

Data Scientist Data Scientist Data Scientist - PPD Raleigh, NC    Around 6 years of experience in Data Science , Data Analyst, Python Developer and Project Management.    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, created ER diagrams and schemas.    Proficient in process research which requires analytic models, data inputs and output, analytic metrics and user interface needs. Developed predictive models using Decision Tree, Random Forest, Naive Bayes, Logistic Regression, Cluster Analysis, and Artificial Neural Networks.    Experienced with machine learning algorithms such as Logistic Regression, Random Forest, KNN, Support Vector Machines, Neural Networks, Linear/Non-Linear Regression and K-means    Expertise in employing techniques for Supervised and Unsupervised learning(Clustering, Classification, PCA, Decision trees, KNN, SVM), Predictive Analytics    Excellent understanding of Agile and Scrum development methodology Experienced in Python to manipulate data for data loading and extraction and worked with python libraries like Matplotlib, Numpy and Pandas for data analysis.    Hands on experience of Data Science libraries in Python such as Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Keras. Equipped with experience in utilizing statistical techniques which include Correlation, Hypothesis modeling, Inferential Statistics as well as data mining and modeling techniques using Linear and Logistic Regression, Clustering, Decision Trees, and K-mean clustering.    Mitigated risk factors through careful analysis of financial and statistical data. Transformed and processed raw data for further analysis, visualization, and modeling.    Experience in Big Data technologies like Spark 1.6, Spark SQL, pySpark, Hadoop 2.X, Hive 1.X    Experience in visualization tools like Tableau 9.X, 10.X for creating dashboards    Automated recurring reports using SQL and Python and visualized them on BI platform like Tableau.    Worked on Artificial Neural Networks and Deep Learning models using Theano and Keras packages using Python.    Worked with No SQL Database including Cassandra and Mongo DB.    Experienced in writing SQL queries, working knowledge of RDBMS like SQL Server 2008, No SQL databases like Mongo DB 3.2    Experience working on Microsoft SQL Server, Oracle, Hadoop/Hive.    Analytical, performance-focused, and detail-oriented

professional, offering in-depth knowledge of data analysis and statistics Utilized complex SQL queries for data manipulation. Excellent Team player and self-starter possess good communication skills. Experience in various phases of Software Development life cycle (Analysis, Requirements gathering, Designing) with expertise in writing/documenting Technical Design Document (TDD), Functional Specification Document (FSD), Test Plans, GAP Analysis and Source to Target mapping documents. Used the version control tools like Git 2.X and build tools like Apache Maven/Ant C, C++, Java, Python, SQL programming skills, with experience in working with functions, packages and triggers. Skilled in Advanced Regression Modeling, Correlation, Multivariate Analysis, Model Building and application of Statistical Concepts. Work Experience Data Scientist PPD - Raleigh, NC February 2018 to Present Responsibilities: Responsible for predictive analysis of CRO trial data to predict patient yield and recruitment timeline for clinical trials across APAC & EMEA regions. Data was extracted extensively by using SQL queries and used python packages for the data mining tasks. Performed Exploratory Data Analysis, Data Wrangling and development of algorithms in Python for data mining and analysis. Extensively used Python's multiple data science packages like Pandas, NumPy, Matplotlib, Scipy, Scikit-learn, Keras and Theano. Ensure complete understanding for the broader team on the datasets and how they tie to clinical trial execution, and thereby modeling & analytics. Identify and resolve contradictions in datasets, develop effective validations and heuristics. Ensure data structure is appropriate for future complex analytics initiatives and not limited to a near-term object. Discuss alternative modeling approaches, considerations, limitations, feature definition, and develop strategy to ensure objectives are met in near-term while also supporting longer-term increased sophistication and performance metrics. Investigational sites chosen for clinical trials rarely meet enrollment goals leading to delays and high costs in conducting clinical trials. We need to identify sites that have the potential to complete the clinical trial on time and under budget. Site performance model outputs and resulting top-line Trial performance summary, based on site selection scenario. Created and designed reports that will use gathered metrics to infer and draw logical conclusions from past and future behavior. Step-by-step instructions for use of new features to facilitate verification of the new

content. Used Python-based data manipulation and visualization tools such as Pandas, Matplotlib and Seaborn to clean corrupted data before generating business requested reports. Developed extension models relying on but not limited to the Random forest, logistic, linear regression, Support Vector Regression and boosting techniques like Gradient Boosting and XGBoost for Site Selection.

Used Python programming language to graphically analyze the data and perform data mining. Performed extensive data mining to find out relevant features in an anonymized dataset using Python. Explored supervised Machine Learning algorithms (Decision Tree Regression, Random Forest Regression, SVM) and used parameters such as Root Mean Squared Error and Mean Absolute Error to select the winning model. Environment: Python 3, SQL, Pandas, Seaborn , Decision Tree, Random Forest, Support Vector Machines, Gradient Boosting, Tableau. Data Scientist 7-Eleven - Irving, TX April 2017 to January 2018 Responsibilities: Implemented end-to-end systems for Data Analytics, Data Automation and customized visualization tools using Python, Hadoop and MongoDB. Used Pandas, NumPy, Seaborn, Matplotlib, scikit-learn in Python for developing various machine learning algorithms. Worked on csv, Json, excel different types of files for the Data cleaning and Data analysis Used Python for statistical operations on the data and Matplotlib for the visualizing the data Ensured that the model has a low False Positive Rate. Managed large datasets using Pandas data frames and MySQL. Built various graphs for business decision-making using Python Matplotlib library. Identified root causes of problems, and facilitated the implementation of cost-effective solutions with all levels of management. Performed Data Cleaning, handled missing data, outliers, feature scaling, features engineering. Application of various ML algorithms and statistical modeling like decision trees, regression models, random forest, SVM, clustering to identify Volume using different packages in python. Performed data visualization with Tableau and generated dashboards to present the findings Created and designed reports that will use gathered metrics to infer and draw logical conclusions from past and future behavior. Worked independently and collaboratively throughout the project lifecycle including data extraction/preparation, design and implementation of scalable machine learning analysis and solutions, and documentation of results. Performed Classification using supervised

algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes. Performed data profiling to merge the data from multiple data sources. Knowledge of other relational database platforms such as Oracle, NoSQL Environment: Python 3, MySQL, Matplotlib, Seaborn, Linear Regression, Logistic Regression, Random Forest, Support Vector Machines, KNN, Tableau. Python Developer Zen3 Info Solutions - Hyderabad, Telangana August 2014 to December 2016 Involved in building database model, APIs and views utilizing Python, in order to build an interactive web-based solution.

Used data types like dictionaries, tuples and object -concepts based inheritance features for making complex algorithms of networks. Designed and managed API system deployment using fast http server and Amazon AWS architecture. Worked on Python Open stack API's. Carried out various mathematical operations for calculation purpose using python libraries. Managed large datasets using Panda data frames and MySQL. Worked with JSON based REST Web services. Performed testing using Django's Test Module. Involved in Agile Methodologies and SCRUM Process. Creating unit test/regression test framework for working/new code. Using Subversion version control tool to coordinate team-development. Developed SQL Queries, Stored Procedures, and Triggers Using Oracle, SQL, PL/SQL. Responsible for debugging and troubleshooting the web application. Supported user groups by handling target-related software issues/service requests, identifying/fixing bugs. Configured the Django admin site, dashboard and created a custom Django dashboard for end users with custom look and feel. Used Django APIs for database access. Used Python for XML, JSON processing, data exchange and business logic implementation. Used Python scripts to update the content in database and manipulate files. Worked through the entire lifecycle of the projects including Design, Development, and Deployment, Testing and Implementation and support. Environment: Python, Django, JSP, Oracle, Java, MySQL, Linux, HTML, CSS. Data Analyst HTC Global Services India Private Ltd September 2013 to July 2014 Responsibilities: Communicated with business users, data architects and developers to identify needs, define project scope and detailed functional and non-functional requirements. Gathered and documented requirements throughout the Software Development Life Cycle (SDLC) Prepared Data Mapping documents, Use Cases, Requirements Traceability Matrix and worked with

the User experience team (Wireframe preparing team) Created requirement documentation, reviews with user representatives, recommended priorities, gave presentations and walkthroughs and obtained user sign-off. Driven daily stand up meetings offshore, iteration planning meetings and requirement gathering meetings. Extracting data from different public data repositories and from different databases and creating datasets for statistical analysis. Identified Multiple Dimensions and Fact tables. Used advance data modeling concepts of degenerated dimension, sub-dimension, Factless fact table, Aggregate fact tables in the Multidimensional model. Extensively used Tab Admin and Tab Cmd commands in creating backups and restoring backups of Tableau repository. Administered user, user groups and scheduled instances for reports in Tableau. Hands-on development in creating and modifying worksheets and data visualization dashboards. Designed & developed various analytical reports from multiple data sources by blending data on a single worksheet in Tableau Desktop. Environment: Windows 7 & 8, MS SQL Data Base, MS Excel, MS Word, MS PowerPoint, MS Outlook, SharePoint, MS Visio, MS Project, Oracle 11g, Tableau Server and Tableau Desktop v8/v9, MS Excel, SQL, ERWIN, SSIS, SSRS.

Java Developer Hyderabad, Telangana May 2013 to August 2013 Responsibilities: Developed web application using Struts, JSP, Servlets, Java beans that uses MVC design pattern Created user-friendly GUI interface and Web pages using HTML, CSS and JSP. Used Eclipse as IDE tool for creating Servlets, JSP, and XML. Implemented the Struts framework based on MVC design pattern. Wrote SQL for JDBC prepared statements to retrieve the data from database. Worked with Database Administrators to solve the problems generated while creating tables for application Developed and implemented customized maintenance plans to meet the needs and keep up to date with the glitches and fixes for performances issues. Developed various UML diagrams like use cases, class diagrams, interaction diagrams (sequence and collaboration) and activity diagrams Involved in build and deploying the application using ANT Environment: Java, SQL, XML, HTML, CSS, JSP, JDBC, Eclipse IDE. Education Bachelors in Computer Science in TOOLS AND TECHNOLOGIES JNTU

Name: Michael Ford

Email: kimberly27@example.net

Phone: +1-229-411-0064x33136