SAIRAJ AMBERKAR

A Data Science professional specializing in data analysis, applied ML and data pipelines, seeking the role of a Data Scientist to leverage AI to bring business value

EDUCATION

2019 — **George Mason University**
M.S. in Computer Engineering (GPA:3.78) — Virginia, USA

2016 — **Mumbai University**
B.E. in Electronics Engineering (GPA: 3.21) — Mumbai, India

PROFESSIONAL EXPERIENCE

2021 | Present — **Engineer 3, Data Science**
Hughes Network Systems. — Maryland, USA

-As a Data Science Engineer at HNS, I have been developing, deploying, and maintaining analytic solutions with objective to minimize mean time to detect by dashboarding, minimize mean time to troubleshoot by developing insights, and to eliminate mean time to repair by integrating self-healing into the services.

-Deploying and hosting end to end cost effective ML solutions on different cloud services such as Kubernetes, Cloud Run, Cloud functions on Google Cloud Platform to take preemptive actions

-Collaborate proactively with product/service owner and managers to determine and address data and reporting needs for various projects as well as for developing the KPI's.As a part of building application suites, creating JIRA stories to track the project progress and documenting the design, methodology and results in confluence

2019 | 2021 — **Data Analyst 2**
Hughes Network Systems. — Maryland, USA

-Data gathering, cleaning, wrangling through different operational and business systems by building custom high volume data pipelines on an ad hoc basis moving the data into Google BigQuery.

-Creating Machine Learning systems with solution as a combination of descriptive, diagnostic, predictive and prescriptive analysis/modelling to improve triage capabilities of equipment's like routers, switches and modems

-To provide insights, recommendations and model inferences through dashboards, real time email alerts or via api's depending on the end-user whether it's a human or another intelligent system that consumes the information.

RESEARCH EXPERIENCE

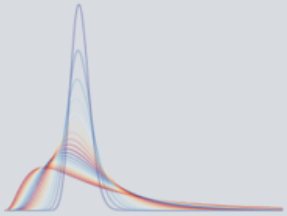
2018 | 2019 — **Graduate Research Assistant, Accelerated-Secure and Energy Efficient Computing Lab (GMU)**
George Mason University — Virginia, USA

-The research focused on the vulnerabilities of deep learning models and their application in domain of Hardware Security and sub-domain of Malware Detection.

-Deployed deep learning models for the task of Image Classification, Malware Detection and Side Channel Attack Detection.

-Generated and studied the impact of adversarial attacks on deep learning models using multiple attack strategies for computer vision tasks for understanding the vulnerabilities of models and applying the same in domain of malware detection for adversarial malware generation.

-Deployed different defense techniques against adversarial attacks with objective of creating robust models with special emphasis on adversarial training.



CONTACT INFO

19800 Crystal Rock Drive,
Germantown, MD-20874
samberka011@gmail.com
website
+1 571-320-5288

STATS

Industry Experience : 3.5+ Yrs
Academia Experience: 1.5 Yrs
ML Models Deployed: 500+
API Consumers: 100000 calls/yr
Data Ingested: 100+ Tbs

SKILLS

Programming Language

Python, R, SQL.

Machine Learning

Classification, Clustering, Feature Engineering, Forecasting, Anomaly & Change Point Detection.

Statistics

Boxplots, Z-Scores, T-tests, A/B Testing, Statistical Distributions and many more.

ML/DL Platforms

Google Cloud Platform (Services: Cloud Run, Cloud functions, GKE, BigQuery, CloudBuild)

Python/R Libraries

Scikit-learn, Numpy, Matplotlib, Pandas, Seaborn, Flask, Dplyr, Tidyverse, Keras and many more

Visualization Tools

Shiny, Flexdashbaord, Google Data Studio, Tableau

PROJECTS

Carrier Signal Strength Monitoring and Automation
Hughes Network Systems — Maryland, USA

Objective: To monitor the signal strength metric for different sites deployed with different carriers for multiple customers and proactively monitor the carrier performance in terms of any degradation and take proactive steps towards a corrective action before formal complaints from the customer

Methodology:Deployed a changepoint detection-based application as a cronjob that runs on a weekly basis, automatically detecting if/when there is a "new normal" established for the signal level and was that normal now better or worse or no different than before

Deployment:The insights are provided via email alerts which say these 125 sites with this carrier and for this enterprise now have a "worse new normal" than before and therefore these sites need to be investigated with possible reasons for drop in signal levels such as change in the carrier provider at the site

Pre-Sales Carrier Recommendation Tool (T-Mobile vs AT&T vs Verizon Wireless)
Hughes Network Systems — Maryland, USA

Objective:The objective is given a new site location for an enterprise customer, need to recommend the best carrier service installation based on multiple signal strength metrics, service uptime and the download speeds in that region from historical data from various previous installations for service improvement

Methodology:Location based clustering per carrier for 16000+ devices and signal metrics summarisation per cluster, further scoring and categorizing each cluster as good, bad, fair based on standard signal quality criteria and given a new site location, locating three different closest clusters (the three cluster represent three different carriers) and recommends the cluster/carrier which is the best among the three based on the earlier scoring/ranking

Deployment:The recommendation is provided to the sales team in form of a user interactive dashboard and to the installer team through installer mobile app

Tools: R, BigQuery, Pubsub, GKE, Cloud Build, Flexdashboard, application suite consists of 3 applications

Anomalous Application Traffic Spotting (Shadow Applications):
Hughes Network Systems — Maryland, USA

Objective:To find anomalous applications that a user visits and application is flagged as anomalous if it has increasing traffic trend in terms of either bytes or packets and is NOT seen in at least 75% of the traffic population for the Enterprise to which that user belongs.

Methodology:A two-stage anomaly detection is deployed, stage one flagging the anomalous users and stage two flagging anomalous applications for the flagged users.

Deployment:Daily email alerts are provided that point out the anomalous user-applications which say this X user working for this Y enterprise visited Netflix.

Modernizing Legacy Application (Legacy Based System to Serverless Cloud System):
Hughes Network Systems — Maryland, USA

Objective:The objective is to move the critical on-premise system hosted on oracle to a cloud-based solution (BigQuery + Kubernetes + Cloud Run) to achieve faster processing, easy maintenance and ease of replication for different customers across the globe without infrastructure concerns.

Methodology: Rearchitected a monolith PL-SQL based system into mesh of microservices (15 different containerized python applications) reducing processing time from 4 hours to 30 mins.

Deployment:The system is integrated into into current Front-End UI Tool.

Tools: Python, R, BigQuery, Cloud Run, GKE, Cloud Build, Datastudio Dashboard

PORTFOLIO

Summary: Time Series Simulation to perform change point analysis to detect anomalous sequences.

Launch App: Time Series Playground

Summary: MNIST Digit Playground with a flavor of Adversarial Machine Learning.

Launch App: Mnist Playground

Summary: Build a website using R and flexdashboard showcasing my data science experience, skills, projects through different visualization.

Launch App: Sherlock of Data

Summary: Build a basic dashboard showcasing impact of vaccination in USA.

Launch App: Covid-19 Exploration

PUBLICATIONS

Paper: Efficient Adversarial Training for Robust Machine Learners.

Conference: International Conference on Computer Aided Design (ICCAD), 2018

Authors: Sai Manoj, **Sairaj Amberkar**, Setareh Rafatirad, and Houman Houmayoun.

Link: Efficient Adversarial Training

Paper: Adversarial Attacks on Microarchitectural Events based Malware Detectors.

Conference: 56th Design Automation Conference, 2019

Authors: Sai Manoj, Abhijitt Dhavlle, **Sairaj Amberkar**, Sahil Bhatt, Setareh Rafatirad, and Houman Houmayoun,

Link: Adversarial Attacks on Microarchitectural Events based Malware Detectors