Sahana Venkatesh

venkatesh.sahana1606@gmail.com | github.com/Sahanave | linkedin.com/in/sahana-venkatesh/

Profile

Senior Machine Learning Engineer with over five years of experience in machine learning products from proof of concept to production. I have a proven track record of success in leading projects and a deep understanding of the machine learning development process. My expertise is in designing, developing, and deploying **Python**-based machine learning pipelines **(Tensorflow, Scikit-learn)** on-premise and on-cloud.

Experience

Senior Machine Learning Engineer 07/2023 - Current
Jaguar Land Rover - Gaydon, Warwickshire *Python, GCP , Scikit-learn, Kubeflow , ApacheBeam*

- Developed and Deployed **kubeflow** ML pipeline in **Vertex AI** on GCP linking **Big Query**. This pipeline processes around 200GB of data from around 51,000 vehicles to infer about battery health. Using Kubeflow, I was able to reduce the cost and latency of streaming results by **40%**. TODO : ADD REAL DATA.
- Led a team of 4 for an in-company NLP AI challenge and created an in-house document retrieval system based on semantic search. I implemented **the Apache Beam** pipeline for text cleaning and running inference of the sentence transformer model. TODO : ADD REAL DATA
- Leading the design of the **MLOps platform** for Central Diagnostic at Jaguar Land Rover, enhancing batch prediction capabilities and refining ML workflows.
- Developed and documented best practices for the Data Science team, emphasizing coding standards, data models, and MLOps workflows, ensuring a cohesive and efficient development environment. Mentored junior data scientists.

Data Scientist 02/2020 - 01/2023
Apeel Sciences - Goleta *Python, Tensorflow, Scikitlearn, Postgres SQL, OpenCV*

- Lead end-to-end delivery of product from proof-of-concept to complete production deployment. Managed the full life-cycle of data science projects and solutions - including ideation, data acquisition, exploration, engineering, prototyping, and deployment.
- Addressed Quality Control for the Field Operation team by implementing a Thermal Camera Integrated edge computer vision pipeline. This reduced human labor costs by at least **30%** and achieved **24 x 7** monitoring of earlier non-feedback systems. This has been filed as a patent Systems and methods for assessment of food item dryness **(US 20230222822)**. This was adopted over **3** different client sites.
- Lead the discussions with the platform team to establish seamless interactions between on-edge machine learning solutions and cloud-based reporting systems **(AWS, Docker, State Machine, SQL and Grafana)** for the internal stake-holders to visualize the performance of deployed solutions.
- Spearheaded the development of a machine learning pipeline **Tensorflow + OpenCV** to identify banana ripening stages, pioneering this approach in the industry. Collaborated with Biological and Formulation Scientists to create image analysis algorithms, oversee data collection, and train models. The accuracy of the model was **80% accuracy** on a **500-image** test dataset. This was captured in the **patent application (US 20230214982)**.
- Conducted a **crowd labeling** experiment, uncovering the ambiguity in human labeling and showcasing the effectiveness of Apeel's fruit coating, contributing to extended client engagements.
- Built a customized object detection training and evaluation pipeline; utilizing state-of-the-art models **(Tensorflow, Transfer Learning)**. It became a vital contributor to the development of a deep-learning object detection engine, as captured in the patent application **(US 20220270269)**. Designed a customized API that significantly reduced model training time, boosting team productivity and efficiency. The API was applied in various projects, showcasing versatility.
- This reduced focused time taken by our data scientists in model development and deployment by **4X (from 2 days to 4 hours)**.

Data Scientist 06/2018 - 01/2020
ImpactVision - San Francisco *Python, Keras, Scikit-learn, C++(Basic)*

- Foreign object or contaminants detection in a salad using Hyperspectral Images for a large salad-kit packaging company.
- Trained SVM model **(sklearn, pandas)** to run in real-time on production lines that handle a multitude of products. The latency was **< 1s** and accuracy was around **60-70%** on real-time testing.
- Used Generative Adversarial Network to generate synthetic data from real data. This aided in understanding underlying data distribution and visualizing the change in distribution over time **(Keras)**.
- MLOps: Supported the maintenance and development of a real-time machine learning system that is integrated with the vision systems. Systems were customized for three different client sites **(C++, sklearn)**

Education

University of Southern California - Los Angeles, USA Jan.2016 - Dec.2017
Masters of Science: Electrical Engineering - Specializing in Data Science and Machine Learning.

Shiv Nadar University - India Aug.2011 - May.2015
Bachelor of Technology in Electronics and Communications Engineering with Minor in Mathematics.

Projects

AnxiAid: AI-Powered Anxiety Management Platform Oct. 2023 - Oct. 2023

wwcode$_{h}ackathon2023_{a}nxiety_{m}anagement_{t}ool$ | youtube.com/L5b5FTNupMM *OpenAI API, Streamlit, MySQL*

- This platform, developed during the WomenWhoCode Hackathon for Social Good 2023, addresses the critical need for accessible mental health support, demonstrating how technology can have a profound social impact.
- I developed this by talking to therapists and we learned from each other how new benchmarks in innovations could be tools they would be used in the future.

AWS Hackathon July 2020: COVID-19 Hackathon July. 2020 - July. 2020

github.com/dimfoo/hackathon | *AWS Sagemaker, Scikit-learn*

- This was a hackathon sponsored by AWS and organized by WomenWhoCode. I worked with a team of diverse skillsets (Subject Matter Experts, Software Engineers).
- We submitted a 68% accurate Logistic Regression model that will tell us which at-risk individuals are likely to test positive for COVID-19 so that health workers can target testing resources and health education toward that population.

Volunteer

Women Who Code - Volunteer Aug.2020 - Present
I volunteered with Women Who Code - Data Science and Women Who Code - London. Women Who Code (WWCode) is an international non-profit organization dedicated to inspiring women to excel in technology careers. I participated in various hackathons and tech talks organized by Women Who Code.

Next Tech Girls - Role Model Aug.2023 - Present
Next Tech Girls (nexttechgirls.com) is an award-winning social enterprise that tackles the skills and gender gaps in the tech industry. They partner with companies and educational institutions to host inspiring tech-focused events and work experience placements for teenage girls, with a focus on those from lower socio-economic or ethnic minority backgrounds.

PyData Global Impact Scholarship Program - Speaker and Mentor Dec.2023 - Present
The PyData Global Impact Scholarship Program is a bridge between experienced data scientists, AI engineers, and open-source contributors to all potential future technology and open source leaders. In 2023, I was one among seventeen contributors (mentor and a speaker) who helped one hundred eleven accepted participants. (global2023.pydata.org).