

SIDDHARTH OZA

Data Scientist/Analyst Enthusiast

Aspiring data scientist/analyst with an engineering degree and expertise in implementing Machine Learning algorithms. Three plus years of experience in data analysis, data visualization while collaborating with all members of the organization to achieve the business objective with good verbal and analytical skills.

Education

- Aug 2017** **MS in Business Analytics (Data Science)** | GPA 3.8
May 2019 The University of Texas at Dallas, Richardson, TX, USA
- Aug 2009** **B.Eng. in Electronics & Telecommunication** | GPA 3.5
Jun 2013 Government college of Engineering, India






Experience

- Jun 2018** **FedEx Services Inc. – Memphis, Tennessee**
Aug 2018 *Data scientist Intern, IT Marketing Team*
- Built Predictive model with 62% accuracy using python to identify the total lbs. of good the client will be sending in the coming six months of his contract
 - Built an ETL pipeline using Abinitio to extract 6.7 million rows of data from different sources and loaded it into Teradata data warehouse for further analysis.
 - Integrated Spotfire with Teradata and developed analytical dashboard using crosstabs, heatmaps, geographic maps, pie charts, data on demand which depicted performance across sales
 - Created Jenkins pipeline using python to deploy spring application to the pivotal cloud foundry for continuous delivery.
- Nov 2013** **Quinnox Consultancy Services Ltd – Pune, India**
Feb 2017 *Data Analyst, Mauritius Banking Client*
- Developed analytical dashboard in tableau to provide actionable customer insights across different customer groups, segments and region
 - Spearheaded the business development for CRM module and increased revenues by 20% in 1 year
 - Designed dashboard reports for bank managers which gave data visualization of monthly performance of the reporting bank staff
 - Designed complex process flows for CRM system which required interfacing with 12 internal banking applications

Academic Projects and Competitions

- Jan 2019** **Natural Language Processing (NLP)- Spam E-mail Classifier** ([Git](#))
Feb 2019
- Built a Naïve Bayes classifier prediction model in python for spam E-mail/SMS classification with 97% precision and 98% recall
 - Created API for the classification model using Flask in python
- Nov 2018** **Applied Machine Learning- Cardiac Arrhythmia Multi Class Classification** ([Git](#))
Dec 2018
- Performed exploratory modeling, standardization, data cleaning using missing value treatment in Pandas on 1.5 million records in Microsoft Azure
 - Built classification algorithms like Logistic Regression, Decision trees, KNN, Random Forest, Linear and kernelized SVM and found the best parameters using scikit learn grid search with cross validation.
 - Improved the classification models by using different bagging and boosting methods thus achieving an accuracy of 89%
- Aug 2018** **Data Visualization Competition- New York taxi trip classification in tableau**
Sep 2018
- Analyzed yellow cab trip data in New York city using visualizations based on the 1.4 million records in Tableau 10.4
 - Derived hourly trends for various location and illustrated periods of the day with highest and lowest demands
 - Determined the top 3 locations with highest cab demand along with time required to reach to their destination

Personal Data

-  Dallas, TX 75252
-  +1(214)533-8209
-  Siddharth.oza@utdallas.edu
-  <http://www.linkedin.com/in/siddharthoza>
-  <https://github.com/siddharthoza>

Skills and Competencies

Analysis Tools and Programming:

Python | R | SQL | Hadoop | Spark | Hive |
UNIX | Tableau | Power BI | Advanced Excel |
Adobe Analytics | AB Initio | MS SQL Server
2012

Cloud Experience:

Google Cloud Platform | Microsoft Azure

Skills:

Feature Engineering | Machine Learning |
A/B Testing | Random Forest | Linear
Regression | Clustering | Logistic Regression |
Statistics | NLP Sentiments/Text Analysis |
Decision Trees | SCIKIT-learn | Ensemble
Modeling | Statistical Modeling | Exploratory
Data Analysis | Time Series | Keras | ARIMA

Certifications and Achievements

- Machine Learning by Stanford University - Andrew ng (Coursera)*
- How Google does Machine Learning on GCP (Coursera)*
- Python for Data Science (DataCamp)*
- SQL for Data Science (DataCamp)*
- Technical Mastery Award – FedEx Services*
- Employee of the month Award Winner at Quinnox Consultancy Services*

Leadership

- Leader | JSOM Dean's Council*
- Audit Officer | IAS-UTD*
- Volunteer | American Red Cross*
- Member | UTD JSOM Toastmasters*