RAHUL VIJAY RAGHATATE

812 325 6756 | rvraghatate10@gmail.com | IN 47408 | linkedin.com/in/rahulraghatate | www.github.com/rahulraghatate

Data Science Professional | Expertise: R, Python, SQL, Matlab, UNIX, Statistics, ML, Deep Learning, Data Engineering

EDUCATION

M.S., Data Science	GPA 3.6/4.0	Indiana University, Bloomington, IN, USA	**Aug 2016 - May 2018**
B.S., Electronics	GPA 3.6/4.0	Veermata Jijabai Technological Institute, INDIA	**Jul 2010 - May 2014**

WORK HISTORY

Associate Instructor - Indiana University   **Aug 2017 - Present**
- Instructing **INFO-I427 [Search Informatics]** class for the Pythonic build of Web crawling, Indexing, Retrieval, Ranking, Evaluation and SEO techniques and Search Engine implementation as a project.

Graduate Research Assistant - Indiana University   **Jun 2017 - Present**
- Under the supervision of Prof. D. Williamson, implementing deep neural networks (DNNs) based speech separation and speaker recognition system using matlab, python, and tensorflow.
- Preprocessing co-channel signals for noise reduction using complex ratio masking, then frame level log-spectral features as input and true speaker identities as training targets.
- Also, using the ideal binary mask for frame level energy calculation of each speaker from the mixture cochleagram and to use their ratio for the soft labels in softmax output layer.

Data Analytics Consultant - The Harlem Bee, NYC   **Jun 2017 - Present**
- Conducting analysis and predictive modeling over NYC public datasets in Python.
- Providing classification, pattern recognition, forecasting based meaningful discoveries at requisite depth for multi-domain business development in Harlem and seed funding improvement by approx. 15%.

Data Specialist - IBM India [Client – Vodafone UK]   **Dec 2014 - Jun 2016**
- Using Agile Methodology developed returned/exchanged orders application in Abinitio and Teradata overcoming 85% order misclassification and 10% data leakage.
- Performed requirement analysis, designed modules, wrapper scripts using UNIX and SQL for BI report generation.
- Conducted root cause analysis to resolve QA issues in multiple applications in coordination with Vodafone's BI analysts.
- Implemented ETL mapping, parallelism, data checks, fault tolerance techniques through Abinitio graphs and database objects (tables, views) in Teradata and Oracle DB ensuring overall EDW integrity.

SKILLS

- **Tools and Languages:** Python, R, Unix, SQL, Matlab, SAS, Tableau, Excel, Abinitio, Teradata, Spark, Hadoop, Github
- **Statistical/Machine Learning Techniques:** Regression/Classification Models, Neural Net, HMM, Deep Learning, PCA, Time-Series, RF, SVM, Anomaly Detection, High Dimensional Data Analysis, Information Retrieval, Signal Processing.

SELECTIVE PROJECTS

- **House Sale Price Prediction [Python, Sklearn, Ridge, Lasso, XgBoost, Stack Model]** model developed over 'Ames Iowa Dataset'. Used PCA and feature selection for dimensionality reduction and achieved 93% prediction accuracy.
- **Analyzed Manhattan Crimes [Python, Sklearn, Heatmap, Trend analysis]** and developed correlation charts and relational visualizations using agglomerative clustering on location, time, day, and month features.
- **Developed Street Sign Detection Model [Spark Cluster, Python, Ansible, Unix]** deployable over cloud cluster for image and video analysis using OpenCV's HOG classifier achieving 96% accuracy for a stop sign. Also, automated Jetstream (OpenStack) and Chameleon cloud-based deployment and benchmarking using Ansible scripts.
- **Analyzed US CO_2 Emission [SAS, Excel, Time-series, Multivariate Regression Analysis]** in relation to human activities. Validated GLM assumptions and established a statistical model for CO_2 emissions.
- **Tumor Malignancy Prediction [Python, k-Means, MySQL, R]** modeled using k-means clustering and PCA on 'Wisconsin Breast Cancer Dataset' in Python with 0.2% false +ve rate.
- **Implemented OLTP/OLAP Based Customer Lifetime Value Evaluation [Python]** useful for analyzing acquisition strategy and estimating marketing cost and returns on customers.
- **Explored Anomalous Behavior [R, time-series, Robust Linear Models, ggplot2]** of Barley Yield during the year 1931-32 using agridata dataset.
- **Time-series analysis of CH_4 emissions [R, time-series, ggplot2, loess fit]** based on the **paper** published in PNAS to study environmental impacts in Boston.