

GitHub & Kaggle Tutorial

GitHub

- 功能: 代码版本控制, 多人协作, 开源资料, 简历
- 界面演示: 创建 repository, branch, fork, pull request(merge), edit, commit, issue, contributions, setting
- [Hello World - GitHub Docs](#)
- Example: Awesome list, [yuxizzz/menu](#)

扩展 (CSC207/209):

- [GitHub Docs](#)
- [github-cheat-sheet/README.zh-cn.md at master · tiimgreen/github-cheat-sheet](#)
- [Git 的安装及使用](#)
- [shiqio3o/GitHub-Tutorial: Getting started tutorial on GitHub](#)

目标: 1. 了解 Git 及 GitHub; 2. GitHub 检索功能; 3. 学会界面演示中的基础操作;

Kaggle

- Datasets
- Compete on Kaggle: Data, Team, Submission
- Example: [Classification of the Alzheimer's Disease | Kaggle](#)
- 扩展: Courses, Model, Code

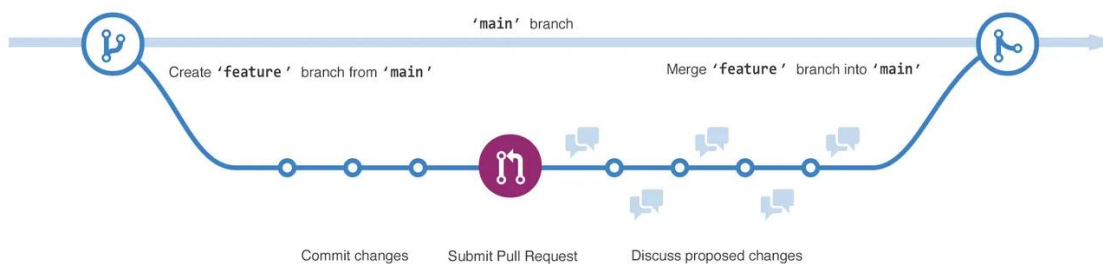
目标: 1. 学会使用 kaggle 获取 dataset; 2. 熟悉 competition 流程及注意事项; 3. 学会利用 Kaggle 上的开源 code 和 model

Git 与 GitHub

- **Git**：一个版本控制系统(Version Control System)，用来记录文件的修改历史，支持多人同时协作开发
- **GitHub**：基于 Git 的代码托管平台，可以把本地的 Git 仓库上传到云端，方便共享和协作。基本的操作包括：（1）创建和使用存储库(repository)（2）启动并管理新分支(branch)（3）对文件进行更改，并将其作为提交(commit)推送(push)到 GitHub（4）打开并合并(merge)拉取请求(pull request)

工作流程:

1. 在本地创建分支并修改文件
2. 使用 Git 提交更改
3. 推送到 GitHub 远程仓库
4. 合作者可以从远程仓库拉取最新版本
5. 借助 Pull Request 等功能实现协作和合并



开源资料(检索功能)

<https://github.com/search?q=machine+learning+awesome&type=repositories>

The screenshot shows the GitHub search interface for the query "machine learning awesome". The search bar at the top contains the text "machine learning awesome". Below the search bar, the results are filtered by "Repositories" (818 results). The left sidebar shows the "Filter by" section with options for Code (1.3M), Repositories (818), Issues (10k), Pull requests (6k), Discussions (1k), and Users (52). The "Languages" section lists Python, Jupyter Notebook, HTML, JavaScript, TypeScript, Markdown, C++, MATLAB, and Shell. The main content area displays four repository results:

- josephmisiti/awesome-machine-learning**: A curated list of awesome Machine Learning frameworks, libraries and software. Python · 69.5k · Updated 14 days ago. Star button.
- EthicalML/awesome-production-machine-learning**: A curated list of awesome open source libraries to deploy, monitor, version and scale your machine learning. machine-learning data-mining awesome deep-learning awesome-list · 19.1k · Updated 4 days ago. Star button.
- jobbole/awesome-machine-learning-cn**: 机器学习资源大全中文版, 包括机器学习领域的框架、库以及软件. 4.3k · Updated on 2024年4月3日. Star button.
- ujjwalkarn/Machine-Learning-Tutorials**: machine learning and deep learning tutorials, articles and other resources. list machine-learning awesome deep-neural-networks deep-learning · 16.6k · Updated on 2024年6月12日. Star button.

On the right side, there are three promotional cards: "Awesome Lists" (An awesome list is a list of awesome things curated by the community. github.com sindresorhus/awesome # View topic), "Sponsor open source projects you depend on" (Contributors are working behind the scenes to make open source better for everyone—give them the help and recognition they deserve. Explore sponsorable projects →), and "How can we improve search? Give feedback". At the bottom right, there is a "ProTip!" section: "Press the / key to activate the search input again and adjust your query."

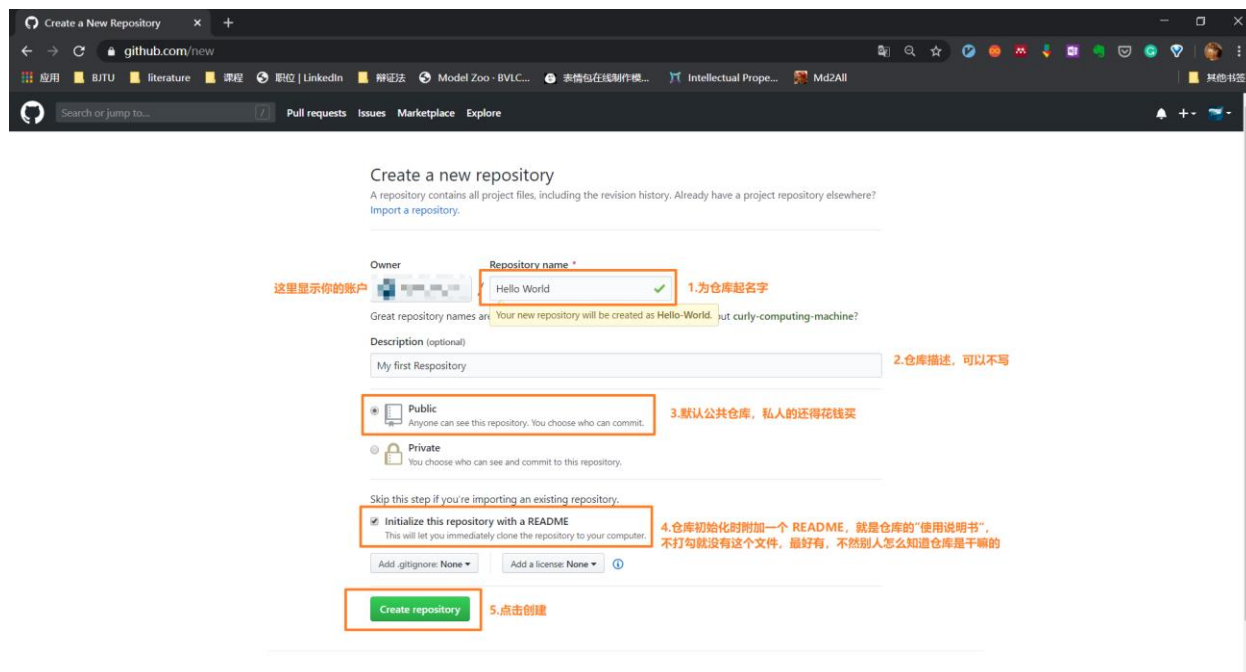
简历

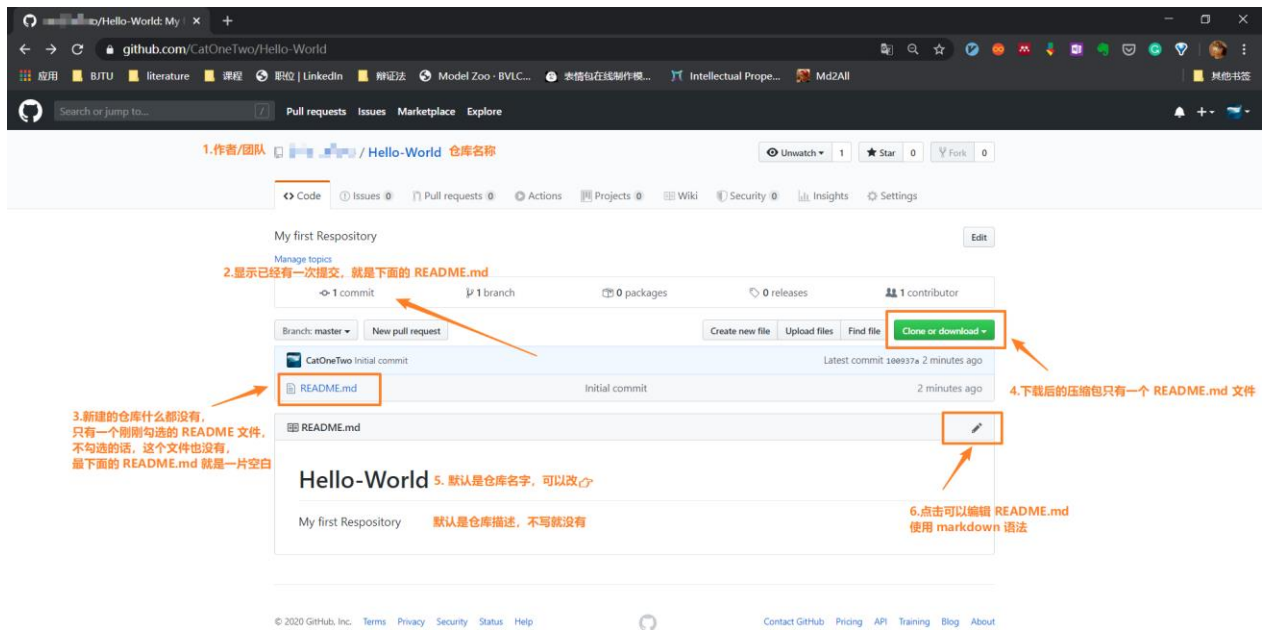
GitHub 主页是对于简历的补充，用实际项目和代码来验证简历中写的技能

界面演示

创建 repository, branch, fork, pull request(merge), edit, commit, issue, contributions, setting

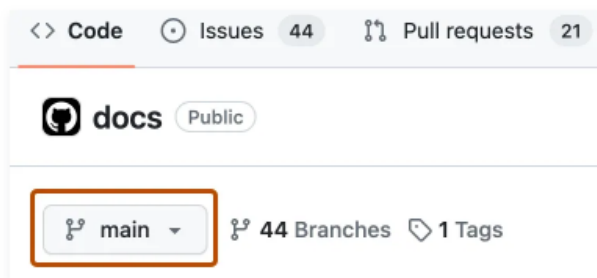
- Hello World - GitHub Docs
- **Pull Request:** 你想参与别人的项目，或想改进别人的项目，改进后需要提交申请，用 pull requests。
- **Repository:** 翻译为仓库，也是你的项目。你可以理解成一个大的文件夹，或者笔记本。一个项目对应一个 Repository。
- **Issue:** Issue 追踪各种想法，增强功能，任务，bug，有点儿像评论留言功能。
- **Fork:** 将 GitHub 的某个特定仓库（所有文件）原封不动地复制到自己的账户下（但不会与原仓库相同）。以此实现分散式版本管理
- **README.md:** 每当创建项目、初始化时，都会帮你自动生成 README.md 文件并显示在仓库首页。一般都是使用 Markdown 语法来描述项目的概要、使用流程、许可协议等





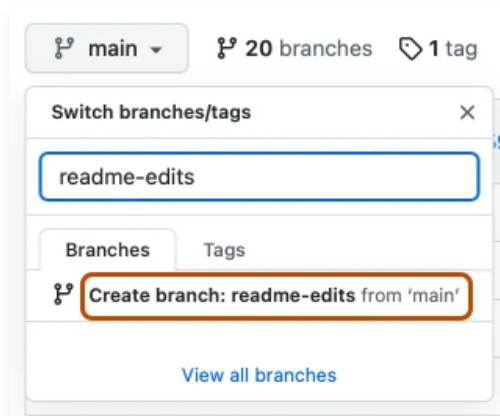
1 单击 `hello-world` 存储库的“代码”选项卡。

2 在文件列表上方，单击“main”下拉菜单。

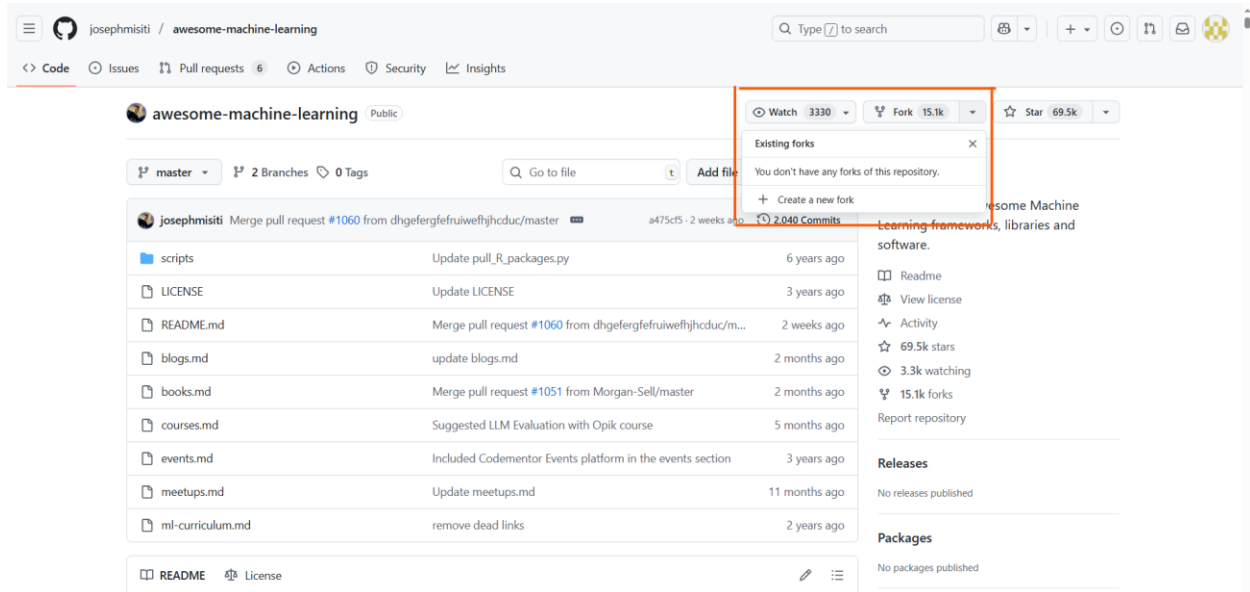


3 在文本框中键入分支名称 `readme-edits`。


4 单击“创建分支: 从 main 创建 readme-edits”。



Fork



Commit

- 1 在你创建的 `readme-edits` 分支下，单击 `README.md` 文件。
- 2 若要编辑文件，请单击 .
- 3 在编辑器中，编写一些关于您自己的内容。
- 4 单击“提交更改”。
- 5 在“提交更改”框中，编写描述更改的提交消息。
- 6 单击“提交更改”。

Pull Request

- 1 单击 `hello-world` 存储库的“拉取请求”选项卡。
- 2 单击“新建拉取请求”。
- 3 在“示例比较”框中，选择你创建的分支 `readme-edits`，与 `main`（原始分支）进行比较。
- 4 在 Compare（比较）页面上的差异中查看您的更改，确保它们是您要提交的内容。




- 5 单击“创建拉取请求”。
- 6 为拉取请求指定一个标题，并写下更改的简要说明。您可以包含表情符号以及拖放图像和 gif。
- 7 单击“创建拉取请求”。

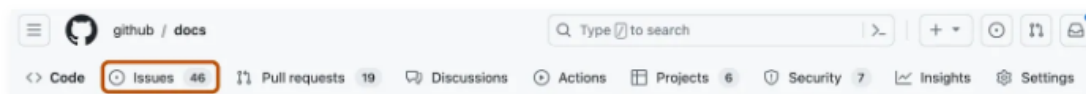
Merge

- 1 在拉取请求底部，单击“合并拉取请求”以将更改合并到 `main` 中。
- 2 单击“确认合并”。您将收到一条消息，指出请求已成功合并且请求已关闭。
- 3 单击“删除分支”。现在你的拉取请求已合并，并且你的更改位于 `main` 上，接下来你可安全地删除 `readme-edits` 分支。如果要对项目进行更多更改，可以随时创建新分支并重复此过程。
- 4 单击返回 `hello-world` 存储库的“代码”选项卡，以在 `main` 上查看已发布的更改。

Issue

1 在 GitHub 上，导航到存储库的主页面。

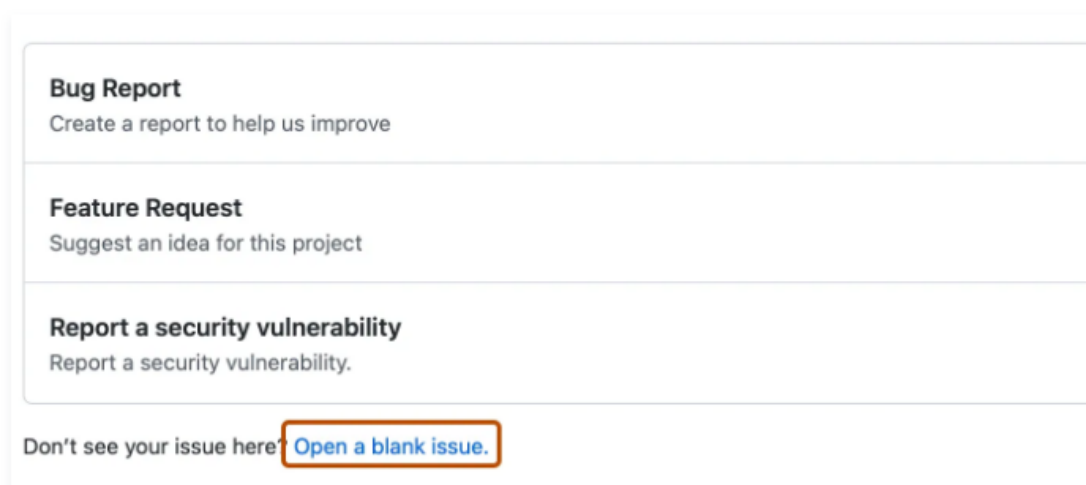
2 在存储库名称下，单击 “问题”。



3 单击“新建问题”。

4 如果你的存储库使用问题模板，请在要打开的问题类型旁边单击“开始”。

如果要打开的问题类型未包含在可用选项中，请单击“打开空白问题”。



5 在“标题”字段中键入问题标题。

6 在注释正文字段中键入问题的说明。

7 如果你是项目维护者，可以将[问题分配给某人](#)、[将其添加到 project](#)、[将其与里程碑关联](#)、[设置问题类型](#)、或[应用标签](#)。

8 完成后，单击“提交新问题”。

Contribution

<https://github.com/josephmisiti/awesome-machine-learning/graphs/contributors>

Setting

Danger Zone

Change repository visibility

This repository is currently public.

[Change visibility](#)

Disable branch protection rules

Disable branch protection rules enforcement and APIs

[Disable branch protection rules](#)

Transfer ownership

Transfer this repository to another user or to an organization where you have the ability to create repositories.

[Transfer](#)

Archive this repository

Mark this repository as archived and read-only.

[Archive this repository](#)

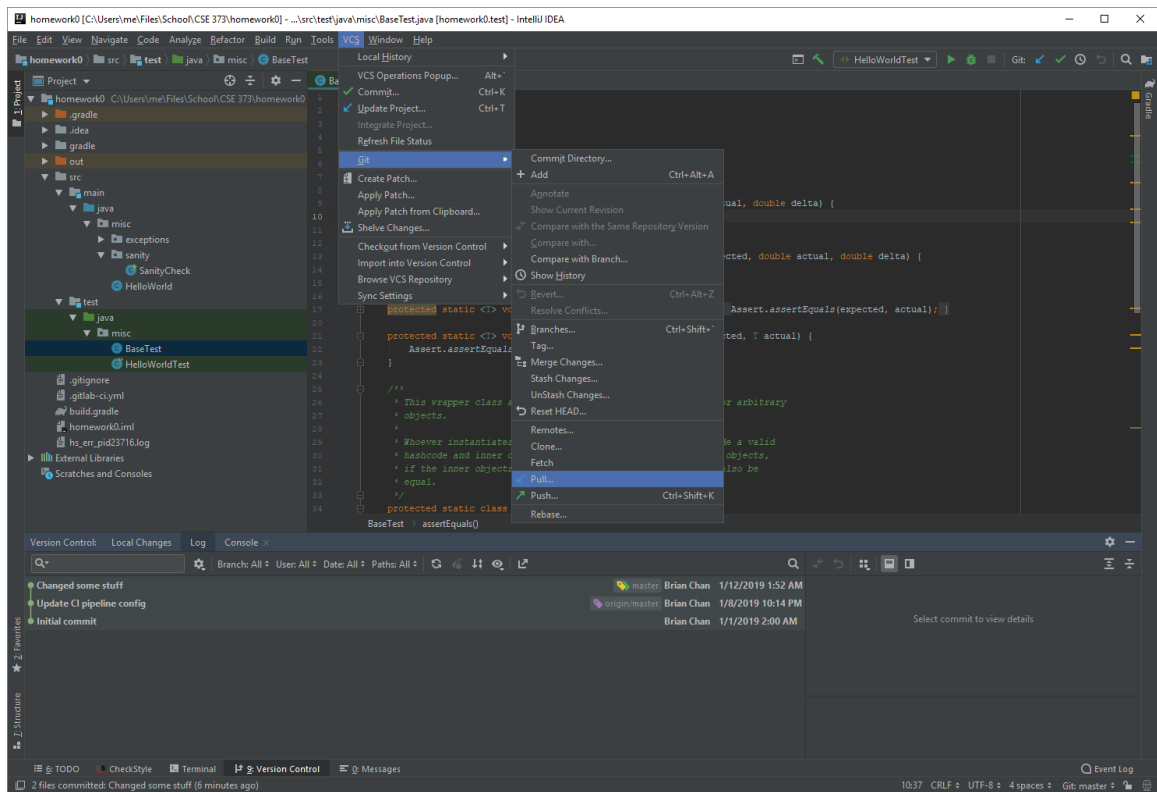
Delete this repository

Once you delete a repository, there is no going back. Please be certain.

[Delete this repository](#)

扩展 (CSC207/209)

- [GitHub Docs](#)
- [github-cheat-sheet/README.zh-cn.md](#) at master · tiimgreen/github-cheat-sheet
- [Git 的安装及使用](#)
- [shiqio3o/GitHub-Tutorial: Getting started tutorial on GitHub](#)



Kaggle

- Datasets: <https://www.kaggle.com/datasets>
- Compete on Kaggle: Data, Team, Submission
- Example: [Classification of the Alzheimer's Disease | Kaggle](#)

Classification of the Alzheimer's Disease

Alzheimer's Disease



Overview Data Discussion Leaderboard Rules Team Submissions

Overview

Alzheimer's disease is a brain disorder that gets worse over time. It's characterized by changes in the brain that lead to deposits of certain proteins. Alzheimer's disease causes the brain to shrink and brain cells to eventually die. Alzheimer's disease is the most common cause of dementia — a gradual decline in memory, thinking, behavior and social skills. These changes affect a person's ability to function.

It is therefore important to detect Alzheimer's disease in an early stage and explore factors associated with Alzheimer's. This dataset contains extensive health information for 2,149 patients, each uniquely identified with IDs ranging from 1 to 2149. The dataset includes demographic details, lifestyle factors, medical history, clinical measurements, cognitive and functional assessments, symptoms, and a diagnosis of Alzheimer's Disease.

Start

Sep 26, 2024

Close

Dec 6, 2024

Competition Host

xbUofT



Prizes & Awards

Kudos

Does not award Points or Medals

Participation

333 Entrants

285 Participants

147 Teams

1,081 Submissions

Tags

Accuracy Score

Table of Contents

Overview

Evaluation



- 扩展:

Courses: <https://www.kaggle.com/learn>

Models: <https://www.kaggle.com/models>

Code: <https://www.kaggle.com/code>