

KRITIK SETH

New York, NY | (551)-344-6726 | kls8193@nyu.edu
linkedin.com/in/kritikseth | github.com/kritikseth | kritikseth.com

EDUCATION

- New York University**, Center for Data Science New York, NY, US
Master of Science in Data Science (GPA 3.45/4.00) May 2024
• Coursework: Machine Learning, Big Data, Computational Cognitive Modelling, Probability & Statistics, Linear Algebra.
- NMIMS University**, Mukesh Patel School of Technology Management & Engineering Mumbai, India
Bachelor of Technology in Data Science with Distinction (GPA 3.88/4.00) May 2022
• Coursework: Statistical Methods 1 & 2, DSA, ML, DL, NLP, Computer Vision, Database Management Systems, etc.

TECHNICAL SKILLS

- **Programming Languages:** Python, R, SQL, C, C++, MATLAB.
- **Tools:** Git, Microsoft Excel, PowerBI, Tableau, Spark, Hadoop, Airflow, Docker, AWS (pursuing)
- **Libraries / Frameworks:** Pandas, Numpy, Scikit-Learn, TensorFlow, PyTorch, NLTK, OpenCV, SciPy, statsmodels.

PROFESSIONAL EXPERIENCE

- Persistent Systems – Academic Intern (Data Science Team)** Jan – Apr 2022
• **Accelerated** manual classification of cells in histopathological images, resulting in **80% increase** in **efficiency**, by building Image Segmentation Models to detect and count different types of cells.
• Engineered pipeline to perform face-matching post-enhancement of government IDs and portrait photos using GANs.
• **Enhanced accuracy** by **15%** and expedited preprocessing with **40% increase** in **speed** to **3 seconds** by streamlining the pipeline to incorporate Deep Learning model for keyword extraction on text, post speech-to-text conversion.
- AkzoNobel – Data Science Intern** Aug – Dec 2021
• **Improved** accuracy of model by **20%**, as measured by its ability to classify colors based on reflection values, by implementing ensemble of Random Forest and Light Gradient Boosting Models using Scikit-Learn.
• **Simplified** color recipe-generating process by building Machine Learning models to generate color recipes using solid colors.
• **Rationalized** relating colors and toners by analyzing large-scale color recipe datasets and performing ETL processes.
- Kenmark ITan Solutions – Junior Data Science Associate** Apr – Jul 2020
• **Led** development of text-cleaning pipeline that **reduced** processing time by **40%** to **7 seconds** and **expedited** integration of data.
• **Improved F1 Score** by **30%** using recurrent neural networks for sentiment analysis on Bert embeddings of cleaned comments.
• Conducted and facilitated knowledge transfer by hosting a **tutoring session** for 11 full-time staff members.
- Sapio Analytics – Data Analyst Intern** Apr – Jun 2020
• **Maximized supply chain efficiency** of delivering Covid-19 vaccines by **designing and publishing** a collaborative dashboard using Tableau and Dash, used AWS to extract key metrics. **Presented** it to the Andhra Pradesh government as a proposal.
• **Analyzed** historical data and market dynamics to predict need of essential supplies at hyper-granular level in India.
• **Managed** SQL database (over 40 tables 100,000 rows each) for COVID-19 Project, used by mobile and web applications.

SELECTED PROJECTS

- Swachhdata** (Regex, Git, PyPi, NLTK, OpenCV, Gensim, NumPy, and Pandas) Aug 2020 – Present
• Developed Swachhdata (open source Python library) that provides simple & efficient data, text and image preprocessing tools.
• It has already been downloaded by over **45,000 users worldwide**.
- Analyzing Optimal Video Game Playing Conditions** (TensorFlow, Scikit-Learn, Scipy, statsmodels, LightGBM) Jan 2023
• Performed predictive and descriptive analysis on Video Game FPS dataset and wrote a report on it (with Dr Pascal Wallisch).
• Collaborated in team of 4 to design & conduct a **statistical test** (Kolmogorov–Smirnov) to check if **Moore's Law** is still valid.
• Trained neural network to predict FPS (**RMSE 0.1025**), built model to recommend optimal settings to play video games on.
- Multi-Disease Detection using Retinal Fundus Images** (PyTorch, TensorFlow, CNN, OpenCV) Oct 2021
• Performed ensemble Deep Learning on Retinal Fundus Images for Multi-Disease Detection (with Prof. Sarada Samantaray).
• Trained **ensemble** Convolutional Neural Network with **0.97 accuracy** to identify 45 diseases from collected images.
- Wherebnb** (Python, Flask, TensorFlow, Scikit-Learn, HTML-CSS, and Tableau) Oct 2020
• Designed website to predict price and popularity of listings using Bayesian decision theory & ML (with Dr Siba Panda)
• Trained an ML model with accuracy of **93.2%**, selected to be displayed as showcase project of the data science department.

ACHIEVEMENTS

- Bajaj Finserv HackRx ML National Hackathon – **Second Place** out of 300 teams in India Jun 2021
- Marsh McLennan Dremio Insights Competition – **Third Place** out of 150 teams in India Jan 2021
- MateLabs Demand Forecasting Challenge – **First Place** out of 100 international participants Sept 2020