

# Gaurav Venkatraman

201-884-3626 | gvenkatr@outlook.com | www.linkedin.com/in/gvraman

Portfolio : gvraman.github.io

## EDUCATION

### M.S. Business Intelligence & Analytics, GPA 3.83

May 2020

Stevens Institute of Technology, Hoboken, NJ

**Coursework :** Multivariate Data Analytics, Data Management, Data Warehousing, Business Intelligence, Knowledge Discovery and Data Mining, Big Data Analytics, Web Mining, Social Network Analysis, Optimization And Process Analytics, Financial Technology

### B.E. Mechanical Engineering

May 2016

Pune University, Pune, India

## TECHNICAL SKILLS

**Certifications :** AWS Certified Cloud Practitioner

**Technologies :** Python (Scikit-learn, Pandas, Numpy, BeautifulSoup, NLTK) R (ggplot2, Tidyverse, igraph), Flask, PySpark, Excel, Tableau, VScode, Jupyter

**Databases :** PostgreSQL, MySQL, Redshift, MS SQL Server, SSIS, SSRS, Neo4J

**Machine Learning:** Regression, Clustering, Decision Trees, Random Forest, SVM, kNN, Naive Bayes

**Big Data and Cloud:** Hadoop, Spark, AWS (EC2, S3, Redshift, Elastic Map Reduce), Heroku

## PROFESSIONAL EXPERIENCE

### Product Analyst Intern, Datakron.ai, Charlotte, NC

Aug 2020 – Present

- Automated data scraping using Webscraper on AWS EC2 to extract labor data and thousands of records from job boards. Pushed all the data to s3 buckets periodically to create a data lake
- Translated business queries into a data model and built a scalable database on Neo4J to create a knowledge graph using the data
- Created a Python script to automate loading the data and creation of nodes and relationships using Cypher QL on a Neo4J instance
- Generated dashboards and visualizations using Tableau on market labor data and delivered insights on job market trends

### Research Assistant, Stevens Institute of Technology, Hoboken, NJ

July 2020 – Present

- Research contact tracing solutions available in the market to tackle the spread of COVID19 on university campus
- Created a POC machine learning model using density based clustering to identify potential infected cases based on a positive case
- Successfully communicated the idea to internal stakeholders to implement a pilot project on the college campus

### Graduate Student Assistant, Stevens Institute of Technology, Hoboken, NJ

July 2019 – May 2020

- TA – Supply Chain Management & Analytics. Provided tutorials on Tableau and Excel Solver to a class of 40 students
- Prepared case studies, Power Point presentations, graded assignments, and assisted students regarding the course material
- Maintained enrollment records of courses and created reports using Salesforce to identify popular courses in each graduate program

### Junior Analyst, Simple & Real Analytics, Mumbai, India

Apr 2017 – June 2018

- Built an E2E scalable Software-as-a-Service product on AWS for an Ed Tech company and other IT training providers
- Included tools such as Hadoop (Pig, Hive, Spark), R server, Jupyterhub and a GUI (Hue) for a better user experience. Eliminated the need for manual configuration of the tools on local systems, instead clients could login to the service and access the tools
- Tested different Hadoop vendors in the market and deployed the Cloudera version over AWS EC2. Handled the AWS Management Console for capacity planning, resource allocation and auto scaling of resources as per demand
- Automated data cleansing of geo location data for a logistics company using Python and reduced processing time by 40%

## PROJECTS

### Analysis and Reporting on e-Commerce sales data : MS SQL Server | Tableau

- Imported a database dump with 1M sales orders to SQL Server and normalized the shipping type column to 6 standardized types
- Performed joins and aggregations on SQL and used heat maps to visualize the No.1 shipping type and average order value by state

### Visualization of Vintage Car Auctions : Tableau

- Analyzed sales data of vintage cars sold in auctions. Used tableau to create visualizations and dashboards to present a story
- Delivered insights on best performing regions, potential for online sales etc. using tree maps, scatter maps and other graph types

### Sentiment Analysis on Amazon Reviews : PySpark | AWS | Elastic Map Reduce

- Analyzed 4M rows of data on AWS EMR, stored data in a Spark cluster and processed and modeled text data using MLlib library
- Experimented with different number of nodes and achieved 50 - 80 % lower processing time with increasing number of nodes

### Predicting Publication House of a News Article : NLP | Python | Sklearn | Flask | Heroku | AWS

- Designed a classifier to accurately predict the publishing house of a news article. Analyzed data containing over 100,000 articles
- Used NLTK for cleaning and transforming text content and evaluated different classifiers. MLP performed best with 86% accuracy
- Created a web application using Flask, hosted it on an EC2 instance and also deployed it on Heroku