

+1-469-655-6141

sushidhar26@gmail.com

<https://www.linkedin.com/in/sushidhar-jayaraman/>

<https://github.com/Sushidhar>



SUSHIDHAR JAYARAMAN



EDUCATION

Master of Science (M.S) in Information Systems | University of Texas – Arlington

AUG 2016 – MAY 2018 GPA: 3.83

- **Big Data Analytics** – Event Recommendation Engine: Built an Event Recommendation engine to recommend events to users based on the events they attended previously. Built Stochastic Gradient Descent classifier and used one-vs-all classifier model to predict the value of the outcome. Ran a 10-fold cross validation with 10 million training set and got an accuracy of 67%. Technologies: PySpark, Scala, NumPy, Plotly, Apache Hadoop.
- **Web and Social Analytics** – Collected 1 million raw tweets for the sentiment analysis of President Donald J. Trump's new administration. Built SVM model based on the extracted feature to predict the sentiment of the new tweets. The feature-based model provided the improved accuracy of 77% compared to the baseline model with 62% accuracy. Technologies: Python (Numpy, Pandas), R, Twitter API, Tableau.
- **Data Science** – Employee attrition Rate: Built logistic regression model with 53 variables. Based on the multicollinearity and significance of the variables and after many iterations, final model was built with 10 variables with improved accuracy of 73%. Technologies: Python, R, MS Excel, IBM Data Science Experience.



EXPERIENCE

Data Scientist | Unique Software Development LLC, Dallas, TX

JAN 2018 – PRESENT

- Formatted, Cleaned, combined and filtered millions of rows and hundreds of variables of The Hub Group, Inc load carrier data.
- Handled importing data from various sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS.
- Responsible for design and development of advanced R/Python programs to transform and harmonize data sets in preparation for modeling.
- Helped creating a clustering machine learning model to rate top 5 carriers suitable for each load. Implemented scoring algorithm to assign scores for each carrier.

Data Engineer | Tata Consultancy Services, Chennai, India

AUG 2013 – JUL 2016

- Developed core ETL logic in the software Product – generates scripts which migrates data from source to target data sources.
- Responsible for performing maintenance activities including database consistency check, data backup and restore strategy, and index defragmentation on SQL Server data store. Developed ETL processes and improved SQL reporting performance by 2x.
- Wrote simple and advanced SQL queries and scripts to create standard and adhoc reports for managers.
- Developed best practice, processes, and standards for effectively carrying out data migration activities.
- Developing automated tools and data pipelines to collect data from disparate source (Files, E-mail, and Website) and store it in the database for Analytics on daily basis.
- Developed Xml Parsers, Static Code Analyzers and migrated data from legacy sources to relational databases.
- Leveraged ETL tools to implement SCD's (Slowly Changing Dimension) and maintain historical data warehouses.
- Designed and developed Metadata Management feature in Java Struts Framework.
- Developed procedure to improve quality of critical tables by eliminating the possibility of entering duplicate data into warehouse.



SKILLS

- Java, Javascript, Python, SQL, Oracle, PostgreSQL, Azure
- Hadoop, Spark, Scala, Kafka, Flume, Linux, MS SQL Server
- R, Tableau, Bluemix, Weka, Nltk, Qlikview, Maven, GIT
- Numpy, pandas, sklearn, SPSS, SAS, Stata, MS Excel

Oracle Certified Professional, Java SE 6 Programmer.