Parul Pandey

parulpandey.com

I bring together data science expertise and developer advocacy to bridge the gap between technology and its users. I am particularly passionate about Machine Learning for High-Risk Applications, where the stakes are significant, and the need for accuracy and ethical considerations is paramount. This led me to co-author a [book](#) in this domain. I am also deeply involved in open source and developer communities, which helps me stay connected and contribute meaningfully. I am a frequent speaker at [conferences](#) and meetups, and I also write on my [blog](#).

Professional Experience

Principal Data Scientist | H2O.ai | October 2022 – November 2024

I rejoined H2O.ai to focus on their Generative AI tools, including **H2O LLM Studio**, **h2oGPT**, and their open-source LLMs like **Danube** and **Mississippi**. My role revolves around community engagement, content creation, and increasing visibility in the research domain. Key contributions include:

- **Community & Content Initiatives:**
  - Launched a company newsletter to keep the community informed and engaged.
  - Highlighted **H2O.ai's research activities** to establish a stronger presence in the research domain.
  - Authored and published insightful technical content on the H2O blog:

    - [Effortless Fine-Tuning of Large Language Models with Open-Source H2O LLM Studio](#)

    - [Explaining Models Built in H2O-3 — Part 1](#)
    - [Winners' Insights: Navigating the Parkinson's Disease Prediction Challenge with AI](#)
    - [Enhancing H2O Model Validation App with h2oGPT Integration](#)
    - [Improving Search Query Accuracy: A Beginner's Guide to Text Regression with H2O Hydrogen Torch](#)
    - [AI for Climate Science: Insights from the Leap Atmospheric Physics Team](#)

Machine Learning Engineer | Weights & Biases | January 2022 – October 2022

My role was to create a strong presence for the Weights & Biases tool within the developer community by working with the product, engineering, and growth teams in unison. I also showcased use cases and best examples of the tool to increase its adoption and create a community of dedicated Weights & Biases champions. I also led the Weights & Biases Kaggle Ambassadors project.

Data Scientist | H2O.ai | July 2019-Jan 2022

I worked and contributed to their open-source product called H2O Wave and their flagship product, Driverless AI. I worked with the customers and academia and acted as a bridge between H2O.ai and

the community through various community events. I have also represented H2O at conferences, events, and panel discussions. I also provided feedback to the product and engineering teams based on my interactions with the community. Here is a list of a few articles I wrote for H2O.ai:

- The Emergence of Automated Machine Learning in Industry
- Speed up your Data Analysis with Python's Datatable package
- Driverless AI can help you choose what you consume next
- Detecting Sarcasm is difficult, but AI may have an answer
- A Deep Dive into H2O's AutoML

Manager | Tata Power | June 2009 – July 2016

My team designed analysis systems to extract meaning from large-scale electronic meter data. We analyzed Consumption and consumer insights to forecast supply and demand peaks. Additionally, we analyzed usage patterns, payment history, and other customer data to detect energy thefts. As part of my job, I worked on SCADA (supervisory control and data acquisition) and Geographical Information System(GIS).

Author | O'Reilly Media

We co-authored a book titled **Machine Learning for High-Risk Applications** for O'Reilly Media, focusing on showcasing technical approaches for Responsible AI at all stages of system development. We have highlighted essential strategies in this book to make our models more robust and secure while being interpretable. I have primarily worked on penalized Generalised Linear Models(GLMs), Generalized Additive Models(GAMs), GA2M and Explainable Boosting Machines, XGBoost with constraints, explaining Pytorch models, adversarial attacks, and many other domain specifics.

Kaggle Grandmaster | Notebooks | www.kaggle.com/parulpandey

I achieved the title of Kaggle Grandmaster in 2020, becoming the first woman in India and the second globally to earn this distinction in the Notebooks category. My passion lies in explaining concepts and exploring datasets, making the Notebooks category a natural fit for me. At my peak, I ranked 7th in this dynamic category. Additionally, my team secured 6th place in the WIDS Datathon.

Notably, my notebook titled Geek Girls Rising: Myth or Reality! won a prize in Kaggle's 2019 ML and DS Survey. My active engagement on Kaggle has led me to speak at various Kaggle Days Meetups and mentor participants in the 2021 Kaggle BIPOC Grant program.

Open-Source Contributions | https://github.com/parulnith

Some of the open-source projects that I'm actively contributing to are:

- **PiML Toolbox:** An integrated Python toolbox for interpretable machine learning model development and validation.
- **DatatableTon:** A collection of 100 Python exercises designed to teach and learn data analysis using the Datatable library.
- **H2O Wave:** An open-source Python development framework that facilitates the creation of real-time interactive AI applications with sophisticated visualizations.

Community Contributions

- **Medium Blog**: Over 20,000 followers with articles featured in official documentation (e.g., Google What-If Tool, EBMs by Microsoft Research).
- **LinkedIn**: Over 100,000 followers; Currently a Top Voice and was a part of LinkedIn's Creator Accelerator Program (2023).
- Served as a judge for **Kaggle's AI Report 2023**, evaluating essays on ethical principles and risk mitigation strategies in **MachineLearning** applications.
- Also judged Mozilla's **Responsible Computing Challenge (RCC)** in India, selecting universities that integrate ethics into technology curricula.
- **Some Notable Talks & Workshops**:

    - *Machine Learning for High-Risk Application -ODSC East(2023)*
    - *https://www.fiddler.ai/webinars/ai-explained-machine-learning-for-high-risk-applications*
    - *Explainable Machine Learning Models* – O'Reilly Workshop (2022).
    - *Understanding Model Predictions* – PyCon Iran (2022).

Education

Bachelor of Technology (B.Tech) in Electrical & Electronics and Engineering,

National Institute of Technology (NIT), Hamirpur, India

Graduated: 2009