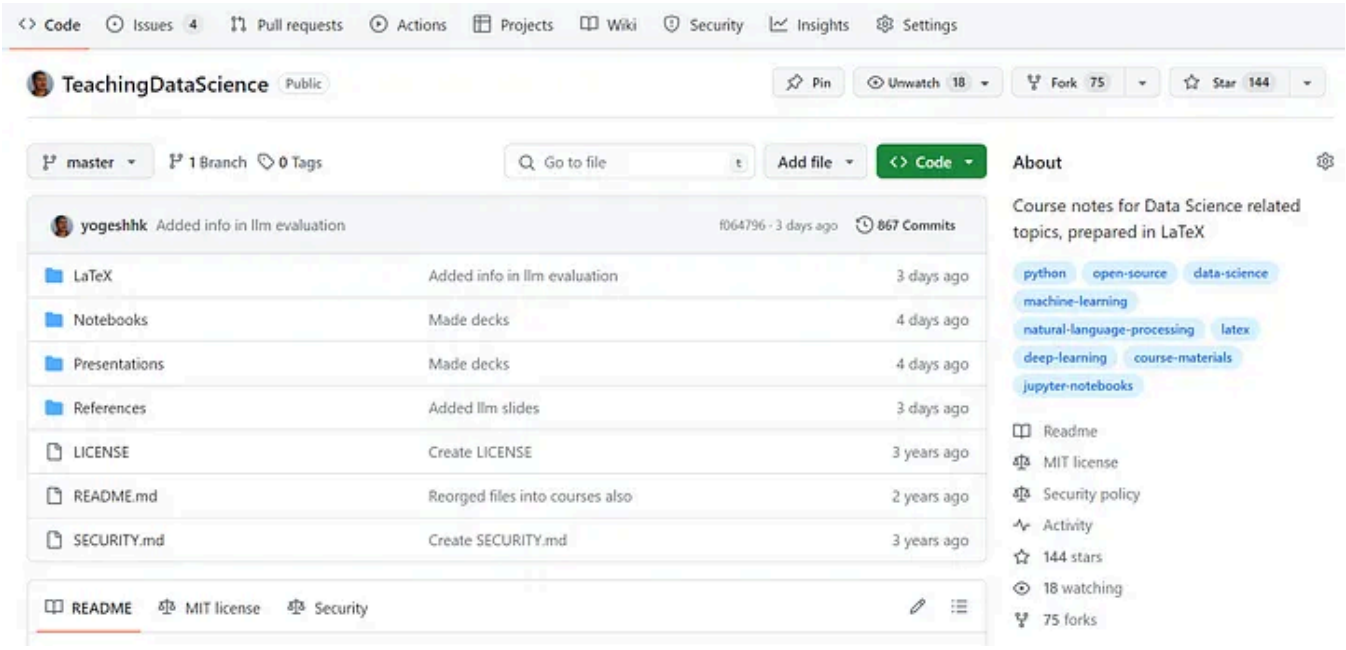
Teaching Data Science

An open-source repository for teaching material, open-free to all

Yogesh Haribhau Kulkarni (PhD)
Published in Technology Hits
4 min read · 1 day ago



Screenshot of the GitHub repository

Are you passionate about data science? Do you dream of unraveling hidden patterns in vast datasets, creating predictive models, and contributing to cutting-edge research? If so, you're in the right place!

Welcome, data enthusiasts, to a treasure trove of knowledge — the "Teaching Data Science" repository by Yogesh H Kulkarni. This GitHub repository is a rich collection of LaTeX course notes covering a spectrum of topics including Python, Machine Learning, Deep Learning, Natural Language Processing, and more. In this blog, we'll explore the purpose, how to use, steps to contribute, and other essential aspects of this invaluable resource.

Purpose: Spreading the Light of Data Science

The Teaching Data Science repository serves a noble purpose: to spread the gospel of data science far and wide. Our mission is simple yet powerful — to make data science accessible to everyone. Whether you're a student, an industry professional, or an enthusiast, we believe that knowledge should flow freely. By sharing our insights, code, and expertise, we hope to empower individuals to harness the power of data. The values driving this endeavor are rooted in giving back to the community and paying knowledge forward. The ultimate goal is to propel the industry from automation to autonomy.

How to Use: Navigating the Maze of Knowledge

The core content is presented as Beamer slides — a dynamic format that combines visuals, text, and equations. These slides cover a wide range of topics, including:

- Python fundamentals
- Machine learning algorithms
- Deep learning architectures
- Natural language processing techniques

But that's not all! We've also transformed these slides into two-column course notes PDFs. Whether you're preparing for a seminar, workshop, or a semester-long course, you'll find valuable material here. The structure of this repository is well thought out, divided into three main directories: LaTeX, Code, and References.

LaTeX Directory

In the LaTeX directory, you'll discover TeX sources alongside essential images. Here's how it's organized:

- **Naming Convention**: Each TeX file follows a consistent naming convention, such as `maths_linearalgebra_matrices.tex`. Clear, concise names make navigation a breeze.
- **Driver Files**: For different event durations, we have driver files like `Main_Seminar_Presentation.tex`, `Main_Workshop_CheatSheet.tex`, and `Main_Course_Notes.tex`. These files compile the relevant sources seamlessly.

Code Directory

Data science isn't complete without hands-on practice. Our code directory houses Python and IPython notebook files. Here's what you'll find:

- **Naming Consistency**: Each code file corresponds to a specific LaTeX topic. We believe in connecting theory with practice.
- **Library-Based TeX Files**: For instance, `sklearn_intro.ipynb` accompanies the `sklearn_intro.tex` lecture. Dive into real-world examples and experiment with libraries.

References Directory

A treasure chest of papers, code, and presentations used as base material for content preparation. This directory acknowledges the importance of building on existing knowledge and resources.

Requirements: Setting Up

To utilize this repository, ensure you have LaTeX installed (tested with MikTex 2.9 on Windows 7, 64bit). Additional LaTeX packages may need installation based on warnings/suggestions. The recommended IDE is TexWorks.

How to Run LaTeX

To embark on your data science journey with Kulkarni's notes, you'll need a few essentials:

- **LaTeX:** We recommend MikTex 2.9 on Windows 7 (64-bit).
- **LaTeX packages:** Install them as prompted to ensure smooth compilation.
- **TexWorks IDE:** This user-friendly editor streamlines your LaTeX workflow.

Running the LaTeX files is straightforward. Driver files are named intuitively, and you can even compile individual files using your preferred LaTeX system. Alternatively, feel free to create your own main files and include the content files for a customized learning experience.

Steps to Contribute

The beauty of open source lies in collaboration, and this repository welcomes contributions with open arms.

1. Navigate to the ‘LaTeX’ folder.
2. Copy your images into the ‘images’ folder and source code to the ‘src’ folder.
3. Sample files are provided for copying and modification: `Main_Sample_Presentation.tex`, `Main_Sample_CheatSheet.tex`, both calling `sample_content.tex`.
4. Fill your material directly in the content file or organize it into multiple files and then `input` them in the content file.

Disclaimer

As with any valuable resource, the Teaching Data Science repository comes with a disclaimer:

- No guarantee of the correctness of the content.
- Notes are built using publicly available material.
- Citing original sources is a priority, but some may be missing.
- Continuous improvements are underway, and feedback is encouraged.
- Suggestions, comments, corrections, and pull requests are not just welcome; they are actively sought.

In conclusion, the Teaching Data Science repository is a beacon of knowledge in the vast ocean of data science. It empowers you to learn, contribute, and grow in this dynamic field. As we collectively strive for “From Automation to Autonomy,” let’s leverage this resource to illuminate our path in the world of data science. Dive in, explore, and let the journey to autonomy begin!

Click pic below or visit LinkedIn profile to know more about the author

Data Science | Teaching | Open Source | Future | Artificial Intelligence

Edit profile

Written by Yogesh Haribhau Kulkarni (PhD)

1.3K Followers · Editor for Technology Hits

PhD in Geometric Modeling | Google Developer Expert (Machine Learning) | Top Writer 3x (Medium) | More at https://www.linkedin.com/in/yogeshkulkarni/

More from Yogesh Haribhau Kulkarni (PhD) and Technology Hits

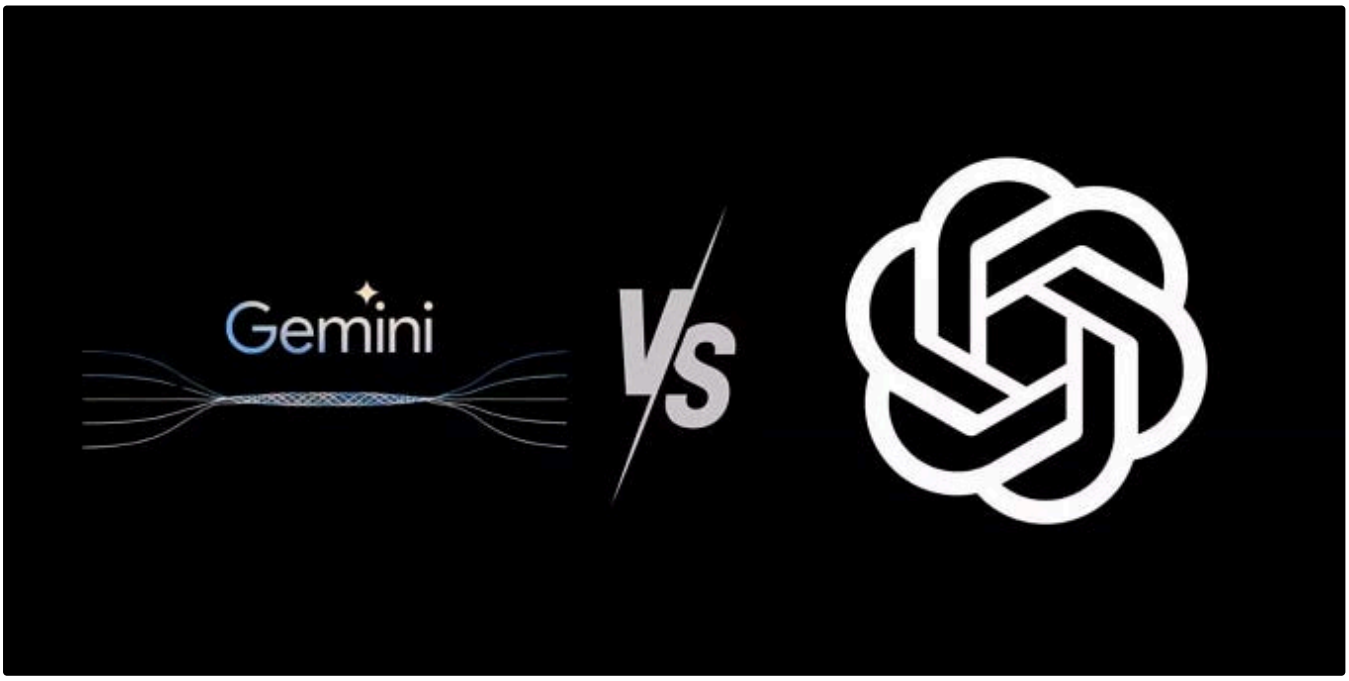
Yogesh Haribhau Kulkarni (PhD) in Technology Hits

Unveiling Manifold learning

What a neural network is really doing?

5 min read · Feb 23, 2024

275 1



Hammad Hassan in Technology Hits

ChatGPT Vs Gemini: Which One I Like As A Writer

As a writer, I know the struggle of digging through mountains of information.

✦ · 4 min read · Feb 8, 2024

1.8K 44

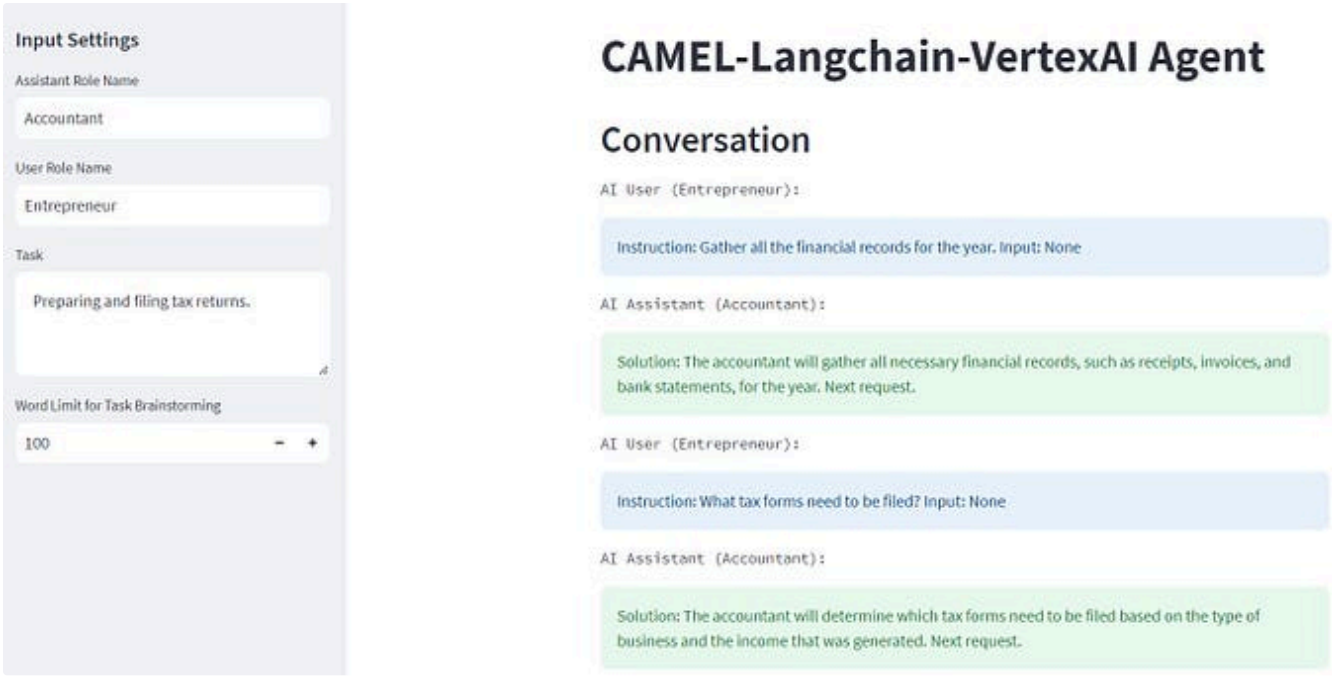
Nuno Campos in Technology Hits

Obsidian Plugins Review—44

Elevate Your Obsidian Experience: Discover the Best New Plugins for Enhanced Note-Taking and Boost Productivity

✦ · 5 min read · Feb 10, 2024

434 4

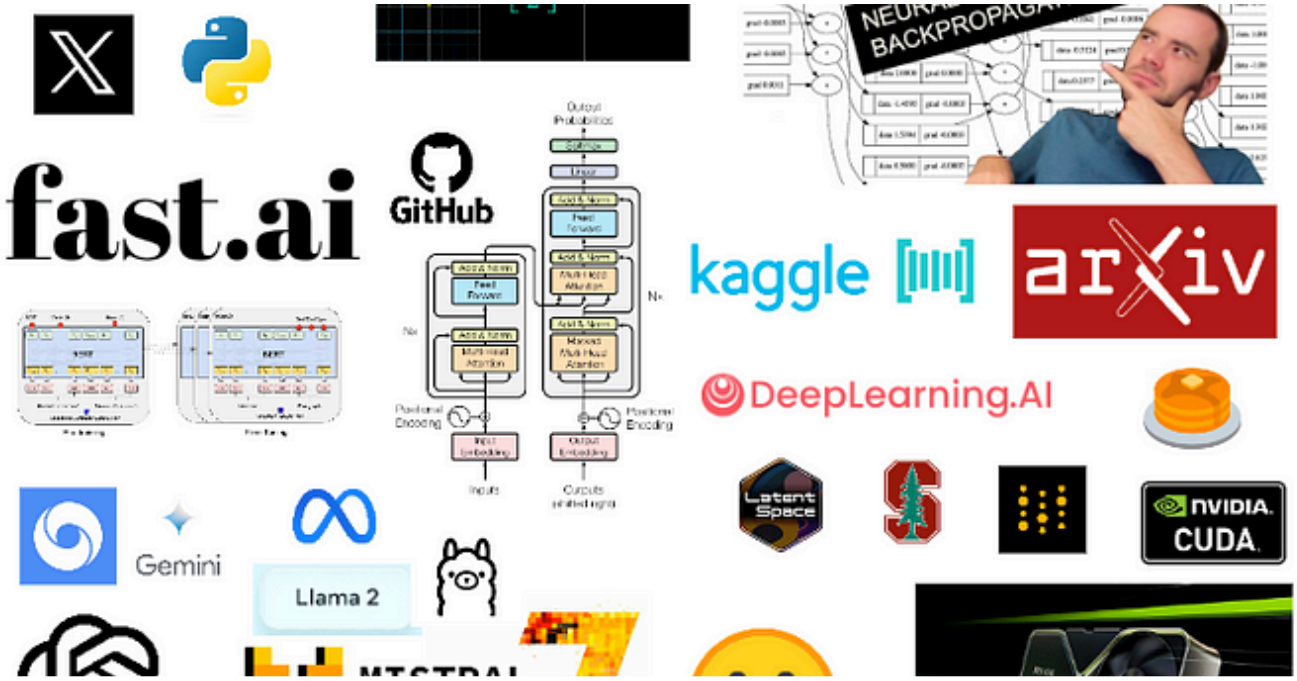


Yogesh Haribhau Kulkarni (PhD) in Google Developer Experts

A CAMEL ride

A Story of AI Role-Playing using CAMEL, Langchain and VertexAI

8 min read · Oct 2, 2023



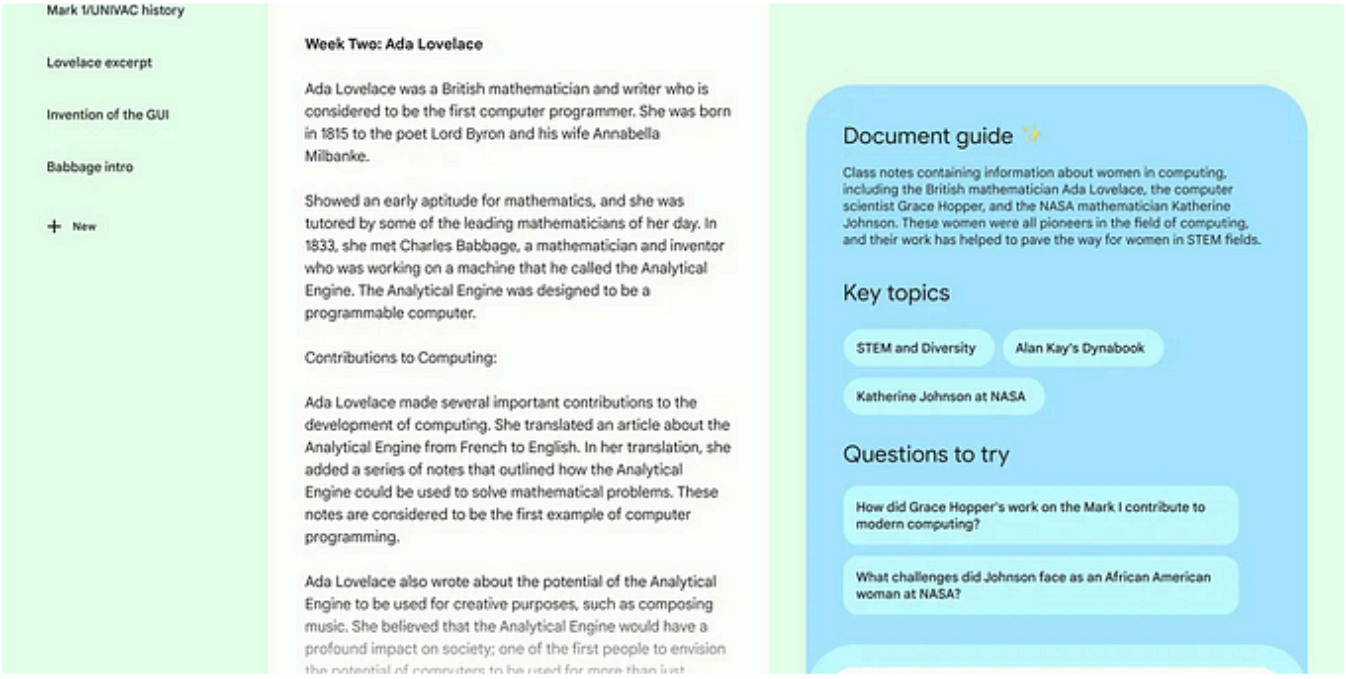
Benedict Neo in bitgrit Data Science Publication

Roadmap to Learn AI in 2024

A free curriculum for hackers and programmers to learn AI

11 min read · Feb 21, 2024

3.9K 45



Jano le Roux in The Startup

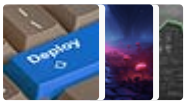
This New AI Tool From Google Is The Closest Thing To A Second Brain I've Seen

Google's NotebookLM is the best thing that has happened to note-taking in the last 1,000 years.

✦ · 5 min read · 3 days ago

575 10

Lists



Predictive Modeling w/ Python
20 stories · 967 saves

AI Regulation
6 stories · 349 saves

ChatGPT prompts
44 stories · 1203 saves

ChatGPT
21 stories · 499 saves

Louis Chan in Towards Data Science

Can You Even __init__.py?

3 things __init__.py can do for you

· 6 min read · 4 days ago

482 2

Christianlauer in CodeX

Google just launched Time Series and Range Functions for BigQuery

How to perform Time Series Analysis with GoogleSQL

· 5 min read · 5 days ago

357 3

Maria Mouschoutzi in Code Like A Girl

Exploratory Data Analysis and Visualization of Maritime Piracy Attacks Dataset

Patterns, trends, and insights in maritime security threats

· 8 min read · 4 days ago

311

Gencay I. in DataDrivenInvestor

5 Pandas Tricks That Will Blow Your Mind As a Data Scientist

You Should Know These Tricks as a Data Scientist

 · 6 min read · Feb 21, 2024

 217