Harshit Krishnakumar

hkk.nitc@gmail.com|(812) 360-8702|107 E John St, Washington, IN-47501
/in/harkrish |/harkrish1|pages.iu.edu/~harkrish

Education

Indiana University — Bloomington, Indiana
M.S. in Data Science | GPA: 3.65 / 4 — May 2018
Relevant Courses: Machine Learning, Big Data, Data Mining, Statistics, Algorithms, Artificial Intelligence, High Dimensional Data Analysis

National Institute of Technology — Calicut, India
Bachelor of Technology in Chemical Engineering | GPA: 6.93 / 10 — May 2013

Technical Skills

Proficient Python, Tensorflow, Keras, MS SQL, ETL (SSIS), SSRS, Power BI, Azure Cloud (IAAS)
Competent R, Tableau, Tableau, Pentaho, Teradata
Familiar C, C++, Hadoop, Apache Pig, QGIS

Work Experience

TrueScripts Management Services — Washington, Indiana
Data Scientist — October 2017 – Present

- Analyse pharmacy claims data to provide proactive and reactive reporting solutions for B2B clients
- Setup seamless integration between different systems for various types of healthcare data like pharmacy claims, member eligibility and MediSpan drug lists using exploratory data analysis techniques in Python, and SSIS/SQL Server as the BI System
- Created KPI Reports to show per member per month cost metrics and trends using Power BI. This is being sent to 170 clients each month as a scorecard
- Used Azure Cognitive Services to build an OCR system to extract information from scanned faxes of "prior-auth" forms. The fax to text conversion system saved up to 4 hours per day in manual data entry

Indiana University — Bloomington, Indiana
Associate Instructor for Machine Learning — Fall 2017

- Teaching assistant for a graduate course Machine Learning (CSCI-B 555), acting as a liaison between the Professor and students

LatentView Analytics — Chennai, India
Senior Analyst — June 2014 – June 2016
Client: Sears

- Designed and automated Agent Performance Dashboard using SSIS and SSRS, which was used by multiple stakeholders to monitor call centre operations
- Devised a new hierarchical attribute "Reason for Call" which was used to tag calls to their reasons. This helped in quantifying the initiatives undertaken by Sears, reported savings of 250K$ per quarter

Client: PayPal

- Identified close to 150 merchants as potential targets for checkout optimization using data mining and classification techniques, this earned a revenue of 3.5M$ over the course of a year
- Analysed web page flows to find variables that affect customer engagement on merchant websites and recommended improvements, this led to an immediate dip in the customer drop rates

Mu Sigma Business Solutions — Bangalore, India
Decision Scientist, Trainee — June 2013 – June 2014
Client: Microsoft

- Worked with Bing Demand Management team to understand requirements and built a data warehouse, from Cosmos DB into SQL Server using SSIS
- Visualized key performance indictors related to demand management in SQL Server Reporting Services. Honoured with spot award for this project

Academic Projects

Voice Activity Detection | Python, TensorFlow, Keras — Fall 2017 – Current
Researching on Voice Activity Detection (VAD) using configurations of Deep Neural Networks. VAD is a binary classifier that detects speech in an audio clip., used in telecom, web conferencing etc. to optimize call quality. Publishing a research paper at IEEE Global SIP 2019 conference

Visualizing User Behaviour on the Places & Spaces Website | Python, SQL Server, QGIS, Tableau — Spring 2017
Analysed Temporal and Geospatial data of user behaviour on the website scimaps.org, and created a presentation using Tableau and QGIS to present our findings

Predicting Malignancy of Tumour | Python, R — Fall 2016
Implemented k-means clustering algorithm on medical data. Analysed the data using R and imported to python for k-means. Got error rates as low as 10% to cluster the data into two groups – benign and malignant

House Price Prediction | Python Scikit Learn, Numpy, matplotlib, pandas — Fall 2016
Predicted house prices in a city based on the houses' attributes using advanced regression techniques like gradient boosting, cross validation and feature reduction by random forests