Sitanshu Kushwaha Data and Software

sitanshu.kushwaha@nyu.edu +1 (929) 643-4480 sitanshu.tech Brooklyn, NY
linkedin.com/in/sitanshukushwaha/ github.com/Sitanshuk

EDUCATION

New York University, MS in Computer Science Sep 2023 – May 2025 | New York
Relevant Courses: Big Data, Data Science, Computer Vision, Blockchain

University of Mumbai, BE in Computer Engineering Aug 2016 – Nov 2020 | Mumbai
9.32/10 CGPA (Rank: Top 10)

PROFESSIONAL EXPERIENCE

IT Technologist Intern, NYU IT Oct 2023 – present | New York
- Implemented **ELT** pipeline with **Snowflake** & **AWS**, integrating diverse data sources (Data Warehouses & Data Marts) for enhanced analysis & decision-making.

Data Engineer, LTIMindtree Jan 2021 – Jun 2023 | Mumbai
Technical Lead, Visioncare MFF Data Engineering team - Johnson and Johnson
- Spearheaded **optimization** efforts in Databricks Spark code, resulting in a **30% reduction** in **execution time** for 50% of Transformation Jobs, significantly improving data timeliness.
- Adopted **event-based triggers** within Azure Data Factory for **ETL** pipelines, strategically **enhancing** the **efficiency** of handling Big Data from multiple sources, **reducing costs by 25%**
- Designed **Monitoring Tableau Dashboard** for the data flow architecture, improving visibility for early identification and resolution of **bottlenecks** and other issues, and **mitigated unforeseen outages by 40%.**
- Optimized deployment processes by implementing **CI/CD** using **Git** and **ARM** templates (**IaC**) in Azure, resulting in a **70% reduction** in **deployment time**.

Machine Learning Engineer, Oniria Creations Mar 2020 – Dec 2020 | Remote, Poland
- Developed a high-accuracy CNN TensorFlow model to precisely identify Pet service provider websites from Bing search results, achieving an impressive **92% accuracy** rate.
- Automated the extraction of valuable CRM Data from service providers' websites using **Natural language processing** and ML techniques, leveraging the powerful **BERT language model (LLM)**.

Intern Machine learning Engineer, Toflo Fintech Consulting Dec 2019 – Mar 2020 | Mumbai
- Engineered a **Recommender System** for their e-commerce platform, leveraging **Big Data** concepts **Increasing click-through rate by 33%** for products displayed on a recommended section of the page.

SKILLS

Big Data (Apache Spark, Kafka, Databricks, Azure Data Factory, AWS, GCP, Data Lake, ETL, NoSQL),
Machine Learning (Scikit Learn, Tensorflow, NLP, Neural Networks, Deep Learning),
Data Analytics (SQL, Pandas, Numpy, Matplotlib, Seaborn, Web Scraping, Tableau),
Languages (Python, JAVA, R.), **Tools** (Git, Docker, Databricks MLOps)

PROJECTS

DineSync - Real-Time Culinary Exploration in NYC, (Big Data, Spark, Kafka, MongoDB, Django)
- Engineered DineSync, a **real-time restaurant recommendation** system leveraging **Kafka** for **processing live** user check-ins, ensuring accurate seat availability data.
- Implemented **live updates** to seamlessly **recommend alternative** restaurants if the user's preferred choice is fully occupied, significantly **enhancing the user experience.**

Subjective Answer Evaluation using Machine Learning, (NLP, Django, TensorFlow)
- Utilized state-of-the-art **NLP** techniques, including **BERT**, **USE**, and **Word2Vec** language models, to assess students' subjective answers by measuring **semantic similarity** against the teacher's answer.
- Published a **research paper** in International Journal for Scientific Research and Development

Classification of Punjabi BBC Articles, (NLP, Keras)
- Developed a generic **Punjabi language model** using a public corpus for effective processing and understanding of Punjabi text, alongside a **sentiment classifier** with **87% accuracy** for categorizing news articles as political or non-political.

Geolocation Data Preprocessing and Clustering, (Unsupervised learning, Geospatial Data Analysis)
- Processed and cleansed **geolocation data**, ensuring data quality and reliability, and applied advanced **clustering** algorithms, including **K-Means** and **DBSCAN**, to analyze **proximity** and **density** patterns.
- Evaluated and **compared** clustering models' **performance**, providing valuable insights for optimal selection, and **visualized** geolocation **data** interactively to uncover and **analyze spatial patterns**.