TAZEB ABERA

8801 Edna Place Rowlett, TX 75089
Phone: (214)430-2241
Email: tazabera@gmail.com
Website: https://mydatascienceenthusiast.com/
Linked Page: https://www.linkedin.com/in/tazeb-abera
GitHub: https://github.com/tazeb6531

Professional Summary

Innovative Senior Data Scientist with extensive experience delivering impactful AI-driven solutions across healthcare, manufacturing, business operations, and engineering materials sectors. At Celanese, successfully applied advanced machine learning techniques to refine search engine accuracy and streamline competitor analysis, achieved a 60% accuracy boost, 40% faster search times, and 15% reduction in development process time. Proficient in integrating advanced tools like Elasticsearch, Azure GenAI, and Snowflake Cortex to foster innovation and enhance decision-making processes. Skilled in NLP, deep learning, predictive modeling, and generative AI, with a focus on improving operational efficiency and driving strategic insights.

Demonstrated leadership in managing cross-functional teams to deliver scalable cloud-based pipelines, Streamlit Dashboards, and custom analytics platforms for organizations like NYCHHC, Samsung, Change Healthcare, and QORVO. Expertise spans EPIC healthcare data, cloud computing, and advanced programming languages, including Python, SQL, and DevOps tools. Holds advanced degrees in Data Science, Machine Learning, and Engineering, and is currently pursuing a PhD. Dedicated to utilizing advanced technologies to solve complex problems, foster innovation, and deliver measurable business outcomes.

Education Background

NorthCentral University -- *PhD in Data Science (not awarded)*
Southern Methodist University – *Master of Science in Data Science.*
Bahir Dar University – *Master of Science in Sustainable Energy Engineering.*
Jimma University – *bachelor's degree in computer science.*

Professional Experience

Senior Data Scientist   06/2024 to Present New *York City Health + Hospitals (NYCHHC) - Remote*

- Led development of Generative AI models to streamline hospital billing (HB) and professional billing (PB) transaction analysis.
- Built Eligibility Prediction models leveraging EPIC healthcare data to improve operational efficiency.
- Developed RAG (Retrieval-Augmented Generation) models to extract and analyze insights from unstructured medical notes.
- Integrated Snowflake Cortex with Snowflake data warehouses to implement scalable and efficient AI-powered pipelines.
- Designed Streamlit Dashboards for real-time data visualization and encounter inspection, enhancing charge capture and identifying anomalies.
- Reduced revenue leakage between HB and PB transactions through semantic analysis.
- Partnered with cross-functional teams to align AI and data solutions with NYCHHC's clinical and financial workflows.
- Automated complex SQL-based data extraction pipelines for performance optimization.

Senior Data Scientist   07/2021 to 05/2024 Celanese – Irving, TX

- Coordinated a skilled team in Agile development of a sophisticated AI-driven search engine, utilizing advanced GenAI, clustering algorithms, and various machine learning models for optimal performance.
- Integration of Advanced Tools: Worked extensively with Elasticsearch and Azure GenAI, integrating them with NER to enhance search capabilities and data analysis.

- Analytics and Workflow Optimization: Enhanced operational efficiency through data-driven strategies and optimized workflows.
- Data Cleaning and Preparation: Conducted thorough data cleaning, feature engineering, and exploratory data analysis (EDA) to ensure high-quality data for analysis.
- Predictive Modeling: Developed advanced predictive ML models to identify competitor producers in the engineering materials sector, providing strategic insights.
- AI-Driven Search Engine Development: Led the creation of an AI-driven search engine with advanced similarity search and entity recognition techniques, improving data retrieval accuracy.
- NLP Implementation: Applied NLP techniques, including entity recognition, to improve search engine operations and enhance data retrieval accuracy.
- Deep Learning Applications: Utilized deep learning methodologies to develop sophisticated models and solutions, advancing the company's AI capabilities.

Data Science Part time 07 /2019 to12/2020 **Change Healthcare**– Dallas, TX
- Python Scripting: Wrote scripts in Python to extract data from HTML files, ensuring efficient data retrieval and processing.
- Business and Medical Procedures Evaluation: Assessed business and medical treatment procedures, incorporating them into the database system for improved management and accessibility.
- Clinical Data Integration: Coordinated with the technical team to integrate clinical data, enhancing the completeness and usability of medical records.
- Statistical Analysis: Utilized various statistical methods and tools to measure performance variables, providing critical insights into operational and treatment efficiencies.
- Custom Reporting: Generated custom reports and created statistical control process charts to monitor and evaluate key metrics.
- Process Improvement: Suggested corrective actions to improve internal procedures and data management practices, driving continuous improvement.
- System Design and Implementation: Designed, documented, tested, and implemented a clinical data collection system, ensuring robust and reliable data capture.

Data Scientist 06/2018 to 06/2019 **QORVO** – Richardson, TX
- Developed and maintained statistical algorithms and models with the focus on big data.
- Core team player in performing data analysis for metrics and decision-making process; directionally goal oriented to bring in defined results.
- Performed EDA and feature engineering to both inform the development of statistical models and improve model performance and flexibility.
- Provide analysis using mathematical modeling tools to improve business processes and decisions.
- Assist with the development of standardized reporting solutions that align metrics and drivers with the associated financials.
- Create dashboards and reports to regularly communicate results and monitor key metrics.

Business Analyst 12/2016 to 05/2018 **Samsung** – Plano, TX
- Understand market volatility, analyze customers' needs and/or market competitiveness to drive accuracy of future mobile phone prices.
- Reduced development costs by 25% by consolidating related products into streamlined solutions.
- Correlation analysis for retailer business.
- Predicting the business performance of the company.
- Prepared interactive reports to organically feed into our reporting tools such as Tableau and R shine.
- Experienced in developing business reports by writing complex SQL queries using views, macros, volatile and global

temporary tables.

- Strong Experience in working on MS Access, MS Excel, and MS Power Point
- Experienced in handling all the domain and technical interaction with application users, analyzing client business processes, documenting business requirements
- Possess strong analytical and problem-solving skills and have quick learning curve. Committed team player and capable of working on tight project delivery schedules and deadlines.
- Proficiency in prioritizing and multi-tasking to ensure that tasks are completed on time.
- Demonstrate a willingness, interest, and aptitude to learn new technologies and skills.
- Good Communication and interpersonal skills.

Data Scientist 01/2015 to 09/2016 **Power Africa|USAID** – Bahir Dar, Ethiopia

- Led a team of over 150 people countrywide over more than 8 regional states.
- Defined work scope procedures for lower-level managers, overcome any technical and legal challenges by partnering with government leaders under the umbrella of Sustainable Development and Poverty Reduction mission.
- Examined Produced Energy Data, arrange results in a presentable and coherent way for upper management and stakeholders.
- Embodied a team of 27 data scientists under my organization who focused on data produced and analyzed from bioenergy, wind, and solar energy.
- Responsible for debugging the project.
- Built user defined function (UDF) and performing various CRUD operations like create, update, read and delete.
- Created Python scripts for merging data into MySQL database.
- Communicating with external venders and high-level officials to resolve queries.
- Report results and SWOT analysis report to high management body.
- Maintain search features/data mapping diagrams.

Power Engineer 01/2015 to 09/2016 Organization and rehabilitation Development in Amhara Region **ORDA** – Bahir Dar,

- Identified and resolved abnormalities with power production equipment and processes.
- Monitored biomass operations for compliance with regulatory requirements.
- Oversaw a well-organized maintenance program for predictive and preventive actions.
- Optimize financial performance.
- Cost - profit analysis for solar panels and other equipment imported from different countries.
- Prepare and present Project performance Report for high level officials.

Skills

- **Programming Languages:** Python, R, JAVA, SAS, SQL, HTML, CSS.
- **Data Science Tools:** PyTorch, TensorFlow, Keras, SciPy, Pandas, NumPy, Jupyter Notebook, Data Bricks.
- **Data Visualization:** Matplotlib, Seaborn, Tableau, Power BI.
- **Cloud Platforms and DevOps:** Azure, AWS, Snowflake, Docker, Kubernetes, GitHub, CI/CD.
- **Machine Learning & AI:** GenAI, LangChain, RAG, NLP, Hugging Face, Llama2, Transformers, BERT, Snowflake Cortex.
- **Big Data & Databases**: Apache Spark, Hadoop, Airflow, MySQL, SQL Server, Snowflake, NoSQL (Cassandra), SSIS, SSRS, ETL.
- **Statistical Tools:** MATLAB, SAS, Statistical Modeling, Predictive Analytics.
- **Soft Skills:** Problem Solving, Project Management, Agile Methodologies.