Bhavish Kumar

315-832-0993 | bhkumar@syr.edu | linkedin.com/in/bhavish-kumar/ | github.com/bhavish2207|Portfolio

EDUCATION

Master of Science in Applied Data Science, Syracuse University, Syracuse, NY (3.97/4.0) May 2021
Relevant Coursework: Statistics, Data Warehouse, Big Data Analytics, Data Mining, Deep Learning, Cloud Management(AWS, GCP)

Bachelor of Technology in Electronics and Communication Engineering, Cochin University, India April 2016

TECHNICAL SKILLS

Languages: SQL, R (caret, ggplot2, tidyverse), Python (Sklearn, PyTorch, keras, matplotlib), RShiny, Spark (MLlib), HTML/CSS
Concepts/Techniques: EDA, Statistics, Data Visualization, Machine Learning & Deep Learning (Regression, Naïve Bayes, Decision Trees, Random Forests, GBTs, SVM, K-Means, Neural Networks), ETL, Data Modeling, Data Warehousing, Dashboarding
Tools/Software: Tableau, Power BI, Alteryx, SSIS, MS Excel, MS SQL Server, Hadoop/Hive, MS Visio, MS Access, AWS, GCP

EXPERIENCE

Data Scientist, *iConsult Collaborative, Syracuse University,* Syracuse, New York Aug 2020 - Present

- Partner with a stakeholder to build Machine Learning models for binary classification of patients utilizing Alteryx
- Develop an end to end Alteryx Data Pipeline to execute ETL & Data Mining operations to predict if a patient has a disease

Decision Scientist, *Mu Sigma Inc.,* Bangalore, India Mar 2018 - June 2019

- Performed Competitor Analysis to generate a 200% rise in customer value proposition score of an eCommerce client
- Formulated a competitor analytics framework comprised of KPI metrics to gauge performance against competitor
- Devised PySpark.SQL queries on survey & Hadoop/Hive data sources for monthly evaluation of KPI metric scores
- Visualized metric scores on an R Shiny dashboard to help Chief Officers and 2500 managers track monthly progress
- Collaborated with client on Big Data exploration, statistical analysis & predictive modeling to improve score accuracy
- Spearheaded a team of 4 that built Partition Tables & Data Pipelines through ETL operations with PySpark.SQL to automate monthly score evaluation process decreasing evaluation & dashboard refresh time by 27-man hours

Trainee Decision Scientist, *Mu Sigma Inc.,* Bangalore, India May 2016 - Feb 2018

- Gathered requirements to develop backend databases of web apps that help fulfill organizational business processes
- Designed normalized databases with MS Visio using Data Modeling techniques and built them on MS SQL Server
- Devised SQL queries & views on MS SQL Server for Data Reporting to aid data-driven decision making
- Developed user interactive web pages using HTML/CSS to serve as frontend for web applications
- Earned Decision Scientist certification issued by Mu Sigma, upon completion of Statistics & Machine Learning courses

ACADEMIC PROJECTS

Retail Sales Forecasting using Recurrent Neural Networks | Python: PyTorch, pandas, matplotlib | Deep Learning | ML

- Aided a retail organization accurately forecast weekly sales at a Store-Department level using Deep Learning models
- Performed Data Preprocessing & developed multivariate time series forecast models by training Long Short Term Memory (LSTM) & Gated Recurrent Unit (GRU) Neural Networks on the processed data, by utilizing PyTorch library
- Trained Linear & Random Forest Regression models to compare performance with LSTM & GRU models

Will Car Be A 'Lemon'? [Webapp Link] **|** Python: sklearn, pandas, matplotlib | R: caret, ggplot2, tidyverse | R Shiny **|** ML

- Recommended good cars to auto dealers to prevent purchase of faulty cars ('Lemons') at auto auctions
- Built Tree Based Ensemble Machine Learning Models to help auto dealers accurately predict if a car will be a 'lemon'
- Analyzed & fine-tuned hyperparameters of Gradient Boosted & Random Forest models to maximize Accuracy & Recall
- Created a web app (R Shiny Dashboard) containing insightful visualizations and a ML model to make predictions

Hospital Readmission Prediction | Python: pandas, matplotlib | Apache Spark: PySpark, MLlib | Machine Learning | Statistics

- Helped hospitals predict if a diabetes patient is going to get readmitted after discharge, to prevent patient readmission
- Trained Tree based Machine Learning models using Apache Spark (PySpark - MLlib) for binary classification of patients

LEADERSHIP & ACCOMPLISHMENTs

Pause & Ponder Nov 2020

- Initiated & led a Data Science project in a team of 4, to win 2nd prize in the "Nov 2020 Altice Innovation Hackathon"
- Built a self-introspection tool [Webapp Link] that applies Text Mining (Topic Modeling) on internet browsing history

Technical Lead, E-Board member of "The Data Science Club", Syracuse University May 2020 – May 2021

- Encourage the Data Science Community by organizing hands on technical events & providing technical support