Data Scientist/Data Science Manager Data Scientist/Data Science Manager Data Scientist/Data Science Manager - American Express/Home Depot Atlanta, GA   Professional qualified Data Scientist/Software Engineer with over 9 years of experience in Data Science and Analytics including Artificial Intelligence/Deep Learning/Machine Learning, Big Data, Data Mining and Statistical Analysis and demonstrated Leadership.   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 schema.   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) and Scala   Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0 Jupiter Notebook 4.X, R 3.0 (ggplot2, Caret, dplyr) and Excel [ ]   Solid ability to write and optimize diverse SQL queries, working knowledge of RDBMS like SQL Server 2008, NoSql databases like AWS DynamoDB, MongoDB   Developed API libraries and coded business logic using C#, XML and designed web pages using .NET framework, C#, Python, Django, HTML, AJAX   Strong experience in Big Data technologies like Spark, Sparksql, pySpark, Hadoop, HDFS, Hive   Experience in visualization tools like, Tableau 9.X, 10.X for creating dashboards   Excellent understanding Agile and Scrum development methodology   Used the version control tools like Git 2.X and build tools like Apache Maven/Ant   Passionate about gleaning insightful information from massive data assets and developing a culture of sound, data-driven decision making   Worked on containerization using Docker and managed the containers in Kubernetes   Ability to maintain a fun, casual, professional and productive team atmosphere   Experienced the full software life cycle in SDLC, Agile, Devops and Scrum methodologies including creating requirements, test plans.   Worked on message queues such as Amazon SNS, SQS, Google Pub Sub and integrated with other services like AWS Lambda, Storage using Bash scripts   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, Hypothetical testing, normal distribution and other advanced statistical and econometric techniques.

Developed predictive models using Decision Tree, Random Forest, Na ve Bayes, Logistic Regression, Social Network Analysis, Cluster Analysis, and Neural Networks. Experienced in Machine Learning and Statistical Analysis with Python Scikit-Learn. Experienced in Python to manipulate data for data loading and extraction and worked with python libraries like Matplotlib, Numpy, Scipy and Pandas for data analysis. Worked with complex applications such as R, SAS, Matlab and SPSS to develop neural network, cluster analysis. Strong SQL and T SQL programming skills, with experience in working with functions, packages and triggers. 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 data parsing, data ingestion, data manipulation,data architecture, data modelling 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. Experienced in Visual Basic for Applications and VB programming languages C#, .NET framework to work with developing applications. Experienced in Big Data with Hadoop, HDFS, MapReduce, and Spark. 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. Authorized to work in the US for any employer Work Experience Data Scientist/Data Science Manager American Express/Home Depot July 2017 to Present Description: American Express delivers predictive analytics as a service, and offers hosted, cloud-based systems for specific business problems, e.g., predicting the behavior of individual consumers, stopping revenue leakage in hospitals, warning of

threats to corporate security or brand health, etc. Home Depot is an American home improvement supplies retailing company that sells tools, construction products, and services. Home Depot is the largest home improvement retailer in the United States, ahead of rival Lowe's   Responsibilities: Demonstrated leadership skills by successfully managing, mentoring and delivering 3 data science projects at Srishti Biz.   Evaluated models using Cross Validation, Log loss function, ROC curves and used AUC for feature selection and elastic technologies like ElasticSearch, Kibana etc Addressed overfitting by implementing of the algorithm regularization methods like L2 and L1. Implemented statistical modeling with XGBoost machine learning software package using Python to determine the predicted probabilities of each model.   Created master data for modelling by combining various tables and derived fields from client data and students LORs, essays and various performance metrics.   Hands on Experience on Spark MLLib using both Python and Scala. Worked on containerization of different modules of the project like Python APIs, Web framework written in Javascript using Docker and managed it using Kubernetes   Formulated a basis for variable selection and GridSearch, KFold for optimal hyperparameters   Utilized Boosting algorithms to build a model for predictive analysis of student's behaviour who took USMLE exam apply for residency.   Used numpy, scipy, pandas, nltk(Natural Language Processing Toolkit),matplotlib to build the model.   Formulated several graphs to show the performance of the students by demographics and their mean score in different USMLE exams.   Application of various Artificial Intelligence(AI)/machine learning algorithms and statistical modeling like decision trees,text analytics, natural language processing(NLP), supervised and unsupervised, regression models. Used Principal Component Analysis in feature engineering to analyze high dimensional data. Performed Data Cleaning, features scaling, features engineering using pandas and numpy packages in python and build models using deep learning frameworks.   Scheduled jobs to run SQL statements to transfer data from Google Bigquery to Mysql using Airflow.   Created deep learning models using Tensorflow and keras by combining all tests as a single normalized score and predict residency attainment of students.   Used XGB classifier if the feature is a categorical variable and XGB regressor for continuous variables and combined it using FeatureUnion and

FunctionTransfomer methods of Natural Language Processing.   Used OnevsRest Classifier to fit each classifier against all other classifiers and used it on multiclass classification problems. Implemented application of various machine learning algorithms and statistical modeling like Decision Tree, Text Analytics, Sentiment Analysis, Naive Bayes, Logistic Regression and Linear Regression using Python to determine the accuracy rate of each model.   Created and designed reports that will use gathered metrics to infer and draw logical conclusions of past and future behavior.   Generated various models by using different machine learning and deep learning frameworks and tuned the best performance model using Signal Hub.   Created data layers as signals to Signal Hub to predict new unseen data with performance not less than the static model build using deep learning framework.   Environment: Python 2.x,3.x, Hive, AWS, Linux, Tableau Desktop, Microsoft Excel, NLP, Deep learning frameworks such as TensorFLow, Keras, Boosting algorithms etc Data Scientist January 2016 to November 2016 Description: The goal of this project is short text summarization on the comment stream of a message from social network services. The users of the social sites always desire to get a brief understanding of a comment stream without reading the whole comment list. Responsibilities: Performed Data Profiling to learn about behavior with various features such as traffic pattern, location, time, Date and Time etc.   Application of various Artificial Intelligence(AI)/machine learning algorithms and statistical modeling like decision trees,text analytics, natural language processing(NLP), supervised and unsupervised, regression models, social network analysis, neural networks, deep learning, SVM, clustering to identify Volume using scikit-learn package in python, Matlab   Categorised comments into positive and negative clusters from different social networking sites using Sentiment Analysis and Text Analytics Analyze traffic patterns by calculating autocorrelation with different time lags.   Sets of comment streams are processed by NLP module in stages like stemming, punctuation removal etc. to categorize into vector model and input into Incremental Short Text Summarization (ISTS) algorithm. Clusters of comments with same pattern will be grouped together and presented visually to the end user.   Ensured that the model has low False Positive Rate and Text classification and sentiment analysis for unstructured and semi-structured data.   Addressed overfitting by implementing of the

algorithm regularization methods like L2 and L1.   Used Principal Component Analysis in feature engineering to analyze high dimensional data.   Performed Data Cleaning, features scaling, features engineering using pandas and numpy packages in python and build models using SAP Predictive Analytics   Created Data Quality Scripts using SQL and Hive to validate successful data load and quality of the data. Created various types of data visualizations using Python and Tableau

Data Scientist First Data/Accenture July 2011 to July 2015 Description: First Data Corporation is a global payment processing company headquartered in Atlanta, Georgia, United States. The company's portfolio includes merchant transaction processing services; credit, debit, private-label, gift, payroll and other prepaid card offerings; fraud protection and authentication solutions. Responsibilities: Provided Configuration Management and Build support for more than 5 different applications, built and deployed to the production and lower environments.   Explored and Extracted data from source XML in HDFS, used ETL for preparing data for exploratory analysis using data munging.   Responsible for different Data mapping activities from Source systems to Teradata, Text mining and building models using topic analysis, sentiment analysis for both semi-structured and unstructured data.   Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS   Write ad-hoc queries using T SQL for data analysis on Microsoft SQL Server   Used R and python for Exploratory Data Analysis, A/B testing, HQL, VQL, Data Lake, AWS Redshift, oozie, pySpark, Anova test and Hypothesis test to compare and identify the effectiveness of Creative Campaigns.   Computing A/B testing frameworks, clickstream   Created clusters to Control and test groups and conducted group campaigns using Text Analytics.   Created positive and negative clusters from merchant's transaction using Sentiment Analysis to test the authenticity of transactions and resolve any chargebacks.   Analyzed and calculated the lifetime cost of everyone in the welfare system using 20 years of historical data. Created and developed classes and web page elements using C# and AJAX. JSP was used for validating client side responses and connected C# to database to retrieve SQL data   Developed LINUXShell scripts by using NZSQL/NZLOAD utilities to load data from flat files to Netezza database.   Developed triggers, stored procedures, functions and packages using cursors and ref

cursor concepts associated with the project using Pl/SQL   Created various types of data visualizations using R, C#, python and Tableau/Spotfire also connected Pipeline Pilot with Spotfire to create more interactive business driven layouts.   Used Python, R, SQL to create Statistical algorithms involving Multivariate Regression, Linear Regression, Logistic Regression, PCA, Random forest models, Decision trees, Support Vector Machine for estimating the risks of welfare dependency.   Identified and targeted welfare high-risk groups with Machine learning/deep learning algorithms.   Conducted campaigns and run real-time trials to determine what works fast and track the impact of different initiatives.   Developed Tableau visualizations and dashboards using Tableau Desktop.   Used Graphical Entity-Relationship Diagramming to create new database design via easy to use, graphical interface.   Created multiple custom SQL queries in Teradata SQL Workbench to prepare the right data sets for Tableau dashboards   Perform analyses such as regression analysis, logistic regression, discriminant analysis, cluster analysis using SAS programming.   Environment: R 3.x, HDFS, C#, Hadoop 2.3, Pig, Hive, Linux, R-Studio, Tableau 10, SQL Server, MS Excel, PySpark. Python Developer Cenvien Technologies - Hyderabad, Telangana January 2011 to June 2011 Description: Cenvien technologies gather the requirements by listening and understanding to the client's business requirement to deliver quality products. It is highly qualified and strongly dedicated developing team that produces unique solutions.   Responsibilities: Developed entire frontend and backend modules using Python on Django Web Framework. Implemented the presentation layer with HTML, CSS and JavaScript.   Involved in writing stored procedures using Oracle.   Optimized the database queries to improve the performance. Designed and developed data management system using Oracle.   Environment: MySQL, ORACLE, HTML5, CSS3, JavaScript, Shell, Linux & Windows, Django, Python Data Analyst Pennar Industries Limited - Hyderabad, Telangana March 2009 to December 2010 Description: As a backend developer of web applications and data science infrastructure. The main area of focus is to come up with comprehensive solutions that need massive capacity and throughput. Responsibilities: Effectively communicated with the stakeholders to gather requirements for different projects   Used MySQL db package and Python-MySQL connector for writing and

executing several MYSQL database queries from Python.   Implemented Client/Server applications using C#, JSP and SQL   Created functions, triggers, views and stored procedures using My SQL. Worked closely with back-end developer to find ways to push the limits of existing Web technology. Involved in the code review meetings. Environment: Python, MySQL, C#. Education Master of Computer Science in Computer Science Lamar University - Beaumont, TX 2015 to 2017 Skills Business Intelligence, SQL, access, testing, Excel, Data Science (9 years), Python (9 years) Links https://www.linkedin.com/in/pavithra-kumar-22b753173 Certifications/Licenses Acadgild - Masters in Data Science February 2019 to Present Assessments Data Analysis   Proficient August 2019 Measures a candidate's skill in interpreting and producing graphs, identifying trends, and drawing justifiable   conclusions   from   data.   Full   results: https://share.indeedassessments.com/share_assignment/5p0ejlgzwei-m6ro   Problem   Solving Expert August 2019 Measures a candidate's ability to analyze relevant information when solving problems. Full results: https://share.indeedassessments.com/share_assignment/dso7visv0vjm2cfc Indeed Assessments provides skills tests that are not indicative of a license or certification, or continued development in any professional field.

Name: Derrick Martin

Email: imercer@example.com

Phone: (971)239-9958x584