Data Scientist Data Scientist Data Scientist - TracFone Wireless, Inc Mary, SC   Professional

Qualified Data Scientist with over 7+ 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 data extraction, data cleaning, statistical

modeling and data visualization with large data sets of structured and unstructured data   Having

knowledge on Apache Spark and developing data processing and analysis algorithms using Python

  Understanding of machine learning ("ML") concepts and application of algorithms in non-academic

environments   Strong software development background in functional and object-oriented

programming   Expertized in developing Machine learning algorithm using Python   Ability to

manipulate, transform, and analyze abstract data structures such as Dataframes   Fundamental

understanding of machine learning concepts including training models, as well as understanding

precision and recall in the real world   Strong programming experience in the following: R, Python,

Matlab.   Performed Collection, cleansing, and verification of structured and unstructured data

Experience in visualization tools like, Tableau 9, 10 for creating dashboards   Excellent

understanding Agile and Scrum development methodology   Used the version control tools like Git

Passionate about cleaning 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 Authorized to work in the US for any employer Work Experience Data

Scientist TracFone Wireless, Inc - Miami, FL October 2016 to Present Responsibilities: Enhancing

data collection procedures to include information that is relevant for building analytic systems

Processing, cleansing, and verifying the integrity of data used for analysis   Doing ad-hoc analysis

and presenting results in a clear manner   Constant tracking of model performance   Excellent

understanding of machine learning techniques and algorithms, such as Logistic Regression, SVM,

Random Forests, Deep Learning etc.   Worked with Data governance, Data quality, Data lineage,

Data architect to design various models.   Independently coded new programs and designed Tables

to load and test the program effectively for the given POC's.   Extending company's data with third

party sources of information when needed.   Designed data models and data flow diagrams using

Erwin and MS Visio. As an Architect implemented MDM hub to provide clean, consistent data for a SOA implementation. Developed Implemented & maintained the Conceptual, Logical & Physical Data Models using Erwin for Forward/Reverse Engineered Databases. Experience with common data science toolkits, such as R, Python, Spark, etc. Good applied statistics skills, such as statistical sampling, testing, regression, etc. Build analytic models using a variety of techniques such as logistic regression, risk scorecards and pattern recognition technologies. Analyze and understand large amounts of data to determine suitability for use in models and then work to segment the data, create variables, build models and test those models. Work with technical and development teams to deploy models. Build Model Performance Reports and Modeling Technical Documentation to support each of the models for the product line. Performed Exploratory Data Analysis and Data Visualizations using R, and Tableau. Perform a proper EDA, Univariate and bi-variate analysis to understand the intrinsic effect/combined. Established Data architecture strategy, best practices, standards, and roadmaps. Lead the development and presentation of a data analytics data-hub prototype with the help of the other members of the emerging solutions team. Involved in analysis of Business requirement, Design and Development of High level and Low-level designs, Unit and Integration testing. Worked with several R packages including knitr, dplyr, SparkR, CausalInfer, spacetime. Interacted with the other departments to understand and identify data needs and requirements. Environment: UNIX, Python 3.5.2, MLLib, SAS, regression, logistic regression, Hadoop, NoSQL, Teradata, OLTP, Random forest, OLAP, HDFS, ODS Data Scientist CNY Asset Management - Syracuse, NY January 2014 to September 2016 Responsibilities: Communicated and coordinated with other departments to collection business requirement Worked on miss value imputation, outliers identification with statistical methodologies using Pandas, Numpy Participated in features engineering such as feature creating, feature scaling and One-Hot encoding with Scikit-learn Tackled highly imbalanced Fraud dataset using undersampling with ensemble methods, oversampling and cost sensitive algorithms Improved fraud prediction performance by using random forest and gradient boosting for feature selection with Python Scikit-learn Implemented machine learning model (logistic regression, XGboost) with

Python Scikit- learn   Optimized algorithm with stochastic gradient descent algorithm Fine-tuned the algorithm parameter with manual tuning and automated tuning such as Bayesian Optimization Validated and select models using k-fold cross validation, confusion matrices and worked on optimizing models for high recall rate   Implemented Ensemble Models with majority votes to enhance the efficiency and performance   Environments: Python (scikit-learn, pandas, Numpy), Machine Learning (logistic regression, XGboost), Gradient Descent algorithm, Bayesian optimization, Tableau Data Scientist PDF Solutions - San Jose, CA October 2012 to December 2013 Responsibilities: Build Analytics systems, data structures, gather and manipulate data, using statistical techniques and predictive modelling to tell people story.   Designing suite of interactive dashboards, which provided an opportunity to scale and measure the statistics of the HR dept. which was not possible earlier and schedule and publish reports.   Provided and created data presentation to reduce biases and telling true story of people by Pulling millions of rows of data using SQL, analysis of Data.   Worked on Machine Learning to compare the Metrics of HR data closely   Technology Stack: R, Machine Learning, SQL server and Tableau.   Building Data capabilities for HR with respect to measure and resolve attrition issues through metrics.   Analyzing recruiting history and employee performance to identify best candidates for hiring. Identify employees for training and succession planning.   Environments: R, Python (pandas, numpy, scikit-learn), Machine Learning (predictive modeling), SQL, Tableau Big Data Developer Applied Materials - Santa Clara, CA July 2011 to September 2012 Responsibilities: Involved in the process of load, transform and analyze data from various sources into HDFS (Hadoop Distributed File System) using Hive, Pig and Sqoop.   Experienced in handling data from different datasets join and preprocess those using Pig join operations.   Worked on Pig script to count the number of times a particular URL was opened in a particular duration.   Developed PIG UDFs for the needed functionality such as custom Pigsloader known as timestamp loader.   Created Hive tables based on the business requirements.   Pig scripts and Hive queries were used to analyze the large data sets.   Scheduling and managing jobs on a Hadoop cluster using Oozie work flow.   Written MapReduce code to process and parsing the data from various sources and storing parsed data into

HBase and Hive using HBase - Hive Integration. Worked on Developing custom MapReduce programs and User Defined Functions (UDFs) in Hive to transform the large volumes of data with respect to business requirement. Involved in creating Hive tables and working on them using Hive QL. Involved in moving all log files generated from various sources to HDFS for further processing through Flume. Implemented Frameworks using Java and python to automate the ingestion flow. Loading and transforming of large sets of structured and semi structured data. Exported filtered data into HBase for fast query. Involved in the installation, configuration and used the Hadoop ecosystem components such as Map Reduce, HDFS, Pig, Hive, Flume, HBase. Environments: Hadoop, HDFS, Hive, Pig, Sqoop, HBase, MapReduce, Flume, UDFs. Python Developer American River Bank - Sacramento, CA March 2010 to June 2011 Responsibilities: Exposure on Multi-Threading factory to distribute learning process back-testing and the into various worker processes. Different testing methodologies like unit testing, Integration testing and web application testing. Design and implemented custom scripts. Extensive use of version controller Team Foundation Server (TFS). Test and validated the custom scripts. Delivered automated solutions for science models. Managed, developed, and designed a dashboard control panel for customers and Administrators using Oracle DB, and VMWare API calls. Implemented configuration changes for data models. Maintained and updated existing automated solutions. Handled potential points of failure through error handling and communication of failure. Anticipated potential parts of failure (database, communication points, file system errors). Troubleshoot the process execution and worked with other team members to correct them. Developed GUI using webapp2 for dynamically displaying the test block documentation and other features of python code using a web browser. Interacted with QA to develop test plans from high-level design documentation Environments: Python, Mysql, HTML, Javascript, HQL, Git, Web services. Education Bachelor's Skills PYTHON (7 years), MACHINE LEARNING (5 years), Hadoop (3 years), HADOOP (3 years), DEEP LEARNING (1 year) Additional Information TECHNICAL SKILLS: Machine Learning: classification, regression, clustering, feature engineering, deep learning, neural networks Programming Languages: Python (pandas, scikit-learn, numpy, scipy, GraphLab Create), R, SQL, Hadoop (MapReduce, Sqoop,

Flume), Scala, Java, Spark (PySpark, MLlib) Operating Systems: Linux (CentOS, Ubuntu, Kali Linux), Windows, MacOS Software Tools: PyCharm, Jupyter Notebook, R studio, Tableau, Microsoft Office, Eclipse IDE

Name: Michael Boyd

Email: jyoder@example.org

Phone: 258-591-4897x99032