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, PySpark(MLlib), HTML/CSS
Concepts/Techniques: EDA, Statistics, Data Visualization, Machine Learning & Deep Learning (Regression, Naïve Bayes, Decision Trees, Random Forests, GBTs, SVMs, Neural Networks), ETL, Data Modeling, Data Warehousing, Dashboarding
Tools/Softwares: Tableau, Power BI, Alteryx, SSIS, MS SQL Server, Hadoop/Hive, MS Excel (Vlookup, Pivot Tables), AWS, GCP

EXPERIENCE

Data Scientist at iConsult Collaborative at Syracuse University, Syracuse, New York August 2020 - Present

- Boosting healthcare quality by partnering with stakeholders of a major hospital and developing Machine Learning models to make patient-level disease predictions
- Developing an end to end Alteryx Data Pipeline to execute CRISP-Data Mining/Machine Learning operation

Mu Sigma Business Solutions Pvt Ltd, Bangalore, India May 2016 - June 2019

Decision Scientist

- Generated an annual 200% increase in customer value proposition score of an eCommerce client by formulating a competitor analytics framework comprising of KPI metrics to help gauge performance against market leader
- Devised PySpark-SQL queries on survey & Hadoop/Hive data sources for monthly evaluation of KPI metric scores
- Visualized metric scores on an RShiny dashboard to help Chief Officers and 2500 managers to 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, which decreased evaluation & dashboard refresh time by 27-man hours

Trainee Decision Scientist

- Collaborated with client & gathered requirements for developing web apps to help fulfil their business processes
- Built normalized databases on MS SQL Server using Data Modeling techniques, to serve as backend for web apps
- Developed user interactive web pages using HTML/CSS to serve as frontend for web applications being constructed

ACADEMIC PROJECTS

Retail Sales Forecasting using Recurrent Neural Networks October 2020 - December 2020

- Spearheaded a team of 2 to aid a major retail organization enhance their inventory planning by accurately forecasting weekly sales at a Store & Store-Department level through separate Deep Learning forecast models for each
- Performed data preprocessing steps of adding 'n' lag value attributes for every store & store-department subset
- 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 Regression & Feed Forward Neural Network models to compare performance with LSTMs & GRUs

Will Car Be A 'Lemon'? April 2020 - May 2020

- Recommended good cars to auto dealers to prevent purchase of faulty cars ('Lemons') at auto auctions
- Led a team of 3 in a strategic direction by implementing CRISP-DM methodology to construct ensemble Machine Learning models (GBM & Random Forest), to help auto dealers accurately predict if a car is going to be a 'lemon'
- Analyzed performance of Machine Learning models & fine-tuned hyperparameters to maximize Accuracy & Recall
- Created a web app (RShiny Dashboard) containing insightful visualizations and a ML model to make predictions

Hospital Readmission Prediction April 2020 - May 2020

- Administered a team of 4 that operated PySpark (MLlib library) & devised Machine Learning models to help hospitals predict if a diabetes patient is going to get readmitted after discharge, and thereby help prevent patient readmission
- Performed Data Wrangling, Data Preprocessing and identified most important features leveraging PCA technique
- Trained a Machine Learning model using 'Gradient Boosted Trees' algorithm & fine-tuned it to maximize 'AUC'

LEADERSHIP

- Won 2nd prize in the 2020 Altice Innovation Hackathon, by initiating & executing a Data Science Project in a team of 4
- Led technical activities as Tech Lead, E-Board member of The Data Science Club at Syracuse University May 2020-21