VIKRANT GAWDE

P: +15083698033, E: vxg7583@rit.edu LinkedIn: www.linkedin.com/in/vikrant06

EDUCATION:

Rochester Institute of Technology, Rochester, NY| current courses (*)
Master of Science – **Information Science and Technology** (GPA 3.6/4.0) **Expected Dec'17**
Courses: Data Mining, Information Retrieval (Java, NLP), Data Warehousing (OLTP, OLAP, ETL, Dimensional Data Analysis, Data Mining, SAS, Pentaho), Analytical Thinking (Machine Learning Algorithms and Tools, Naïve Bayes, K-means, Neural Networks).
Machine Learning & Big Data (SVM, Decision Trees, Neural Networks, Hadoop, Mahout, MapReduce) *, Time Series Analysis & Forecasting*, Geographic Information Systems (ArcGIS, ArcGIS online, Collector for ArcGIS)*

Narsee Monjee Institute of Management Studies, Mumbai, India
Bachelor of Technology -- **Computer Science** **May 2016**
Database Technology, Distributed Computing, Intelligent Systems, Big Data Analytics, **Software Engineering** and Data management.

TECHNICAL SKILLS:

Languages:	**Java**, **Python (Pandas, scikit-learn), R, SQL**	**Web Technologies:**	JavaScript, HTML 5/ CSS 3, JQuery, XML
Databases:	MySQL, MongoDB, Neo4j	**Tools:**	Pentaho, PowerBI, Tableau, Weka, MS Excel, Qlik

WORK EXPERIENCE:

Health Catalyst, SLC, Utah| Data Architect Intern| Site Utilization & Cost Optimization for UPMC (Predictive Analytics)
(June 2017 – September 2017)
(**R**, Azure SQL, Dimensional Modeling, Qlik, ETL, SSMS, Visualization, PowerBI)

- Fetch and manipulate healthcare and finance data using R and SQL to perform Dimensional modeling and exhaustive analytics to make informed decision for UPMC EPIC healthcare departments. Design, develop and deploy data marts for Health Catalyst where tasks involve providing comprehensive data warehousing solutions for healthcare systems. Integration of clinical, financial and patient data into a single repository for business intelligence, analytics and to further build a model for predictive analysis. Statistical machine learning activities like Lasso, Random Forest and Linear Mixed Models.

NMAH & Associates, Mumbai, India| Data Analyst Intern [Part Time]
(May 2013 – Jul 2016)
(MySQL, V look ups, Dimensional Modeling, **SSIS**, Regression, Classification)

- Responsible for designing data marts (star schema) for clients to satisfy varied business processes. Created statistical models to generate insights to drive business decisions.

PROJECTS:

Crime Classification in San Francisco (In Progress)

- Currently working on a classification model that would predict and provide hidden insights on the type of crime classification in San Francisco and the factors that govern this classification output. Various machine learning models like Gradient Boosting Trees, SVMs, and Logistic Regression are being implemented and compared. After this analysis, changes to the data and mathematical transformations will be performed to get the model that best fits the data and is capable of predicting and providing valuable insights.

Design and implementation of Data warehouse

- Performed dimensional analysis and designed a star schema for a data warehouse model for a fictional product company using MySQL Workbench. Performed data integration using Pentaho and performed OLAP operations using SQL to build aggregated Data Marts for accessibility of the end users.

Predictive Analysis of Medical No Shows Using Machine Learning Algorithms

- Predictive model to analyze whether a given patient would show up or not for the appointment scheduled based on parameters like age, sex, occupation and appointment details.

Search Engine for Movie Database Using MongoDB

- Developed a search engine which could filter and retrieve movies based on genre, actor name, keywords, movie name and partial search mechanism. This application was built using php drivers for MongoDB where information could be added, deleted or modified in the database. Involved features like shopping cart and comments for a specific movie.

Community Information Engine| SearchRIT

- A web application that behaves like a smart search engine to retrieve data that is scraped from the Reddit discussion board pertaining to Rochester Institute of Technology. This is primarily a distributed system that uses the machine learning model which is generated using the simple K-means clustering algorithm to cluster documents based on a similarity score. This engine utilizes the OKAPI BM25 algorithm to search for terms and phrases in the documents.

OTHER PROJECTS:

- A defense algorithm implemented on the chess platform using look-ahead principle and SDLC methodology | **Java**.
- A project to display three recipes using concepts of XML like ID/IDREF, enumerated attributes, parameter entities and DTDs
- **Correlation and Regression analysis on the annual rainfall vs annual rice production in India over a span of fifty years**