

# VIRAJ SHETTY

Analytics Practitioner | Data Enthusiast | HIPAA Certified  
vvs200004@utdallas.edu · [linkedin.com/in/virajshetty47/](https://www.linkedin.com/in/virajshetty47/) · [github.com/VirajVShetty](https://github.com/VirajVShetty)  
Richardson, TX · +1 (469)388-7873

## EDUCATION

**The University of Texas at Dallas, Richardson, TX** (GPA: 3.5/4.0)  
MS in Information Technology Management (*Business Analytics Specialization*) August 2021 - Present  
*Relevant coursework: Machine Learning, Advance Statistics for Data Science, Database Foundations, Big Data*

**The University of Mumbai, Mumbai, Maharashtra, India** (GPA: 3.2/4.0)  
BE in Computer Engineering June 2016 - November 2020  
*Relevant coursework: Analysis of Algorithms, Cloud Computing, Natural Language Processing, Data Warehousing*

## EXPERIENCE

**Loopback Analytics, Real World Data Analyst Intern** June 2022 - Present

- Assembled Client Feasibility requirements and extracted data using **Snowflake** to optimized data fetching by **37%**
- Wrote optimized SQL Queries in Snowflake on **100+ GB** of Data to answer client research question.
- Worked on **Data Clustering** of Socio-Economic Data using Census API on **Databricks & Snowflake** which reduced null values in columns from **24%** to **5%**.

**m-Uni Campus, Data Analyst Intern (Part Time)** December 2018 - February 2020

- Produced a Python based **YouTube-scraper** which scraped **10K** links and title of a specific keyword.
- Performed **ETL** using **pandas** to store the data with optimized **SQL scripts** which increased access speed by **67%**.
- Formulated a recommendation system and integrated the scripts in the student portal which led to **30%** increase in student retention.

## PROJECTS

**Truck Risk Factor Analysis, Big Data Project,** August 2022 - Present

- Extracted truck fleet data from **10+** sources, cleaned the data and loaded the data in a **HDFS**.
- Imported this dataset into **Hive Tables** using optimized **Hive Loading Queries**.
- Connected Hive table to **Tableau** using **Hive ODBC** and analyzed data to get truck with highest risk factor.

**Payroll Management System, Database Project, [Code]** January 2022 - May 2022

- Gathered employee details to formulate and developed SQL scripts to fetch employee data from databases
- Constructed database objects, performing Data Normalization, Joins, Relationships, **ERD** using **MySQL Workbench**

**Fraud Transaction Detection Well Fargo, ML Project, [Code]** January 2022 - April 2022

- Processed **1 million** rows of transaction data from Wells Fargo using Prep Builder to extract **300K** rows of data.
- Built a robust model using a **xgboost** with RMSE of **0.459** to detect fraud with an accuracy of **93%**.

**Chicago Crime Data Analysis, Tableau Visualization, [Viz]** January 2021 - March 2021

- Processed **1 million** rows of data from Data.gov Chicago Crime using Prep Builder to extract **500K** rows of data.
- Connected the data source with Tableau to give a detailed analysis of Crime in different parts of Chicago.

## SKILLS

**Programming Languages & Databases:** Python, R, SQL, MongoDB, MS Access.

**Tools & Services:** AWS S3, Git, JIRA, Visual Studio, Tableau, Snowflake, DataBricks, PowerBI, MS Excel

**Framework & Libraries:** Pytorch, Tensorflow, Pandas, Matplotlib, Numpy, Keras, Scikit-Learn.

**Machine Learning Algorithms:** Random Forest, KNN, Naïve Bayes, SVM, Decision Tree.

**Statistical Techniques:** Hypothesis Testing, Statistical Analysis, Inferential Statistics, Logistic Regression

**Big Data Tools:** Hadoop, Spark, Hive, Impala, SQOOP, Flume

## EXTRACURRICULAR & CLUBS

- Communications Officer** at **Project Management Club** - University of Texas at Dallas
- Music Committee Member** at **SUAAB** - University of Texas at Dallas
- Member of **TapTheTech** - NGO that advocated equal education featured on Coursera.org [\[Link\]](#)