BASIL VARGHESE

Boston, MA | +1 (617) 480 1870
varghese.b@northeastern.edu | linkedin.com/in/bkvarghese | github.com/BasilKVarghese

SUMMARY

A curious and analytical **problem solver** with **3 years of experience** in the analytics industry with expertise in **Python** and **SQL**, having worked with cross functional teams on **data science projects** building **machine learning and statistical models** to solve complex business problems and understanding how it translates to value for the organization. Committed to helping companies make data-driven decisions by building robust data pipelines and extract intelligence using state-of-the-art techniques in machine learning and deep learning.

EDUCATION

Northeastern University, Khoury College of Computer Sciences — Boston, MA
Master of Science in Data Science | Current GPA: 3.83/4 — Expected May 2023
Relevant Courses: Algorithms, Natural Language Processing, Supervised Machine Learning, Foundations of Artificial Intelligence

Vellore Institute of Technology — Vellore, India
Bachelor of Technology in Computer Science Engineering with specialization in Bioinformatics | GPA – 8.48/10 — May 2018
Relevant Courses: Data Structures and Algorithms, Linear Algebra, Probability and Statistics, Data Warehousing and Data Mining

TECHNICAL SKILLS

Programming languages	Python (NumPy, Scikit-learn, TensorFlow, Keras), **R** (dplyr, Rshiny), SQL, PySpark
Data Science	Regression, Classification, Clustering, Feature Selection & Engineering, Model Training, Deployment
Deep Learning	Recurrent NN, Convolutional NN, LSTM, GRU, Sequence Models
Statistics	Hypothesis Testing, Statistical Inference, Bayesian Inference
Databases	PostgreSQL, MySQL, Azure Synapse Analytics, Hive, Hadoop
Cloud Technologies and others	Azure Databricks, Azure DevOps, Azure Data Factory, GIT

PROFESSIONAL EXPERIENCE

Mu Sigma Business Solutions — Bengaluru, India
Decision Scientist — **Jan 2020- Jul 2021**

- Designed and developed an internal temperature **prediction system** for electrolytic cells for an Aluminum manufacturing client, using and **ensemble framework** of **classification and regression models** along with heuristic engines in **Python**, which addressed customer pain points by increasing the availability of temperature measurements from once in 48 hours to every hour
- Built an **Anomaly Detection** framework detecting process failures 8 hours in advance in an industrial electrochemical cell, leveraging **statistical inference** and **classification frameworks** (Random Forest, Gradient Boosting)
- Led a team of 3 as the project **Deployment Lead** and handled the **deployment and monitoring** of 20+ **machine learning** models through Azure CI/CD pipeline leveraging Azure Databricks for job scheduling and Azure Synapse Analytics as backend
- Created interactive **visualizations** using **Plotly** and **Seaborn** to understand model performance and plan for improvement
- Mentored 4 new recruits and delivered knowledge transfer sessions to enable quick adoption into the team and deliver fast results
- Analyzed and presented a cost optimization solution for Azure Databricks to the client, reducing total cost incurred from Azure Databricks by approx. 50% by optimizing cluster utilization across users

Mu Sigma Business Solutions — Bengaluru, India
Sales Ambassador — **Oct 2018 – Dec 2019**

- Implemented strategic goals for sales reach-out as an end-to-end process including lead identification, targeting and negotiation, with a 25% response ratio in cold mails
- Managed account growth for deals closed through farming, relationship management etc, for 4 global accounts
- Strategized with leadership team & prospective customers from various verticals & horizontals to structure value propositions for real-world business problems

XCODE Life Sciences — Chennai, India
Data Analytics Intern — **Dec 2016 - Jan 2017**

- Performed **cleaning** and integration of raw **genotypic** and **phenotypic data** from various sequencing platforms
- Conducted **statistical analysis** on raw genomic data to understand various trends in the Indian gene pool

ACADEMIC PROJECTS

Driver Drowsiness Detection System (***Python, OpenCV, TensorFlow***) Apr 2022 – Present

- Currently working on building a Driver Drowsiness Detection System using OpenCV and TensorFlow in python
- The goal of the project is to train a deep learning model and detect the drowsiness on people's faces in video data from a webcam

Netflix Movie Classification System (***Python, NLP, Classification, Clustering, Deep learning***) Mar 2022 – Apr 2022

- Built a Movie Classification System based on the movie plot using an ensemble of **NLP**, **clustering**, and **classification** methods
- The movie plots were clustered to different groups using **hierarchical/k-means clustering** and an **LSTM** model was trained on the clustered data to classify new movies to one of the clusters with a **precision score** of 0.75

Dots and Boxes game using Artificial Intelligence Agents (***Python, AI, Adversarial Search***) Mar 2022 – Apr 2022

- Built an AI powered game using search strategies such as minimax with alpha/beta pruning
- Compared the performance of the game agent when trained with other techniques such as Expectimax and Monte Carlo tree search

Time Series Sales Forecasting (*Python, Time-Series, Regression*) Oct 2021 – Dec 2021

- Developed a **sales forecasting model** on **time-series** data for a retailer, leveraging statistical models (Holt-Winters exponential smoothing) and regression ensemble models (XGBoost) achieving a MAPE of 7.4

Parkinson's Disease Prediction System (*R, Neural Networks*) Jan 2018 – Apr 2018

- Built a Parkinson's Disease Prediction System on speech data using **resilient neural networks** with a UPDRS RMSE of 8.95
- The project compared the accuracies of various machine learning models to a neural network model in predicting Parkinson's disease using various voice features from speech data of patients

CERTIFICATIONS

- Deep Learning Specialization, Coursera
- Microsoft Azure AI Fundamentals certification (*In Progress)*