**Rajeev Jadda**

**jaddarajeevreddy@outlook.com**

**(682) 302-0941**

**PROFESSIONAL SUMMARY:**

* Professional qualified **Data Scientist/Data Analyst** with **over 8 years** of experience in **Data Science** and **Analytics** including Machine Learning, Data Mining and Statistical Analysis
* Involved in the entire data science project life cycle and actively involved in all the phases including **dataextraction, data cleaning, statistical modeling** and **data visualization** with large data sets of structured and unstructured data
* Experienced with machine learning algorithm such as logistic regression, **random forest, XGboost, KNN, SVM, neural network, linear regression, lasso regression** and k-**means**
* Implemented **Bagging** and **Boosting** to enhance the model performance.
* Strong skills in statistical methodologies such **as A/B** test, experiment design, **hypothesis** test, **ANOVA**
* Extensively worked on **Python 3.5/2.7** (Numpy, Pandas, Matplotlib, NLTK and Scikit-learn)
* Experience inimplementing data analysis with various analytic tools, such as **Anaconda 4.0JupiterNotebook 4.X, R 3.0** (ggplot2, Caret, dplyr) and **Excel**
* Solid ability to write and optimize diverse SQL queries, working knowledge of **RDBMS** like **SQLServer2008**, **NoSql**.
* Experience in visualization tools like, **Tableau9.X, 10.X** for creating dashboards
* Excellent understanding **Agile** and Scrum development methodology
* Used the version control tools like **Git 2.X**
* Passionate about gleaning insightful information from massive data assets and developing a culture of sound, data-driven decision making
* Ability to maintain a fun, casual, professional and productive team atmosphere
* Experienced the full software life cycle in **SDLC**, **Agile** and Scrum methodologies.
* Skilled in Advanced Regression Modeling, Correlation, Multivariate Analysis, Model Building, Business Intelligence tools and application of Statistical Concepts.
* Proficient in Predictive Modeling, Data Mining Methods, Factor Analysis, **ANOVA**, **Hypotheticaltesting**, normal distribution and other advanced statistical and econometric techniques.
* Developed predictive models using Decision **Tree**, **RandomForest**, **NaïveBayes**, **LogisticRegression**, **ClusterAnalysis**, and Neural Networks.
* Experienced in Machine Learning and Statistical Analysis with **PythonScikit**-Learn.
* Experienced in **Python** to manipulate data for data loading and extraction and worked with **python** libraries like **Matplotlib**, **Numpy**, **Scipy** and **Pandas** for **dataanalysis**.
* Worked with complex applications such as **R**, **SAS**, **Matlab** and **SPSS** to develop neural network, cluster analysis.
* Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data.
* Skilled in performing **dataparsing**, data manipulation and data preparation with methods including describe data contents, compute descriptive statistics of **data, regex, split and combine, Remap, merge, subset, reindex, melt and reshape**.
* Strong **SQL** programming skills, with experience in working with functions, packages and triggers.
* Experienced in Visual Basic for Applications and **VB** programming languages to work with developing applications.
* Experienced in Data Integration Validation and Data Quality controls for **ETL** process and **Data** Warehousing using **MS Visual Studio SSIS, SSAS, SSRS**.
* Proficient in **Tableau** and **R**-**Shiny** data visualization tools to analyze and obtain insights into large datasets, create visually powerful and actionable interactive reports and dashboards.
* Automated recurring reports using **SQL** and Python and visualized them on **BI** platform like **Tableau**.
* Worked in development environment like **Git** and **VM**.
* Excellent communication skills. Successfully working in fast-paced multitasking environment both independently and in collaborative team, a self-motivated enthusiastic learner.

## EDUCATION:

Bachelor’s in Electrical and Electronics from JNTU, India

## TOOLS AND TECHNOLOGIES:

|  |  |
| --- | --- |
| **Languages** | Java 8, Python, R |
| **Packages** | ggplot2, caret, dplyr, Rweka, gmodels, RCurl, tm, C50, twitter, NLP, Reshape2, rjson, plyr, pandas, numPy, seaborn, sciPy, matplot lib, scikit-learn, Beautiful Soup, Rpy2, sqlalchemy. |
| **Web Technologies** | JDBC, HTML5, DHTML and XML, CSS3, Web Services, WSDL, |
| **Tools** | Erwin r 9.6, 9.5, 9.1, 8.x, Rational Rose, ER/Studio, MS Visio, SAP Power designer. |
| **Databases** | SQL, Hive, Impala, Pig, Spark SQL, Databases SQL-Server, My SQL, MS Access, HDFS, HBase, Teradata. |
| **Reporting Tools** | MS Office (Word/Excel/Power Point/ Visio), Tableau,Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x, Cognos7.0/6.0. |
| **ETL Tools** | Informatica Power Centre, SSIS. |
| **Version Control Tools** | SVM, GitHub. |
| **Project Execution**  **Methodologies** | Ralph Kimball and Bill Inmon data warehousing methodology, Rational Unified Process (RUP), Rapid Application Development (RAD), Joint Application Development (JAD). |
| **BI Tools** | Tableau, Tableau server, Tableau Reader, SAP Business Objects, OBIEE, QlikView, SAP Business Intelligence, Amazon Redshift, or Azure Data Warehouse |
| **Operating System** | Windows, Linux, Unix, Macintosh HD, Red Hat. |

**PROFESSIONAL EXPERIENCE:**

|  |  |
| --- | --- |
| **Client: CVS Health, Irving, TX** | **May 2017 – Till date** |
| **Role: Data Scientist** |  |

**Responsibilities:**

* Responsible for gathering requirements from business analysts and operational analysts and identifying the data sources required for the reports.
* Used Python programs automated the process of combining the large SAS datasets and Data files and then converting as Teradata tables for Data Analysis.
* Developed Python programs for manipulating the data read from Teradata data sources and convert them as CSV files.
* Worked on numerous ad-hoc data pulls for business analysis and monitoring.
* Designed and developed various monthly and quarterly business monitoring excel reports by writing Teradata SQL and using in MS Excel pivot tables.
* Performed in depth analysis on data and prepared ad-hoc reports in MS Excel and SQL scripts.
* Performed verification and validation for accuracy of data in the monthly/quarterly reports.
* Developed Teradata SQL scripts using various characters, numeric and date functions.
* Created multi-set tables and volatile tables using existing tables and collected statistics on table to improve the performance.
* Extract up to date accounts data from Teradata database by using SQL, BTEQ.
* Developed the Interfaces in SQL, for data calculations and data manipulations.
* Used MS Excel and Teradata for data pools and adhocs reports for business analysis
* Performed in depth analysis in data & prepared weekly, biweekly, monthly reports by using SQL, MS Excel.
* Experience in automation scripting using shell and Python.
* Automated Windows SAS Scripts on UNIX SAS platform Used Python programs to automate the process of combining the large SAS datasets and Data files and then converting as Teradata tables for Data Analysis.
* Designed stunning visualization dashboards using tableau desktop and publishing dashboards on tableau server and desktop reader.
* Developed programs with manipulate arrays using libraries like NumPy and Python

**Environment:** Teradata, SQL Assistant, MS office, MS Excel, Agile, Windows, UNIX, SAS EG, SQLPutty.

|  |  |  |
| --- | --- | --- |
| **Client: Visa Inc, Austin, Tx** | | **Sep 2016– April 2017** |
| **Role: Data Scientist** |  | |

**Responsibilities:**

* Extracted data from SQL and prepared data for exploratory analysis using NumPy & Pandas.
* Participated in all phases of data mining, data cleaning, data collection, developing models, validation, and visualization and performed Gap analysis.
* Worked on different data formats such as JSON, SQL, CSV & Excel Datasets and performed machine learning algorithms in Python.
* Programmed a utility in Python that used multiple packages (NumPy, Scipy, pandas).
* Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, Naive Bayes, KNN.
* Conducted K-Means clustering algorithm for segmentation of various customers.  
  Performed time series analysis for forecasting of sales for next fiscal year using ARIMA based modeling.
* Data transformation from various resources, data organization, features extraction, feature engineering and feature preprocessing from data.
* Conducted exploratory data analysis using descriptive statistics & inferential statistics.
* Build statistical plots using seaborn and matplotlib to perform the distribution analysis such as CDF, PDF, PMF as well as regression plot and boxplot etc. to get insights.
* Performed feature preprocessing such as missing value imputation, outlier treatment, scaling the data and creating dummy variables for machine learning algorithms.

## Client: AT&T, Chicago, IL Jan 2016 – Aug 2016

## Role: Data Analyst

**Responsibilities:**

* Worked with business analysts, senior project managers, and programmers to gather business requirements and specifications.
* As part of this team I was involved in Retrieve data from Database and developing reports based on the business requirements.
* Generating the reports for Teradata and Oracle Database to analyze the customer behavior and plan strategies to improve the response rates of marketing Campaigns.
* Created Numerous Volatile, Set, Multiset, Derived, and Global temp tables.
* Extensively used joins while extracting data from multiple tables.
* Developed Teradata SQL scripts using various characters, numeric and date functions.
* Improved performance while using performance tuning techniques like primary Index and collection statistics on index columns.
* Hands on Experience in Writing Python Scripts for Data Extract and Data Transfer.
* Strong Experience in writing Python programs for manipulating the data reading from various Teradata and convert them as one CSV Files.
* Strong experience in using Excel & MS Access to dump the data & analyze based on business needs.
* Worked with Excel Pivot Tables for various business scenarios
* Performed Verification, Validation and Transformations on the Input data (Text files) before loading into target database.
* Designed and developed weekly, monthly reports by using MS Excel Techniques (Charts, Graphs, Pivot tables) and Power point presentations.
* Used MS Excel and Teradata for data pools and adhocs reports for business analysis
* Extensively used Base SAS programs to convert data into Teradata table into CSV and other flat files.
* Responsible for analyzing business requirements and developing Reports using PowerPoint, Excel to provide data analysis solutions to business clients.

## Client: Bank of America -Wilmington, DE July 2013 – Dec 2015

## Role: Data Scientist

**Description:** Bank of America is a multinational banking and financial services corporation. It is ranked 2nd on the list of largest banks in the United States by assets. As of 2016, Bank of America was the 26th largest company in the United States by total revenue.

**Responsibilities:**

* Participated in all phases of research including data collection, data cleaning, data mining, developing models and visualizations.
* Collaborated with data engineers and operation team to collect data from internal system to fit the analytical requirements.
* Redefined many attributes and relationships and cleansed unwanted tables/columns using SQL queries.
* Utilized Spark SQL API in PySpark to extract and load data and perform SQL queries.
* Performed data imputation using Scikit-learn package in Python.
* Performed data processing using Python libraries like Numpy and Pandas.
* Worked with data analysis using ggplot2 library in R to do data visualizations for better understanding of customers' behaviors.
* Visually plotted data using Tableau for dashboards and reports.
* Implemented statistical modeling with XGBoost machine learning software package using R to determine the predicted probabilities of each model.
* Delivered the results with operation team for better decisions.

**Environment:** Python, R, SQL, Tableau, Spark, Machine Learning Software Package, recommendation systems.

## Shore Infotech India Pvt Ltd, Hyderabad, India July 2012 – June 2013

## Role: Data Analyst

**Responsibilities:**

* Interacted with Business Users for gathering, analyzing, and documenting business requirements and data specifications.
* Worked on numerous ad-hoc data pulls for business analysis and monitoring.
* Designed and developed various monthly and quarterly business monitoring excel reports by writing Teradata SQL and using in MS Excel pivot tables.
* Performed in depth analysis on data and prepared ad-hoc reports in MS Excel and SQL scripts.
* Performed verification and validation for accuracy of data in the monthly/quarterly reports.
* Developed Teradata SQL scripts using various characters, numeric and date functions.
* Created multi-set tables and volatile tables using existing tables and collected statistics on table to improve the performance.
* Developed Teradata SQL scripts using OLAP functions like rank and rank () Over to improve the query performance while pulling the data from large tables.
* Utilized ODBC connectivity to connect Teradata from MS Excel to automate the data pull and refresh graphs for weekly and monthly reports.
* Experience in performing Dual Data Validation on various Businesses critical reports working with another Analyst.
* Developed UNIX shell scripts to run batch jobs and communicate logs to the users.
* Automated Teradata SQL scripts in UNIX by using shell scripting.
* Developed SAS programs to import the data from Mainframe text file into Teradata Tables.
* Wrote SAS Programs to convert Excel data into Teradata tables.
* Experience in converting the SQL scripts into SAS SQL scripts.
* Used Python programs to automate the process of combining the large SAS datasets and Data files and then converting as Teradata tables for Data Analysis.
* Developed Python programs for manipulating the data reading from various Teradata and convert them as one CSV Files.
* Developed programs with manipulate arrays using libraries like NumPy and Python.

## Client: IBM, Hyderabad, India May 2010 – June 2012

## Role: Jr. Data Analyst

**Responsibilities:**

* Responsible for gathering requirements from Business Analysts and Operational Analysts and identifying the data sources required for the requests.
* Importing/exporting large amounts of data from files to Teradata and vice versa
* Using Python programs automated the process of combining the large SAS datasets and Data files and then converting as Teradata tables for Data Analysis.
* Used Teradata and spreadsheets as data sources for designing Tableau Reports and Dashboards.
* Distributed Tableau reports using techniques like Packaged Workbooks, PDF etc.
* Designed Physical and Logical Data Models using Visio.
* Designed and developed Ad-hoc weekly, monthly Tableau reports as per business analyst, operation analyst, and project manager data requests.
* Created dashboards and data visualizations using Action filters, Calculated fields, Sets, Groups, Parameters etc., in Tableau
* Responsible for collecting the data and loading it into the Data base.
* Extensively used ETL methodology for supporting data extraction, transformations and loading processing, in a complex EDW using Informatica.
* Developed reports with Custom SQL and views to support business requirements.
* Worked on Set, Multiset, Derived and Volatile Temporary tables.
* Extracted data from existing data source and performed ad-hoc queries.
* Performance tuned and optimized various complex SQL queries.