SHUMETIE TEFERA

DATA SCIENTIST

Tel: 213-352-9188 **E-Mail**: tshumetie5@gmail.com **GitHub**: https://github.com/shum05
LinkedIn: https://linkedin.com/in/tshumetie5 **Website**: s3shumetie23.com

Skills

Programming Languages: Python, R, SQL, JavaScript
Data Visualization: Power BI, Tableau, Excel, Looker Studio, Matplotlib, Seaborn
Machine Learning: Clustering, Predictive Modeling, Unsupervised Learning, Supervised Learning, Reinforcement Learning, Bagging and Boosting, Linear Models, Support Vector Machines (SVM), Decision Trees, Random Forest, K-Nearest Neighbors (KNN), Gradient Boosting, Classification, Regression, Clustering, Feature Selection, Model Evaluation
Deep Learning: CNN, RNN, LSTM, GAN, Transfer Learning, Autoencoders
Natural Language Processing (NLP): Text Mining, Sentiment Analysis, Word Embeddings, Tokenization, Topic Modeling, Text Classification, Word2Vec, SpaCy, NLTK
Data Analysis: data collection, cleansing, transformation, data modeling, mining, data warehousing, wrangling
Statistical Analysis: hypothesis testing, Time Series Analysis, Descriptive Statistics, Inferential Statistics, Regression Analysis, statistical models
Data Tools and Cloud Technologies: Flask, Docker, CI/CD, Jupyter Notebooks, VSCode, PyTorch, NoSQL(MongoDB), TensorFlow, Keras, Pandas, NumPy, snowflake, AWS, Microsoft Azure
Communication: Excellent verbal and written communication

Education and Certifications

Addis Ababa University, **M.Sc. Applied Mathematics- Numerical Analysis** -Sep 2010 – Jul 2012
Mekelle University, **B.Sc. Civil Engineering** -Sep 2013 – Feb 2016
Careerera, **Post Graduate Program Certificate in Data Science** -Jun 2022 – May 2023
AWS Certified **Solutions Architect-Associate**
Salesforce Certified **MuleSoft Developer** (API)

Professional Experience

Careerera– Herndon, VA - Data Science Intern (*Jun 2023 – Nov 2023)*

- Conducted NLP sentiment analysis, leading to scaled up customer feedback and analyzed Time Series data, reducing forecasting errors by 5%
- Crafted visually stunning data visualizations using Power BI, Tableau; facilitated clear communication of complex financial data, enabling strategic decisions resulting in 10% increase in company profitability
- Enriched data quality with 15% fewer errors by building and evaluating ML models for customer churn prediction, achieving a 20% reduction in customer attrition

McKesson Corporation – Chino, CA- Data Analyst (*Jul 2023 – Jun 2023)*

- Resolved data inconsistencies at 98% data accuracy, developed automated data cleaning scripts using Python, saving 10+ hours per week
- Spearheaded advanced statistical techniques to identify trends, patterns, and correlations to maximize customer retention and a 10% reduction in churn rate

DigiSure Insurance Solutions Inc –Burlingame, CA- Data Analyst (*Sep 2021 – Jun 2022)*

- Executed market research on customer preferences and industry trends for informed product development strategies, leading to a $1.5M revenue growth
- Engineered visual reports, KPI scorecards, filters, parameters using Power BI desktop for claim analysis and leveraged DAX formulas to transform raw data into actionable insights, driving revenue growth by 15%

First General Bank- Los Angeles, CA- SQL/ETL Developer (*Sep 2021 – Jun 2022)*

- Exploited Shell Scripting techniques to automate data migrations, prompting a 50% cut down in migration time, 30% reduction in maintenance time and minimal data loss
- Optimized SQL queries, achieving a 15% increase in database efficiency and 40% brought down query response times with 25% improvement in report generation time

Projects

Question-Answering Tool for Bridge Design Manuals - Python, Haystack, Mistral, ChainLit

- Designed a robust Question Answering tool for civil engineering Bridge Design Manuals using Python, Haystack, Mistral, and ChainLit to extract answers from a dataset, enhancing access to critical information and streamlining the design process

Forecasting Economic Evolution- Time Series Analysis (Python, Facebook Prophet)

- Utilized Python, Facebook Prophet, and Seaborn to conduct a comprehensive time series analysis of currency exchange rates, forecasting future trends and uncovering valuable insights to support informed financial decisions for a 5-year growth and transformation plan

Insurance Claim Analysis – Python, Pandas, Power BI

- Utilized Python and Pandas for data analysis, alongside Power BI for data visualization, to analyze insurance claim data, leading to a 15% performance improvement and a cost reduction of $2 million in the insurance claims process