Keshav Karai Lakshmipathy

Boston, MA | 617-792-1381 | karailakshmipathy.k@northeastern.edu | LinkedIn | GitHub

MOTIVATION

I am passionate about solving business problems using Data Science, Statistics, Machine Learning and AI. I systematically and creatively use my skillset to add tangible value to the team, the business, and the end-user. I am constantly learning and always looking to improve and create an impact.

TECHNICAL SKILLS

Languages:	SQL, Python, R, C++, Java
Frameworks:	Pandas, NumPy, Data Structures, Matplotlib, Seaborn, Scikit-learn, NLTK, PyTorch, Keras, Tensorflow
Concepts:	Statistical Modelling, Stochastic Modelling, Machine Learning, Deep Learning, Reinforcement Learning, Causal Inference, Hypothesis Testing, Regression, Classification, A/B Testing, Bayesian Techniques, Anomaly Detection, Computer Vision, Data Validation and Cleaning, Feature Selection, Modeling, Model Interpretability and Error Analysis, Reports and Visualization, Data Centric AI**,** Big Data, Hadoop, Clustering, Data Mining, ETL
Tools:	Github, Anaconda, Jupyter Notebooks, RStudio, Google Co-lab, Anaconda, PyCharm, AWS, Cloud Technologies, Tableau, Power BI, Rest APIs
Other Skills:	Business acumen and intuition, problem solving, communication skills

EDUCATION

Northeastern University* | *Boston, MA

Master of Science in Information Systems-Data Science — *Expected May 2022*

Gopalan College of Engineering | Bangalore, India

Bachelor of Engineering in Computer Science — *June 2016*

ACADEMIC AND RESEARCH PROJECTS

Morpho: Open-Source Toolkit for Generative AI from social media — Feb 2022 - *Present*

- This project is aimed at creating an open-source toolkit for generative AI from social media
- A set of Python tools that allows one to create synthetic data using social media scraping tools and Generative Deep Learning Models such as GANs, VAEs and diffusion models to create Novel Images.
- The idea is to build a tool by creating high quality images of animals, sneakers, butterflies starting with a keyword list

Deep Generative Models with Autoencoders, Variational Autoencoders and GANs — *November 2021*

- Image generative models were implemented using some popular approaches like Autoencoders, Variational autoencoders and GANs.
- It was further experimented with to see how each of them were used to generate new images, work for different datasets and different architectures to see how it influences image generation
- Synthetic Images were generated using AE, VAEs and GANs

Ads Optimization and the Multi-Armed Bandit Problem – Reinforcement Learning — *September 2021*

- Real-world Binomial Trial data like Ads were simulated and modeled with reinforcement learning agent to maximize Ad click rewards
- Data was modeled with Thompson Sampling with Beta Distribution along with other algorithms like Epsilon-Greedy, Upper Confidence Bound and Random Sampling
- Epsilon of 0.1 for epsilon greedy, C=1 in UCB was observed to achieve best results least regret for least number of episodes
- UCB's performance reduced in higher action space values whereas Thompson Sampling and E-Greedy was unaffected.
- 5 bandits were simulated, and Ad 1 was observed to be best Ad to use with conversion rate of 31.56%

LSTM-Based Auto-Encoder for Multi-Sensor Anomaly Detection — *December 2020*

- Aimed to detect anomalies in machines which need to be removed to maintain seamless performance and efficiency. Dataset contained machine data from 15 identical machines
- Data preprocessing was done using Gaussian Mixture model where data was split into 3 components – Normal, Faulty and Failed. Isolation Forests were implemented to separate normal data points from outliers
- MAE loss for training and validation set averaged to 0.16 and 0.19 and MAE loss of 0.19 and 0.20 was recorded for the test set
- AE_LSTM method was recorded to have good stability and great anomaly detection accuracy

Home Depot Product Search Relevance Prediction — *October 2020*

- This project was aimed at improving Home Depot Customers' shopping experience by developing a model that can accurately predict relevance of search results
- Used NLP, Statistical Learning Methods, Linear Models, Time Series and Forecasting to study Correlation between features, detect linear relationship between features and pre-processed the data
- Built a Random Forest Regressor and Tuned Hyperparameters to get the best prediction accuracy with R2 Score of 0.19 and MSE of

0.22, studied feature importance

Prediction of Cardiovascular Disease using Ensemble Modelling *September 2020*

- Applied Ensemble Modelling and built a Binary Classifier to predict the cardiovascular disease in an individual by studying the features
- Used Statistical Learning Methods and Data Visualization to study Correlation between features, detect multicollinearity using VIF, Pre-Processed the data using feature selection and feature extraction
- Built Random Forests, Logistic Regression and Gradient Boosting Classifier ML Methods which gave similar F1 Score of 0.91, 0.93 and 0.95, tuned hyperparameters and studied feature importance

WORK EXPERIENCE

Northeastern University *January 2022 – Present*
Graduate Teaching Assistant | *Boston, MA*

- Teaching Assistant for the course CSYE7370 Deep Learning and Reinforcement Learning in Game Engineering
- Tutored students on course topics, concepts, questions regarding research, assignments, and code

EMD Serono Inc. *January 2021 – August 2021*
Data Scientist Co-Op | *Billerica, MA*

- Pre-Processed and Visualized Digital pathology data from human subjects using R-Shiny, RNA-Sequencing Data from Pre-Clinical Trials
- Performed Data cleaning and wrangling on large-scale clinical and genomics data, loaded it to AWS S3 data lake and queried it using relevant S3 SELECT and RSQL queries
- Performed Differential Expression analysis on pre-clinical data which involved data cleaning, preprocessing and validation in a controlled and restricted environment handling sensitive data
- Extensively used git, CRAN, Bioconductor, Gene data profiler, cloud platforms.
- Developed projects in R that are published and used by cross-functional scientific teams across the company.
- Extended current object-oriented code base and learned new technologies and methods
- Presented results to Cross Functional Team of Scientists, stakeholders, and incorporated feedback

Google | *Bangalore, India* *April 2019 - December 2019*
Data Analyst – Digital Marketing

- Analyzed website traffic data, ad campaigns on business websites using data analytical techniques, data visualizations, EDA, derived insights and helped drive Business Strategies like Dynamic Re-Marketing, Ad Campaign optimization, Website Optimization, Keyword Optimization with Company CEOs looking to market their business on Google's Digital Marketing Platform like Google Ads and Google Shopping
- Presented results and analysis to business CEOs and small business owners
- Advised and implemented marketing Strategies, worked on WebTech tools like Google Analytics, Google Merchant Center, Google Tag Manager

HP inc.| *Bangalore, India* *September 2016 - Jan 2019*
Data Analyst

- Analyzed consumer data for regression and demand forecasting to drive business plans. Forecasted number of potential consumers for different quarters for better decision making