# 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, Data Science (NLP Specialization) (GPA: 3.73/4.00)** — **May 2024**
- Relevant Courses: Machine Learning, Big Data, Cognitive Modeling (Reinforcement Learning), Probability & Statistics, Natural Language Processing, Optimization & Computational Linear Algebra, Natural Language Understanding NLU, Advanced Python.

**NMIMS University,** MPSTME — **Mumbai, India**
**Bachelor of Technology, Data Science (GPA: 3.88/4.00)** — **May 2022**
- Relevant Courses: Data Structures & Algorithms, Machine Learning ML, Deep Learning DL, Computer Vision CV, NLP, Financial Engineering Risk Management, Business Visualization, Cloud Computing, Statistical Modeling, Artificial Intelligence AI.

## TECHNICAL SKILLS

- Programming Languages: Python, Cython, C, C++, R, SAS, MATLAB, SQL, PL-SQL, NoSQL, PostgreSQL, MySQL, Excel VBA.
- Tools: PyTorch, TensorFlow, Scikit-Learn, Langchain, Tableau, PowerBI, Snowflake, Teradata, Hadoop, MapReduce, PySpark, Spark, Google Cloud Platform GCP, Amazon Web Service AWS, PowerPoint, Excel, Docker, Databricks, Kafka, Hive, Jira, Agile.
- Algorithms: Timeseries, Linear & Logistic Regression, Supervised & Unsupervised Learning

## RELEVANT EXPERIENCE

**NYU Stern School of Business – Data Science Project Lead** (New York, US) — **May 2023 – Present**
- Led development of ***publicly operational*** Carbon Compass tool for 'NYC Local Law 97,' championing energy efficiency in compliant buildings. Spearheaded end to end project management, ensuring seamless execution from ideation to deployment.
- Designed and published Tableau dashboard using data visualization tools merging energy benchmarking & mortgage data from top banks, offering comprehensive view of NYC's LL97 carbon emissions' major financiers for sustainable finance & compliance.

**Memorial Sloan Kettering Cancer Center – Graduate Student Researcher** (New York, US) — **Sept – Dec 2023**
- ***Led a cancer research initiative,*** employing Large Language Models and Named Entity Recognition (NER) to automate gene annotation in research articles. Streamlined updating process of OncoKB database by accelerating gene annotation through the development of a BioMed BERT powered model, ***mitigating manual efforts and reducing time intensive process.***

**Logitix – Data Science Intern** (Florida, US) — **June – Dec 2023**
- Trained an ensemble machine learning model (XGBoost and SVM) to predict ticket tiers with 94% accuracy, securing lucrative partnerships with multiple prestigious sports venues and directly ***generating $100K in revenue*** through ticket sales.
- Leveraged continuous integration and continuous development practices, including test automation and monitoring, to ensure successful deployment of ML models and application code, while ensuring communication with app development team.
- Formulated dynamic pricing problem as price forecasting problem and developed custom analytical explainable models that generated insights to help the pricing team, ***reduced the pricing decision making time by 15 minutes.***
- Collaborated with data analytics team to enhance clustering algorithms, focusing on business objectives and model accuracy. Developed business solutions dashboard to convey technical insights to non-technical stakeholders through data storytelling.

**Persistent Systems – Machine 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 data pipeline to incorporate Deep Learning model for keyword extraction on text, post speech-to-text conversion.

**AkzoNobel – Data Science Intern** (Mumbai, IN) — **Aug 2021 – Mar 2022**
- ***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 (classification) using Voting classifier.

**Kenmark ITAN – Junior Data Science Associate** (Mumbai, IN) — **April – July 2022**
- 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***. Authored end to end documentation.

**Sapio Analytics – Data Analyst Intern** (Mumbai, IN) — **April – June 2022**
- ***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 3 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).

## SELECTED PROJECTS

**Suspicious Clause Detection in T&C Documents** (TensorFlow, HuggingFace, NLTK) — **Nov – Dec 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** (OpenAI, LangChain) — **May – June 2023**
- Utilized LLMs and LangChain for backtesting, extracting key information and timelines for outcome detection.

**Moving Target Interception – Multi-Agent Reinforcement Learning (MARL)** (Python, NumPy, OpenCV) — **Mar – May 2023**
- Engineered and published a MARL framework, training agents to make coordinated decisions to capture an evasive thief.

**Music Recommendation System** (Spark, Dask, Python, Hadoop, NumPy) — **Mar – April 2023**
- Developed ***collaborative filtering*** based music recommendation system on large-scale interactions data (50GB+), achieving 3 fold improvement in mean average precision (MAP) over baseline. Performed data mining to improve model.

**Analyzing Optimal Video Game Playing Condition** (PyTorch, sklearn, Scipy, statsmodels, statistical testing) — **Nov – Dec 2022**
- Collaborated with a cross-functional team to execute Kolmogorov-Smirnov ***statistical test***, validating Moore's Law.
- Trained a neural network model with 2x improvement in predicting FPS compared to traditional ML approaches (regression).

**Swachhdata – *60,000 downloads*** (Regex, Git, PyPi, NLTK, OpenCV, Gensim, NumPy and Pandas) — **May 2021 – Present**
- Programmed 3000+ lines to develop this library, delivering modular preprocessing & pipeline tools for data, text and image ETL.

**Wherebnb** (Python, Flask, TensorFlow, Scikit-Learn, HTML-CSS, JavaScript and Tableau) — **Aug – Oct 2020**
- Built Airbnb clone and used Deep Learning for price predictions of real listings and provided data analysis using real Airbnb dataset.
- Implemented state-of-the-art text generation LSTMs to analyze listings and provide hosts with tailored title recommendations.