

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, USA

Master of Science in Data Science – NLP Concentration (GPA: 3.62/4.00)

May 2024

- **Relevant Courses:** Machine Learning, Big Data, Computational Cognitive Modelling, Probability & Statistics, Natural Language Processing with Representation Learning, Optimization and Computational Linear Algebra, Natural Language Understanding.

NMIMS University, MPSTME

Mumbai, India

Bachelor of Technology in Data Science (GPA: 3.88/4.00)

May 2022

- **Relevant Courses:** Programming in C++, Machine Learning, DL, CV, NLP, Financial Engineering & Risk Management, Financial Institutions & Markets, Business Visualization, Cloud Computing, Statistical Methods, AI, Applied Mathematics.

TECHNICAL SKILLS

- **Programming Languages & Frameworks:** Python, SQL, R, C, C++, MATLAB, PL-SQL, Hadoop (Map-Reduce), Excel, Docker
- **Tools:** Tableau, PowerBI, Snowflake, Airflow, Spark, PyTorch, TensorFlow, scikit-learn, Langchain, SAS, Google Cloud, AWS

RELEVANT EXPERIENCE

Logitix – Data Scientist (Florida, US)

June 2023 – Present

- Trained a machine learning model (XGBoost) to predict ticket tiers with 94% accuracy, added CI/CD pipelines, securing lucrative partnerships with multiple prestigious sports venues and directly **generating \$100K in revenue** through ticket sales.
- Streamlined categorization by performing unsupervised machine learning on ticket sales data, implementing BIRCH and K-Means clustering algorithms to create 5 tiers hence enabling efficient analysis.
- Collaborated with the strategy analyst team to enhance clustering algorithms, boosting model accuracy and reliability, and developed a business solutions dashboard to convey technical insights to non-technical stakeholders through data storytelling.

New York University – Data Science Research Assistant (New York, NY, US)

May 2023 – Present

- Designed and deployed Tableau Dashboard merging energy benchmarking & mortgage lienholder (econometrics) data, providing comprehensive data visualizations of largest financiers of NYC's LL97 carbon emissions to promote sustainable finance.
- Conducted extensive data mining to extract large datasets, facilitating the analysis of sustainable investment trends.

Persistent Systems – Deep Learning Intern (Mumbai, IN)

Jan – April 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.
- Enhanced accuracy by 15% and expedited preprocessing with **40% increase in speed to 3 seconds** by streamlining pipeline to incorporate Deep Learning model for keyword extraction on text, post speech-to-text conversion.

Kenmark ITAN – Junior Data Science Associate (Mumbai, IN)

April – July 2020

- Led development of text-cleaning pipeline, **reducing processing time by 40% to 7 seconds** and expediting integration of data.
- Implemented a baseline recommendation system using sentiment analysis for a client's social media application, leading to an **increase in user retention time by 3 minutes** as validated through **A/B testing**.
- Conducted and facilitated knowledge transfer by hosting a **tutoring session for 11 full-time staff members**.

Sapio Analytics – Data Analyst Intern (Mumbai, IN)

April – June 2020

- **Maximized supply chain efficiency** of COVID-19 vaccine deliveries by spearheading the development of a collaborative dashboard (Tableau & Dash), leveraging AWS to extract key metrics. Presented it to Andhra Pradesh government leaders.
- Analyzed historical data and market trends to predict need of essential supplies at hyper-granular level in India (ad hoc queries).
- Managed SQL database (over 40 tables with 100,000 rows) for COVID-19 Project, integrated by mobile and web applications.

SELECTED PROJECTS

Suspicious Clause Detection in T&C (PyTorch, HuggingFace, NLTK)

Sept – Oct 2023

- Built NLP web app which detected suspicious clauses in lengthy T&C documents by fine-tuning large language models (GPT).

Backtesting Financial Analysts' Future Predictions (Open AI, LangChain, OpenCV, AI)

May – June 2023

- Utilized pre-trained LLMs and Langchain for backtesting, extracting key information and timelines for outcome forecasting.

Moving Target Interception - Multi-Agent Reinforcement Learning (MARL) (Python, Numpy, OpenCV)

Mar – May 2023

- Engineered an innovative MARL framework, training agents to make co-ordinated decisions to capture an evasive thief.

Music Recommendation System (Spark, Dask, Python, Hadoop)

Mar – April 2023

- Developed **collaborative filtering** based music recommendation system on large-scale interactions data (50GB+), achieving 3 fold improvement in mean average precision over baseline.

Analyzing Optimal Video Game Playing Conditions (TensorFlow, sklearn, Scipy, statsmodels, LightGBM)

Nov – Dec 2022

- Collaborated with a cross-functional team to execute a Kolmogorov-Smirnov **statistical test**, validating Moore's Law.
- Trained a neural network model with **2x improvement** in predicting FPS compared to traditional ML approaches.

Swachhdata - 50,000 downloads (Regex, Git, PyPi, NLTK, OpenCV, Gensim, NumPy, and Pandas)

May – Aug 2021

- Programmed **3,000+ lines** to develop Swachhdata library, delivering modular preprocessing tools for data, text, and images.

Wherebnb (Python, Flask, TensorFlow, Scikit-Learn, HTML-CSS, and Tableau, Cloud)

Aug – Oct 2020

- Built an Airbnb clone leveraging Deep Learning for precise price and popularity predictions of real listings.
- Implemented state-of-the-art **text-generation** RNNs to analyze listings and provide hosts with tailored title recommendations.