

Samir Thota

Dallas, TX | (972)-215-9063 | thota@berkeley.edu

EDUCATION

University of California, Berkeley, Berkeley, CA

Expected: May 2024

Bachelor of Science in Data Science, Concentration in Business and Industrial Analytics

Certifications: Amazon Cloud Practitioner, Databricks Certified Data Engineer, Snowflake Snow Pro, Certified Python Associate

Programming Coursework: Structure and Interpretation of Computer Programs, Data Structures, Probability for Data Science, Industrial Data Systems, Data Mining and Analytics, Machine Learning, Corporate Finance

PROFESSIONAL EXPERIENCE

Nova Energy Capital, Remote

January 2024 – Present

Data Engineer Intern

- Implemented predictive pricing APIs for the energy trading market processing over 1,000,000 data points
- Developed real-time and batch Data Pipelines, resulting in a 8% increase in data accuracy

DataTron, Dallas, TX

June 2023 – July 2023

Data Engineer Intern

- Worked on model catalog to monitor models for bias, drift, & performance anomalies in real-time
- Revised source code to help facilitate JupyterHub-integration for improved data exploration

SourceLead, Frisco, TX

June 2022 – August 2022

Technology Intern

- Optimized search results on the Application Tracking system to decrease run time by 10%
- Conducted testing on alpha software to identify bugs and provide recommendations

Dallas Venture Capital, Dallas, TX

June 2021 – October 2021

Analyst Intern

- Sourced and evaluated key-metrics in 2000 companies within the SaaS sector to narrow down candidates
- Invited and managed 50 investors and Fund Managers to collaborate on fund administration CRM

LEADERSHIP & AFFILIATIONS

University of California, Berkeley Track and Field, Berkeley, CA

August 2020 – May 2024

- Dedicated 20-25 hours weekly for conditioning, meetings, travel, and practice as a scholarship athlete
- Gained and demonstrated valuable leadership and communication skills as part of a team of 100+

Delta Consulting, Berkeley, CA

August 2020 - May 2022

- Drove market research efforts for WikiEducation with an in-depth analysis of prevailing market trends
- Administered over 20 expert testimonials for a UK EdTech firm aiming to procure a 50k government grant

TECHNICAL PROJECTS

Cloud-based IMDB Data Warehouse - AWS(SQL)

August 2023

- Performed ETL operations within the buckets and processed within a Redshift data warehouse
- Designed and managed an end-to-end data pipeline initialized with S3 buckets

NBA Career Predictor- Python(Pandas,Sklearn,Matplotlib)

August 2023

- Cleaned data and manipulated multiple datasets to create to facilitate a comprehensive analysis
- Constructed an optimized regression model to forecast 1000s of specific player career outcomes
- Designed visualizations to provide insights and influence decision-making with data driven metrics

Spam & Ham- Python(Pandas,Sklearn)

April 2023

- Developed a binary classifier based on feature engineering and logistic regression to identify spam emails
- Validated performance classifiers through testing to reduce error and mitigate overfitting

SKILLS

Languages - Python, Java, MySQL/SQL, Tableau, R

Tools/Frameworks - Pyspark, Pandas, Sklearn, Numpy, Django

Technologies - Kubernetes, Terraform, AWS, Git