SANDEEP POLAVARAPU

DATA SCIENTIST

+1(240)906-7966 | Maryland Heights, USA | LinkedIn
bit.ly/pvnsandeep | sandeeppvn@gmail.com | GitHub

Data Scientist with 4+ years of experience, proficient in developing data-intensive applications and end-to-end pipelines, leveraging expertise in data analysis, machine learning, NLP, deep learning, data engineering, feature engineering, model development, deployment, monitoring, and continuous training using Python, R, SQL, PyTorch, and AWS.

WORK EXPERIENCE

Data Scientist – Bayer R&D

July 2023 - Present

- Engaged in the pre-commercial plant breeding of the Crop Science department, I developed and maintained prescription models. By harnessing statistics, machine learning, and numerical optimization, these models significantly propelled our research projects.
- As part of the MLOps initiative, I was pivotal in integrating Docker setups over VPN configurations and implementing Airflow for improved workflow automation. My contributions enabled teams in deploying tailored AWS SageMaker environment with a seamless CI/CD process using GitHub Actions
- As a pivotal figure in the Analytics and Product Design team, I consistently delivered critical technical support during high-pressure phases of North America Soy Advancements.
- Collaborated globally across Germany, US, Brazil, Netherlands, and India under tight timelines, working alongside diverse domain experts such as breeders, lab scientists, and data engineers, adeptly handling the complexities of catering to each stakeholder's unique requirements.

Data Scientist - Craftle, Inc

Jan 2023 – July 2023

- Lead data science initiatives in designs, development, and deployment, collaborating with UI and business using Agile practices and MLOps practices leading to 25% reduction in Time to Market (TTM) of new models.
- Orchestrated an end-to-end modularized Generative AI pipeline using pre-trained ControlNet that enables users to generate a redesigned room from an image and text prompt using python and HuggingFace achieving an FID score of 15.5 and a 20% increase in user engagement.

ML Researcher - NASA LCLUC, UMD

Jan 2022 - Dec 2022

- Developed Time series crop classification on multispectral satellite images from large data sources (>10TB) using stacked Bi-directional LSTM to study CO2 emissions of sugarcane burning in Thailand; achieving a producer accuracy of 96% and user accuracy of 92%, which was 8% above the baseline. (Project Link)
- Implemented a Computer Vision Semantic Segmentation model using DeepLabV3+ to map burnt area in sugarcane residue images, resulting in an IOU (Intersection over Union) score of 77.7% and presented the findings of the above study as a report at the NASA LCLUC conference. (Project Link)
- Devised a sophisticated statistical model using SciPy numerical optimization on soil and weather data for winter wheat Phenology dates prediction achieving an R-squared score of 0.85. The model was used for farming and logistic decision making leading to 15% increase in crop yield and 20% improvement in identifying crop cycles for logistic planning. (Project Link)

ML Engineer II - Akamai Technologies

July 2018 - July 2021

- Constructed and evaluated Anomaly Detection ML pipelines using A/B testing on Decision Trees, SVM, XGBoost, CNN-LSTM, and Autoencoders to detect and mitigate DDOS (Distributed Denial of Service) attacks. Deployed models with Docker, AWS ECR, SageMaker, Apache Spark (PySpark) and Kinesis streaming. Monitored data drift and performed continuous training, reducing critical event reaction time by 30% and achieving an acceptable 82% recall score.
- Engineered ETL pipelines, developed and deployed a secure NLP-based Topic modeling Document Information Retrieval system using LDA (dimensionality reduction), K-Means clustering and knowledge graphs on hybrid cloud (AWS S3). This resulted in: (a) Semantic Similarity score of 0.71, (b) MAP score of 0.5, (c) Increase Click Through Rate of 0.1.

EDUCATION

University of Maryland – College Park, MD Aug 2021 - Dec 2022
Masters in Machine Learning GPA: 3.93/4.0
Relevant coursework: *Communications for Data Science, Optimization, Cloud Computing, Deep Learning, NLP, Computer Vision*

PES University - Bangalore, India Aug 2014 - July 2018
Bachelors in Computer Science and Engineering GPA: 8.31/10.0
Minor: Management Studies

SKILLS

Languages:
- Python
- R
- C++
- Linux (Bash)
- SQL

Machine Learning:
- Scikit-learn
- Pandas
- NumPy
- SciPy
- Keras
- TensorFlow
- PyTorch

DevOps:
- Jira
- Docker
- CI/CD Jenkins
- GitHub (Actions)
- REST API

Data Engineering:
- AWS Sagemaker
- Apache Airflow
- Big Query
- PySpark
- MySQL
- MongoDB

Cloud Computing:
- AWS
- GCP
- Azure
- Salesforce

PROJECTS

Text to Image Generation (Project Link)
Designed a Generative AI Multi-modal learning model using all-MiniLM-L6-v2 transformer-based sentence embeddings and a custom counting loss function on a Conditional GAN (CGAN) to create Images from Text.

Generative AI for Financial Engagement (Project Link)
Through mock demographic and Twitter data, customer behavior and purchasing patterns were analyzed. OpenAI API of GPT-4 and advanced prompt engineering Generative AI was employed to produce tailor-made marketing emails evaluating on key metrics of Email Open Rate and the Conversion Rate for Stock Portfolio .

AWARDS & RECOGNITION

- Akamai Innovation Award
- One Akamai collaboration Award
- Akamai Urgency and Persistence Award
- Research grant for Drone Stabilization, PES University
- Academic Distinction Award, PES University

LANGUAGES

- English (C2)
- Telugu (C2)
- Kannada (C2)
- Hindi (C1)
- German (A1)