VEDANT SAHAI

+1 (814)-769-0201 vedantsahai18@gmail.com linkedin.com/in/vedantsahai18 github.com/Vedantsahai18

EDUCATION

Pennsylvania State University May 2024
Master of Science in Computer Science & Engineering ***(GPA: 3.75 / 4.00)***
- Teaching Assistant: CMPSC 132: Programming and Computation II: Data Structures

University of Mumbai May 2021
Bachelor of Engineering in Computer Engineering ***(CGPA: 9.58 / 10.00)***

EXPERIENCE

Julep AI Inc. Oct 2024 – Present
Machine Learning Engineer
- Increased User Engagement by 15% within a month of launch by developing the ”Browser Use” feature, enabling automation of browser tasks through AI-driven workflows.
- Optimized the workflow performance to accelerate executions from hours to under 45 minutes for 2.5k runs by migrating the database from CozoDB to TimescaleDB.
- Developed Julep Python-based CLI and 10+ AI workflow templates, improving developer participation to 5k GitHub stars.
- Enhanced platform features by 20% by programming a FastAPI-Docker service enabling interaction between AI workflows and External Tools.

JP Morgan Chase & Co. Jun 2023 – Aug 2023
AIML Summer Associate
- Improved the Transaction Detection Rate (TDR) by 100-150 basis points by implementing the Online ML XGBoost algorithm for the Transaction Risk model (TRS).
- Attained 90% accuracy by programming a TabNet-based Deep Neural Network as a challenger for the TRS XGBoost model.
- Ensured 95% code coverage by implementing a PyTest-based testing framework for the TRS Feature Engineering codebase.

Plexflo Oct 2021 – Jul 2022
Machine Learning Engineer
- Developed Evidence, a Meter Data Management & Analytics (MDMS) software with a latency of less than 90ms, by leveraging ITRON, Sensus Xylem, and Siemens data streams, powered by AWS, Apache Flink, and a custom ML model.
- Achieved an F1 score of 85% for a non-intrusive load monitoring model by leveraging a Variational Autoencoder (VAE) model.
- Scaled a Flask + AWS Timestream backend to support up to 20,000 IoT devices for MDMS over Grafana.

Sync Energy AI Jul 2020 – Sept 2021
ML Research Intern
- Optimized the power outage extraction, resulting in a 40% faster response, by deploying an AWS Lambda-Python-REST API.
- Improved accuracy to 83% in estimating the locations of utility poles from Google Street View images by employing Mask R-CNN and Image Processing.
- Constructed a knowledge graph of 500+ wildfire papers, utilizing Neo4j to identify 150+ interconnected variables.

TECHNICAL SKILLS

Languages: Python, C, JavaScript, Java, LaTeX

ML Frameworks: PyTorch, TensorFlow, RASA, Scikit-learn, XGBoost, HuggingFace, NLTK, SpaCy, LlamaIndex, Langchain, Autogen

Databases: MySQL, PostgreSQL, MongoDB, Timescale

Cloud: Amazon Web Services (AWS), EC2, S3, Lambda, API Gateway, SageMaker, IAM

Technologies: React.js, Flask, FastAPI, PyTest, GitHub Actions, Elasticsearch, Git, Docker, PySpark, Grafana, Sphinx, Temporal

PROJECTS

Datacertus May 2022
Reacts. JS, AWS APIs, Python, Docker, AWS ELB, AWS Lambda, DynamoDB, Scikit, RASA, PyTorch, Keras, Scikit-Learn
- Trained a BERT summarization model to summarize updates on disaster information with 75% ROGUE-L.
- Increased update frequency by over 50% by developing a pipeline using Lambda, Selenium, and the BERT model to track natural calamities.
- Reduced data processing time by over 10% by integrating AWS Lambda, S3, and API Gateway-based trigger events.

Conversational AI for Secure Healthcare Assistance May 2021
Docker, Node.JS, MongoDB, RASA, Blockchain, Python, Express.JS, jQuery, AJAX, Vault, Nginx
- Processed over 10,000 Electronic Health Records concurrently by architecting a microservice-based software.
- Attained the intent prediction confidence up to 95% by developing and integrating a RASA-powered therapy chatbot.
- Integrated the BigchainDB + IPFS blockchain database with the RASA for behavior analysis and secure medical record storage.

PUBLICATION

Vedant S., et al. (2021) Leveraging Deep Learning and IoT for Monitoring COVID19 Safety Guidelines Within College Campus. In: Garg D., Wong K., Sarangapani J., Gupta S.K. (eds) Advanced Computing. IACC 2020. Communications in Computer and Information Science, vol 1367. Springer, Singapore.