ANKITESH S. RAO

52 Callodine Ave, Upper
Buffalo, NY 14226
Phone No: +1 716 (563)-8128
Email: ankitesh@buffalo.edu
GitHub: https://github.com/AnkiteshR
LinkedIn: https://www.linkedin.com/in/ankiteshsinghrao/

SUMMARY

Software Developer, skilled in Python, SQL and ETL with an experience of 3 years in the field of information technology with a comprehensive background in SQL querying and optimization. Possesses extensive knowledge in the field of data analysis, visualization, business intelligence, database management and machine learning. Currently seeking internship/full-time opportunities in data Science/data analytics/business intelligence from summer 2019.

EDUCATION

MS in Management Information Systems, University at Buffalo, June 2019 (Expected), 3.83/4.0

- **Coursework:** Database Management Systems, Distributed Computing and Big Data Technologies, Advanced Statistics for Managers, Data Visualization using Tableau, Predictive Analytics, Web Analytics for E-Commerce

Bachelor of Technology in Information Technology, Jaypee Institute of Information Technology (India), May 2015

- **Coursework:** Machine Learning and Big Data Analytics, Database Systems, Data Mining, Information Systems

TECHNICAL SKILLS & CERTIFICATIONS

- **Programming Languages:** Python, SQL
- **Libraries and Frameworks:** NumPy, Pandas, Matlplotlib, Scikit-learn, Seaborn, Keras, OpenCV, TensorFlow
- **Databases:** SQL Server, Oracle 11g
- **Statistics:** Hypothesis Testing, Exploratory Data Analysis, ANOVA, Descriptive Statistics, Probability Distribution, Sampling Distribution, Regression
- **Machine Learning:** Pattern Recognition and Predictive Modeling using Classification, Regression, Clustering, Linear Regression, Logistic Regression, Decision Trees, Random Forest, Association Rule Mining, Naive Bayes, Neural Network, K-Nearest Neighbors, Recommendation Systems, Support Vector Machines, Gradient Descent
- **Hadoop Ecosystem:** HDFS, MapReduce
- **Tools:** Jupyter Notebook, PyCharm, Microsoft Power BI, Tableau, IBM Watson, MS Excel (Filters/Slicers, VLOOKUP, VBA, Macros etc.), MS Office, Confluence, Microsoft Team Foundation Server, GIT
- **Certifications:** Tableau Desktop Specialist, Querying Microsoft SQL Server 2012/2014 (70-461), Data Warehousing for Business Intelligence, a 5 course specialization, University of Colorado, Coursera

PROFESSIONAL EXPERIENCE

Senior Member Technical, CDK Global www.cdkglobal.com **Mar 2018 – Jul 2018**

- Analyzed and optimized existing SQL queries, improving effectiveness by 20%.
- Wrote functions and stored procedures, and updated existing ETL to reduce overall running time of a report from 2 hours to 3 minutes.

Software Engineer 2, MAQ Software www.maqsoftware.com **Aug 2015 – Jan 2018**

- Centralized data pull for more than 15 upstream sources by creating an automated data refresh application through SSIS.
- Processed data collected at data staging layer by writing stored procedure and storing data as facts and dimensions in SQL Data Warehouse.
- Created OLAP Multidimensional Cube (SSAS) to assist business clients in Sales Trend Analysis through Metrics such as MoM, YoY etc.
- Interacted with clients to gather business requirements, modeled them into user stories and negotiated sprint plan for structured delivery process.
- Responsible for researching, replicating, performing root cause analysis and providing solution to the data issues reported by customer.

ACADEMIC PROJECTS

Logistic Regression Model for Autistic Spectrum Disorder (ASD) Prediction **Jan 2018**

- Classified autistic patients based on the screening results with logistic regression.
- Obtained metrics depicting the prevalence of ASD across gender, race and country.
- Built logistic regression model and obtained the ROC Curve and F1 Score. F1 Score after 10-Fold Cross Validation 99.73% (+/- 0.02%).

Utilized: Machine Learning, Python, Numpy, Pandas, Seaborn, Matplotlib, Sklearn, Jupyter Notebook

Air Quality Prediction based on Relative Humidity **Dec 2018**

- Predicted the relative humidity at a given point in time based on the all other attributes affecting the change in RH.
- Cleaned the data and replaced the NaN values with the monthly average of the particular hour.
- Split the data into test and train sets and applied linear regression, decision tree, random forest and support vector machine for designing the model for predicting RH.
- Predicted the RMSE (Root Mean Square Error) for the various algorithms and concluded that the random forest algorithm is the best choice for predicting the RH because of the lowest RMSE value.

Utilized: Machine Learning, Python, Numpy, Pandas, Seaborn, Matplotlib, Jupyter Notebook

Exploratory Data Analysis and Reporting **Fall 2018**

- Used MS Excel and Tableau Prep to create an employment dataset from the United States Department of Labour, Bureau of Labour Statistics.
- Exported the dataset created to Tableau to create a dashboard (Civilian Labour Force Reporting) to understand the overall employment scenario in the US from different perspectives.
- Performed region, gender, age and race wise analysis, created time series visualizations and did forecasting using Tableau to understand the dynamics of people contributing to the growth.

Utilized: Tableau, Tableau Prep, MS Excel