

SHREYANS JAIN Email ID: jshrey8@gmail.com Phone: +91 9166868184 | DOB- 15th May 1993

Independent ML Interpretability Researcher with 8+ years of experience as a Data Scientist, specializing in the development and deployment of OCR systems, search engines, recommender systems, and time series forecasting models.

Also, I'm an independent contributor to an ML Interpretability research project working along with a team from NYU and a contributor to the Open-Source library Minari - Farama Foundation.

Side Quests

- Volunteer Contributor to **Minari - Framework for Offline Reinforcement Learning Research**, A Farama Foundation Project (<https://github.com/Farama-Foundation/Minari>)
- Moderator of a **Reinforcement Learning Paper Reading** Group for the **Machine Learning Collective(MLC)** Community

Work Experience

INTERPRETABILITY RESEARCHER @ INDEPENDENT, BENGALURU (Jun'2024 – Present)

- Replicating [Toy Models of Superposition Paper](#) from Anthropic.
- Working on a problem set analyzing the impact of **Non-Uniform Sparsity among features on Superposition in Toy Model settings**.
- Building Visualize Training Package i.e. open-source package for **Visualizing Training Dynamics & Latent States of a Neural Network** (<https://github.com/shreyansjainn/visualizing-training>).

SENIOR DATA SCIENTIST @ GEP, MUMBAI (Jul'2021 – May'2024)

- Leading a P2P Pod including multiple products (**Invoices OCR, Search, Requisition, Catalog**) involving end-to-end product lifecycle management, engineering, software architecture design, machine learning, deployment, product roadmap design & people management which serves lakhs of users per day.
 - **INVOICE OCR:**
 - Extraction of 17 fields (accuracy ranging from 85-95%) from Invoices across **100,000 different supplier templates**
 - Document Detection for incoming documents (Invoice, Credit Memos, Advanced Invoice)
 - Validation of extracted values against system data using **Fuzzy Text Matching & Text Similarity** methods
 - Line Items Material Classification using **LLMs (google-flan-t5) with an accuracy around 85%**
 - Entity & Intent Detection using **prompt engineering on GPT-3.5 models** used to detect the document type and associated entities from the user query and redirect to the respective document team.
 - Domain Specific and Generic Taxonomy recommendation using **prompt engineering on GPT-3.5 models** with the item name as an input.
 - **SEARCH:**
 - Powering the global **Item catalog, Category & Supplier Search** across the whole platform, including integration with external catalogs
 - **CATALOG:**
 - Catalog Item Recommendation against a similar non-catalog item
 - Catalog Creation Recommendation using historical purchasing patterns
- Developed a generic framework for **API testing (both Synchronous and Asynchronous APIs)** currently used across the AI team.

DATA SCIENTIST - II @ BOOKMYSHOW, MUMBAI (Nov'2019 – Jul'2021)

- Created the **Next-Gen Real Time Centralized Personalization System** which captured the **Affinity (GLOVe), Collaborative (ALS), Content Similarity (Doc2Vec) and Exploration (Reinforcement Learning)** aspects, to power all the recommendations to the end user of the platform across various products.

- Built a **Pricing Platform** to provide **End-to-End solutions** to all Pricing needs including **Dynamic Pricing, Demand Forecasting, Trend Simulations, and Base Price Estimations** enabling the **Democratisation of the overall Strategic Process of Pricing**.
- Conceptualized **User Behaviour Personas** using Behavioural Data which enabled the Product Design team to create **Dynamic and Personalized UI/UX Designs**.
- Created & Deployed an **Engagement Score Metric** to quantify and measure the ongoing **User Engagement** on the platform which was used as an input to various **CRM, Ad-tech and User Profiling** activities.

DATA SCIENTIST : Grade – ASSISTANT MANAGER @ HOTSTAR, MUMBAI (Jan'2019 – Oct'2019)

- Created an **Email Categorization Model** to assist the **Customer Support Team**, allowing them to handle queries in a fast and optimised way.
- Build a **Topic Identification Model** to identify **top issues/concerns** faced by users using email, phone & social media data.
- Created a **User Segmentation Model** of our **customer base** to understand their traits & behaviour and deploy targeted marketing & advertising campaigns suited to each segment.
- Created a **Churn Prediction Model** for our **Cricket user base** which predicts the churn probability of the customer at the end of the month/subscription period.

CONSULTANT @ FRACTAL ANALYTICS, MUMBAI (Mar'2016 – Jan'2019)

- Created a **Recommender System in Python using Collaborative Filtering** for a Major CPG Company for recommending various Products & their Volume to Stores across India.
- Refined the **Consumption Forecasting Model using R** to predict future monthly consumption based on historical trends of the data for around 700 different brand-forms combinations in North America & France.
- Created a **Text Mining Model** to automate the **product mapping** procedure based on the similarity and historical mapping of the products.
- Devised an **advanced business model** to estimate the **daily shipments** required to meet monthly shipments target for a **Major Global** Consumer Packaged Goods Firm, based on historical shipment trends for 9 countries.
- Created & Refined the **Consumption Forecasting Model using R** to predict future monthly consumption based on the past 5 years of data for 400 different brand-forms combinations.
- Designed a **Scenario Testing functionality using Spotfire**, which accepts inputs from users and estimates future inventory levels based on Consumption & Shipment Forecasts.

Skills

- **Technical:** Python, Azure, MLOps (Docker, Kubernetes), ElasticSearch, Unit Testing & API Testing, Azure OCR & Form Recognizers, SQL, Pyspark, Airflow, GCP, AWS
- **Machine Learning:** SVM, Glove, Doc2Vec, Reinforcement Learning, Thompson Sampling, Collaborative Filtering, ALS, LLMs, Google Flan t5, GPT-3.5

Education

- B.Tech | Civil Engineering | MNIT-Jaipur (2011-2015) | CGPA : 7.75/10 | AIEEE Rank- 9295
- 12th | Maheshwari Public School, Jaipur (CBSE Board, 2011) | 88.00 %
- 10th | Maheshwari Public School, Jaipur (CBSE Board, 2009) | 91.00 %