

はじめてのAIプログラム学習キット5 Kaggle編

kaggle



スペクトラム・テクノロジー株式会社

<https://spectrum-tech.co.jp>

sales1@spectrum-tech.co.jp

学習キット 目次

	ページ
• 利用にあたっての注意事項	<u>3</u>
• kaggle	
1. 概要	<u>4</u>
2. 設定	
① 新規登録	<u>5</u>
② プロフィール・アカウント作成	<u>6</u>
3. メニュー説明	
① ホーム	<u>8</u>
② コンペ	<u>9</u>
③ データセット	<u>10</u>
④ コード	<u>11</u>
⑤ 議論	<u>12</u>
⑥ コース	
4. Code使用方法	
① Notebook作成	<u>13</u>
② Notebook使用例	<u>15</u>
5. Competition使用方法	<u>19</u>
6. dataset使用方法	<u>22</u>
7. Discussions使用方法	<u>24</u>
8. 事例	<u>25</u>
9. コンペに出展	<u>27</u>

利用にあたっての注意事項

① 記載内容

- Kaggleについて、弊社独自で翻訳、解釈しているため、間違いがあるかもしれません。正しい情報については、kaggle.comを参照ください。

② 免責事項

- 本マニュアルの操作により、重大な故障等が発生した場合においても、弊社は責任を負いません。内容を利用者が把握して操作するものとします。
- 本マニュアルの操作により、個人情報が出た場合、利用者の責によるものとします。

③ 問い合わせ

- 本マニュアルにおける、問い合わせについては、お受けできません。各利用者が調査して対応してください。但し、有料サポートを利用の場合は、この限りではありません。

④ 弊社AIプログラム学習キット5紹介

- https://spectrum-tech.co.jp/products/ai_tensorflow.html
- Pi版、グラフィック版など各種あります。事例は、その一部のものであります。本格的python, tensorflow, pytorchなど学習したい場合は、是非ご利用ください。

kaggle

1. 概要

① 概要

- Kaggle(カグル)は企業や研究者がデータを投稿し、世界中の統計家やデータ分析家とその最適モデルを競い合う、予測モデリング及び分析手法関連プラットフォーム及びその運営会社である。2017年googleが買収

② 利用方法

- Github: <https://github.com/Kaggle/kaggle-api>
- 公式サイトにアクセスして、Webで利用。使い方はjupyter notebookと同じ。また、GPU(時間制限あり)が準備されている。RTX2080よりは遅い。
- 英語環境のみ
- また、テストデータをダウンロードしてローカル環境でも動作

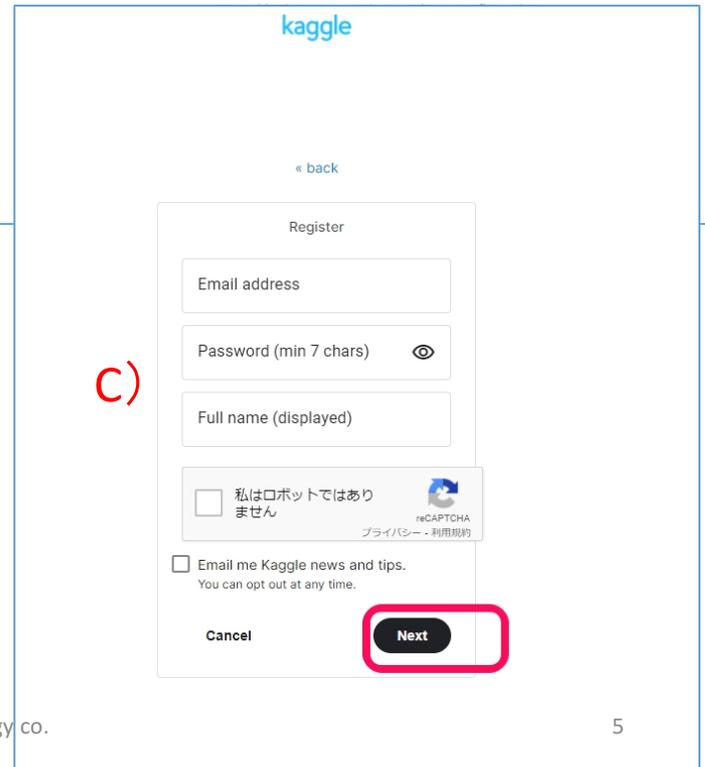
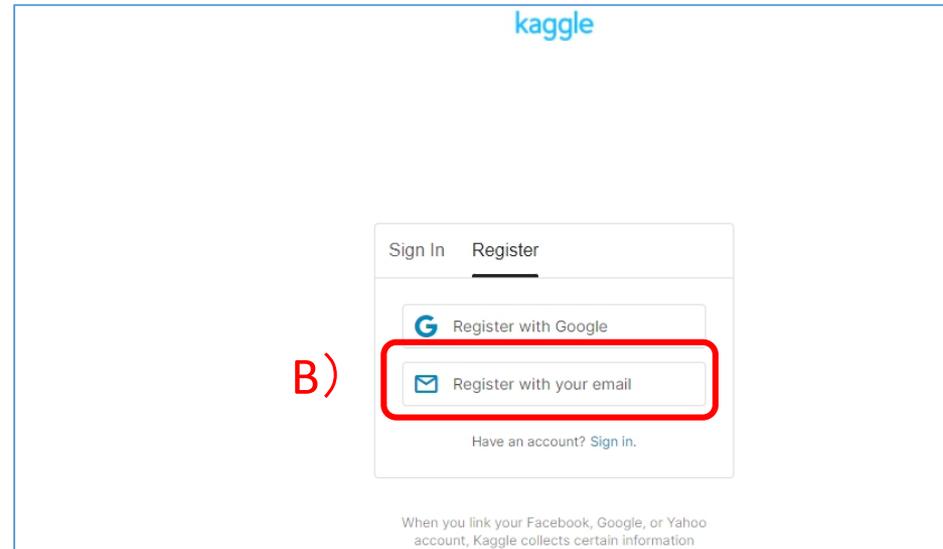
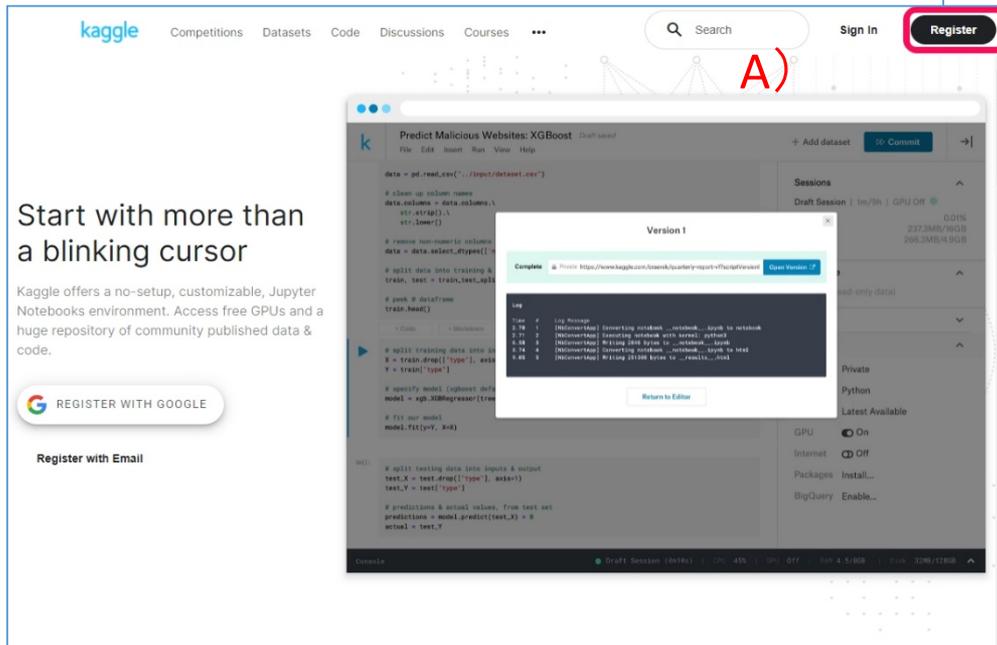
③ 必要なスキル

- Pythonのプログラムを作成できるスキル(中級)
- TensorflowなどのML(Machine learning)の基礎的なスキル(初級)
- 技術英語を読み書きできるスキル(翻訳ツールの活用)

kaggle 2. 設定

① 新規登録

- A) Register: アカウント登録
- B) Id:メール
- C) パスワード、ユーザ名
- D) メール認証、完了(写真省略)



kaggle 2. 設定

② プロフィール・アカウント作成

A) プロフィールなどのメニュー

B) プロフィール

- 写真、会社、住所、ホームページなど公開してもよい情報を記入

The image consists of three overlapping screenshots from the Kaggle website. The leftmost screenshot shows a notebook titled 'Heart Disease Report' by Ching-Chun Yeh, featuring a pie chart with four segments: 'sex_1,target_0' (37.6%), 'sex_0,target_1' (23.8%), 'sex_1,target_1' (30.7%), and 'sex_0,target_0' (7.9%). The middle screenshot shows a user profile menu with 'Your Profile' highlighted by a blue arrow. The rightmost screenshot shows a user profile for 'mas', a developer at spectrum technology co. in Tokorozawa, Saitama, Japan. A red arrow points to the 'profile' link in the top right corner of the profile page.

kaggle 2. 設定

② プロフィール・アカウント作成

A) プロフィールなどのメニュー

B) プロフィール

- 写真、会社、住所、ホームページなど公開してもよい情報を記入

C) アカウント

- GPUを利用するためのSMS認証があります。

The screenshot shows the Kaggle user interface. On the left, a sidebar menu is visible with the following items: 'Your Profile', 'Account', 'Sign Out', 'Your accelerator quota', and 'Your notifications'. A red arrow points from the 'Account' menu item to the 'Account' tab in the main content area. The main content area displays the user's profile information, including their name 'ma 08', bio 'developer at spectrum technology co.', location 'Tokorozawa, Saitama, Japan', and a link to their website 'https://spectrum-tech.co.jp'. Below the profile information, the 'Account' settings are shown, including 'User Name', 'Email Address', 'Phone Verification', and 'Email Preferences'. A red arrow points from the 'Account' tab to the 'Account' settings section. A red arrow also points from the 'Account' settings section to the 'Phone Verification' status, which is 'Verified'. A red arrow points from the 'Phone Verification' status to the text 'gpu使用する場合にSMS認証があり。'. A red arrow points from the 'Account' settings section to the text 'account'. A red arrow points from the 'Account' settings section to the text 'GPUの残り時間'. A red arrow points from the 'Account' settings section to the text 'メニュー'. A red arrow points from the 'Account' settings section to the text 'Edit Public Profile'. A red arrow points from the 'Account' settings section to the text 'all rights reserved 2021 spectrum technology co.'. A red arrow points from the 'Account' settings section to the text '7'.

ma 08

Account

GPUの残り時間

メニュー

account

gpu使用する場合にSMS認証があり。

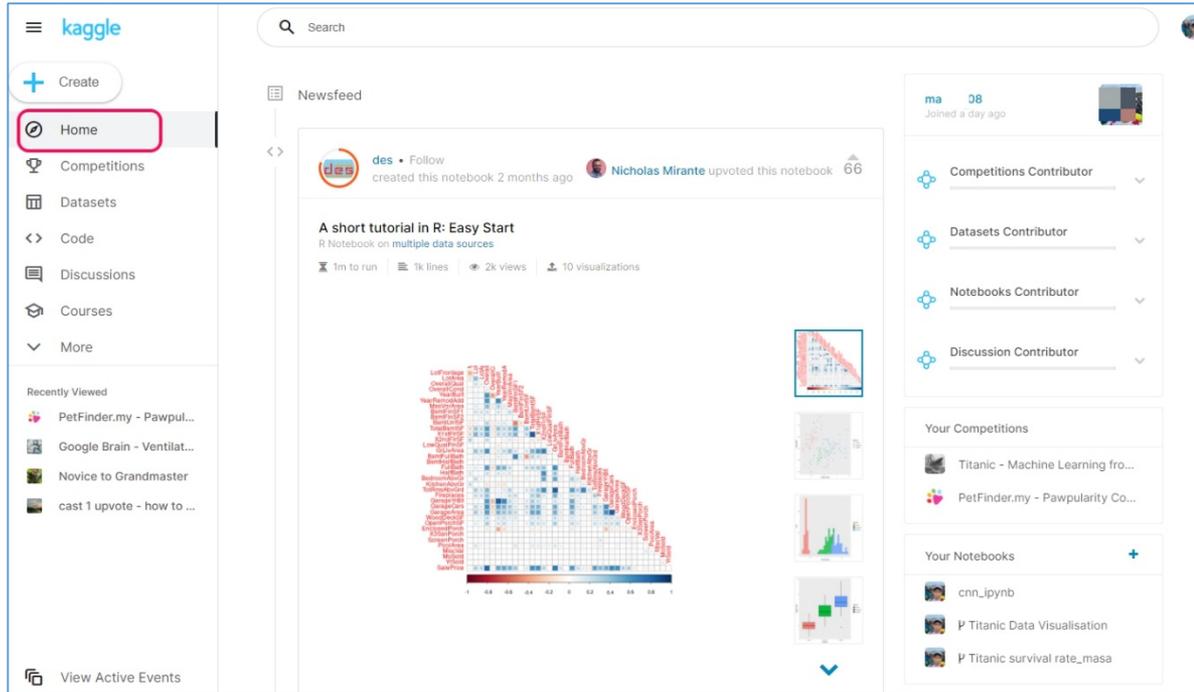
all rights reserved 2021 spectrum technology co.

7

kaggle

3. メニュー説明

- ① ホーム
 - 右側に個人のメニュー
- ② コンペ
- ③ データセット
- ④ Code
- ⑤ 議論
- ⑥ コース



kaggle

3. メニュー説明

① ホーム

② コンペ

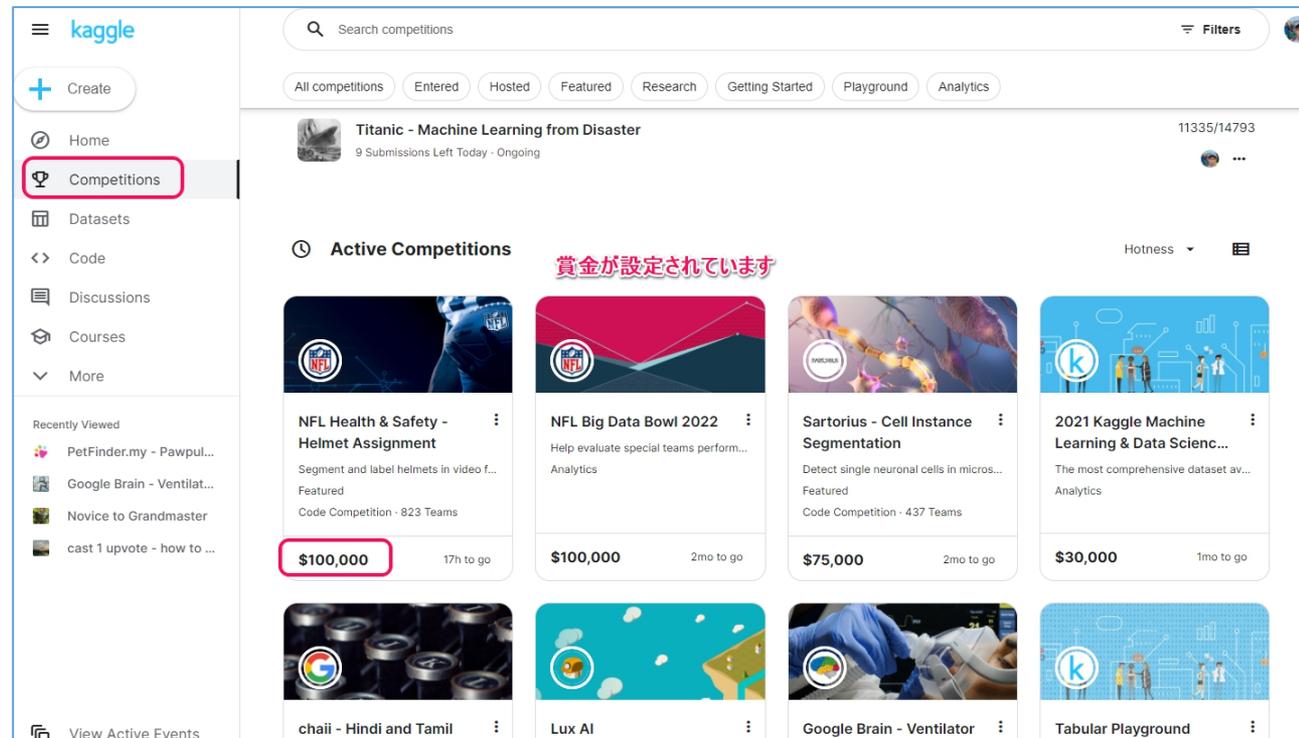
- 企業から賞金付きの課題が出されて、正確さなどでランク付けされます。モデルの作成がカギを握ります。

③ データセット

④ Code

⑤ 議論

⑥ コース



kaggle

3. メニュー説明

① ホーム

② コンペ

③ データセット

- 有名なmnistなどのデータが準備されています。個別にアップロードする必要なし

④ Code

⑤ 議論

⑥ コース

The screenshot shows the Kaggle website interface. On the left sidebar, the 'Datasets' menu item is highlighted with a red box. The main content area displays the 'Datasets' page, which includes a search bar, a 'New Dataset' button, and a 'Your Work' button. Below these are filters for various categories like 'Tasks', 'Computer Science', 'Education', 'Classification', 'Computer Vision', 'NLP', and 'Data Visualization'. A 'Trending Datasets' section is visible at the bottom, featuring four dataset cards: 'WTF Visualizations Dataset', 'NBA All Star Games 1951-2021', 'Forbes Celebrity 100 1999-2020', and 'HRV OFFICE ENVIRONMENT'. Each card shows the dataset name, creator, update time, usability score, size, and file format. At the bottom of the page, there is a footer that reads 'all rights reserved 2021 spectrum technology co.' and a 'View Active Events' button.

kaggle

3. メニュー説明

- ① ホーム
- ② コンペ
- ③ データセット
- ④ Code

- Jupyter notebookのようなものが準備されており、pythonのコードを入力します。
- また、ほとんどのモジュールはインストールされているのでpipなどの操作不要

- ⑤ 議論
- ⑥ コース

The screenshot shows the Kaggle website interface. On the left is a sidebar with navigation options: Home, Competitions, Datasets, Code (highlighted with a red box), Discussions, Courses, and More. Below these are 'Recently Viewed' items. The main content area is titled 'Code' and includes a search bar, a 'New Notebook' button, and a 'Your work' button. Below this is another search bar for public notebooks and a row of filter buttons for various topics like Python, R, Beginner, NLP, Finance, Random Forest, GPU, TPU, and Competition notebook. At the bottom, there is a 'Trending' section with several notebook thumbnails, including one titled '[TPS-11] NN Baseline (Keras)' and another 'Titanic Prediction - ANN'.

kaggle

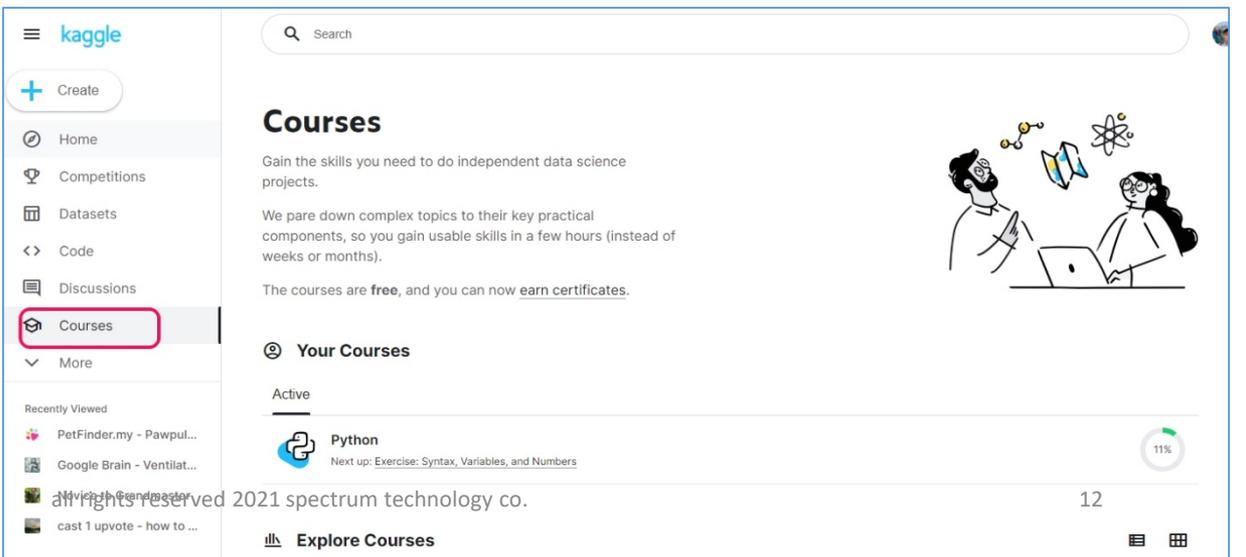
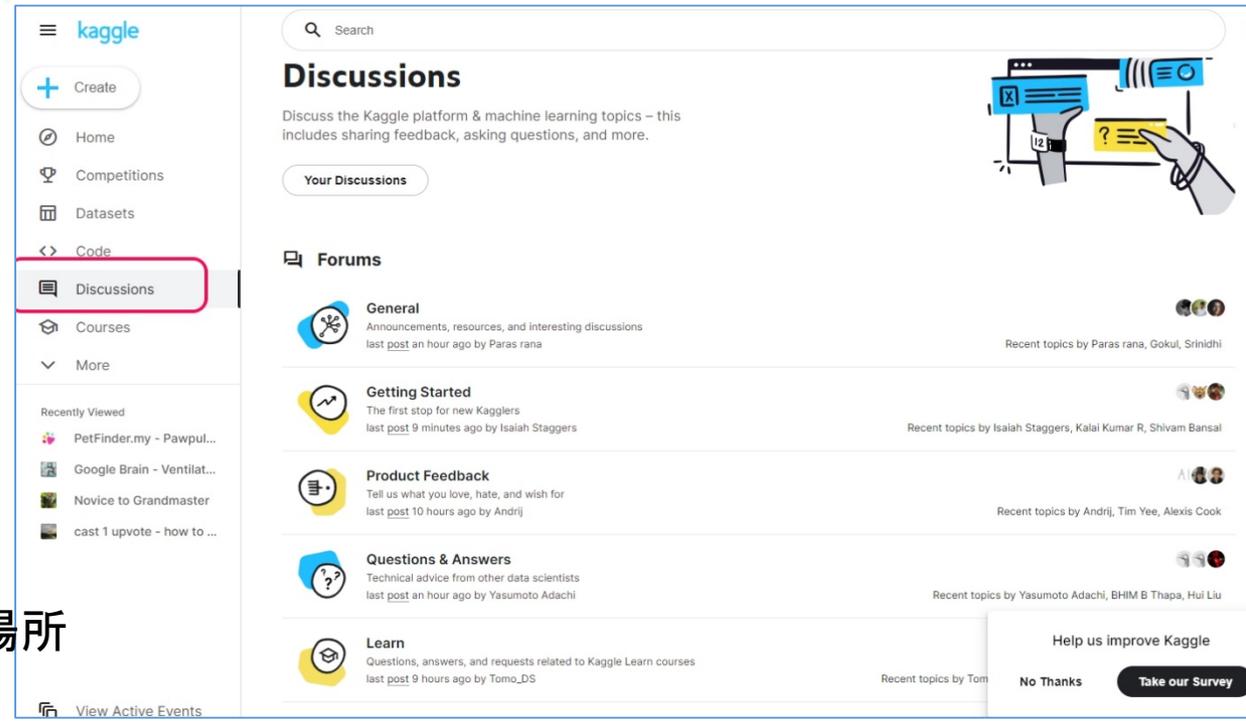
3. メニュー説明

- ① ホーム
- ② コンペ
- ③ データセット
- ④ Code
- ⑤ 議論

• コミュニティのような場所

- ⑥ コース

• 学習コースが準備

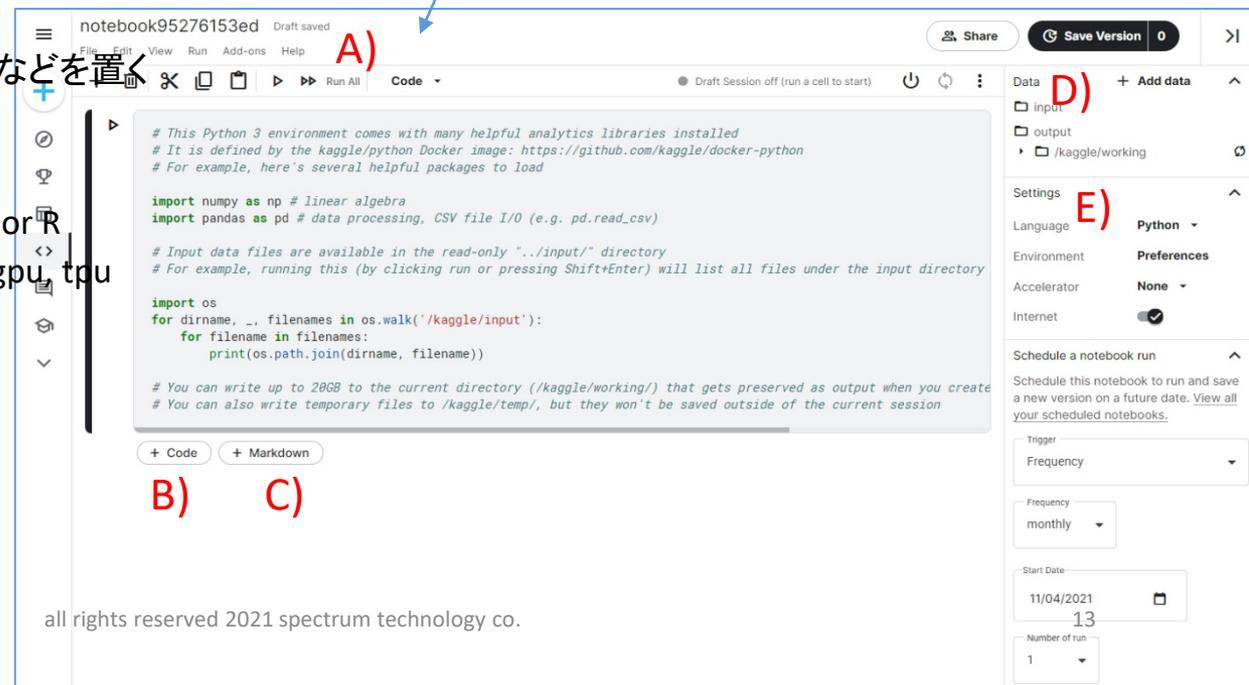
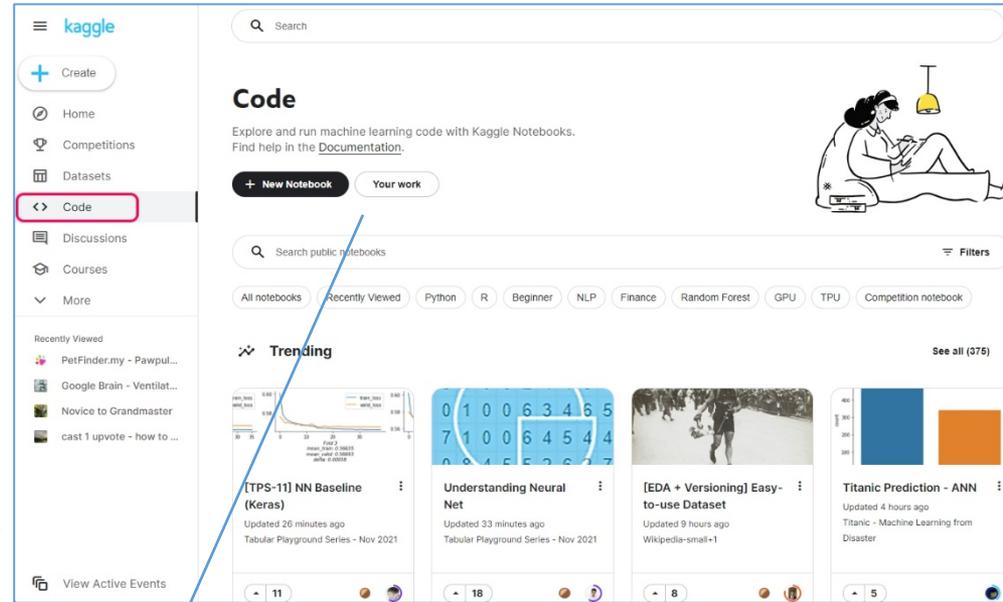


kaggle

4. Code使用方法

① Notebook作成

- 新規Notebook
 - Code > New notebook
- 使い方
 - A) Jupyter notebookと同じ
 - B) +code: pythonのプログラムを記入
 - C) +Markdown: コメント記入
 - D) Data
 - Input: 使用するcsvなどを置く
 - Output: 出力
 - E) Setting
 - Language: Python or R
 - Accelerator: None, gpu, tpu

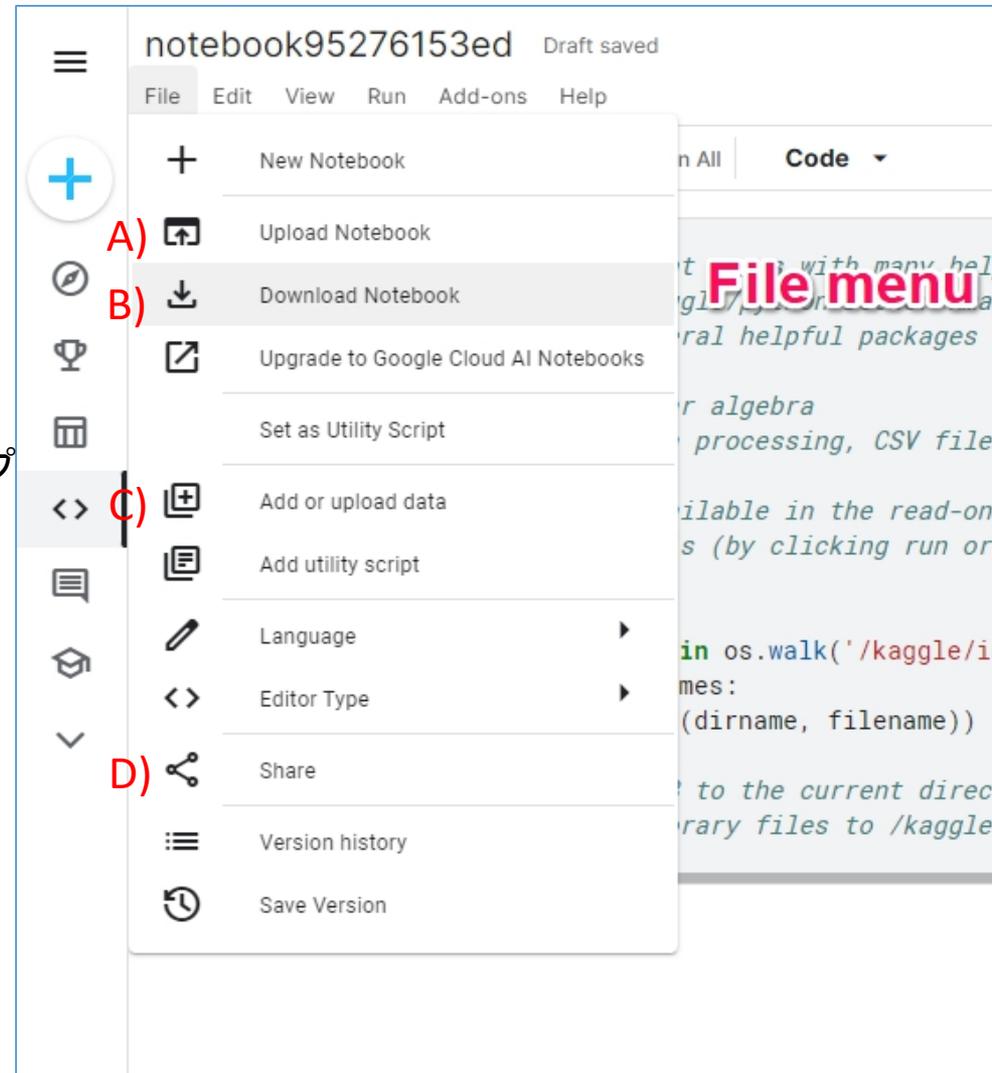


kaggle

4. Code使用方法

① Notebook作成

- 新規Notebook
- 使い方
- File > menu
 - A) Upload notebook: ローカルから
 - B) Download notebook: ローカルへ
 - C) Add or upload data: datasetのアップ
 - D) Share: Publicに公開できる



kaggle

4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例
 - A) Course > Python選択
 - B) Hello, Pythonを選択

The image displays two screenshots from the Kaggle website. The first screenshot shows the 'Courses' page with the 'Courses' menu item highlighted in red (A). Below it, the 'Python' course is highlighted in red (B). The second screenshot shows the 'Python' course page with the 'Hello, Python' lesson highlighted in red (B).

Screen 1: Kaggle Courses Page

- Search bar: Search
- Text: The courses are free, and you can now earn certificates.
- Section: Your Courses
- Active: Python (Next up: Exercise: Syntax, Variables, and Numbers)
- Section: Explore Courses
- Intro to Machine Learning: Learn the core ideas in machine learning, and build your first models.
- Pandas: Solve short hands-on challenges to perfect your data manipulation skills.
- Intermediate Machine Learning: Handle missing values, non-numeric values, data leakage, and more.
- Data Visualization: Make great data visualizations. A great way to see the power of coding!

Screen 2: Kaggle Python Course Page

- Section: Lessons
- Lesson 1: Hello, Python (A quick introduction to Python syntax, variable assignment, and numbers) - 50% progress
- Lesson 2: Functions and Getting Help (Calling functions and defining our own, and using Python's builtin documentation)
- Lesson 3: Booleans and Conditionals (Using booleans for branching logic)

all rights reserved 2021 spectrum technology co.

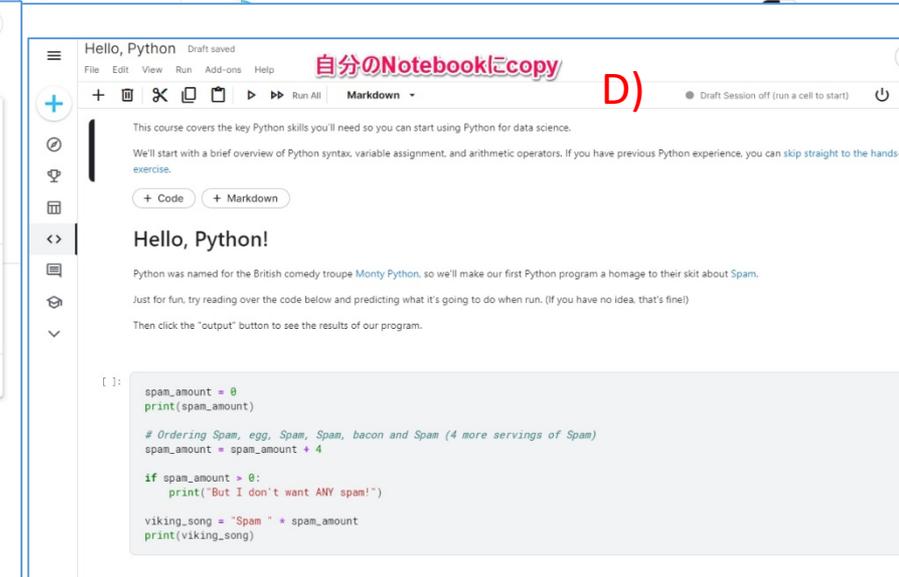
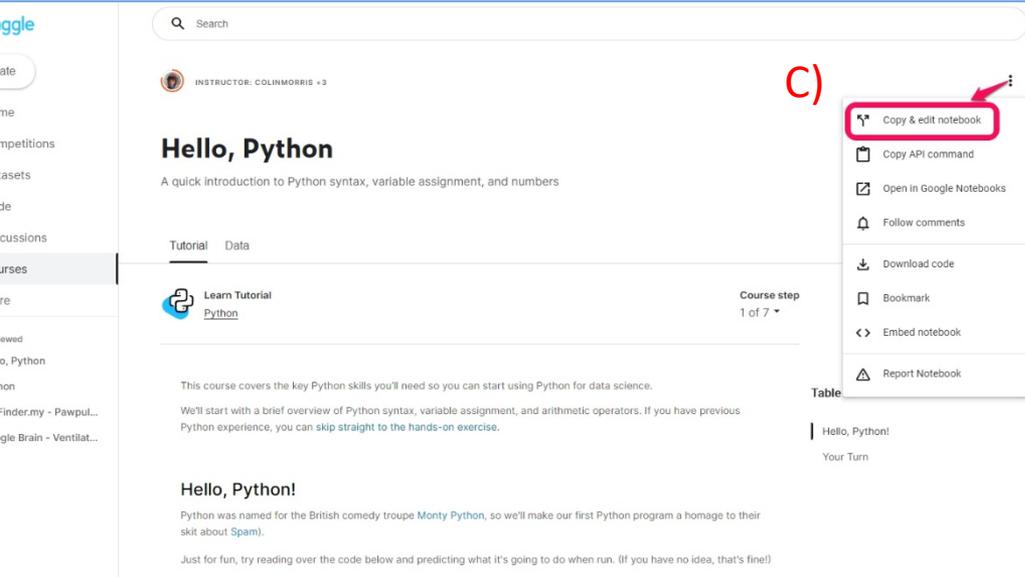
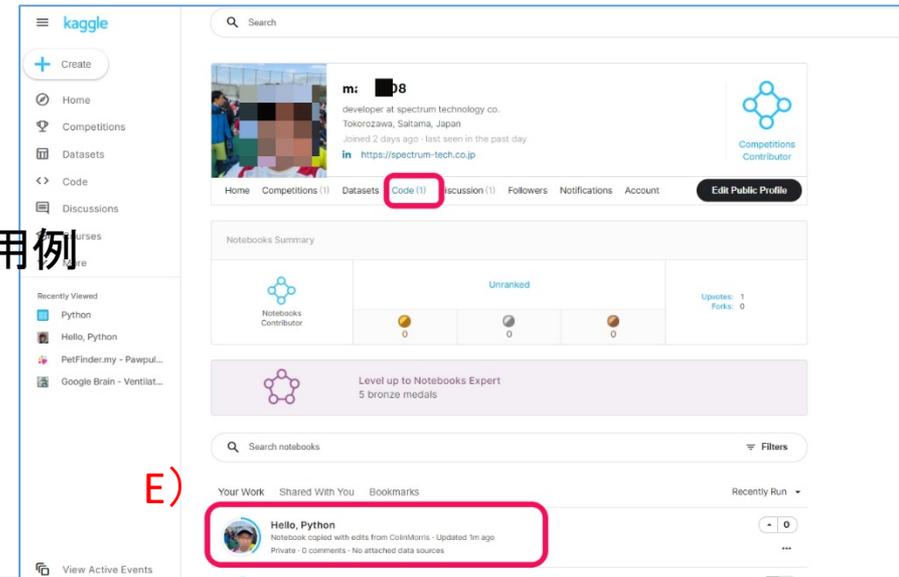
kaggle

4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例

- Course > Python 選択
- Hello, Python を選択
- Copy notebook
- 自分のnotebookにコピーされます。
- Home > my page > codeにあることを確認



kaggle

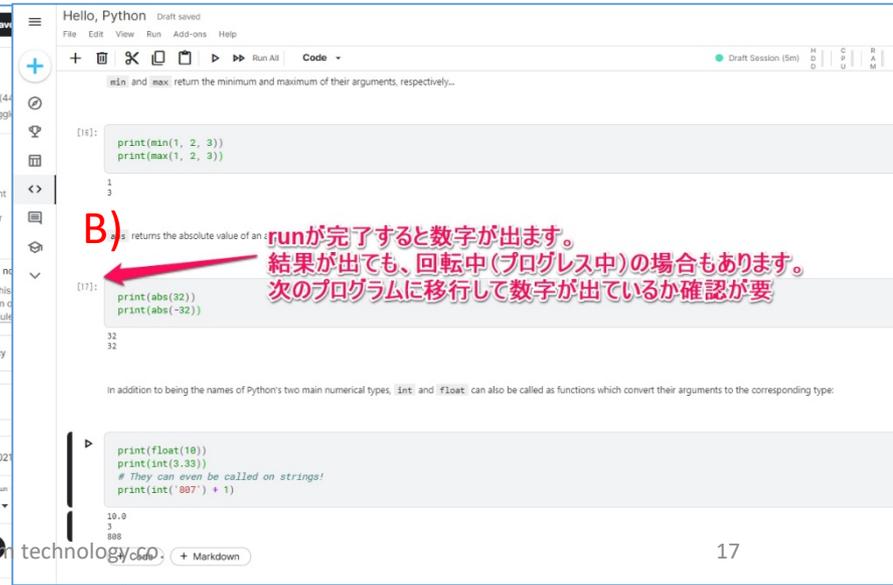
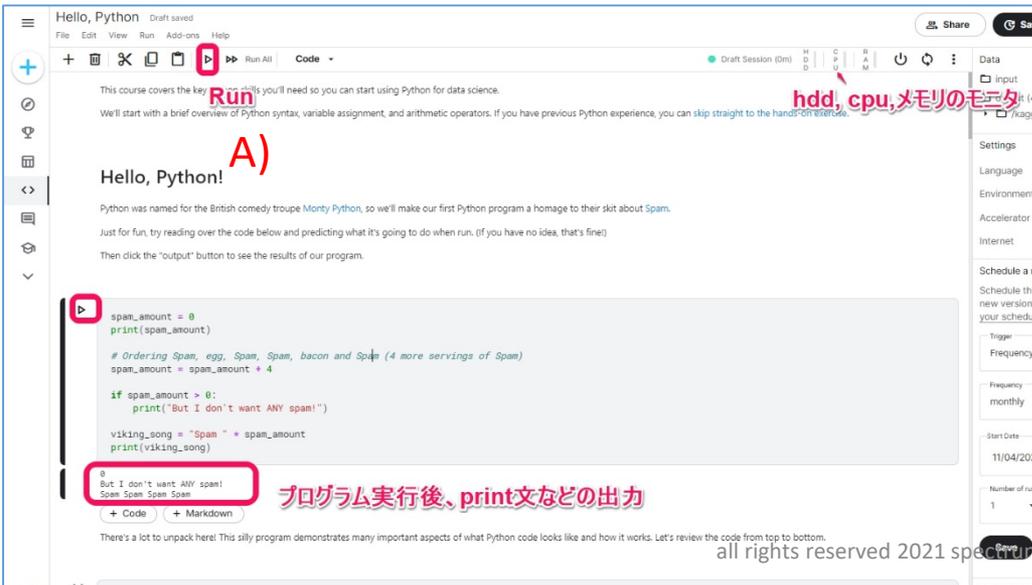
4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例
- Hello,Python使用例

A) Run: codeのある場所でRunを押すと実行。print文がある場合は出力される

B) 数字の確認: Runが完了すると数字がでます。プログレス中の場合もあるので注意。



kaggle

4. Code使用方法

② Notebook使用例

- Courseからpython学習をコピーした使用例
- Hello,Python使用例
 - Hello,pythonは全てのプログラムが完了すると自動保存。又は別のバージョンでSave verion
 - My page>code: hello pythonを選択するとこれまでの状況が分かります

A) Notebook: 動作時間、実行結果

B) Logs: ログが表示

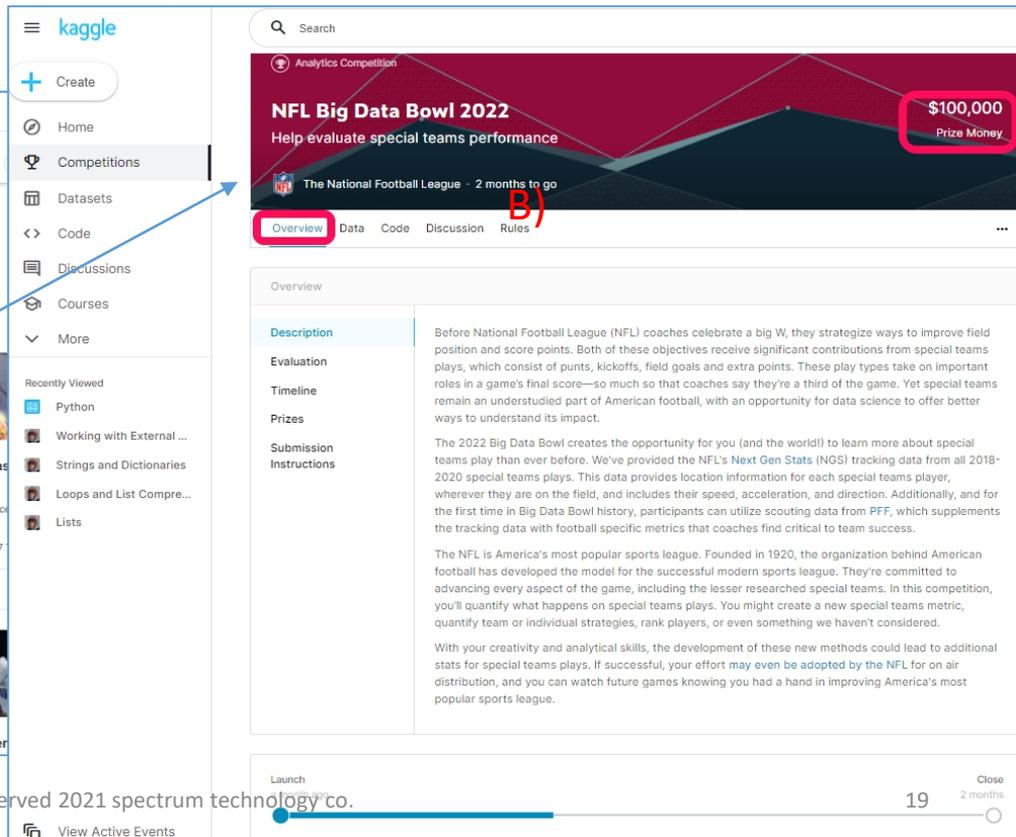
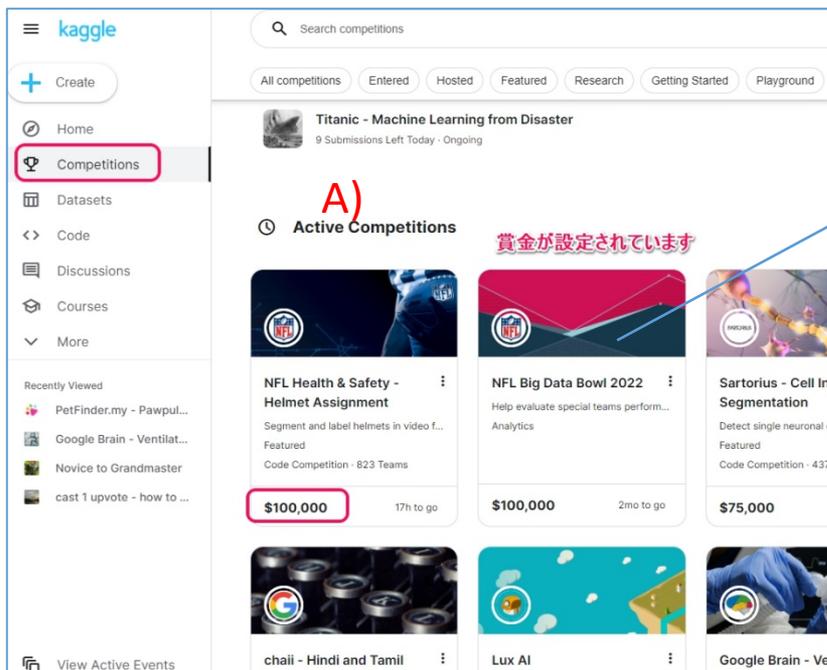
kaggle

5. Competition使用方法

① コンペ使用方法

A) 活動中で参加したいコンペの選択

B) コンペ内容確認: 概要をよく確認。賞金が設定しているのは、難しい課題です。初心者は、賞金のないものから



kaggle

5. Competition使用方法

① コンペ使用方法

- A) 活動中で参加したいコンペの選択
- B) コンペ内容確認
- C) データ: 提供されてるcsvデータなどを確認
- D) ルール: 賞金、チームの人数などのルール

This screenshot shows the 'Data' tab of the 'NFL Big Data Bowl 2022' competition. The page title is 'NFL Big Data Bowl 2022' with a prize money of '\$100,000'. The description is 'Help evaluate special teams performance'. The 'Data' tab is highlighted with a red box. Below the navigation bar, there is a 'Data Description' section with a red 'C)' next to it. The 'Summary of data' section states: 'The 2022 Big Data Bowl data contains Next Gen Stats player tracking, play, game, player, and PFF scouting data for all 2018-2020 Special Teams plays. Here, you'll find a summary of each data set in the 2022 Data Bowl, a list of key variables to join on, and a description of each variable.' The 'File descriptions' section lists 'games.csv' and 'plays.csv'. At the bottom, a terminal window shows the command: `kaggle competitions download -c nfl-big-data-bowl-2022`. A 'Data Explorer' window is open, showing a list of files including 'PFFScoutingData.csv' (2.06 MB), which is highlighted with a red box and a red arrow pointing to it with the text '提供されているデータ' (Data provided).

This screenshot shows the 'Rules' tab of the 'NFL Big Data Bowl 2022' competition. The page title is 'NFL Big Data Bowl 2022' with a prize money of '\$100,000'. The description is 'Help evaluate special teams performance'. The 'Rules' tab is highlighted with a red box. Below the navigation bar, there is a section with a red 'D)' next to it. The text reads: 'By clicking on the "I understand and accept" button, you indicate that you agree to be bound with the rules outlined below.' Below this text is a button labeled 'I Understand and Accept'. The 'Rules' section contains the following text: 'Updated: November 1, 2021. Section A.3 was removed. Section A.3 prohibited the use of external data. External data is allowed following the guidance in Section B.7.C.' and 'One account per participant' with the subtext: 'You cannot sign up to Kaggle from multiple accounts and therefore you cannot submit from multiple accounts.'

kaggle

5. Competition使用方法

① コンペ使用方法

- A) 活動中で参加したいコンペの選択
- B) コンペ内容確認
- C) データ: 提供されてるcsvデータなどを確認
- D) ルール: 賞金、チームの人数などのルール
- E) コード: 参照になる他の人のNotebook

The screenshot shows the Kaggle interface for the 'NFL Big Data Bowl 2022' competition. The main header includes the competition title, a subtitle 'Help evaluate special teams performance', and a prize money amount of '\$100,000'. Below this, there are navigation tabs: 'Overview', 'Data', 'Code' (highlighted with a red box), 'Discussion', and 'Rules'. A 'New Notebook' button is also visible. Below the navigation, there is a search bar for notebooks, and a list of notebooks is displayed. The first notebook is titled 'One weird flying ball' and is highlighted with a red box and the text '参考のnotebookが出てます' (Reference notebook is out). The second notebook is titled 'NFL | Prediction of result draw a ball after punt'.

kaggle

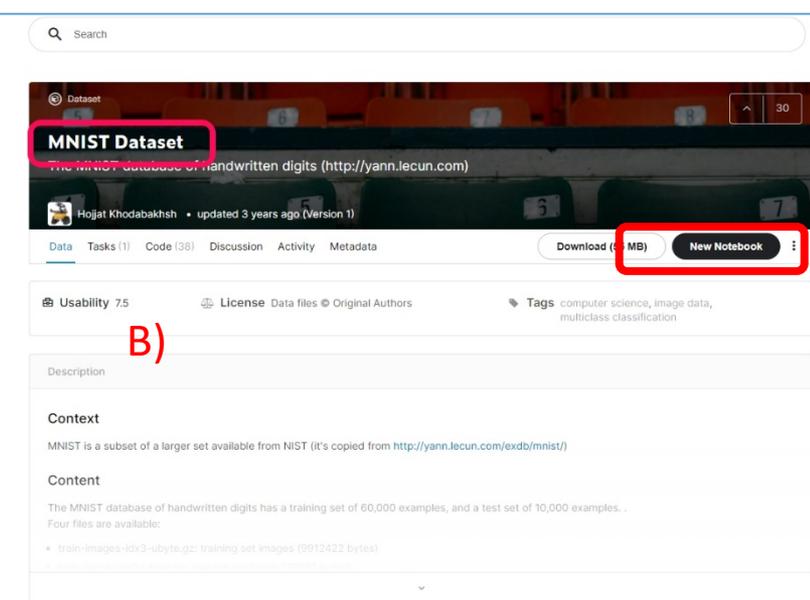
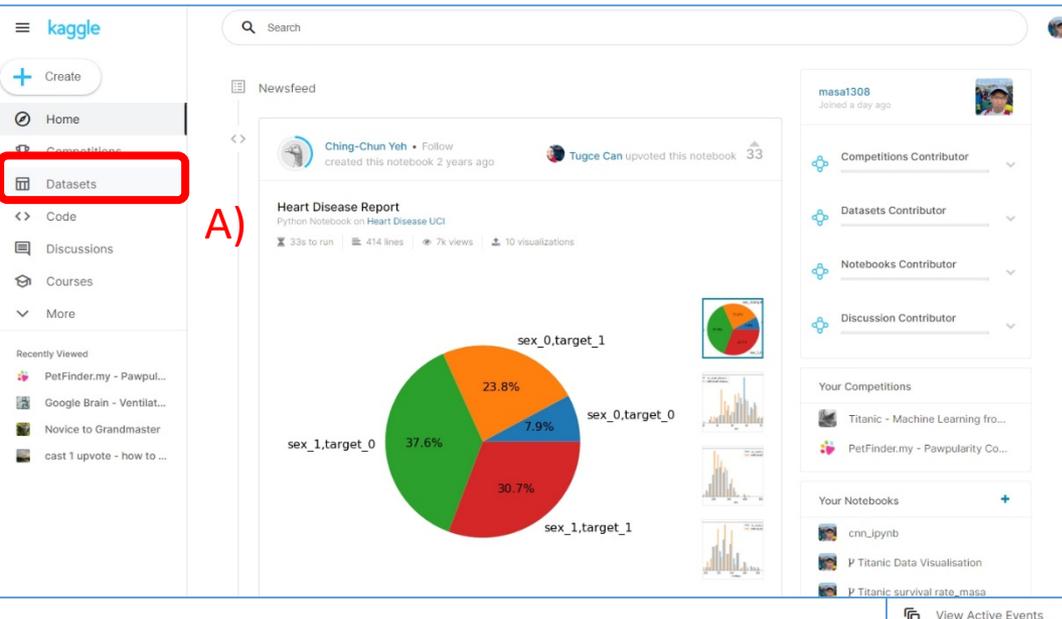
6. dataset使用方法

① dataset追加方法1

A) Dataset

B) Mnistを検索して、mnist datasetを選んで、notebook作成。

C) Notebookでコードを記入

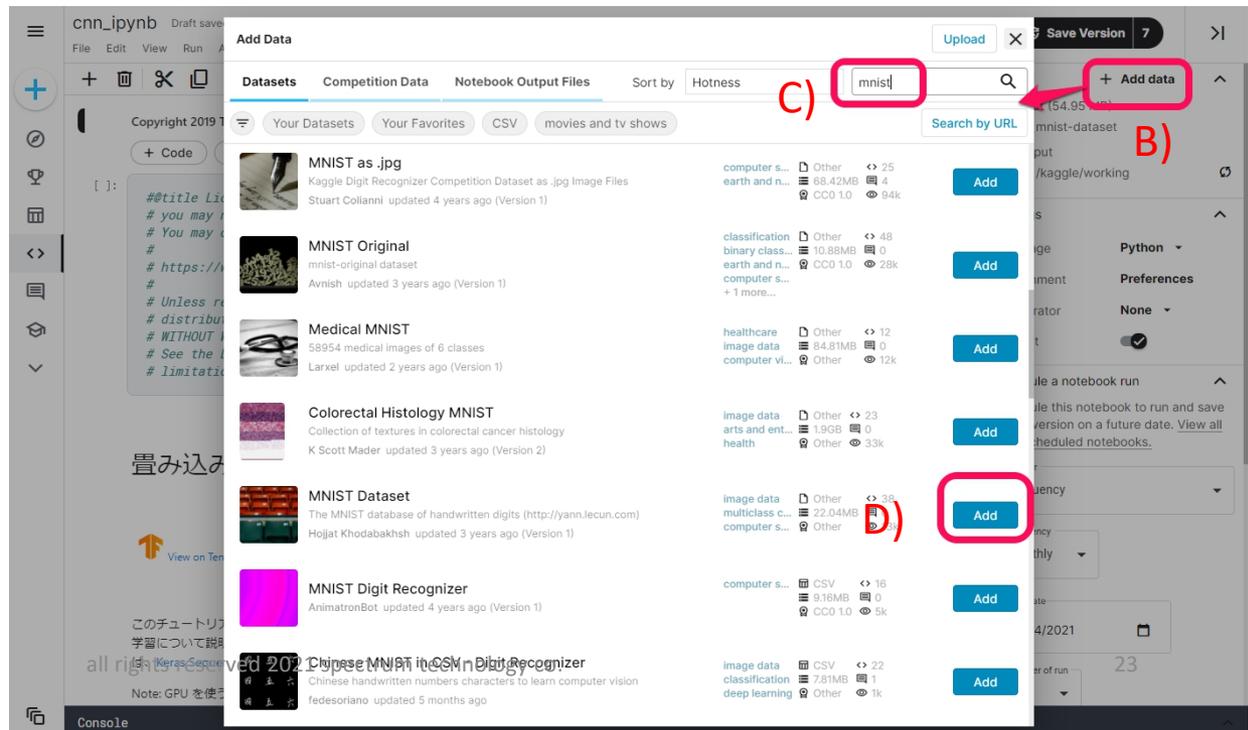


kaggle

6. dataset使用方法

② dataset追加方法2 (一般的にはこちらか)

- A) Notebook作成またはアップロード
- B) Add data.
- C) mnistで検索
- D) Mnist datasetのAdd
- E) Notebookのinputにデータを確認

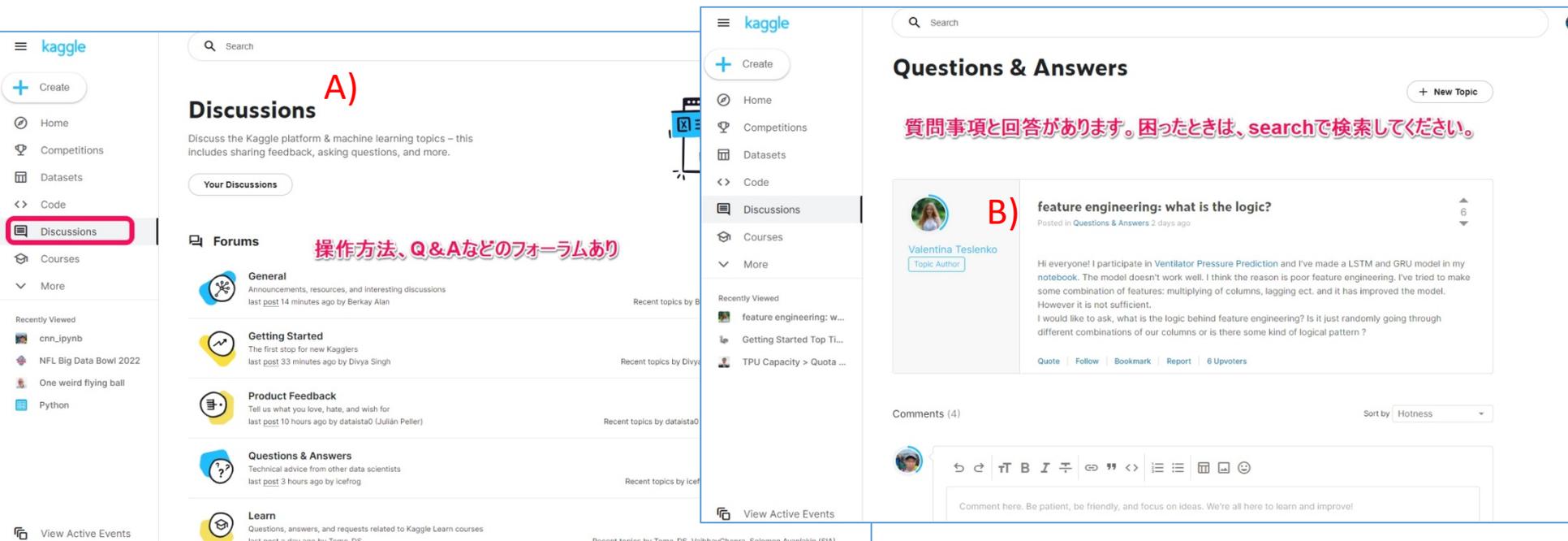


kaggle

7. Discussions使用方法

① 議論使用方法

- A) Discussions > 操作方法、Q&Aの内容が記載
- B) 詳細: 個別の内容を見ると解決策があります。



kaggle

8. 事例

[https://spectrum-](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

[tech.co.jp/products/ai_tensorflow.html](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

弊社ホームページのAIプログラム学習キットも参照ください。

① CNN.ipynb

- 弊社で作成したtensorflow v2のcnn.ipynbを公開しています。
- <https://www.kaggle.com/masa1308/cnn-ipynb>
- Mnist dataset(手書き文字のベンチマーク用)を使って、Convolutional Neural Networksを学習するものです。
- Cpuとgpuの学習速度の違いを体感してください。本gpuはRTX2080より少し遅いくらいです。十分使えます。但し時間制限38hours/weekがあります。

The screenshot shows the Kaggle interface for a notebook titled 'cnn_ipynb'. The notebook title is '畳み込みニューラルネットワーク (Convolutional Neural Networks)'. Below the title, there are buttons for 'View on TensorFlow.org', 'Run in Google Colab', 'View source on GitHub', and 'Download notebook'. The main content area contains introductory text in Japanese: 'このチュートリアルでは、MNIST の数の分類をするための、シンプルな畳み込みニューラルネットワーク (CNN: Convolutional Neural Network) の学習について説明します。このシンプルなネットワークは MNIST テストセットにおいて、99%以上の精度を達成します。このチュートリアルでは、Keras Sequential APIを使用するため、ほんの数行のコードでモデルの作成と学習を行うことができます。' Below this text, there is a note: 'Note: GPU を使うことで CNN をより早く学習させることができます。もし、このノートブックを Colab で実行しているならば、編集 → ノートブックの設定 → ハードウェアアクセラレータ → GPU から無料の GPU を有効にすることができます。' At the bottom, it says 'TensorFlowのインポート'.

kaggle 8.事例

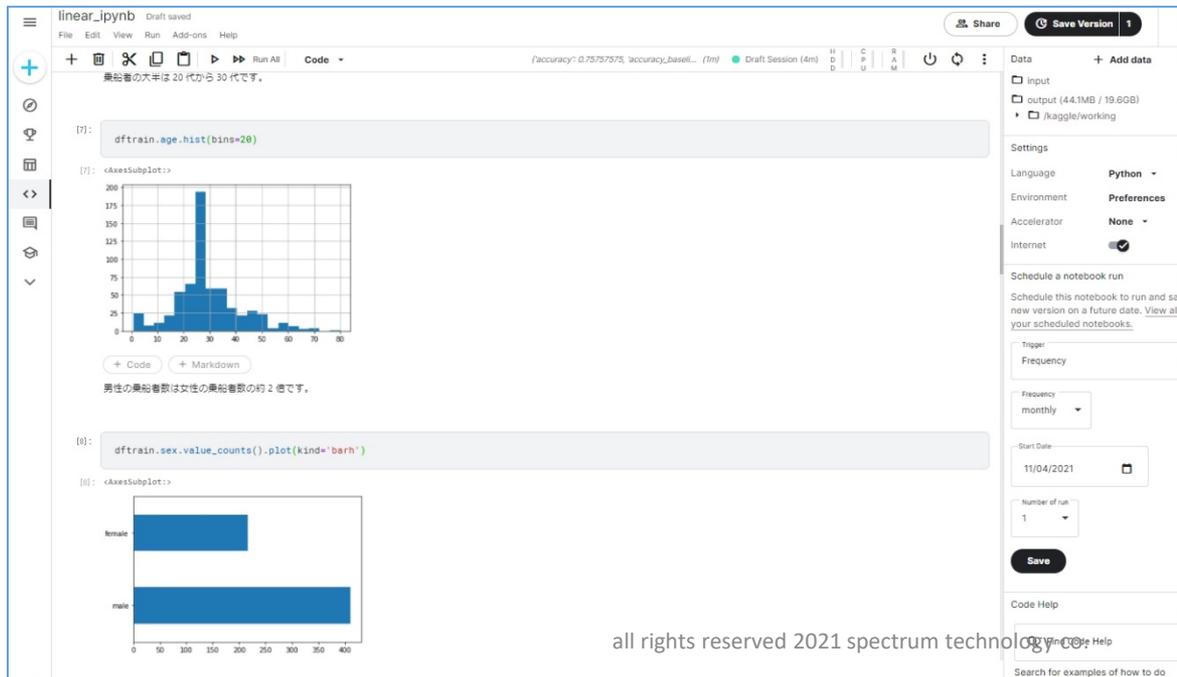
[https://spectrum-](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

[tech.co.jp/products/ai_tensorflow.html](https://spectrum-tech.co.jp/products/ai_tensorflow.html)

弊社ホームページのAIプログラム学習キットも参照ください。

② linear.ipynb

- 弊社で作成したtensorflow v2のlinear.ipynbを公開しています。
- <https://www.kaggle.com/masa1308/linear-ipynb>
- titanic号の生存状況を予測モデルです。
- コンペにも同様なものがあります。以下も参照ください。
- <https://www.kaggle.com/c/titanic>



kaggle

9.コンペへ出展

① 出展に必要なこと

- スポンサーとしてコンペへの出展も可能です。
- プロフィール、解決したい内容を記述し、kaggleにコンタクトします。
- More > host a competition: get started
- プロフィール作成

The screenshot shows the Kaggle homepage. On the left sidebar, the 'Host a Competition' link is highlighted with a red box. In the top right corner, a 'Get Started' button is also highlighted with a red box. The main content area features a header with the Kaggle logo and a search bar, followed by a main heading: 'Kaggle makes it easy to connect your data with data scientists.' Below this, there is a sub-heading: 'Kaggle is a platform for data science competitions. We help you solve difficult problems, recruit strong teams, and amplify the power of your data science talent.' A call-to-action button says 'Interested in hosting a classroom competition? Visit Kaggle InClass >'. At the bottom, there is a section for 'For Business Problems' with a 'Learn more' button and a diagram showing people connected by lines.

The screenshot shows the 'Host a Competition' form on Kaggle. The text 'お客様情報入力' (Customer Information Input) is overlaid in red. The form includes fields for Name, Organization, Job Title, E-mail, and Human?. Below the form, there is a 'Recently Viewed' section with a list of items: linear_ipynb, MNIST Dataset, Titanic - Machine Lear..., Titanic survival rate, m..., and feature engineering: w... The bottom of the page shows a footer with 'all rights reserved 2021 spectrum technology.co' and 'Active Events'.

kaggle

9.コンペへ出展

① 出展に必要なこと

- スポンサーとしてコンペへの出展も可能です。
- プロフィールを作成し、kaggleからのコンタクト待ち。
- More > host a competition: get started
- プロフィール作成

② コンペ内容作成

- 課題解決したい内容を作成し、必要なdataを添付し、賞金などを設定し提出
- 既にある課題内容をよく確認し、提出しましょう。

