Data Scientist Data Scientist Data Scientist - Home Depot Atlanta, GA   Professional qualified Data Scientist/Data Analyst with over 9 years of experience in Data Science and Analytics including Artificial Intelligence/Deep Learning/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, 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)   Experience in implementing 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 databases like MongoDB3.2   Developed API libraries and coded business logic using C#, XML and designed web pages using .NET framework, C#, Python, Django, HTML, AJAX   Strong experience for over 5 years in Image Recognition and BigData technologies like Spark 1.6, Sparksql, pySpark, Hadoop 2.X, HDFS, Hive 1.X   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 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   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.   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, Na

veBayes, LogisticRegression, Social Network Analysis, 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, Python, Theano, H20, SAS, Matlab and SPSS to develop neural network, cluster analysis. Strong C#, 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 dataparsing, 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. Worked with NoSQL Database including Hbase, Cassandra and MongoDB. Experienced in Big Data with Hadoop, HDFS, MapReduce, and Spark. 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,Adobe Analytics 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. Work Experience Data Scientist Home Depot - Atlanta, GA March 2017 to Present Description: The Home Depot Inc. or Home Depot is an American home improvement supplies retailing company that sells tools, construction products, and services. The company is headquartered at the Atlanta Store Support Center in unincorporated Cobb County, Georgia (with an Atlanta mailing address). The MRO company Interline Brands is also owned by The Home Depot with 70 distribution centers across the

United States.The Home Depot is the largest home improvement retailer in the United States. Responsibilities: Articulate and understand a business problem, identify challenges, formulate the machine learning problem or NLP problems and provide/prototype solutions Collect and manipulate large volumes of data; build new and improved techniques and/or solutions for data collection, management, and usage Performed Data Profiling to learn about behavior with various features of Merchant and competitor details using Python Matplotlib. Using SQL to extract data from EDW in BigQuery and customize according to project needs and built data scheduling jobs in Airflowto transfer data into MySQL Used numpy, scipy, pandas, nltk(Natural Language Processing Toolkit),matplotlib to build the model. Extracted data from HDFS using Hive, Presto and performed data analysis using Spark with Scala, pySpark, Redshift and feature selection and created nonparametric models in Spark Build several regression techniques to predict price recommendations for Home depot products based on Competitor data and Home Depot data Performed Data Cleaning, features scaling, features engineering using pandas and numpy packages in python and build models using deep learning frameworks. Used XGB classifier if the feature is an categorical variable and XGB regressor for continuous variables and combined it using Feature Union and Function Transformer methods of Natural Language Processing. Used One vs Rest 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. Environment: Python 2.x,3.x, Hive, Linux, Tableau Desktop, Microsoft Excel, NLP, Airflow, GCP, BigQuery, Kubernetes, Docker, Spark, PySpark, Boosting algorithms etc. Data Scientist Opera Solutions, New Jersey January 2016 to February 2017 Description: Opera Solutions, LLC is a technology and analytics company mainly focused on big data. The firm uses a combination of machine learning science, advanced predictive analytics, technology, large-scale data management, and human expertise. Opera Solutions 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. Responsibilities: Performed Data Profiling to learn about behavior with various features of USMLE examinations of various student patterns using Tableau, Adobe Analyticsand Python Matplotlib. 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. 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. Extracted data from HDFS using Hive, Presto and performed data analysis using Spark with Scala, pySpark, Redshift and feature selection and created nonparametric models in Spark Application of various Artificial Intelligence(AI)/machine learning algorithms and statistical modeling like decision trees,text analytics, Image and Text Recognitionusing OCR tools like Abbyy, natural language processing(NLP),supervised and unsupervised, regressionmodels. 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. Created deep learning models using Tensor flow and Keras by combining all tests as a single normalized score and predict residency attainment of students. Used XGB classifier if the feature is an categorical variable and XGB regressor for continuous variables and combined it using Feature Union and Function Transformer methods of Natural Language Processing. Used One vs Rest 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 with cloud based products like Azure ML Studio and Dataiku.   Generated various models by using different machine learning and deep learning frameworks and tuned the best performance model using Signal Hub and AWS Sage maker/Azure Data bricks.   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 Tensor FLow, Keras, Boosting algorithms etc Data Scientist FirstData - Atlanta, GA January 2015 to December 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.   Implemented public segmentation using unsupervised machine learning algorithms by implementing k-means algorithm using PySpark.

Using Air Flow to keep track of job statuses in repositories like MySQL and Postgre databases.

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   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 and time spent databases using Airflow   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 Analysisto 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 LINUX Shell 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, Tensorflow to create Statistical algorithms involving Multivariate Regression, LinearRegression, LogisticRegression, PCA, Image Recognition,Random forest models, Decision trees, Support Vector Machine for estimating the risks of welfare dependency. Identified and targeted welfare high-risk groups with Machinelearning/deep learningalgorithms. 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 SQLqueries in TeradataSQLWorkbench 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. Data Scientist Bank of America - Wilmington, DE August 2012 to December 2014 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 and also used C# connector to perform SQL queries by creating and connecting to SQL engine. 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. 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. Python Developer Cenvien Technologies - Hyderabad, Telangana January 2011 to July 2012 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 Programmer 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++, 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 Bachelor of Computer Science in TOOLS AND TECHNOLOGIES Informatica Power Centre Skills Android. (Less than 1 year), Application development (Less than 1 year), Bi (Less than 1 year), Data warehouse (Less than 1 year), Linux (6 years)

Name: Ronald Weeks

Email: lisatorres@example.org

Phone: 8558167416