) 🎉🔥**
Jonathan Besomi - in RSNA-MICCAI Brain Tumor Radiogenomic Classification · Last comment 1h ago by Abhishek Prajapati
104 upvotes, 50 comments (New comments)
- Commonly used Machine Learning Algorithms**
vardhan SIRAMDASU - in General · Last comment 14m ago by EmBahr
26 upvotes, 33 comments
- About the targets**
Ahmet Erdem - in MLB Player Digital Engagement Forecasting · Last comment 6h ago by something4kag
22 upvotes, 4 comments
- Competition Relaunch - New Data!**
inversion - in SETI Breakthrough Listen - E.T. Signal Search · Last comment 14h ago by Mani Sarkar
71 upvotes, 31 comments
- Companies and Data Science Interview questions**
sai pavan saketh - in Getting Started · Last comment 1h ago by Cristiano de Magalhães, Eng.
97 upvotes, 115 comments
- Did we miss something ?**
Lukasz Borecki - in Tabular Playground Series - Jul 2021 · Last comment 2h ago by Lukasz Borecki
15 upvotes, 26 comments

RECENCY

Last 30 Days Last 7 Days Today

MY ACTIVITY

Commented Published Viewed

AUTHOR

Admin

TYPES

Site Forums Competition Dataset

TAGS

Search for tags

Clear **Apply**

Discussion from across Kaggle

Clear filters to enable search Filters

Beginner X Hotness ▾

-  Resources of the Week : #12
Pranjal Verma · in General · Last comment 1h ago by Shashank Srivastava 11 comments ...
-  Best Data Science ↗ Youtube Channels !
Jonas Neri · in Getting Started · Last comment 4h ago by Elias Elfarri 42 comments ...
-  More than 250 Awesome Public Datasets
PAVAN KUMAR D · in General · Last comment 10h ago by hashmi 42 comments ...
-  ►►Resources of the Week : #12
Pranjal Verma · in Getting Started · Last comment 2h ago by laxman kusuma 10 comments ...
-  ⚡ Everything about : GAN - Generative Adversarial Networks ⚡ #3Resources 🎉🎉
Shivani Rana 63 · in General · Last comment 42m ago by Kaan BOKE 11 comments ...
-  What is the secret behind random seed 42?
Old Monk · in RSNA-MICCAI Brain Tumor Radiogenomic Classification · Last comment 7h ago by Manu Siddhartha 13 comments ...
-  Commonlit : Some Good things to learn
Shahules · in CommonLit Readability Prize · Last comment 1h ago by Araik Tamazian 42 comments ...

Discussion from across Kaggle

Clear filters to enable search

Filters

Computer Vision X

What is the secret behind random seed 42?

Old Monk · in RSNA-MICCAI Brain Tumor Radiogenomic Classification · Last comment 7h ago by Manu Siddhartha

21 comments

Recent Comments

Resources of the Week : #11

Pranjal Verma · in Getting Started · Last comment 9h ago by hashmi

27 comments

Recently Posted

Pretrained 3D-CNNs and other resources

Alvaro Francesc Budria Fernández · in RSNA-MICCAI Brain Tumor Radiogenomic Classification · Last comment 9h ago by LucaMTB

2 comments

Most Votes

Resources of the Week : #11

Pranjal Verma · in General · Last comment 2d ago by sarahpop小蜗牛

22 comments

Most Comments

18 comments

Featured Code Competition

Optiver Realized Volatility Prediction

Apply your data science skills to make financial markets better

\$100,000 Prize Money

Optiver · 1,051 teams · 2 months to go (2 months to go until merger deadline)

Overview Data Code Discussion Leaderboard Rules Team My Submissions New Topic ...

Discussions

Following

Search discussions

Filters

All Owned Bookmarks Hotness

Pinned topics

Competition data FAQ

Jiashen Liu · Last comment 2h ago by Gunes Evitan

80 comments

New comments · 73 comments

seconds_in_bucket not starting from zero in some parts of the filler data

Jiashen Liu · Last comment 3d ago by Jiashen Liu

16 comments

23 comments

Featured Code Competition

Optiver Realized Volatility Prediction

Apply your data science skills to make financial markets better

Optiver · 1,051 teams · 2 months to go (2 months to go until merger deadline)

\$100,000 Prize Money

$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{(S_i - \bar{S})^2}{\bar{S}}}$

Overview Data Code Discussion Leaderboard Rules Team My Submissions New Topic Bookmark



Jiashen Liu
Competition Host
836th place

Competition data FAQ

Posted in [optiver-realized-volatility-prediction](#) 21 days ago

Hi all,

Again, welcome to our competition! Very excited to see so many Kagglers participating already!

We have received a lot of questions about the competition data, here we would like to provide a list of FAQ to make everyone up-to-speed. Most of below is already covered in the [data page](#) but good to mention here as well if helpful.

Q: Is the same stock_id representing same stocks in all competition dataset?
A: Yes, stock_id is a unique identifier of a stock in real life, and the group of 112 unique stock_id will be present in all datasets.

Q: What does time_id mean and how the target-feature data are structured?



Konrad Banachewicz

I interrogate data for a living at TNG Quant Consultancy
Haarlem, North Holland, Netherlands
Joined 11 years ago · last seen in the past day

[competitions](#)

[Home](#) [Competitions \(113\)](#) [Datasets \(48\)](#) [Code \(62\)](#) [Discussion \(1,285\)](#) [Followers \(807\)](#) [...](#) [Edit Public Profile](#)



Discussion Grandmaster

Total Posts: 1285
Comments: 1186
Topics: 99
Net Votes: 3357
Votes/Post: 2.61

Discussion Summary

Discussion Grandmaster	Current Rank	Highest Rank	Total Posts: 1285 Comments: 1186 Topics: 99 Net Votes: 3357 Votes/Post: 2.61
	103 of 269,397	7	
71	91	609	

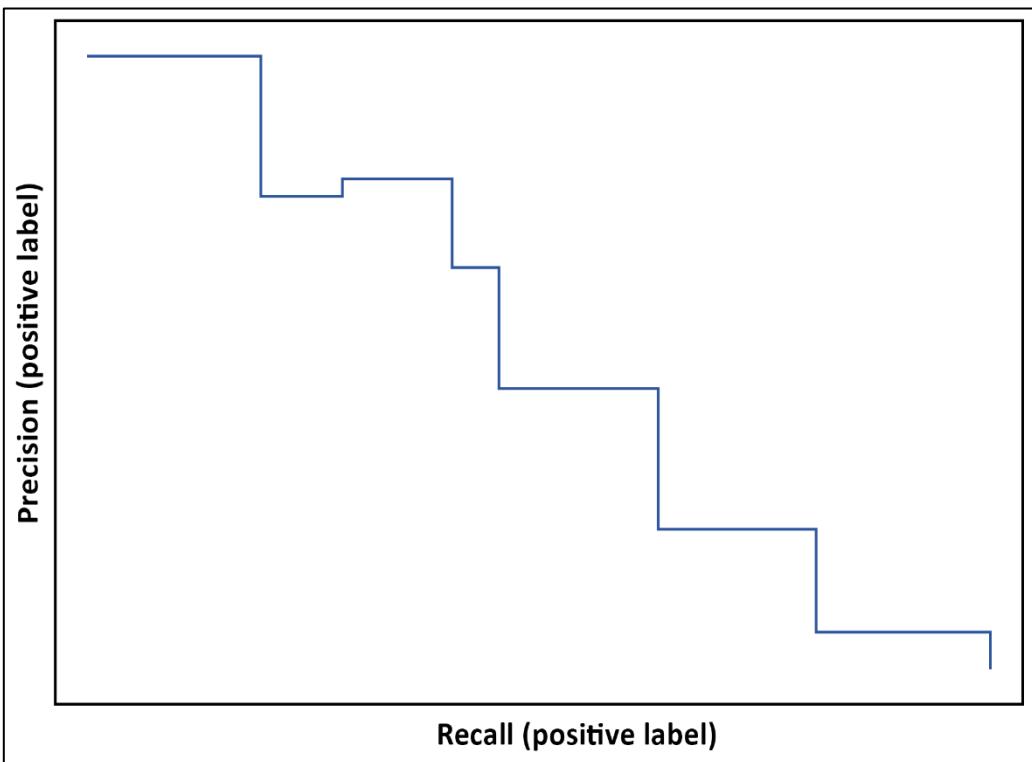
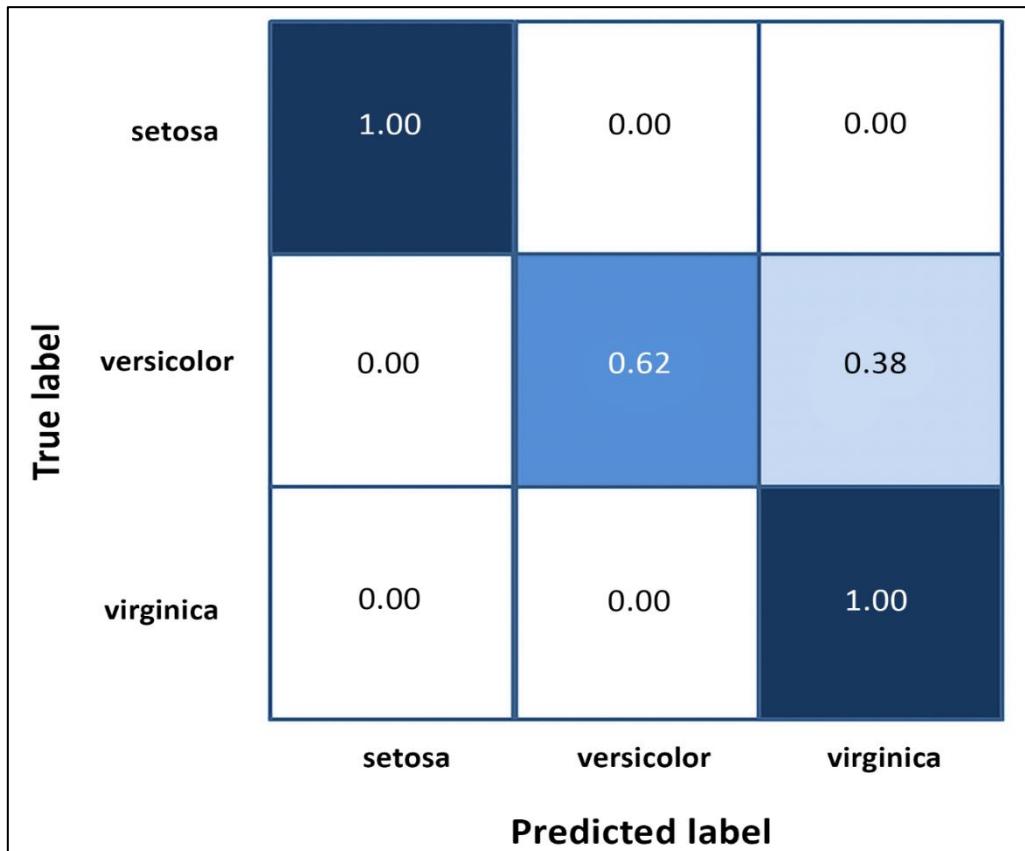
3,357 upvoted Sort by [Most Votes](#)

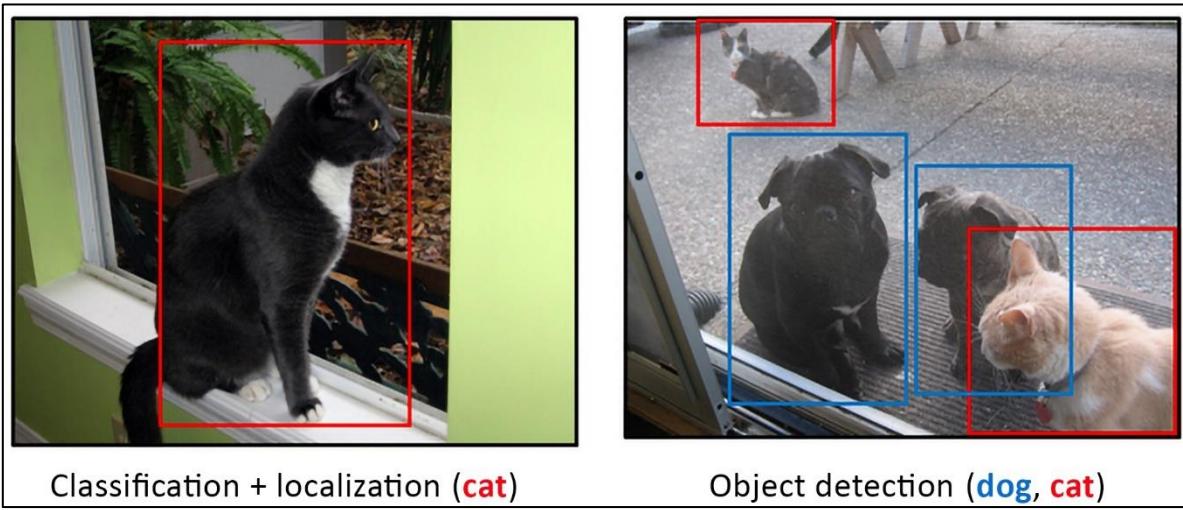
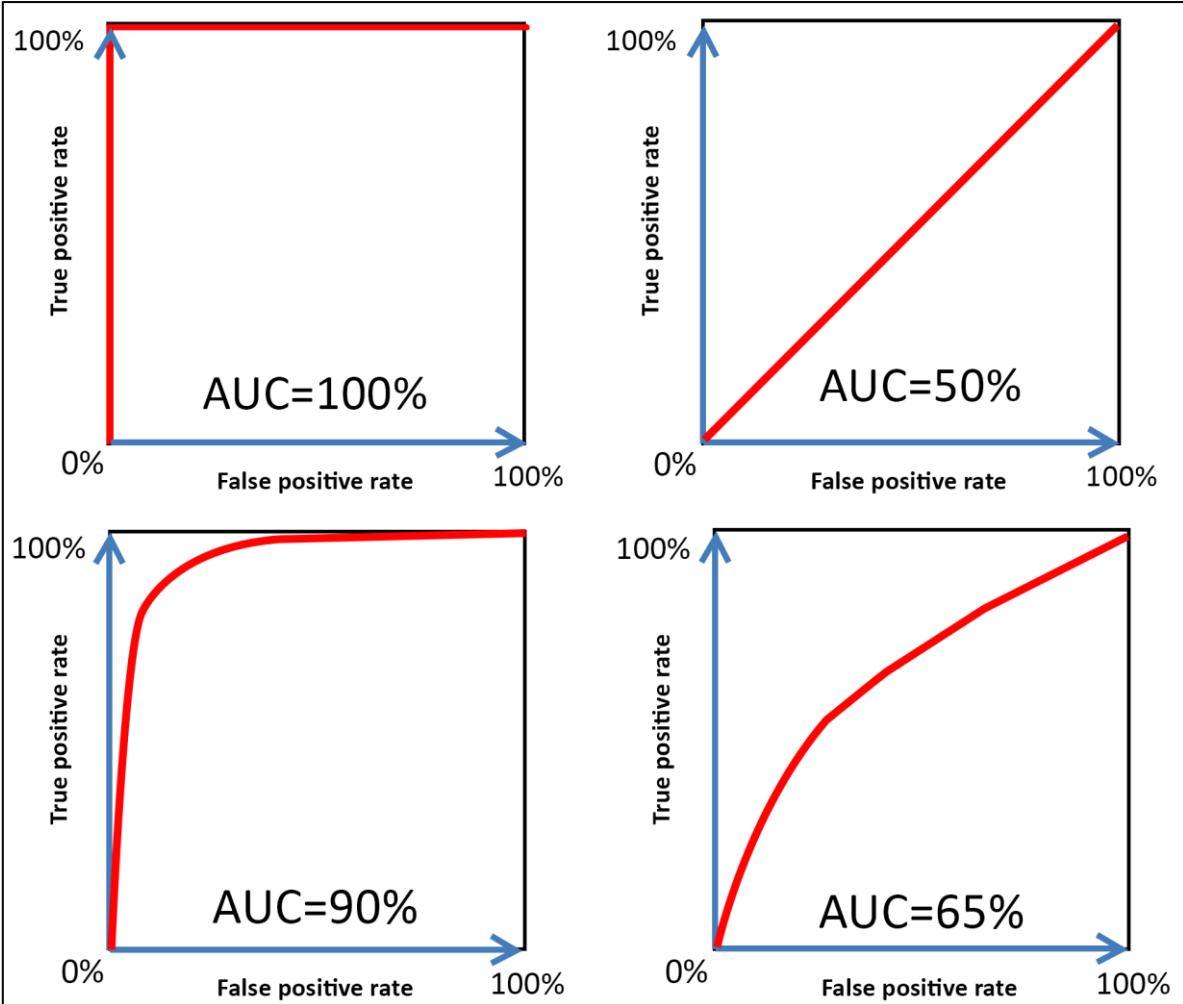
[Comments and topics](#) [Comments](#) [Topics](#) | [Bookmarks](#)

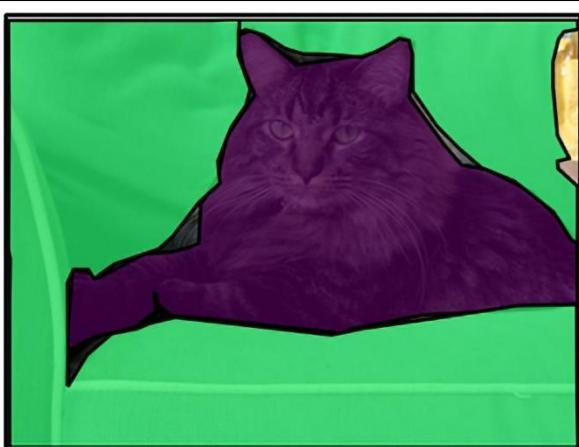
 [Embeddings, Cosine Distance, and ArcFace Explained.](#) 472
Topic 10 months ago in Shopee - Price Match Guarantee

What are These Strange Words?

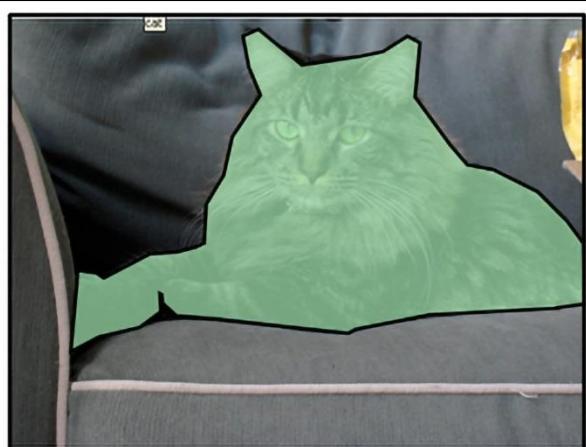
Chapter 5: Competition Tasks and Metrics



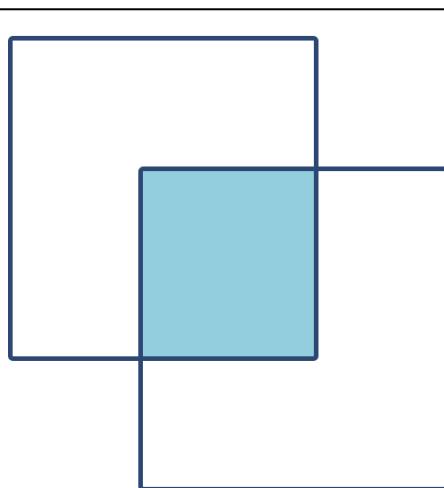




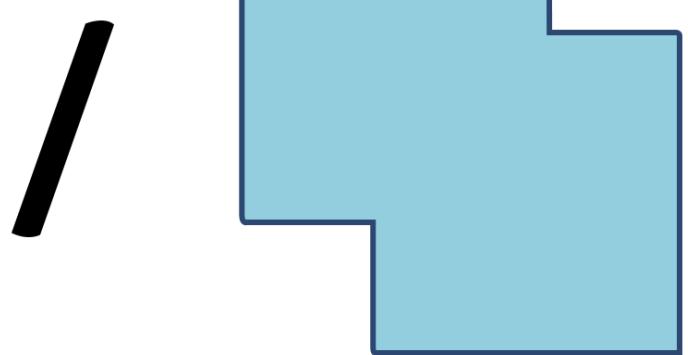
Semantic segmentation (**cat**, **sofa**)



Instance segmentation (**cat**)



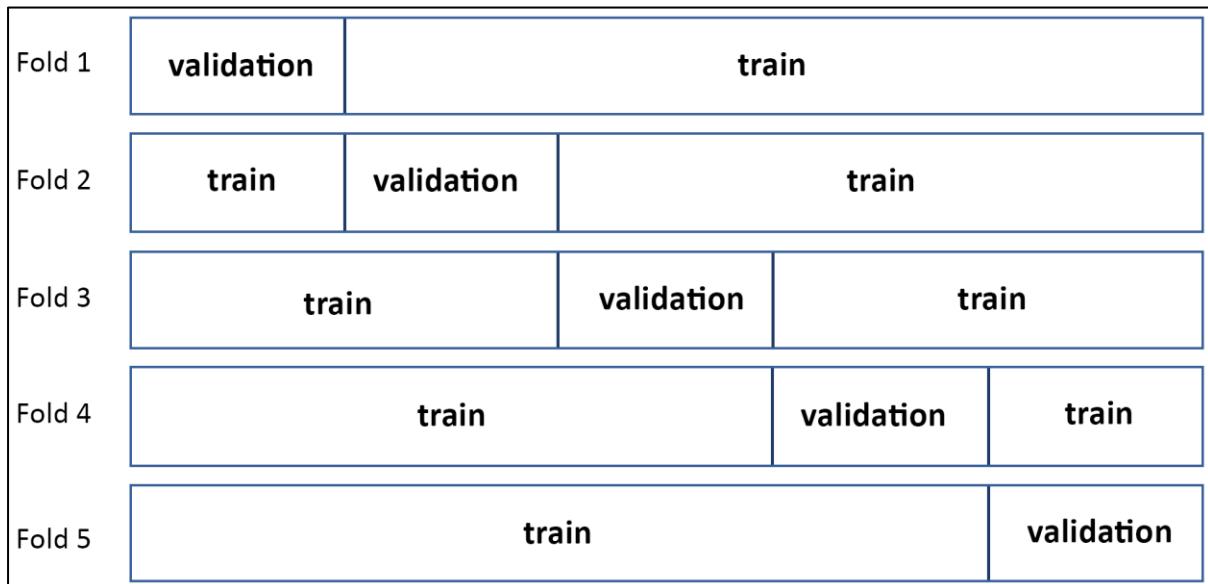
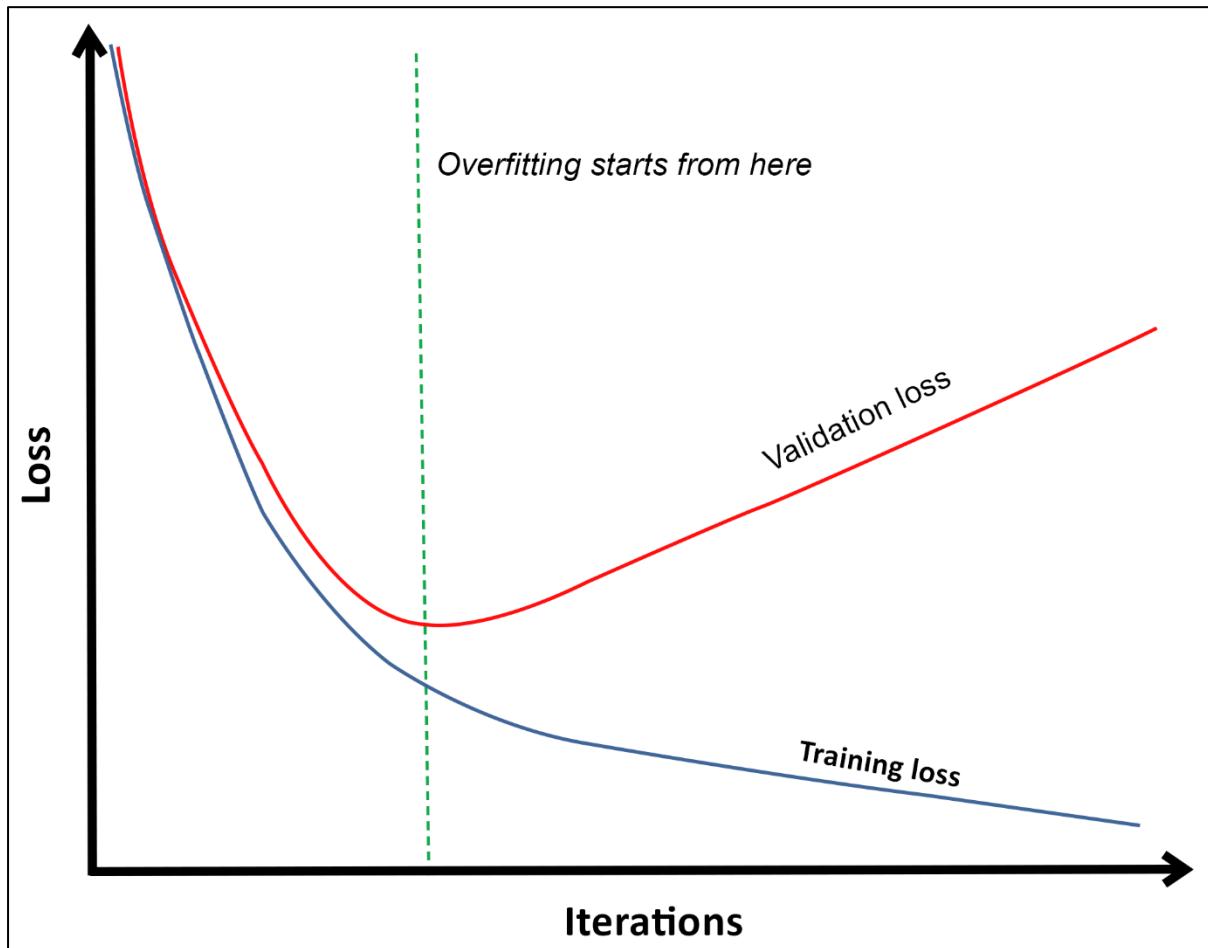
Area of overlap

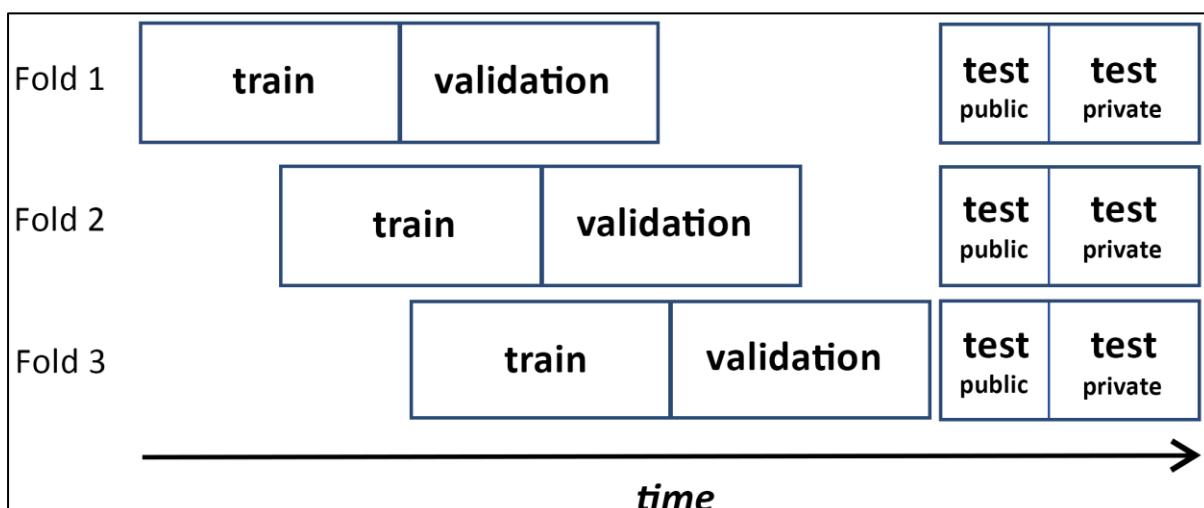
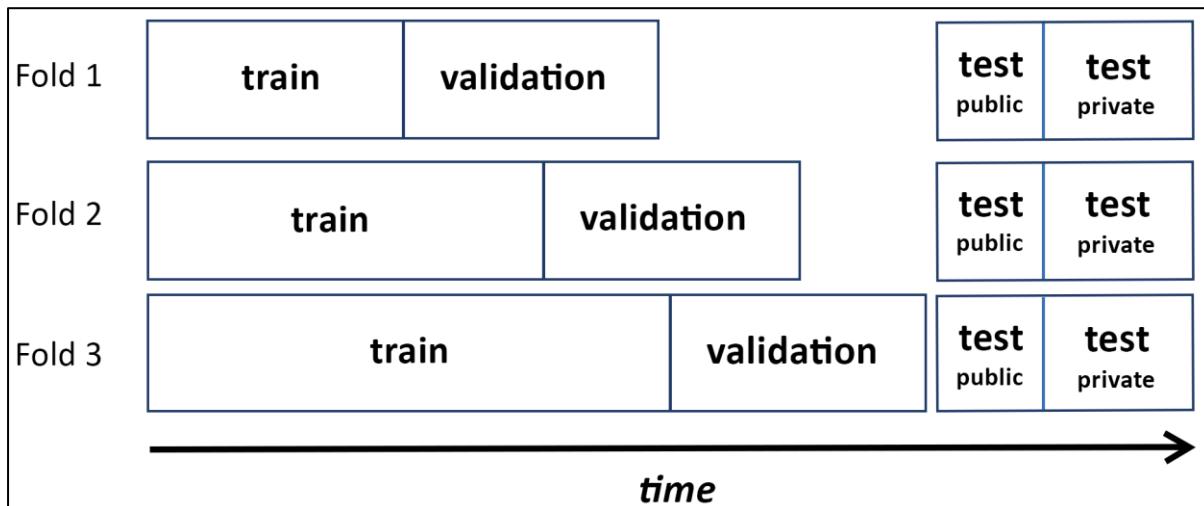
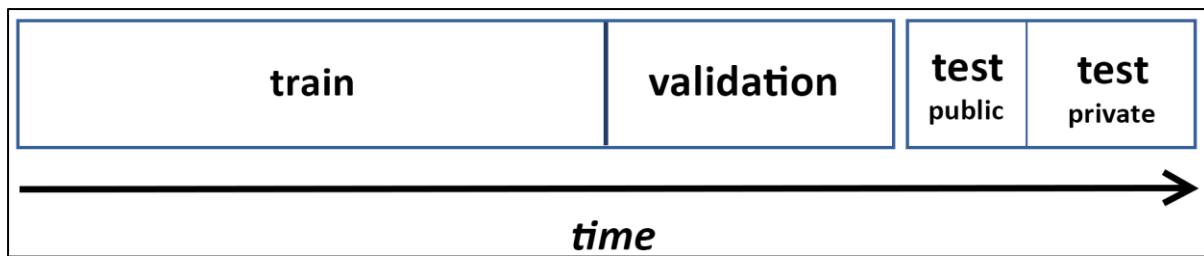


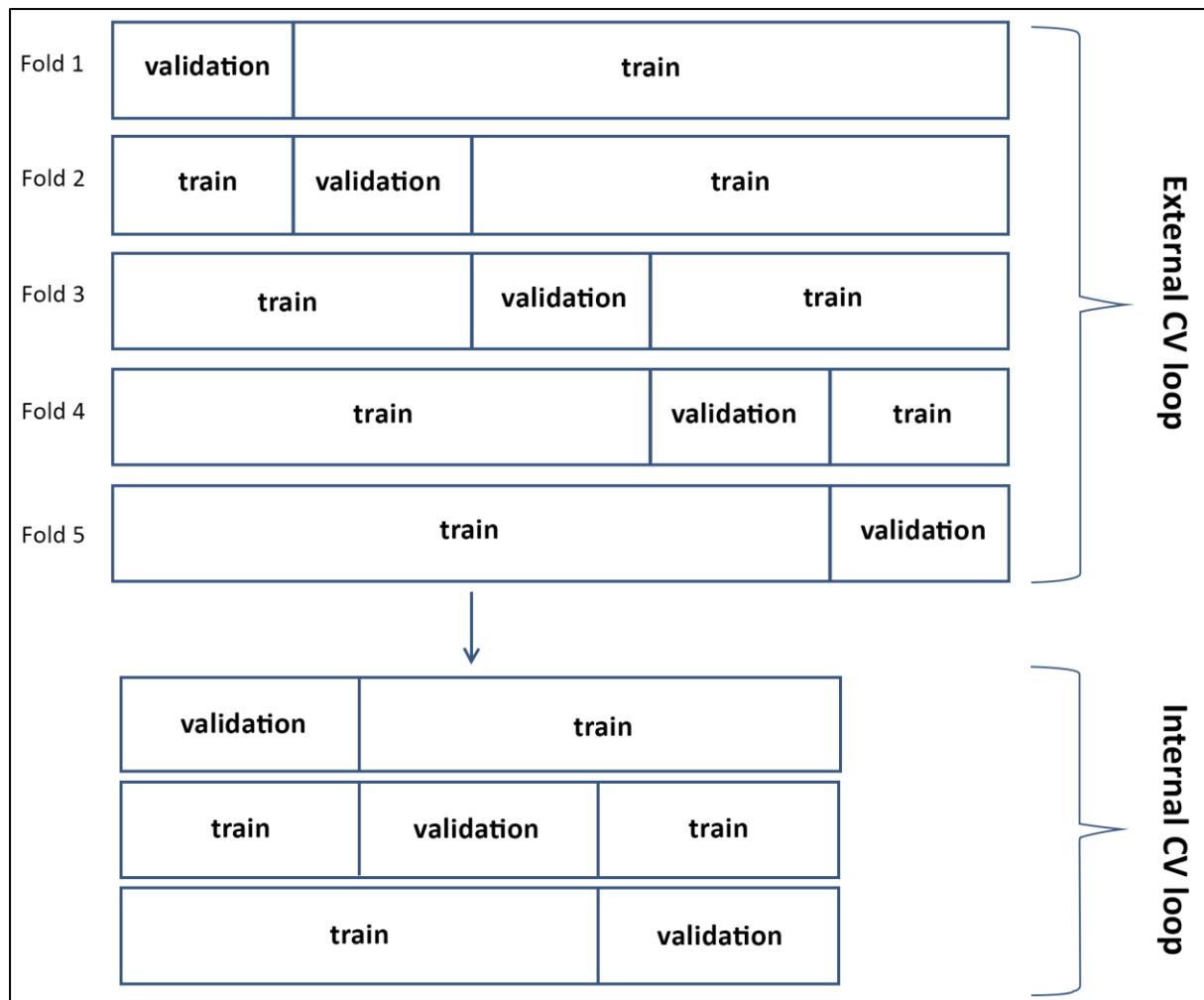
Area of union

$$2 * \frac{\text{Area of overlap}}{\text{Sum of areas}}$$

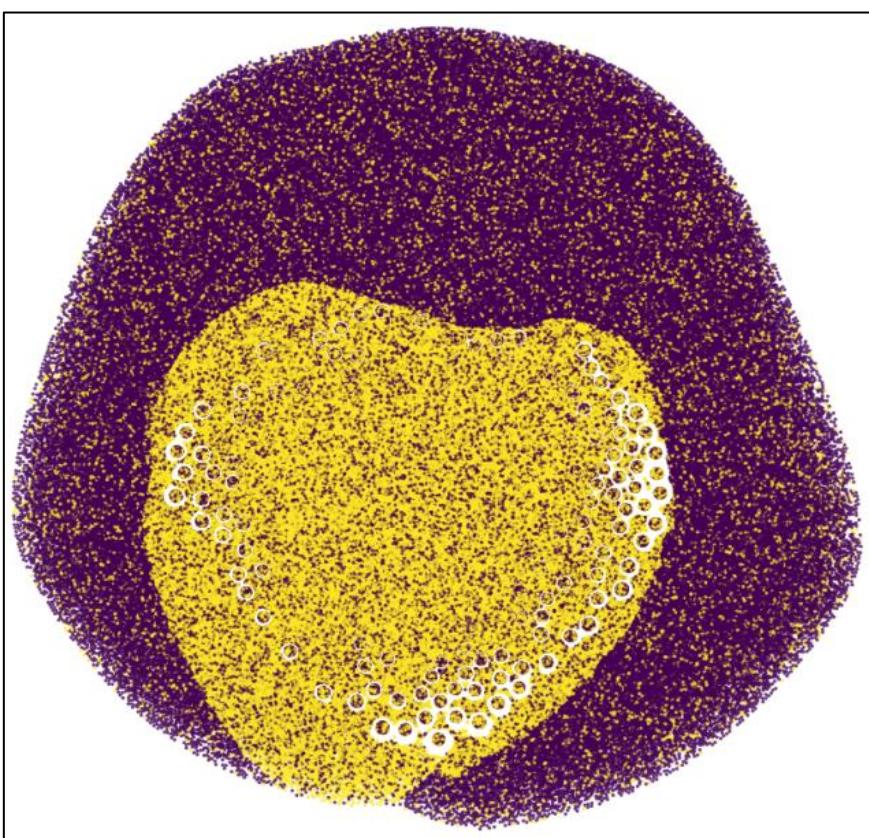
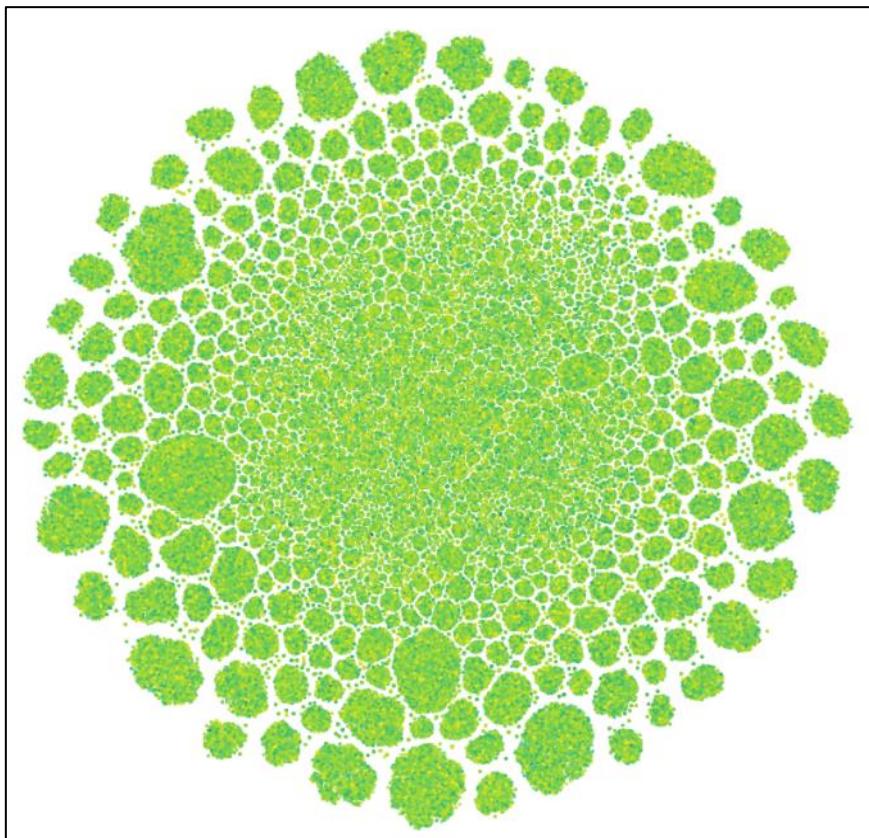
Chapter 6: Designing Good Validation

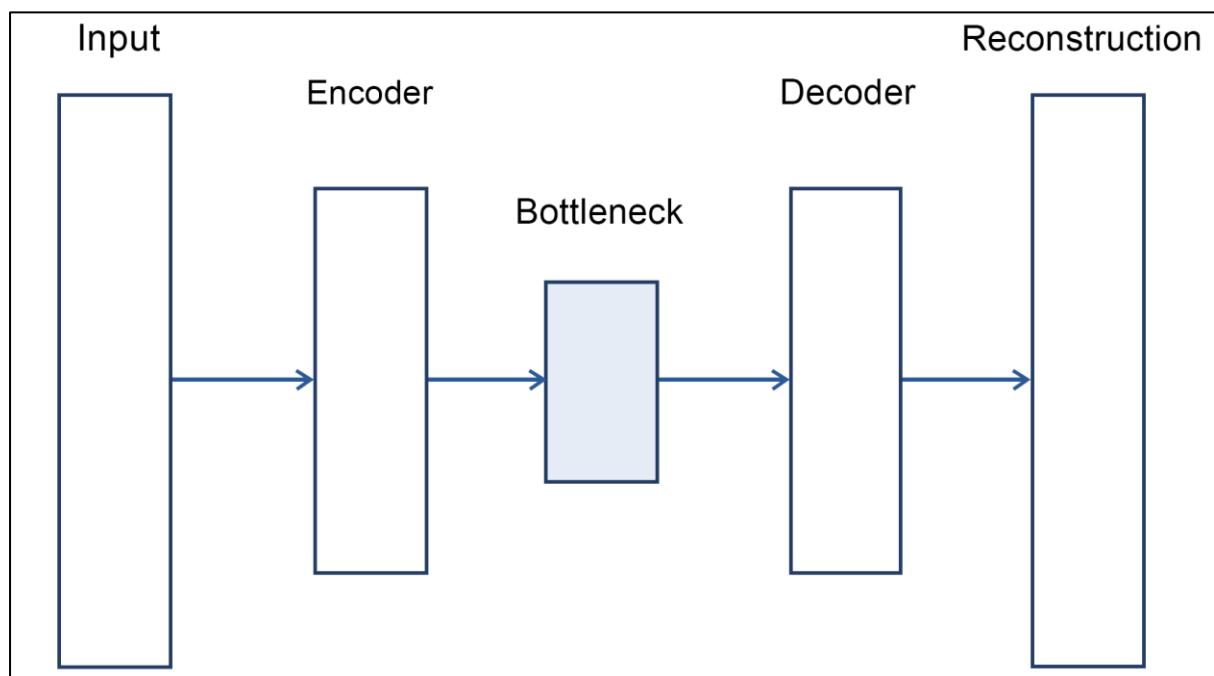
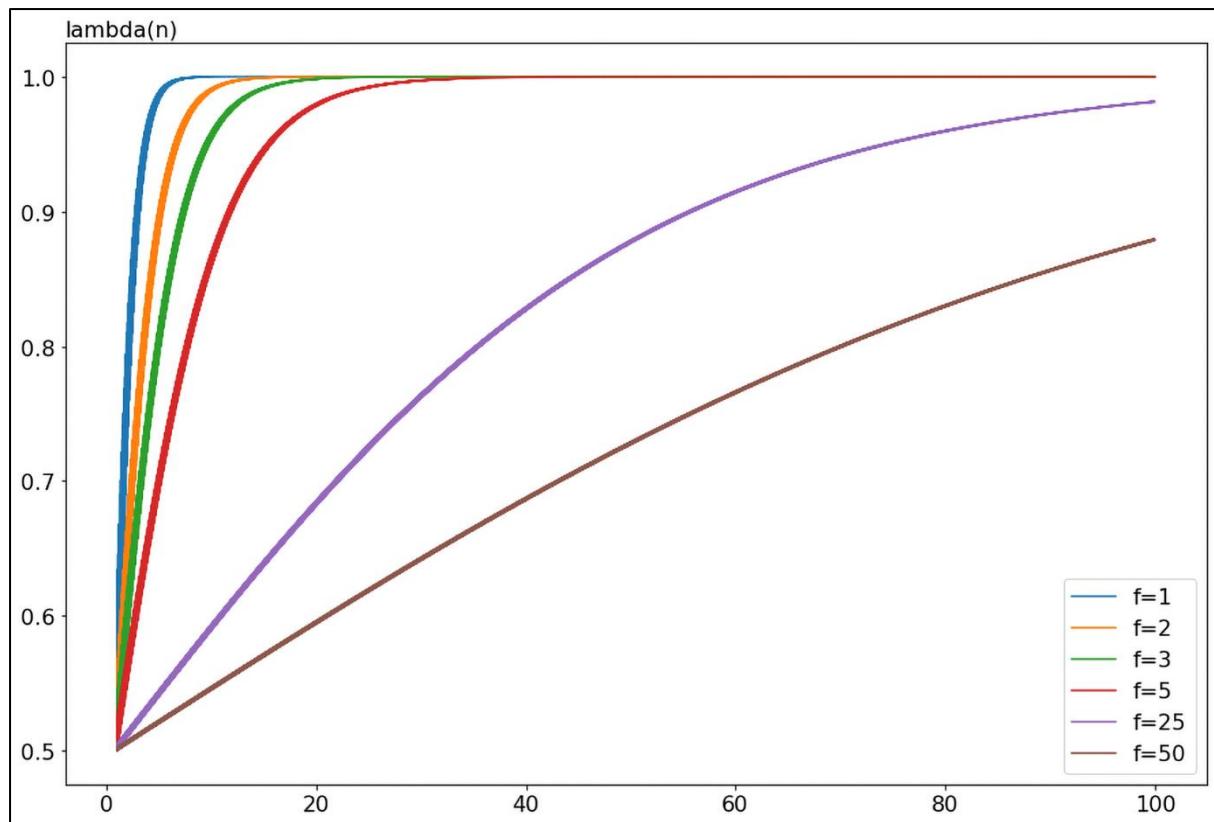


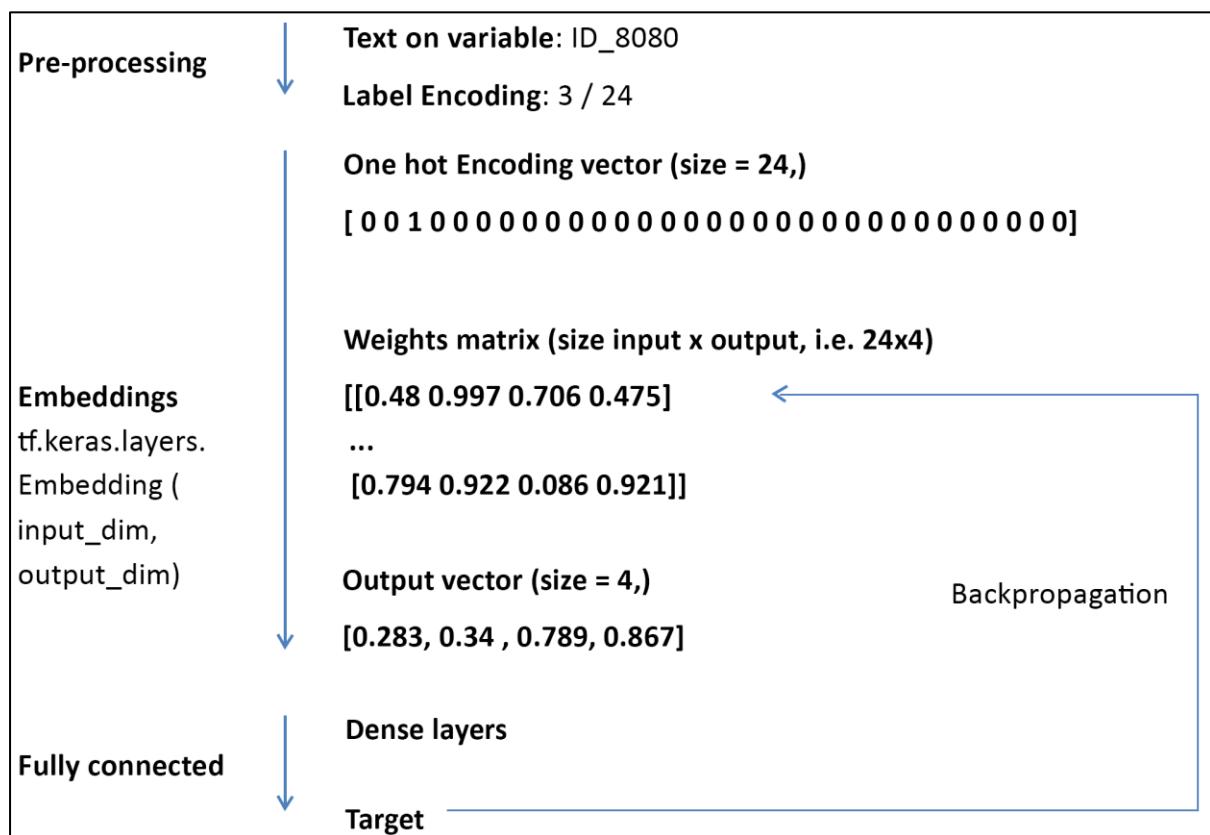
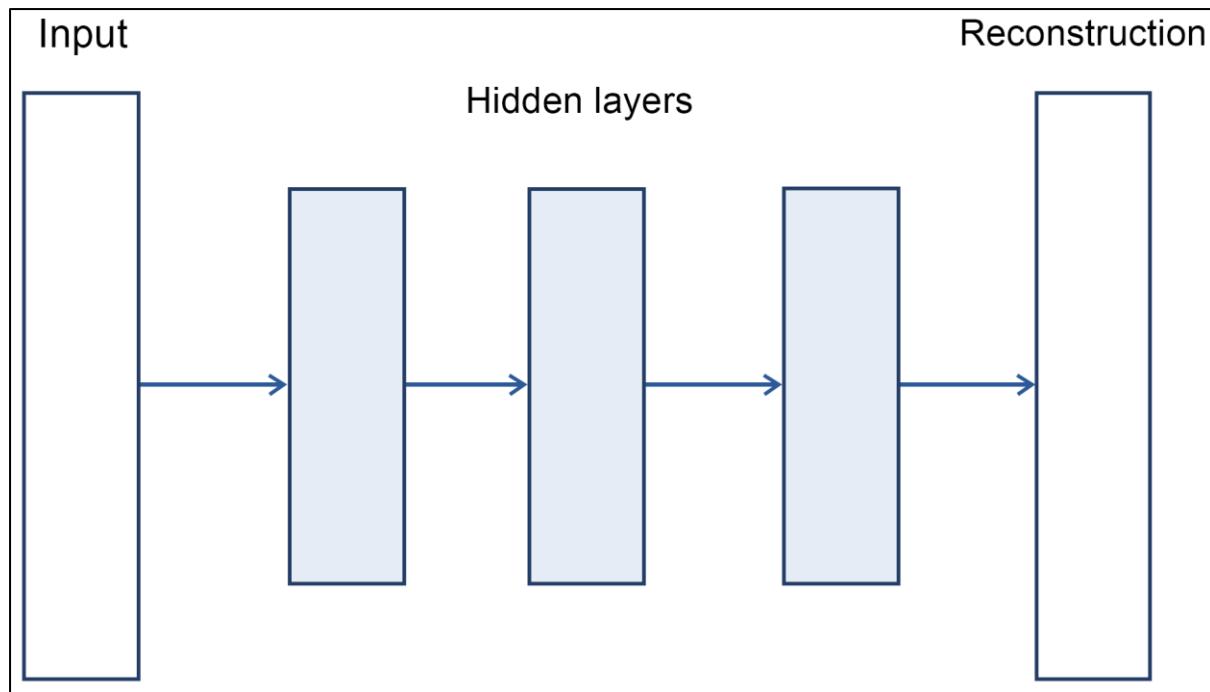




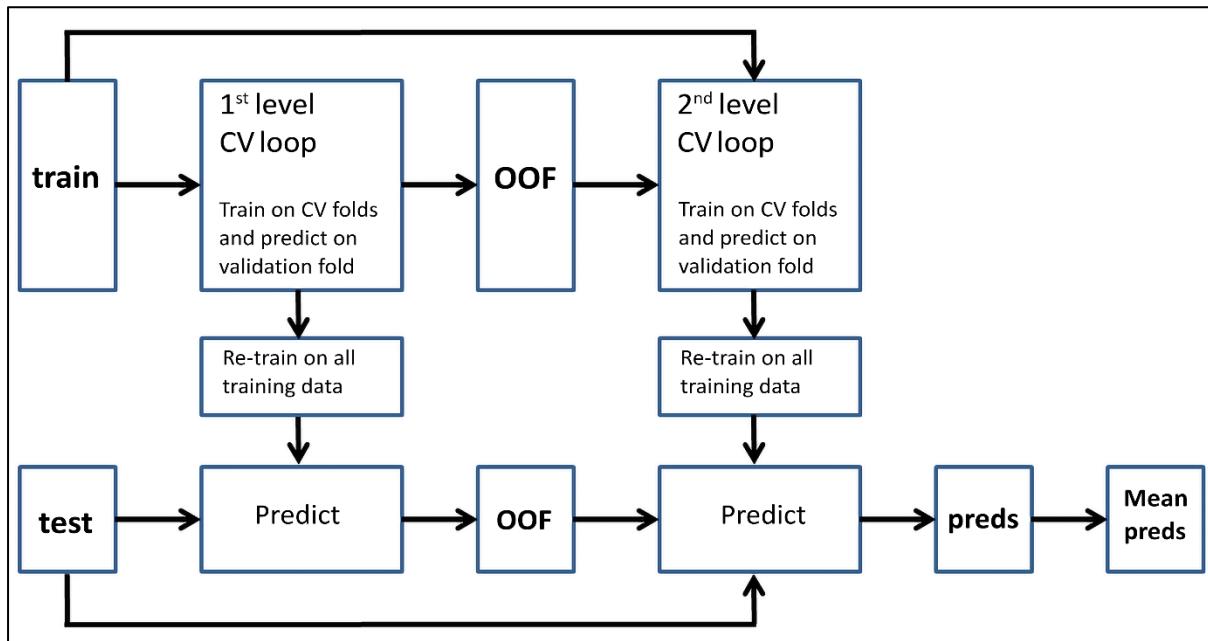
Chapter 7: Modeling for Tabular Competitions





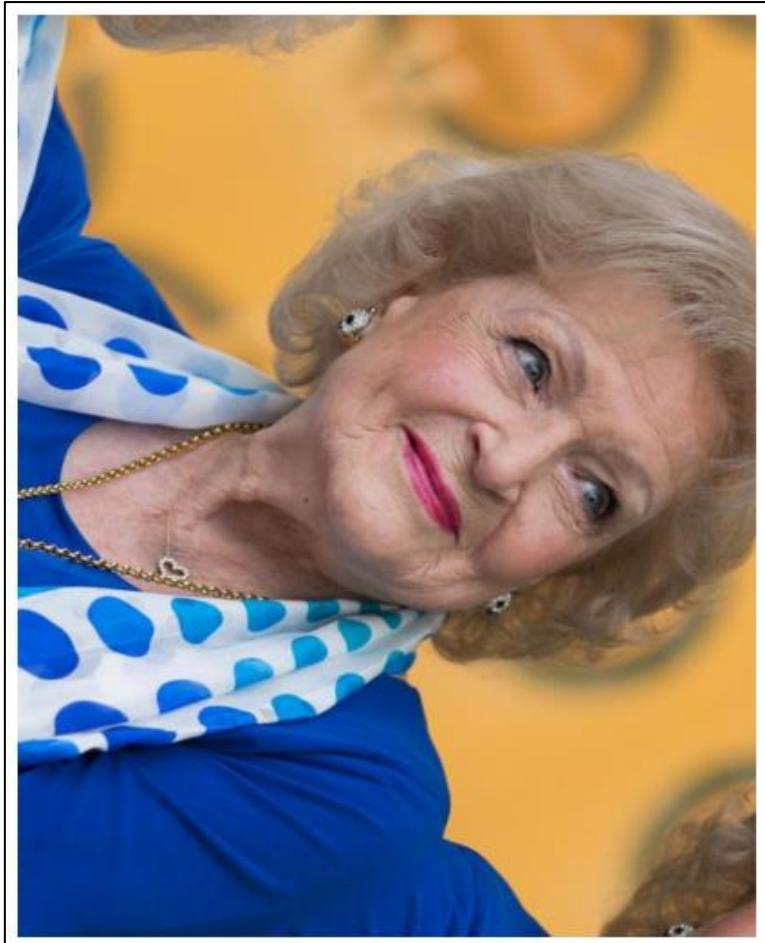


Chapter 9: Ensembling with Blending and Stacking Solutions



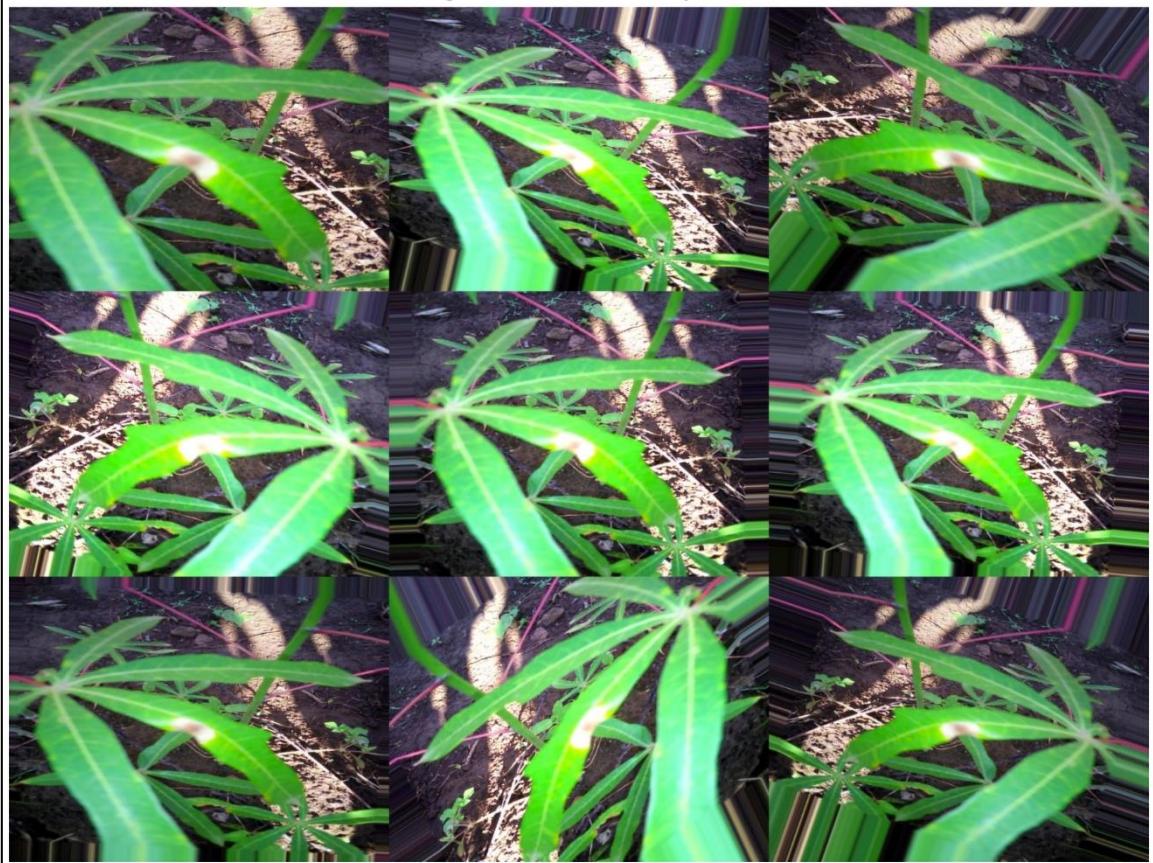
Chapter 10: Modeling for Computer Vision

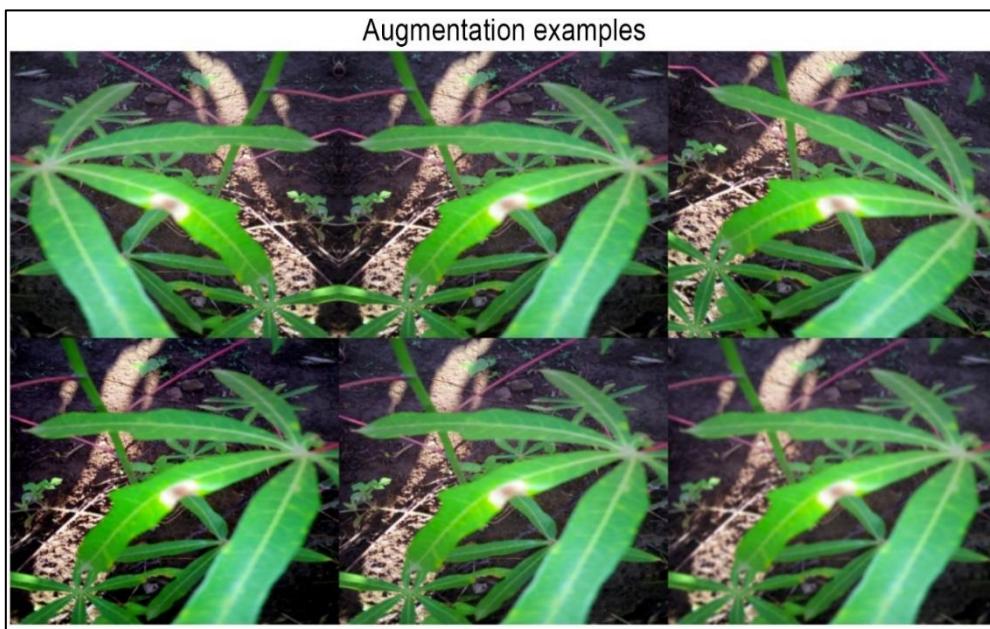
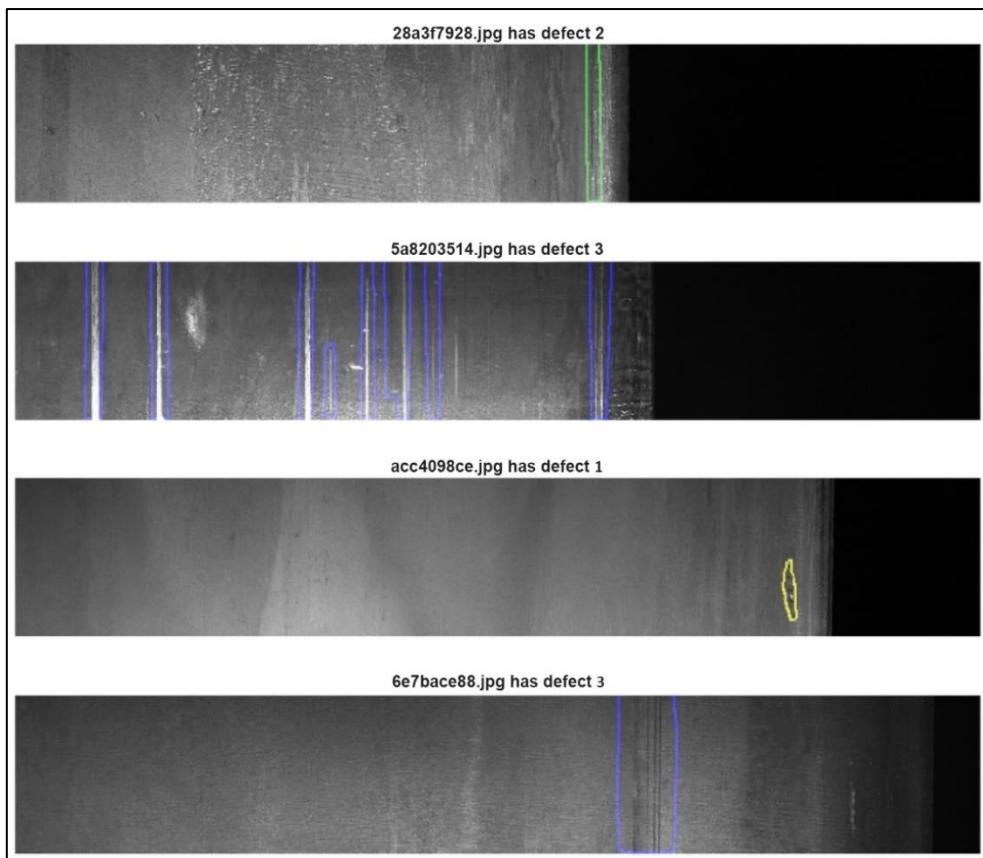


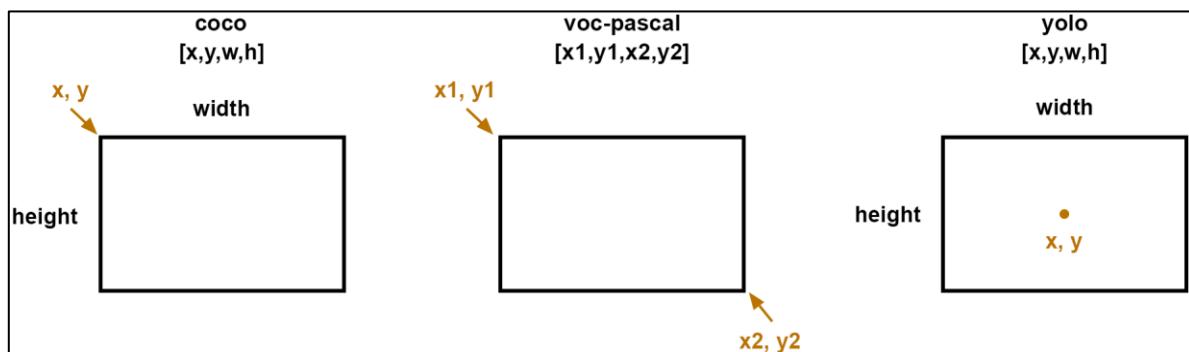
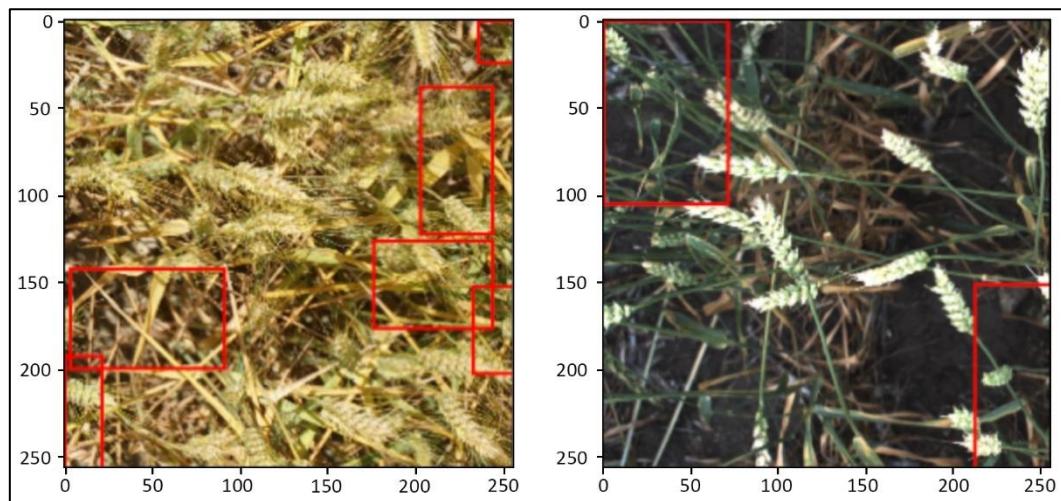
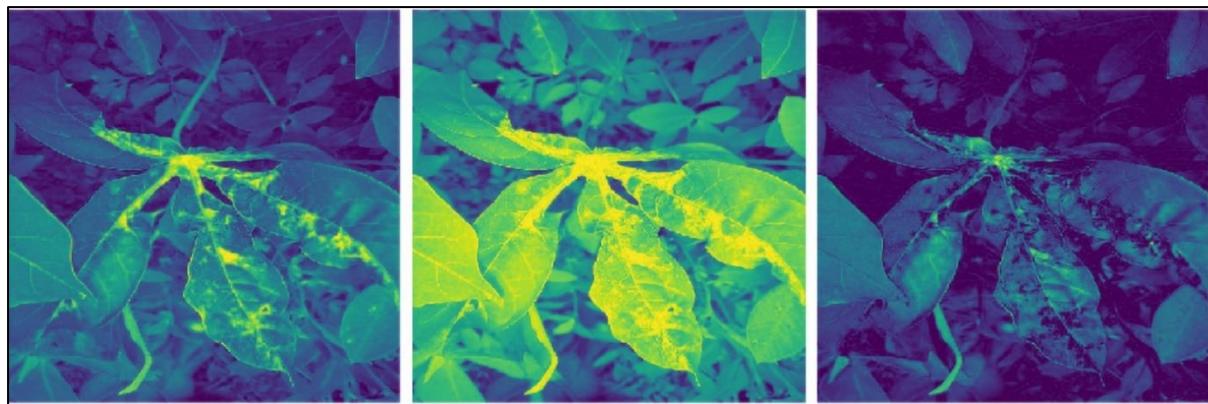


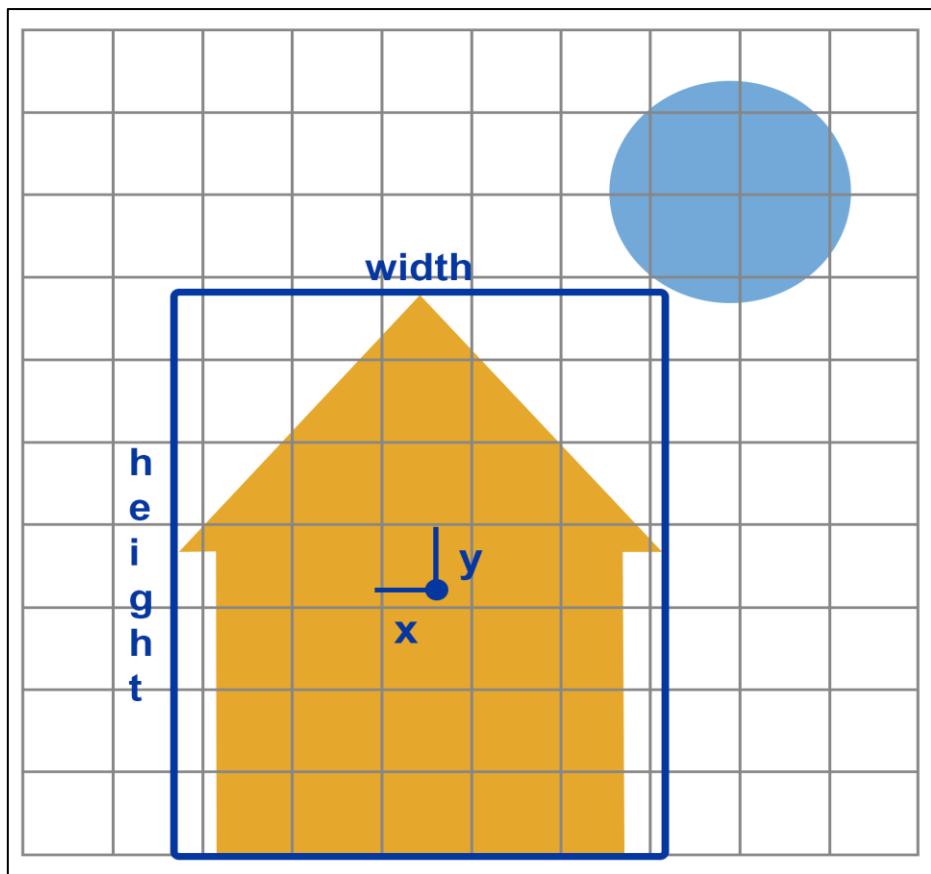


Augmentation examples



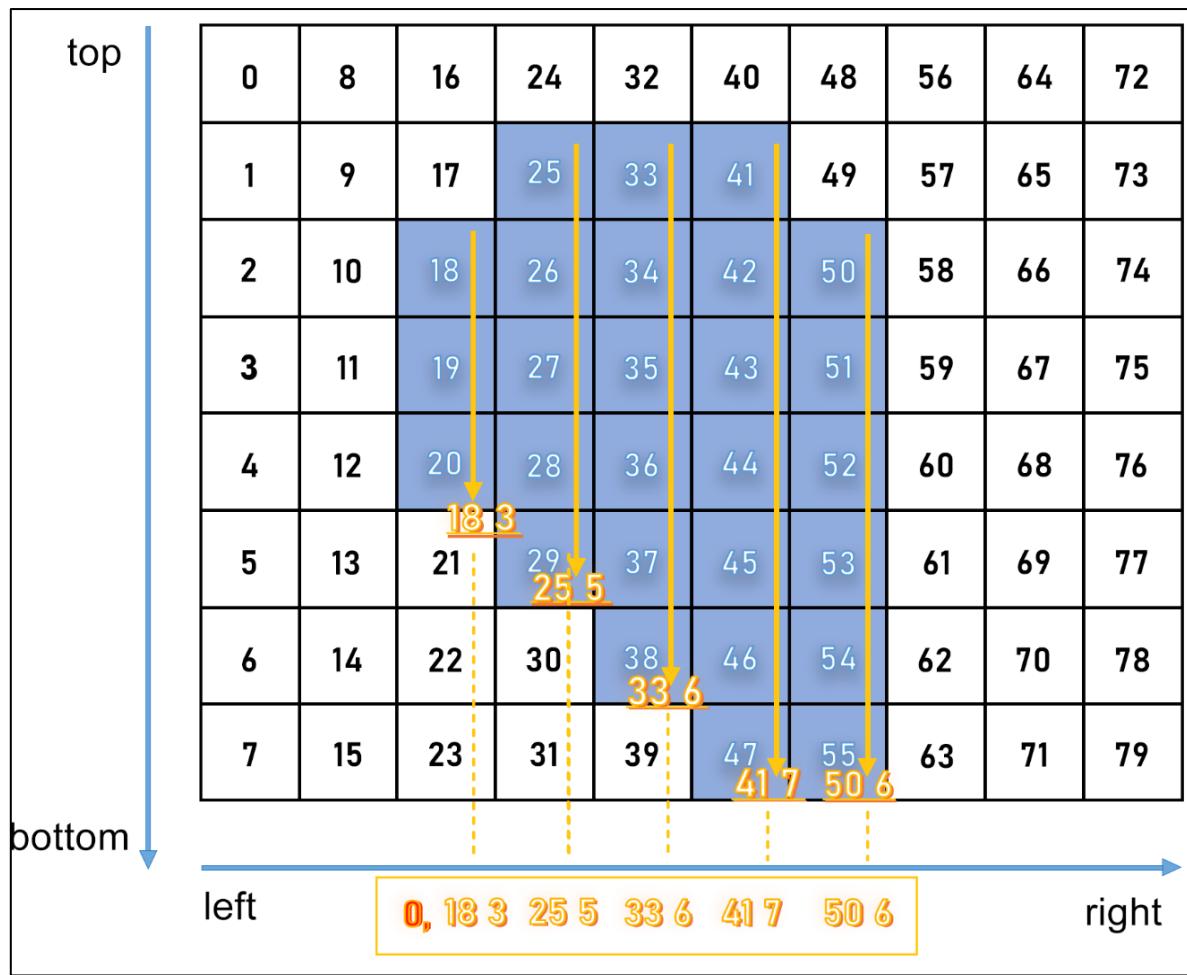






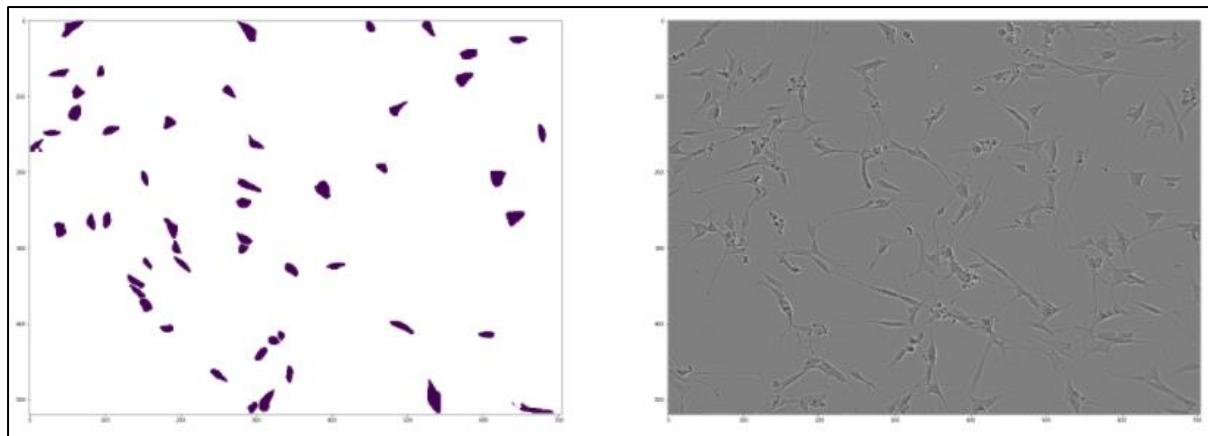
	image_id	width	height	bbox	source
0	b6ab77fd7	1024	1024	[834.0, 222.0, 56.0, 36.0]	usask_1
1	b6ab77fd7	1024	1024	[226.0, 548.0, 130.0, 58.0]	usask_1
2	b6ab77fd7	1024	1024	[377.0, 504.0, 74.0, 160.0]	usask_1





```
[01/06 22:26:36 d2.data.datasets.coco]: Loading ../input/sartorius-annotations/annotations_train_f4.json takes 1.16 seconds.
[01/06 22:26:36 d2.data.datasets.coco]: Loaded 485 images in COCO format from ../input/sartorius-annotations/annotations_train_f4.json
[01/06 22:26:38 d2.data.build]: Removed 0 images with no usable annotations. 485 images left.
[01/06 22:26:38 d2.data.build]: Distribution of instances among all 3 categories:
| category | #instances | category | #instances | category | #instances |
|:-----:|:-----:|:-----:|:-----:|:-----:|:-----|
| shsy5y | 41952 | astro | 8360 | cort | 8556 |
| | | | | | |
| total | 58868 | | | | |
[01/06 22:26:38 d2.data.dataset_mapper]: [DatasetMapper] Augmentations used in training: [ResizeShortestEdge(short_edge_length=(640, 672, 704, 736, 768, 800), max_size=1333, sample_style='choice'), RandomFlip()]
[01/06 22:26:38 d2.data.build]: Using training sampler TrainingSampler
[01/06 22:26:38 d2.data.common]: Serializing 485 elements to byte tensors and concatenating them all ...
[01/06 22:26:38 d2.data.common]: Serialized dataset takes 6.79 MiB
```

model_final_f10217.pkl: 178MB [00:04, 35.8MB/s]



	id	predicted
0	7ae19de7bc2a	139541 4 140244 7 140948 8 141652 8 142356 9 1...
1	7ae19de7bc2a	96418 4 97121 6 97825 7 98529 8 99233 8 99937...
2	7ae19de7bc2a	26627 14 27329 17 28031 19 28733 21 29435 23 3...
3	7ae19de7bc2a	148230 2 148931 6 149633 9 150336 11 151039 13...
4	7ae19de7bc2a	224918 2 225620 7 226324 9 227027 12 227731 13...

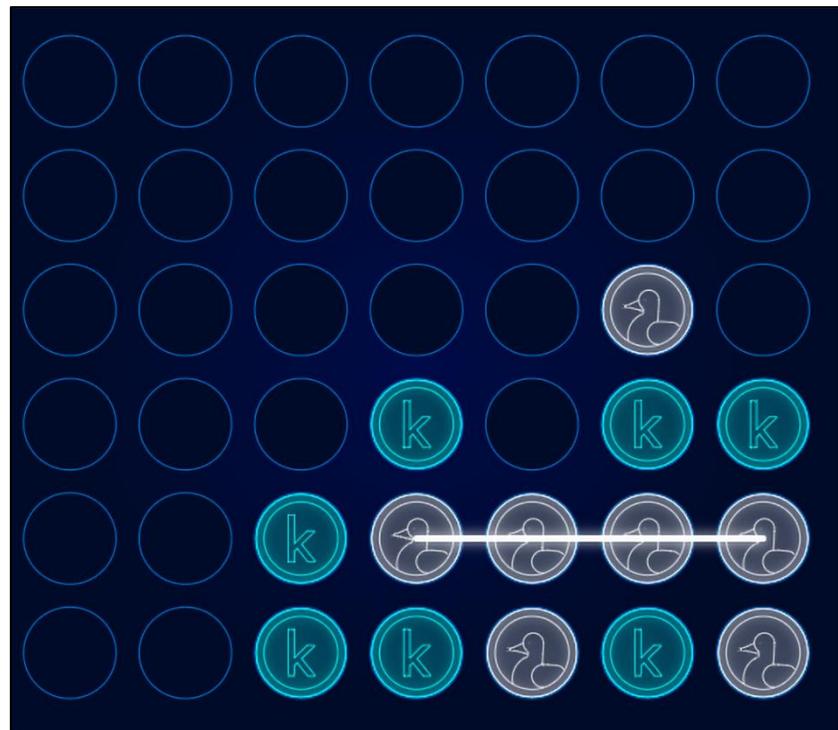
Chapter 11: Modeling for NLP

	textID	text	selected_text	sentiment
0	cb774db0d1	I'd have responded, if I were going	I'd have responded, if I were going	neutral
1	549e992a42	Sooo SAD I will miss you here in San Diego!!!	Sooo SAD	negative
2	088c60f138	my boss is bullying me...	bullying me	negative
3	9642c003ef	what interview! leave me alone	leave me alone	negative
4	358bd9e861	Sons of ****, why couldn't they put them on t...	Sons of ****,	negative

	textID	text	sentiment	predicted_negative	predicted_neutral	predicted_positive
0	f87dea47db	Last session of the day http://twitpic.com/ezh	neutral	0.022949	0.967165	0.009886
1	96d74cb729	Shanghai is also really exciting precisely s...	positive	0.000075	0.012165	0.987760
2	eee518ae67	Recession hit Veronique Branquinho she has to ...	negative	0.993622	0.006364	0.000014
3	01082688c6	happy bday	positive	0.000020	0.005859	0.994122
4	33987a8ee5	http://twitpic.com/wp I like it	positive	0.006184	0.119946	0.873870
5	726e501993	thats great weee visitors	positive	0.000165	0.019434	0.980401
6	261932614e	I THINK EVERYONE HATES ME ON HERE lol	negative	0.916203	0.081649	0.002148
7	afa11da83f	so wish i could but im in school and myspace ...	negative	0.877504	0.116624	0.005871
8	e64208b4ef	and within a short time of the last clue all ...	neutral	0.116272	0.859304	0.024424
9	37bcad24ca	What did you get My day is alright haven't do...	neutral	0.223977	0.756474	0.019550

question_title	question_body	question_user_name	question_user_page	answer
What am I losing when using extension tubes in...	After playing around with macro photography on...	ysap	https://photo.stackexchange.com/users/1024	I just got extension tubes, so here's the skin...
What is the distinction between a city and a s...	I am trying to understand what kinds of places...	russellpierce	https://rpg.stackexchange.com/users/8774	It might be helpful to look into the definitio...
Maximum protusion length for through-hole comp...	I'm working on a PCB that has through-hole com...	Joe Baker	https://electronics.stackexchange.com/users/10157	Do you even need grooves? We make several pro...
Can an affidavit be used in Beit Din?	An affidavit, from what i understand, is basic...	Scimonster	https://judaism.stackexchange.com/users/5151	Sending an "affidavit" it is a dispute between...

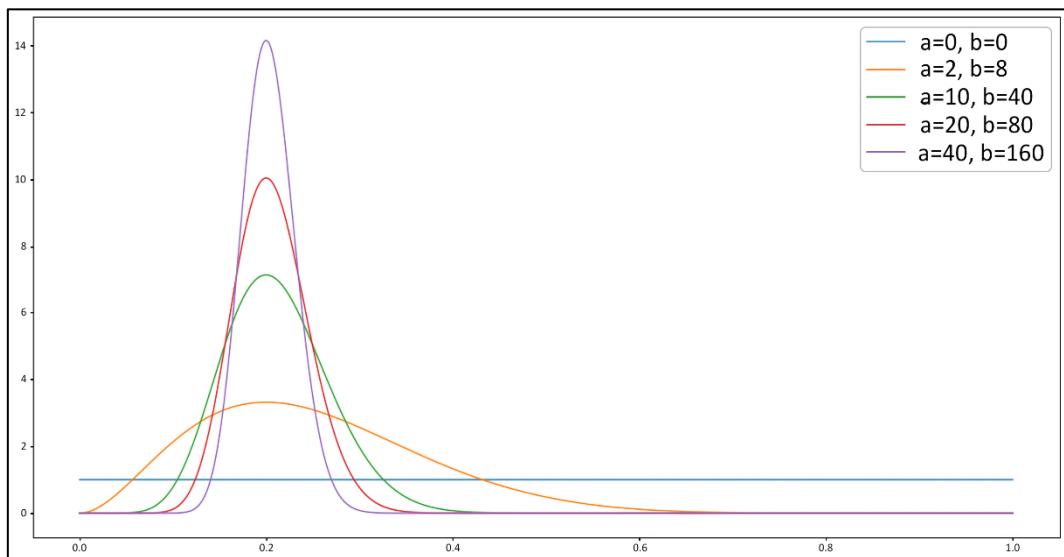
Chapter 12: Simulation and Optimization Competitions



		Rock	Paper	Scissors	
		Rock	0, 0	-1, 1	1, -1
		Paper	1, -1	0, 0	-1, 1
		Scissors	-1, 1	1, -1	0, 0

	Player 1	Player 2
Action:	2	2
Name:	Scissors	Scissors
Icon:	✂️	✂️
Result:	Tie	
Reward:	5	-5

▶ ⏴ 160 / 1000



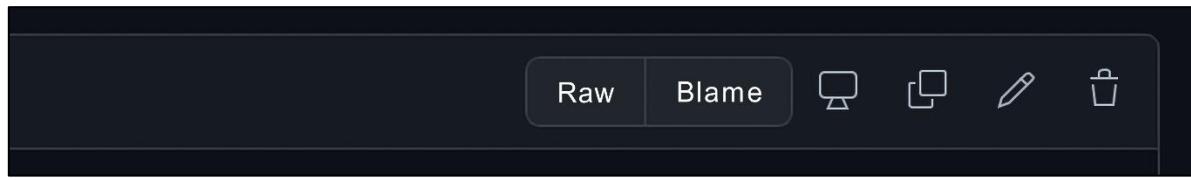
kaggle Left / Right Arrow: Increase / Decrease Step Row Keys: Playback Speed Space: Pause / Play Adversarial Multi-armed Bandit

	Player 1	Player 2
Action:	2	42
Result:		Win
Reward:	261	342

II 681 / 2000



Chapter 13: Creating Your Portfolio of Projects and Ideas



The background of the main content area is a solid blue color with a faint, light-blue grid pattern radiating from the center. In the center of this grid, the word "kaggle" is written in a large, white, lowercase sans-serif font.

Hi lucamassaron!

Calculating word frequency just scratches the surface of natural language processing. In this Snapshots video, Product Manager Meg Risdal walks us through her analysis of [Animal Crossing reviews](#) while exploring the Shifterator package's word shift graphs, an alternative to word clouds. She also provides an overview of the [Quick Save and Version Naming features](#) in Notebooks!