

Viraj Shetty

Dallas, TX - 75252 | (469) 388-7873 | virajvshetty47@gmail.com | <https://www.linkedin.com/in/virajshetty47/>

EDUCATION

THE UNIVERSITY OF TEXAS AT DALLAS

Master of Science, Information Technology and Management

Aug 2021 – May 2023

3.67 GPA

THE UNIVERSITY OF MUMBAI

Bachelor of Engineering, Computer Engineering

Jun 2016 – Nov 2020

3.2 GPA

TECHNICAL SKILLS

- **PROGRAMMING:** R, SQL, API, JavaScript, Python ([Pytorch](#), [TensorFlow](#), [Pandas](#), [Matplotlib](#), [NumPy](#), [Keras](#), [Scikit-Learn](#))
- **B.I TOOLS & DATABASES:** SAS, PowerBI, SPSS, WPS, Tableau, Alteryx, MS Excel, MySQL, MongoDB, MS Access, SQLite
- **MACHINE LEARNING:** Random Forest, KNN, Naïve Bayes, SVM, Decision Tree, Gradient boosting, Dimensionality Reduction
- **STATISTICS:** Hypothesis Testing, Statistical Analysis, Inferential Statistics, Logistic Regression
- **BIG DATA:** Hadoop, Sqoop, Spark, Hive, Impala, Flume Kafka, Scala, PySpark
- **OTHER:** Linux, Git, Snowflake, Databricks, JIRA, Visual Studio, Git, Jupyter Notebook, VMWare, Streamlit

WORK EXPERIENCE

LOOPBACK ANALYTICS, DALLAS, TX

Jun 2022 – Dec 2022

[Real World Data Analyst Intern](#)

- Performed **Ad-Hoc** analysis and wrote optimized **SQL** queries for **100+ GB** data to improve data fetching efficiency by **37%**.
- Improved data pipeline refresh rate by **25%** through maintenance and **PowerBI** report implementation
- Worked on data clustering of Socio-Economic Data using Census-API on **Databricks & R**, resulting in a significant decrease in null values in column from **24% to 5%**.
- Spearheaded assembly of feasibility requirements using **Snowflake & Excel** resulting in an increase in efficiency of **15%**.
- Collaborated across different functions, identified, and communicated new insights through **cross-functional** teams.

M-UNI CAMPUS, MUMBAI, INDIA

Dec 2018 – Feb 2020

[Data Analyst Intern](#)

- Built data pipelines to collect, process, and store datasets from multiple sources using **Apache Spark & AWS Glue**.
- Automated the **data ingestion** process, reducing manual errors and processing time by **60%**.
- Conducted **root cause analysis** for the campus business model and implemented a Python-based YouTube scraper that scraped **10K** links and title of a specific search term keyword.
- Performed **ETL** using pandas to store the data with optimized **SQL scripts** which increased data access speed by **67%**.
- Implemented a recommendation system in the student portal, resulting in a **30%** increase in student retention.

ACADEMIC PROJECTS

Truck Risk Factor Analysis: [Big Data Project](#)

[HDFS](#) | [TABLEAU](#) | [PYSPARK](#)

- Extracted and loaded **100,000** rows of truck fleet data in **HDFS** and imported this data in **Hive Table**.
- Formulated a visualization using **Tableau** which resulted in mitigating the risk factory by **28%**.
- Trained a **Logistic Regression Model** with an accuracy of **83%** and an ROC of **0.78**.

Payroll Management System: [Database Project](#) [\[Code\]](#)

[MYSQL WORKBENCH](#) | [VISIO](#) | [SQL](#)

- 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: [Machine Learning Project](#) [\[Code\]](#)

[TABLEAU PREP](#) | [PYTHON](#) | [JUPYTER NOTEBOOK](#)

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

Quora Duplicate Question Identification: [Machine Learning Project](#) [\[Code\]](#)

[PYTHON](#) | [NLP](#) | [SKLEARN](#)

- Processed **10,000 rows** of similar questions asked in Quora and added contracted words dictionary.
- Developed and implemented a machine learning model using callbacks such as EarlyStopping and ReduceLROnPlateau to improve the training process.
- Improved previous model resulting in better similar question pair detection with an accuracy of **83%**.

Real Time Bridge Health Management System: [Capstone Project](#) [\[Code\]](#)

[PYTHON](#) | [PYQT5](#) | [FIREBASE](#)

- Researched **10+** papers to create a framework for a Real Time Bridge Health Management System
- Formulated a web-based application that aggregated multiple data sources which resulted in ease of access by **40%**
- Created an automated Bridge Health Monitoring to reduce human intervention by **100%** using Audio and Image Processing

HONORS AND CLUBS

- [Member of TapTheTech](#) - NGO that advocated equal education featured on Coursera.org [\[Link\]](#)

CERTIFICATIONS

- [HIPAA Certification](#) (2022-2023), [Google Ads Certification](#) (2022-2023), [JIRA Fundamentals](#)