Introduction to Kaggle

Presenter:

Hira Nawaz (002)

Roman Ahsan (003)

Aliza Farooq (004)

Eman Ahsan (019)

Instructor: Dr. Usman Ali

Course Title: TOPL

Department of Computer Science, MSCS (SEM I)
University of Gujrat, Main Campus, Gujrat , Pakistan

Introduction

What is Kaggle?

- Kaggle is the world largest online platform for Data Science, ML and AI.
- Owned by **GOOGLE.**

Used by:

1. Students
2. Researchers
3. Professionals

Why Use Kaggle Platform?

- Provide Free Real world datasets
- Hands on Practice with ML
- Access to High quality Notebooks
- GPU support (free)
- Learn from top experts
- Build a strong portfolio
- Helps in research and thesis work
- Competitions and Challenges
- Join Discussions & Ask questions

How to use Kaggle?

Step 1: Create a Kaggle Account

Step 2: Navigate to datasets

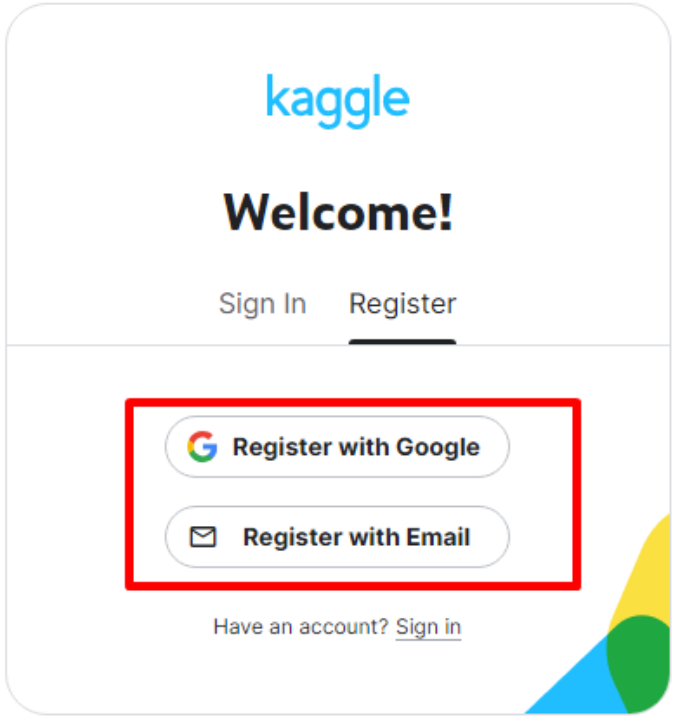
Step 3: Import dataset into a notebook

Step 4: Write basic ML code

Step 5: Train and Test model

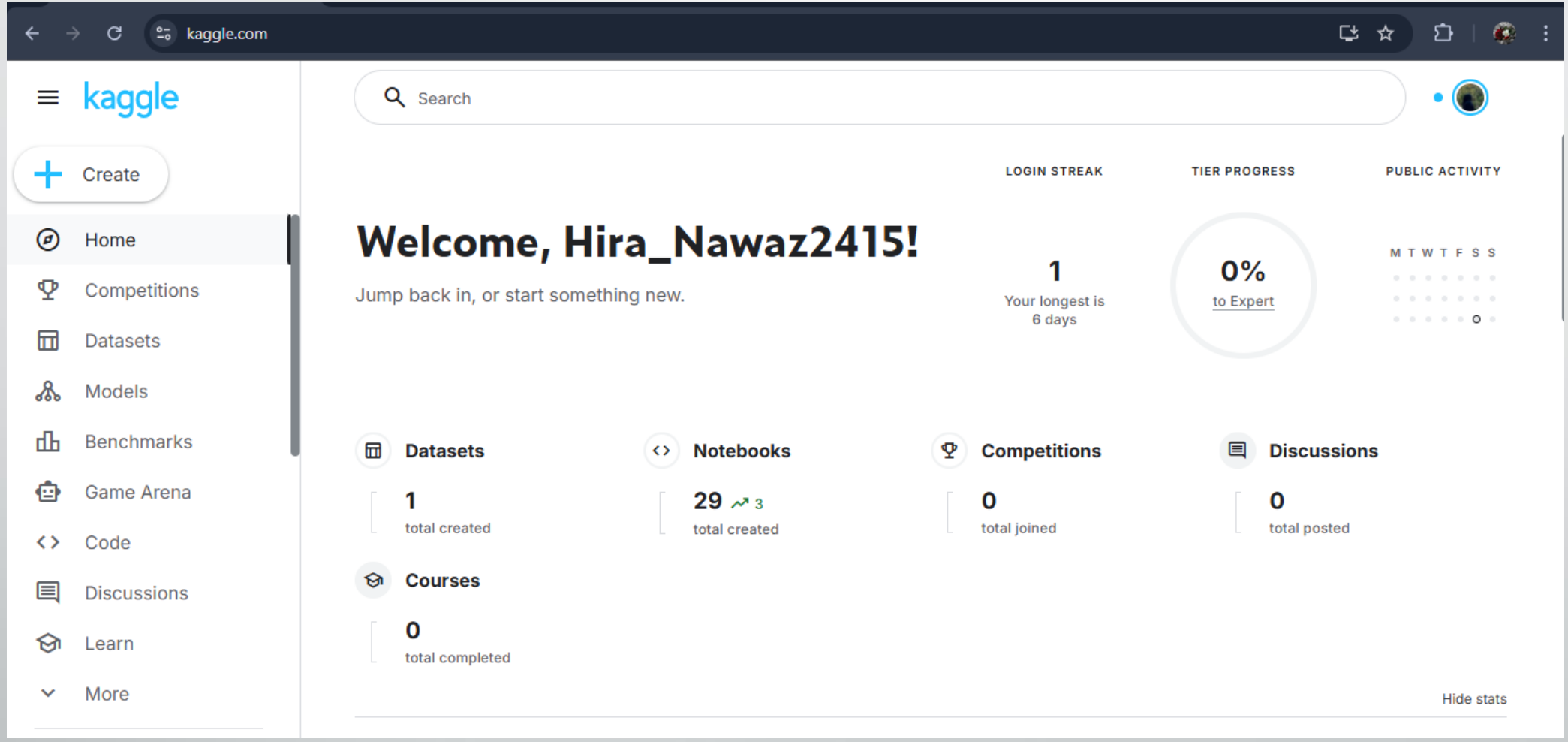
Step 6: Submit Predictions

Step 7: View leaderboard



When you link your Facebook, Google, or Yahoo account, Kaggle collects certain information stored in that account that you have configured to make available. By linking your accounts, you authorize Kaggle to access and use your account on the third party service in connection with your use of kaggle.com.

Contact Us / Support


kaggle.com
kaggle
Search
Create
Home
Competitions
Datasets
Models
Benchmarks
Game Arena
Code
Discussions
Learn
More
LOGIN STREAK
TIER PROGRESS
PUBLIC ACTIVITY
Welcome, Hira_Nawaz2415!
Jump back in, or start something new.
1
Your longest is 6 days
0%
to Expert
M T W T F S S
Datasets
1
total created
Notebooks
29 3
total created
Competitions
0
total joined
Discussions
0
total posted
Courses
0
total completed
Hide stats

What Kaggle Provides?

Kaggle Ecosystem/ Main Components

Datasets	Code/ NoteBooks	Courses	Competitions
Discussion Forums	Profile Progression System	User Ranking System	Models
Benchmarks			

Kaggle Datasets

- Datasets are collections of unstructured and structured data.
- Kaggle hosts a wide variety of public datasets for free
- 100,000 plus datasets across domains healthcare ,Finance, NLP, CV, Iot Sports etc.
- CSV, Images, Text, Audio, Video formats
- Pre labeled, high quality, community shared
- You can also contribute by uploading your own collected datasets

How to use a Dataset?

1. Browse Kaggle Dataset
2. Search/filter by **topic,file type, size, popularity**
3. Download or add dataset to Kaggle Notebook
4. Explore , Preprocess and train Model

Examples:

- Titanic
- MNIST Digits
- Chest X Ray Images
- House Prices Dataset

kaggle

Search

Create

Home

Competitions

Datasets

Models

Benchmarks

Game Arena

Code

Discussions

Learn

More

Datasets

Explore, analyze, and share quality data. Learn more about data types, creating, and collaborating.

+ New Dataset

Your Work

Search datasets

Filters

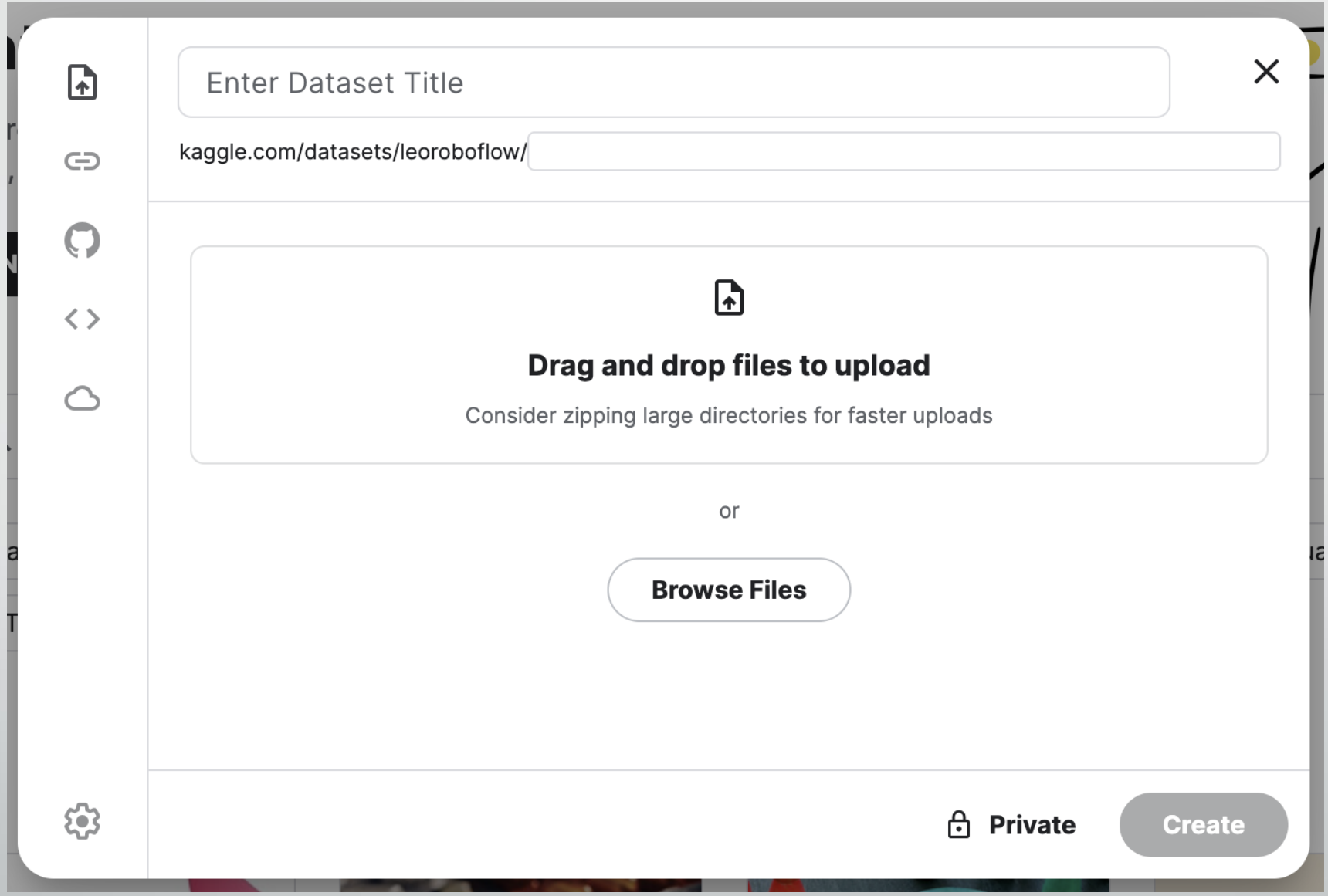
All datasets | Computer Science | Education | Classification | Computer Vision | NLP | Data Visualization | Pre-Trained Model

Trending Datasets

See All

How to upload your own dataset own Kaggle?

1. Click on **Create** Button.
2. Click on " Plus New Dataset' on Kaggle
3. Upload files (CSV,JSON, images..etc)
4. Add **title, description, tags, license**
5. Publish
6. Click on **Make Public** to make it public.


Enter Dataset Title
kaggle.com/datasets/leoroboflow/
Drag and drop files to upload
Consider zipping large directories for faster uploads
or
Browse Files
Private
Create

Meta Kaggle

- Official **dataset of Kaggle itself**
- Contains complete data about:
  - Competitions
  - Datasets
  - Notebooks
  - Users & Rankings
  - Medals and achievements
  - Discussion posts
- Useful for **research, analysis, and visualization**
- Updated regularly
- Great for studying **Kaggle trends & user behavior**
- Accessible as a **public dataset** on Kaggle

Code/NoteBooks

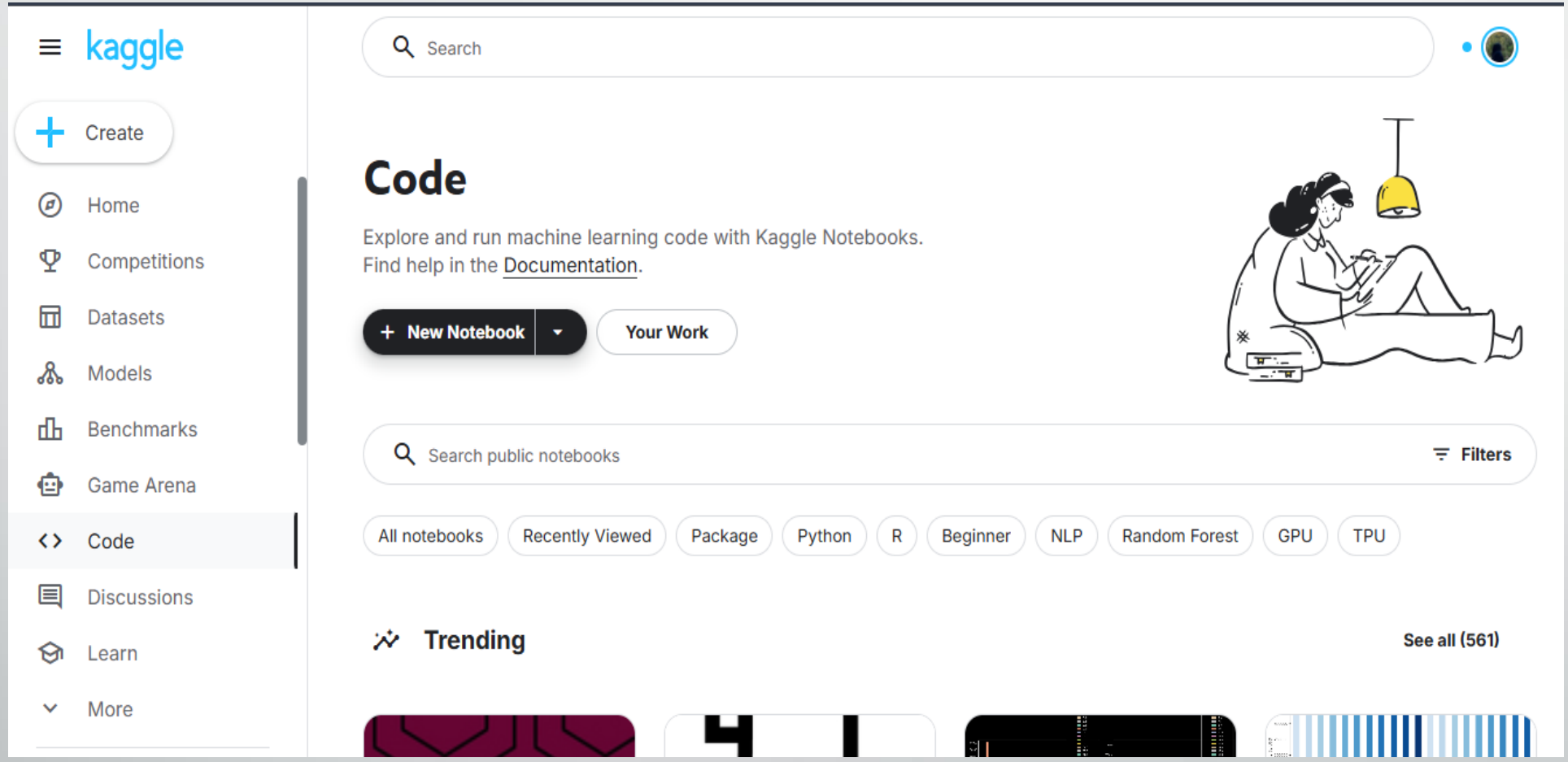
- Cloud based **Jupyter Notebook Environment**
- Supports **Python & R**
- Free **GPU/TPU** for deep learning
- Run, edit and share code online
- Import dataset directly from Kaggle
- Easy Collaboration: fork ,comment and version control
- No installation needed

Features:

- Notebook templates
- Auto save & versioning
- Add datasets, models and script with one click

How to create Notebook?

- Go to "**Code**" → **New Notebook**
- Attach dataset → start coding


kaggle
Search
Create
Home
Competitions
Datasets
Models
Benchmarks
Game Arena
Code
Discussions
Learn
More
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
Package
Python
R
Beginner
NLP
Random Forest
GPU
TPU
Trending
See all (561)

Courses

- Free, beginner-friendly **micro-courses**
- Short, practical lessons with hands-on coding
- Covers essential ML & data skills
- Learn at your own pace
- Certificates available after completion

Benefits:

- Practical exercises
- Real datasets included
- Easy for students & researchers
- Great for building foundational skills

Popular Courses:

- Python
- Pandas
- Data Visualization
- Machine Learning
- Deep Learning
- Computer Vision
- Natural Language Processing (NLP)
- SQL
- Intro to AI/ML Ethics

Kaggle Competitions

- Compete with global data scientists
- Participate and it is a way to validate your knowledge
- Team can be formed and compete with other teams
- Real-world machine learning challenges
- Submit predictions → get ranked on leaderboard
- Win medals (Bronze, Silver, Gold)
- Improve skills through hands-on practice

Types of Competitions

1. **Featured:** Sponsored, high rewards
2. **Research:** Academic + experimental tasks
3. **Recruitment:** Win prizes + get job opportunities
4. **Getting Started:** Beginner-friendly (Titanic, Digit Recognizer)
5. **Playground:** Fun, intermediate-level tasks

How to Join a Competition

1. Go to **Competitions** tab
2. Select a challenge
3. Read problem → download data
4. Train model in Notebook
5. Submit predictions (CSV)
6. Track position on leaderboard

Competition flow diagram

Problem → Model → Submission → Leaderboard

Kaggle Benchmarks

- A **benchmark is a reference model or score**
- Shows **how well a model should perform** on a dataset
- Helps you **compare your model** with others
- Provides a **starting point** for improvement
- Often included in Kaggle competitions as a **baseline**

Models

- Kaggle Models is a repository of pre-trained models that are deeply integrated with Kaggle's platform making them easy to use in Kaggle Competitions and Notebooks
- A **model is a machine learning program** that learns from data
- Takes input (data) → predicts output (answers)
- Trained on **Kaggle datasets** in Notebooks
- Can be **shared** or **used for competitions**
- Supports popular tools like **Python, TensorFlow, PyTorch, scikit-learn**

Dataset → Model → Prediction → Evaluation

Kaggle Discussion Forum (Community)

- Largest data science discussion community
- Post your doubts
- Ask questions & get help from experts
- Share ideas, insights and solutions
- Learn from top Kagglers' explanations
- Participate in topic-wise threads (ML, DL, NLP, CV, Data Cleaning)
- Get guidance on competitions and datasets
- Collaborate with peers worldwide
- Upvote, comment, and follow discussions
- Very Active Community
- Boost your profile if you help others

Benefits of Kaggle Discussions

- Learn from real-world ML conversations
- Get instant help when stuck
- Discover new techniques
- Build your network & reputation

User Ranking System

Kaggle Ranks (Levels):

1. **Novice**
2. **Contributor**
3. **Expert**
4. **Master**
5. **Grandmaster** (highest level)

How Ranking Works:

- Earn **medals** by performing well
- Medals increase your **Kaggle profile score**
- Higher ranks = more achievements + community reputation

Medals Categories:

- **Competitions:** Gold, Silver, Bronze
- **Notebooks:** Gold, Silver, Bronze
- **Datasets:** Gold, Silver, Bronze
- **Discussion:** Gold, Silver, Bronze

What Affects Rank:

- Quality of your notebooks
- Useful datasets you upload
- High-ranking competition submissions
- Helpful answers in discussions

Kaggle Performance Tiers

Novice

You've joined the community.

- [x] Register!

Contributor

You've completed your profile, engaged with the community, and fully explored Kaggle's platform

- [x] Run 1 notebook or script
- [x] Make 1 competition or task submission
- [ ] Make 1 comment
- [ ] Give 1 upvote

Competitions

Experts

- 2 bronze medals

Master

- 1 gold medal
- 2 silver medals

Grandmaster

- 5 gold medals
- Solo gold medal

	0-99 Teams	100-249 Teams	250-999 Teams	1000+ Teams
Bronze	Top 40%	Top 40%	Top 100	Top 10%
Silver	Top 20%	Top 20%	Top 50	Top 5%
Gold	Top 10%	Top 10	Top 10 + 0.2%*	Top 10 + 0.2%*

Kaggle Performance Tiers

Discussions

Bronze	5 Votes
Silver	20 Votes
Gold	50 Votes

Experts

You would start getting ranked

50 bronze medals

Master

You can participate in master-only competition

50 silver medals

200 medals in total

Grandmaster

You are best of the best

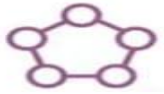
50 gold medals

500 medals in total

Kaggle Performance Tiers

Datasets

Bronze	1 Vote
Silver	5 Votes
Gold	10 Votes



Experts

You would start getting ranked

3 bronze medals

Master

You can participate in master-only competition

1 gold medal

4 silver medals

Grandmaster

You are best of the best

5 gold medals

5 silver medals

Discussion Medals

Discussion Medals are awarded to popular topics and comments posted across the site, as measured by net votes (upvotes minus downvotes). Not all upvotes count towards medals: votes by novices and votes on old posts are excluded from medal calculation.

Bronze	1 Vote
Silver	5 Votes
Gold	10 Votes

Profile Progression System

How Your Profile Grows:

- Earn **medals** (competitions, notebooks, datasets, discussions)
- Gain **ranking points** that boost your profile level
- Improve **skills** by completing courses
- Increase visibility through high-quality contributions
- Build reputation via comments, votes & collaborations

Progress Depends On:

- Quality of shared work
- Consistency of activity
- Engagement with community discussions
- Performance in competitions

Kaggle Profile

Awesome Data Scientist

Awesome Data Scientist at Kaggle

San Francisco, California, United States

Joined 2 years ago · last seen in the past day

Followers 2268

Following 6434

Notebooks Grandmaster

Home | Competitions (49) | Datasets (19) | Code (189) | Discussion (613) | Followers (2,268) | Notifications | ••• | Edit Profile

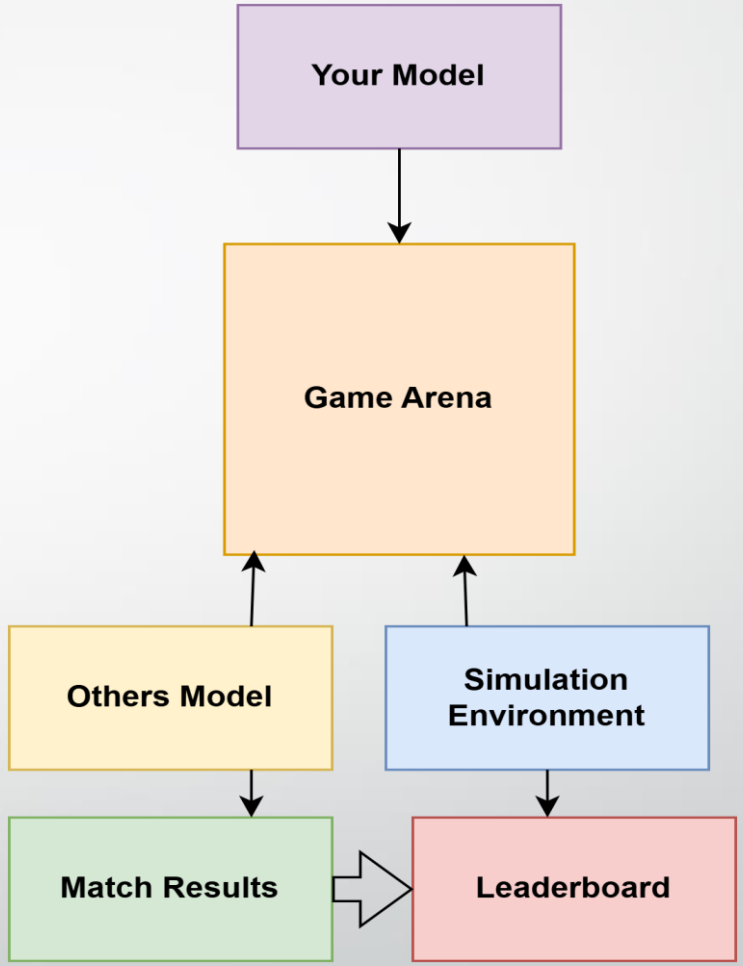
Competitions Master	Datasets Expert	Notebooks Grandmaster	Discussion Expert
Current Rank: 6263 of 159,199	Current Rank: 44 of 35,417	Rank: 2 of 163,650	Current Rank: 558 of 190,474
Highest Rank: 94	Highest Rank: 21		Highest Rank: 132
1 · 3 · 2	2 · 2 · 10	42 · 3 · 27	9 · 17 · 204

Game Arena

- Kaggle **GameArena** is where **AI models compete against each other**
- Used in **reinforcement learning** and AI competitions
- Participants submit **agents or models** that play games or solve challenges
- Tracks results on **leaderboards**
- Helps students **learn AI strategies and improve models**

Agent → Game → Leaderboard

- Put your ML model in a **competitive playground**.
- Compete **against other models** or a **simulation**.
- See **real-time wins/losses**, not just static scores.
- Great for **learning, testing, and improving models dynamically**.



How to import kaggle Notebook to Colab?

Method 1: Download Notebook from Kaggle

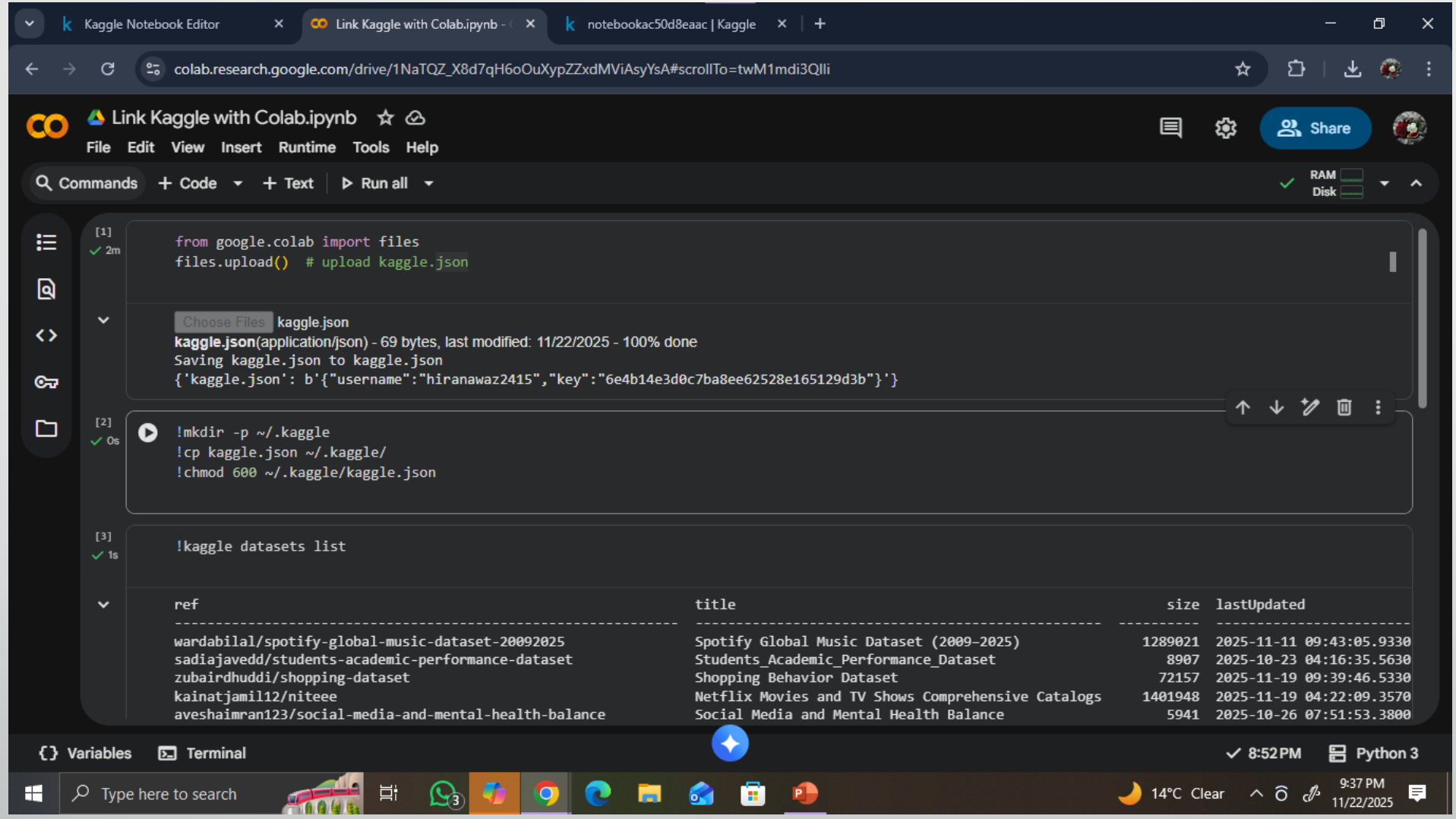
- Open your Kaggle Notebook
- Click **File** → **Download** → **.ipynb**
- Upload the .ipynb file directly to **Google Colab**

Method 2: Use GitHub as a Bridge

- Save Kaggle notebook to **GitHub**
- Open the .ipynb file in Colab using **Open in Colab** option

Method 3: Load Kaggle Dataset in Colab

- Upload your kaggle.json key
- Install Kaggle API in Colab
- Download datasets directly inside Colab


Kaggle Notebook Editor
Link Kaggle with Colab.ipynb -
notebookac50d8eaac | Kaggle
colab.research.google.com/drive/1NaTQZ_X8d7qH6oOuXypZZxdMViAsyYsA#scrollTo=twM1mdi3QIIi
Link Kaggle with Colab.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text Run all
RAM
Disk
[1]
2m
from google.colab import files
files.upload() # upload kaggle.json
Choose Files kaggle.json
kaggle.json(application/json) - 69 bytes, last modified: 11/22/2025 - 100% done
Saving kaggle.json to kaggle.json
{'kaggle.json': b'{"username":"hiranawaz2415","key":"6e4b14e3d0c7ba8ee62528e165129d3b"}'}
[2]
0s
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle/
!chmod 600 ~/.kaggle/kaggle.json
[3]
1s
!kaggle datasets list
ref title size lastUpdated
wardabilal/spotify-global-music-dataset-20092025 Spotify Global Music Dataset (2009-2025) 1289021 2025-11-11 09:43:05.9330
sadiajavedd/students-academic-performance-dataset Students_Academic_Performance_Dataset 8907 2025-10-23 04:16:35.5630
zubairdhuddi/shopping-dataset Shopping Behavior Dataset 72157 2025-11-19 09:39:46.5330
kainatjamil12/niteee Netflix Movies and TV Shows Comprehensive Catalogs 1401948 2025-11-19 04:22:09.3570
aveshaimran123/social-media-and-mental-health-balance Social Media and Mental Health Balance 5941 2025-10-26 07:51:53.3800
Variables Terminal
8:52 PM Python 3
Type here to search
14°C Clear
9:37 PM
11/22/2025

How to link kaggle with Github?

Linking **Kaggle with GitHub** can mean two things either

1. **Pushing Kaggle notebooks to GitHub**

2. **using GitHub code/data in Kaggle notebooks**.

How to link Kaggle Notebook with Github repository?

Direct Method:

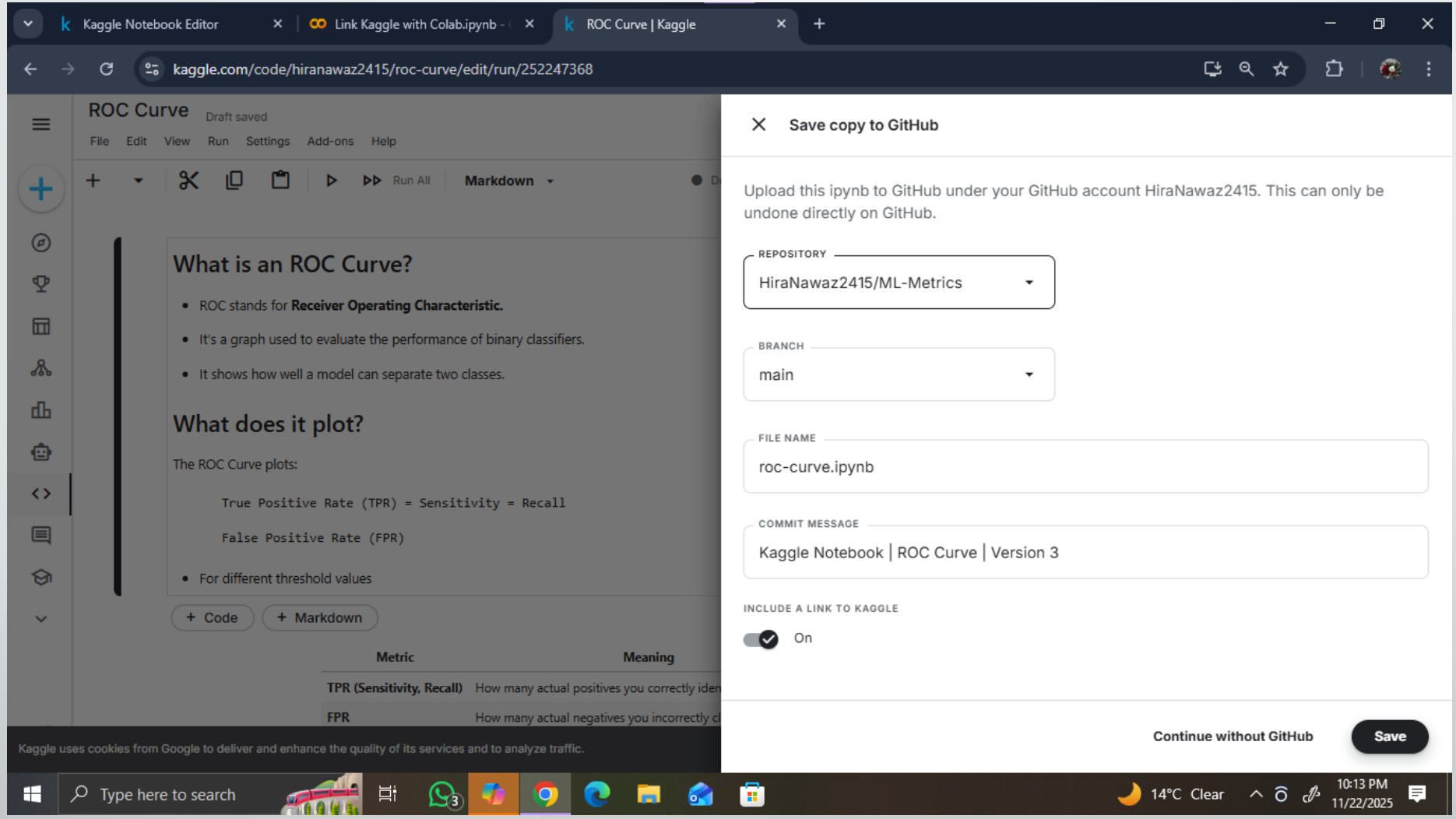
- Open your Kaggle Notebook
- Click **File → Save Version**
- Under "Save Version", choose **"Copy to GitHub"**
- Connect your GitHub account (first time only)
- Select repository + branch
- Click **Publish**

What You Can Upload:

- Notebook (.ipynb)
- Dataset files used in the notebook
- Output files
- Environment metadata

Benefits:

- Automatic version control
- Share work easily
- Collaborate with teams
- Backup your notebook safely


Kaggle Notebook Editor
Link Kaggle with Colab.ipynb -
ROC Curve | Kaggle
kaggle.com/code/hiranawaz2415/roc-curve/edit/run/252247368
ROC Curve
Draft saved
File Edit View Run Settings Add-ons Help
Run All
Markdown
What is an ROC Curve?
ROC stands for Receiver Operating Characteristic.
It's a graph used to evaluate the performance of binary classifiers.
It shows how well a model can separate two classes.
What does it plot?
The ROC Curve plots:
True Positive Rate (TPR) = Sensitivity = Recall
False Positive Rate (FPR)
For different threshold values
+ Code
+ Markdown
Metric
Meaning
TPR (Sensitivity, Recall)
How many actual positives you correctly iden
FPR
How many actual negatives you incorrectly c
Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic.
Save copy to GitHub
Upload this ipynb to GitHub under your GitHub account HiraNawaz2415. This can only be undone directly on GitHub.
REPOSITORY
HiraNawaz2415/ML-Metrics
BRANCH
main
FILE NAME
roc-curve.ipynb
COMMIT MESSAGE
Kaggle Notebook | ROC Curve | Version 3
INCLUDE A LINK TO KAGGLE
On
Continue without GitHub
Save
Type here to search
14°C Clear
10:13 PM
11/22/2025

Clone the GitHub repo in Kaggle Notebook

Step 1: Open Kaggle → Notebooks

- Go to: https://www.kaggle.com/code
- Click **"New Notebook" → "Import Notebook"**

Step 2: Paste GitHub URL

- Paste your repository's **raw .ipynb URL** or the full repo URL
- Kaggle will list all notebooks in that repo.
- Choose the one you want and click **Import**.

Alternative: Clone the GitHub repo in Kaggle Notebook

- If you want the **entire repository,** you can use git clone:
- !git clone https://github.com/HiraNawaz2415/Assignment-1-TOPL-Roll-002.git
- This will create a folder in Kaggle's working directory:
- /kaggle/working/Assignment-1-TOPL-Roll-002/

Load files from the repo

- For example, if there's a CSV in the repo:
- import pandas as pd
- df = pd.read_csv('/kaggle/working/Assignment-1-TOPL-Roll-002/data.csv')
- df.head()

Conclusion

- **Kaggle** is the world's largest platform for **data science and machine learning**
- Provides **datasets, competitions, notebooks, and courses**
- Offers **hands-on practice** for students, researchers, and professionals
- Community and discussion forums **enhance learning and collaboration**
- **User ranking & profile system** motivates growth and skill development
- Useful for **thesis, research, and portfolio building**
- Free access to **GPU/TPU** and practical ML resources

Final Thought:

Kaggle = **Learn + Practice + Compete + Collaborate + Research**

Questions & Answers?

Thanks to all of You for listening

God bless all of you more and keep success.

“Just work hard and put your trust on Allah Almighty.”