Sai Surya Varshith Nukala

Austin, TX| (857) 423-4444| saisuryavarshithnukala@gmail.com

EDUCATION

Master of Science, Computer Science   Jan 2024
Boston University Graduate School of Arts and Sciences, Boston, MA
Relevant Courses: Machine Learning, Data Science Tools and Applications, Computer Vision, Natural Language Processing

Bachelor of Technology, Computer Science and Engineering   June 2022
Vellore Institute of Technology, Vellore, India

WORK EXPERIENCE

Data Analyst| *Great American Insurance Group***|** *Austin,TX*   Mar 2024 –Present
- Monitored and maintained AWS-based ETL pipelines using AWS Glue, Lambda, and S3 to preprocess and integrate policyholder and claims datasets for analysis.
- Analyzed insurance claims data using Python and AWS Athena, identifying trends and providing actionable insights for underwriting and claims management teams.
- Collaborated in developing a proof-of-concept predictive model using AWS SageMaker to forecast claims severity, supporting enhanced risk evaluation processes.
- Developed interactive dashboards in Tableau to visualize policy renewal rates, claims performance, and customer segmentation, enabling stakeholders to make data-driven decisions.

Machine Learning Engineer Intern| *Atfal Mafkoda* **|** *BU Spark!* **|** *Boston, MA*   Sep 2023 – Dec 2023
- To identify missing individuals in anonymous image collections, a facial detection and matching pipeline was developed using MTCNN and VGG-Face within the **DeepFace** framework, streamlining the identification process.
- Leveraged **Streamlit** to create an interactive interface for viewing potential matches at various thresholds, which significantly cut down manual photo reviews and sped up the identification of missing individuals.
- Constructed an automated system to process and organize image data, leveraging API-sourced JSON files, for 2000 cases.
- Achieved **92%** accuracy by evaluating various validation metrics and chose cosine metric for its ability to precisely identify matches, narrowed the review pool size by **80%.**

Teaching Assistant, Advanced Analytics II | *Boston University* **|** *Boston, MA*   Sept 2023 – Dec 2023
- Identified actionable insights to improve the curriculum by analyzing trends in post lecture surveys combined with assessment performance of the students, enhancing the learning of the course.
- Performed quality assurance tasks, such as evaluating Python assessments and projects, and provided targeted recommendations to enhance students' conceptual understanding and improve the quality of their projects during office hours.

Data Science Intern | *Flutch Technologies Pvt Ltd.* **|** *Bangalore, India*   April 2021 – Jun 2022
- Built AWS-based ETL pipelines using AWS Glue, S3, and Lambda to ingest, process, and transform large influencer datasets, streamlining campaign analytics workflows.
- Applied clustering and regression models to segment YouTube influencers and predict engagement metrics, leveraging Amazon SageMaker for model development and deployment.
- Conducted sentiment analysis on audience comments using NLP techniques with Amazon Comprehend, improving alignment between promotional efforts and audience sentiment.
- Designed and automated recommendation systems using PySpark, AWS Glue, and Step Functions, enabling scalable data workflows and performance insights.

CERTIFICATIONS

AWS Certified: Data Engineer – Associate : https://www.credly.com/badges/2d24cd2a-3fe1-4188-9a59-20c4279e7021/linked_in_profile
Snowflake SnowPro Core :https://achieve.snowflake.com/63c5c68f-d251-4c32-9ed1-ec2a59a029b1?record_view=true#gsbnne79

PROJECT EXPERIENCE

An Interactive Chatbot to Navigate Reddit Discussions| *Boston University*   Nov 2023 – Dec 2023
- Architected a Rasa-based chatbot, with the BERTopic model, adept at engaging users and delivering subreddit discussions tuned to their interests.
- Employed topic modeling to categorize around 150k subreddit posts, enabling the chatbot to provide context-aware results, adapting to real-time factors such as season and user engagement of each post.
- Enhanced model performance by using appropriate text preprocessing strategies, including contextual emoji translations and outlier noise elimination, to standardize the data.

RELEVANT SKILLS

Key Skills: Modelling (Regression, Tree based, K-Mean, CNN, Timeseries), Trend analysis, Sentiment analysis, Data scraping, Data Mining, Data Analysis, Data Preprocessing, A/B testing, Hypothesis testing, Advance EDA, AWS (Sage Maker,Lambda Functions,S3)
Tools and Languages: Python (SciKit-Learn, keras, TensorFlow, PyTorch, SpaCy, Gensim, NLTK etc.), Git
Databases and Data Visualization Tools: Tableau, SQL