Sanjan Vijayakumar

Boston, MA | +1(857) 971- 0821 | vijayakumar.sa@northeastern.edu
Available: January 2022 | linkedin.com/in/sanjanvijayakumar

EDUCATION

Northeastern University, Khoury College of Computer Sciences — Boston, MA
Master of Science in Data Science, **GPA: 3.84** — (Expected) December 2021
Related courses: Supervised and Unsupervised Machine learning, Data Mining, Algorithms, Deep learning, NLP

Visveswaraya Technological University, RNS Institute of Technology — Bangalore, India
Bachelor of Technology in Electronics and Communication — June 2015 – June 2019

TECHNICAL KNOWLEDGE

Programming Languages:	Python, SQL, R, Mathematica, Matlab, Java
Tools/Frameworks:	Sci-kit, Keras, TensorFlow, PyTorch, Spark, Hadoop, Docker, Git, Tableau, Seaborn, Plotly
Databases:	MySQL, PostgreSQL, MongoDB, DynamoDB, Amazon RDS, Redshift, Cloud SQL, Bigtable
Machine Learning Algorithms:	Linear Regression, Logistic regression, SVM, K-Means, Random Forest, CNN, RNN, LSTM, Deep Neural Networks, Ensemble Learning, XG-Boost, Light-GBM, GAN
Publications/Achievements:	Semi-finalist - EY Data Science Challenge 2021, Paper on Breast Cancer Classification (IJRES), Paper on Sequential and Parallel Alpha-Beta Pruning algorithm (IJIET)

PROFESSIONAL EXPERIENCE

LabCorp, Burlington, NC — May 2021 – August 2021
Data Science Intern

- Created a pip installable package using AWS Sagemaker Studio that reduces data cleaning and pre-processing time by **40%**
- Built and deployed a collaborative-filtering test recommendation system for tested users (AWS Sagemaker and Personalize)
- Devised a framework using AWS Athena, Glue and S3 for direct storage, retrieval, and addition of test reference ranges
- Integrated an allergy recommender system with a model production-deployment package to make **97%** accurate predictions

eMotionRx Inc, Cambridge, MA — May 2020 – December 2020
Data Scientist

- Orchestrated the migration of ~**3Tb** of on-prem sensor data onto cloud storage buckets on Google Cloud Platform
- Streamlined ETL data warehousing using Cloud Dataflow that reduced data pre-processing and latency by **30%**
- Designed an ensemble tree-based machine learning model with **94%** accuracy to identify pressure distribution on a foot-sole
- Performed data analysis using Google Cloud BigQuery and built **5+** dashboards using Tableau and Cloud Data Studio

VI Solutions, Bangalore, India — June 2018 – September 2018
Data Science Summer Intern

- Accelerated run-time on an Object detection application by **30%** by improving the region of interest selection using OpenCV
- Deployed an image classification model using modified ImageNet Convolutional Neural Networks with an accuracy of **87%**

ACADEMIC PROJECTS

Attention based Image-Captioning system, Northeastern University — February 2021 – March 2021
An image caption generator using attention mechanism and deep learning

- Generates image captions using a local attention model on top of a VGG-16 model as an application of NLP and deep learning
- Contextual representation is generated by combining the CNN softmax scores into a vector representation passed into the GRU (RNN) which generates a BLEU score of 67.32. GitHub: https://github.com/sanjsvk/Attention-based-image-captioning

Restaurant Business Analysis and Optimization Using Yelp Review Data, Northeastern University — March 2020 – April 2020
Built a system for users and restaurants to gain insights from reviews using unsupervised learning

- Topic modeling (LDA, LSA) and clustering methods (K-means, GMM) were used to identify the top 5 review points and cuisines enjoyed by customers. This helps restaurants dynamically make changes based on real-time reviews.
- Created a recommender system which provides users with recommendations based on an input text (cuisine, location, etc), thus streamlining the user experience. GitHub: https://github.com/sanjsvk/Yelp-Restaurant-Business-Analysis

Time Series Forecasting for energy consumption prediction, Northeastern University — October 2019 – December 2019
Developed a project to predict energy consumption in buildings using time-series energy meter readings for 1 year

- Carried out EDA in Tableau and RStudio to provide insights on efficiency of meters during various times of the year
- Built models based on LSTM, Decision Tree Regressor, and Light-GBM to predict energy consumption and compare RMSE (lowest 1.21) to incentivize energy conservation investments. GitHub: https://github.com/sanjsvk/Time-Series-Forecasting