Harsh Shah
San Jose, California - 95113 • (216)612-9582 •shahharsh178@gmail.com
LinkedIn: - https://www.linkedin.com/in/harsh-shah-783723a6/

- Business Analytics and Machine Learning expert with deep knowledge in the field of Artificial Intelligence using Python, R and MATLAB. Involved in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling and data visualization using Tableau with large data sets of structured and unstructured data. Having knowledge on Apache Spark and developing data processing and analysis algorithms using Python. Effective collaborator with excellent communication and interpersonal skills, ability to work individually and as a part of team, fast and keen learner and the energy to align diverse project interests with project management expectations.

EDUCATION:

Master of Science in Computer and Information Science, Cleveland State University, Cleveland, Ohio **May 2019**
Bachelor of Technology in Computer Science and Engineering, Gujarat Technological University, India **May 2017**

TECHNICAL SKILLS:

- **Machine Learning:** classification, regression, clustering, feature selection, deep learning, neural networks, Artificial intelligence
- **Programming Knowledge:** Python (pandas, Scikit-learn, Numpy, Scipy, GraphLab Create), R, SQL, Java, Spark (PySpark, MatplotLlib), MATLAB, C#, C++
- **Software Tools:** PyCharm, Apache Jupyter Notebook, **R studio**, **Palantir**, **Tableau,** Microsoft Office, Eclipse IDE, JIRA, Microsoft Azure, SAS, **SPSS**, **MS SQL**, Oracle DB, IBM DB2

Work Experience:

Business Analyst at FCA (Fiat Chrysler Automotive), Auburn Hills, MI **December 2019 – June 2020**

- Collaborated and negotiated with users, IT team, managers, and stakeholders to gather requirements, analyze and create reports using Excel, Palantir, Dashboards & Scorecards (KPI) and created analysis with Palantir environment to develop Python, Java, SQL programming using database connection and tables associated with it. Have created Contour analysis for different programs and projected it within the company.
- Helped FCA to reduce cost of 2 million dollar per quarter using Palantir and **Tableau** by creating visualization and financial reports for per quarter basis and vehicle reports reported to NHTSA. Created data pipeline using machine learning and Python and created data testing environment on Palantir and Tableau for report generation.
- Providing Business Solutions to the management using analytical knowledge of programming and visualization. Created and optimized stored procedures and queries to extract, transform, validate and load ETL process and flow diagrams. Develop dashboards using Excel, Palantir, PowerPivot, R, Power Query and reduced manual reports by 70%. Used Palantir gateways to keep the reports updated. Optimized the process for fiscal year reporting by IBM DB2 database.
- Coordinated, prepared and presented analysis to the management and to the team using Palantir for VSRC project within FCA.
- Mentored a team of 5 employees on dashboard development for 5 months.

Data Scientist at Ag-Analytics LLC, Ithaca, NY **September 2019- November 2019**

- Assisted in research of big data in Farming industry using ArcGIS (ESRI), Python and Tableau development.
- Collected, studied and provided modelling for large dataset, conducted reports and data management. Used OAuth 2.0 to fetch data from Client's Web API and used it for analysis of data and modeling. Created Web Application using C# and SQL database for client's needs and added Python functionality associated with it. Managed Front End Web development.
- Created website for detailed crop information with graph analysis for crop generation and crop planting with detailed analysis for soil testing for Profit layers, FarmScope and Fieldlayers using C# and MSSQL. Converted JSON data to SQL data using Python.
- Imported data from Azure SQL DB, Azure Blob, Excel, SQL Server DB, MS Access to generate and automate reports by 90% to get real-time data.

Data Analyst at Samay Software Ltd, Vadodara, India **June 2015- May 2017**

- Created data libraries and developed software based on customer requirements and needs and providing technical support for customers using Java and Python.
- Gathered requirements from the client for the development of a website, using JIRA and MS Office suite. Conducted analysis on pricing, leads, customer behavior to help drive long-term strategy using SQL Server, Python, MS Excel, and Tableau.
- Design and developed reports in SQL Server reporting Services. Manage the development of design and procedure for metrics reports. Developed client reports, KPI for the website with Tableau that include data cleaning, data mining, analysis, profiling, user behavior metrics, regression analysis, and ETL strategies.
- Proposed solutions to improve system efficiencies and reduce total cost. Analyzed accuracy of data migration and an increase in client enrolment by using EXCEL.
- Worked closely with the marketing teams to ensure product sustainability and digital marketing.

Math Emporium Data Analyst and Student Tutor at Cleveland State University **August 2018- May 2019**

- Assisted students about specific topics or subject areas that are giving them trouble and give them further explanation, take notes during sessions so student can follow up later.
- Perform testing requirements for students and updating databases. Helped total 200 student's complete mathematical analysis and evaluate them based on the results.
- Assisted peers and colleagues in fundamental understanding of calculus and comforted to complete the elementary course.

Certifications:

Data Science
Data Science Methodology
Python
Deep learning
Python Data Analysis
Text analytics
R analytics
Machine learning - Dimensionality Reduction
Project Management
Tableau Advanced Training

Academic Projects:

Audio Clustering using Machine Learning

- Developed a machine learning based project which can identify song's type like happy, sad, pop or rock based on song's frequency and bandwidth using **R language, Octave and Audacity** using clustering unsupervised algorithm.
- Implemented recursive algorithm and analytics system using any type of song's frequency and wavelength.

Document Clustering using Python

- Gathered information from Portland police department and create an Arc-GIS based map for the areas which have seen maximum crime as well as minimum crime during a year time using **Python** PyCharm, SPSS and **R**. Used Apache Jupyter notebook to develop the clusters and did analysis using **Tableau** by building interactive dashboards.

Real Estate Valuation analysis using Python

- Gathered information from UCI database for Taiwan's total valuation prices per square foot and successfully identified house price per unit of area in Taiwan based on given information for past 10 years with maximum impact. Have used Logistic Regression, Neural Network, Decision Tree and Gradient boosting algorithm to identify nearby house price from given Longitude and Latitude and operated **Google Analytics** and utilized the analysis and plotted it on Google maps.

Predicting 30-Day Hospital Readmissions using Python

- Built a machine learning model that uses the UCI machine learning repository dataset of hospital re admissions and perform predictive analytics using **Python** and **R** on the data to forecast the highest 10% patients at maximum risk of readmission within a 30-day time period. We used **ROC curve**, precision-recall curves to evaluate the results.