

Data Scientist/Machine Learning Data Scientist/Machine Learning Data Scientist/Machine Learning
Norcross, GA 8+years of experience in DataScience and Analytics including MachineLearning,
DataMining, DataBlending&StatisticalAnalysis. Over 5+Experience with MachineLearning
techniques and algorithms (such as k-NN, NaiveBayes, etc). Experience in AWS
(AmazonWebServices)EC2, VPC, IAM, IAM, S3, CloudFront, CloudWatch, CloudFormation, Glacier,
RDSConfig, Route53, SNS, SQS, ElasticCache. AzureCloud Extensive full cycle CloudAzure
experience with full BigData, Elasticsearch and SOLR, MachineLearning and DeepLearning
development and deployment. Experienced with machinelearningalgorithms such as
logisticregression, randomforest, XGboost, KNN, SVM, neuralnetwork, linearregression,
lassoregression,andk-means. Expertise in synthesizing Machinelearning, PredictiveAnalytics and
BigData technologies like Hadoop, Hive, Pig. Strong skills in statistical methodologies such as
A/Btest, experimentdesign, hypothesisistest, ANOVA. Experience in implementing data analysis
with various analytic tools, such as Anaconda 4.0,R 3.0 (ggplot2, Caret, dplyr) and Excel. Solid
ability to write and optimize diverse SQLqueries, working knowledge of RDBMS like SQLServer2008

Experience in BigData technologies like Spark1.6, Sparksql, pySpark, Hadoop2.X, HDFS,
Hive1.X. Experience in visualization tools likeTableau9.X, 10.X, DataBlendingfor creating
dashboards. Proficient in PredictiveModeling, DataMiningMethods, FactorAnalysis, ANOVA,
Hypotheticaltesting, normal distribution and other advanced statistical and econometric techniques.

Developed predictive models using DecisionTree, RandomForest, NaiveBayes, LogisticRegression,
ClusterAnalysis, and NeuralNetworks. Excellent knowledge of Machine Learning, Mathematical
Modeling and Operations Research. Comfortable with R, Python, SAS , MATLAB,
Relationaldatabases. Deep understanding & exposure of BigDataEco-system. Expert in creating
PL/SQLSchemaobjects like Packages, Procedures, Functions, Subprograms, Triggers, Views,
MaterializedViews, Indexes, Constraints, Sequences, ExceptionHandling, DynamicSQL/Cursors,
NativeCompilation, CollectionTypes, RecordType, ObjectType using SQLDeveloper. Hands on
Experience in implementing ModelViewControl (MVC) architecture using Spring, JDK, CoreJava
(Collections, OOPSConcepts), JSP, Servlets, Struts, springs, Hibernate, JDBCand provided

ServerAdministrator duties LogicalPosition. Worked with complex applications such as R, SAS, Matlab, and SPSS to develop a neural network, cluster analysis. 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 R and visualized them on BI platform like Tableau. Worked in a development environment like Git and VM. Excellent communication skills. Successfully working in fast-paced multitasking environment both independently and in a collaborative team, a self-motivated enthusiastic learner. Authorized to work in the US for any employer

Work Experience

Data Scientist/Machine Learning FleetCor Technologies Inc - Norcross, GA May 2017 to May 2017

Responsibilities: Analyzed Trading mechanism for real-time transactions and build collateral management tools. Compiled data from various sources to perform complex analysis for actionable results. Utilized machine learning algorithms such as linear regression, multivariate regression, naive bayes, Random Forests, K-means, & KNN for data analysis. Measured Efficiency of Hadoop/Hive environment ensuring SLA is met. Developed Spark code using Scala and Spark-SQL/Streaming for faster processing of data. Prepared Spark build from the source code and ran the PIG Scripts using Spark rather using MR jobs for better performance. Analyzing the system for new enhancements/functionalities and perform Impact analysis of the application for implementing ETL changes. Imported data using Sqoop to load data from MySQL to HDFS on regular basis. Developed Scripts and Batch Job to schedule various Hadoop Program. Used TensorFlow to train the model from insightful data and look at thousands of examples. Designing, developing and optimizing SQL code (DDL / DML). Building performant, scalable ETL processes to load, cleanse and validate data. Expertise in Data archival and Data migration, ad-hoc reporting and code utilizing SAS on UNIX and Windows Environments. Tested and debugged SAS programs against the test data. Processed the data in SAS for the given requirement using SAS programming concepts. Imported and Exported data files to and from SAS using Proc Import and Proc Export from Excel and various delimited

text-based data files such as .TXT (tab delimited) and .CSV (comma delimited) files into SAS datasets for analysis. Expertise in producing RTF, PDF, HTML files using SAS ODS facility. Providing support for data processes. This will involve monitoring data, profiling database usage, trouble shooting, tuning and ensuring data integrity. Participating in the full software development lifecycle with requirements, solution design, development, QA implementation, and product support using Scrum and other Agile methodologies. Collaborate with team members and stakeholders in design and development of data environment. Learning new tools and skillsets as needs arise. Preparing associated documentation for specifications, requirements and testing. Optimizing the Tensorflow Model for an efficiency. Used Tensorflow for text summarization. Used Spark API over Cloudera Hadoop YARN to perform analytics on data in Hive. Wrote Hive queries for data analysis to meet the business requirements. Developed Kafka producer and consumers for message handling. Responsible for analyzing multi-platform applications using python. Used storm for an automatic mechanism to analyze large amounts of non-unique data points with low latency and high throughput. Developed MapReduce jobs in Python for data cleaning and data processing. Environment: Machine learning, AWS, MS Azure, Cassandra, SAS, Spark, HDFS, Hive, Pig, Linux, Anaconda Python , MySQL, Eclipse, PL/SQL, SQL connector, SparkML. Data Scientist/Machine Learning Ford Motor - Dearborn, MI December 2015 to April 2017 Responsibilities: Provided the architectural leadership in shaping strategic, business technology projects, with an emphasis on application architecture. Utilized domain knowledge and application portfolio knowledge to play a key role in defining the future state of large, business technology programs. Participated in all phases of datamining, datacollection, datacleaning, developingmodels, validation, and visualization and performed Gapanalysis. Developed MapReduce/Spark, R modules for machine learning & predictive analytics in Hadoop on AWS. Implemented a R-based distributed randomforest. Utilized Spark, Scala, Hadoop, SparkStreaming, MLLib, Rabroadvariety of machinelearning methods includingclassifications, regressions, dimensionallyreduction etc. and utilized the engine to increase user lifetime by 45% and triple user conversations for target categories. Well versed with: CloudlaaS and PaaS implementations in

both private and public clouds like VMware, Openstack, AmazonAWS and Cloudfoundry (Pivotal and HPStackato). Used R and h2o.ai libraries for developing various machine learning algorithms and utilized machine learning algorithms such as linear regression, multivariate regression, naive Bayes, Random Forests, K-means, & KNN for data analysis. Worked on database design, relational integrity constraints, OLAP, OLTP, Cubes, and Normalization (3NF) and De-Normalization of the database. Worked on customer segmentation using an unsupervised learning technique clustering. Utilized Spark, Scala, Hadoop, HBase, Kafka, Spark Streaming, MLlib, Python, a broad variety of machine learning methods including classifications, regressions, dimensionality reduction etc. Data analysis, reporting using Tableau. Perform numerous data pulling requests using SQL Server 2012. Designed and implemented system architecture for Amazon EC2 based cloud-hosted solution for the client. Tested Complex ETL Mappings and Sessions based on business user requirements and business rules to load data from source flat files and RDBMS tables to target tables. Hands on experience in Hadoop ecosystem with components Hadoop MapReduce, HDFS, Oozie, HiveQL, Sqoop, HBase, MongoDB, Zookeeper, Pig, and Flume. Environment: R, SQL, Oracle 12c, Netezza, SQL Server, Informatica, Java, SSRS, PL/SQL, T-SQL, Tableau, MLlib, regression, Cluster analysis, Scala NLP, Spark, Kafka, MongoDB, logistic regression, Hadoop, Hive, Teradata, random forest, OLAP, Azure, MariaDB, SAP CRM, HDFS, ODS, NLTK, SVM, JSON, Tableau, XML, Cassandra, MapReduce, AWS. Data Scientist/R Developer L'Oréal - Shanghai, CN June 2014 to November 2015 Responsibilities: The conducted analysis in assessing customer consuming behaviors and discover the value of customers with RF analysis, applied customer segmentation with clustering algorithms such as K-Means Clustering and Hierarchical Clustering. Collaborated with data engineers to implement the ETL process, wrote and optimized SQL queries to perform data extraction and merging from Oracle. Used R and Spark to develop a variety of models and algorithms for analytic purposes. Performed data integrity checks, data cleaning, exploratory analysis and feature engineering using R. Developed personalized product recommendation with Machine Learning algorithms, including Gradient Boosting Tree and Collaborative filtering to better meet the needs of existing customers and acquire new customers. Used R and Spark to

implement different machine learning algorithms, including Generalized Linear Model, Random Forest, SVM, Boosting and Neural Network. Evaluated parameters with K-Fold Cross Validation and optimized performance of models. A highly immersive Data Science program involving Data Manipulation and Visualization, Web Scraping, Machine Learning, GIT, SQL, UNIX Commands, R programming, NoSQL. Identified risk level and eligibility of new insurance applicants with Machine Learning algorithms. Utilized SQL and HiveQL to query, manipulate data from variety data sources including Oracle and HDFS, while maintaining data integrity. Performed data visualization and Designed dashboards with Tableau and D3.js and provided complex reports, including charts, summaries, and graphs to interpret the findings to the team and stakeholders. Environment: R, MATLAB, MongoDB, exploratory analysis, feature engineering, K-Means Clustering, Hierarchical Clustering, Machine Learning), Python, Spark (MLlib, PySpark), Tableau, Micro Strategy, SAS, Tensor Flow, regression, logistic regression, Hadoop 2.7, OLTP, random forest, OLAP, HDFS, ODS, NLTK, SVM, JSON, XML and MapReduce.

Data Analyst C Client - Chicago, IL June 2012 to August 2012

Responsibilities: Participated in JAD sessions, gathered information from Business Analysts, end users and other stakeholders to determine the requirements. Hands on Experience in Cloud Computing such as AWS storage, Compute, Databases SQL. Designed the Data Warehouse and MDM hub Conceptual, Logical and Physical data models. Performed Daily Monitoring of Oracle instances using Oracle Enterprise Manager, ADDM, TOAD, monitor users, tablespaces, memory structures, rollback segments, logs, and alerts. Used Normalization methods up to 3NF and De-normalization techniques for effective performance in OLTP and OLAP systems. Generated DDL scripts using Forward Engineering technique to create objects and deploy them into the databases. Worked on database testing, writing complex SQL queries to verify the transactions and business logic like identifying the duplicate rows by using SQL Developer and PL/SQL Developer. Used Teradata SQL Assistant, Teradata Administrator, PMON and data load/export utilities like BTEQ, FastLoad, MultiLoad, FastExport, TPump on UNIX/Windows environments and running the batch process for Teradata. Developed SQL queries to fetch complex data from different tables in remote databases using joins, database links and Bulk collects. The migrated

database from legacy systems, SQL server to Oracle and Netezza. Used SSIS to create ETL packages to validate, extract, transform and load data to pull data from Source servers to staging database. Worked on SQLServerconceptsSSIS (SQLServerIntegrationServices), SSAS (SQLServerAnalysisServices) and SSRS (SQLServerReportingServices). Python Developer Aspect Software July 2010 to May 2011 Environment: ER Studio, Teradata13.1, SQL, PL/SQL, BTEQ, DB2, Oracle, MDM, Netezza, ETL, RTF UNIX, SQL Server2010, Informatica, SSRS, SSIS, SSAS, SAS, Aginity. Client: Aspect Software - INDIA July 2010 - May 2011 - July 2012 Role: Python Developer Responsibilities: Worked on the project from gathering requirements to developing the entire application. Worked on Anaconda Python Environment. Created, activated and programmed in Anaconda environment. Wrote programs for performance calculations using NumPy and SQLAlchemy. Wrote python routines to log into the websites and fetch data for selected options. Used python modules of urllib, urllib2, Requests for web crawling. Experience using all these ML techniques: clustering, regression, classification, graphical models. Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data mining solutions to various business problems and generating data visualizations using R, Python and Tableau. Used with other packages such as Beautiful Soup for data parsing. Involved in development of Web Services using SOAP for sending and getting data from the external interface in the XML format. Used with other packages such as Beautiful Soup for data parsing. Worked on development of SQL and stored procedures on MYSQL. Analyzed the code completely and have reduced the code redundancy to the optimal level. Design and build a text classification application using different text classification models. Used Jira for defect tracking and project management. Worked on writing and as well as read data from CSV and excel file formats. Involved in Sprint planning sessions and participated in the daily Agile SCRUM meetings. Conducted every day scrum as part of the SCRUM Master role. Developed the project in Linux environment. Worked on resulting reports of the application. Performed QA testing on the application. Held meetings with client and worked for the entire project with limited help from the client. Environment: Python, Anaconda, Synder (IDE), Windows 7, Teradata, Requests, urllib,

urllib2, BeautifulSoup, Tableau, python libraries such as NumPy, SQL Alchemy, MySQLdb, Education Bachelor's Skills APACHE HADOOP HDFS (3 years), Oracle (3 years), python. (3 years), SQL (4 years), XML (3 years) Additional Information Skills: Big Data/Hadoop Technologies Hadoop, HDFS, YARN, MapReduce, Hive, Pig, Impala, Sqoop, Flume, Spark, Kafka, Zookeeper, and Oozie Languages C, C++, HTML5, DHTML, WSDL, css3 XML, R/R Studio, SAS Enterprise Guide, SAS R, R (Caret, Weka, ggplot), Python (NumPy, SciPy, Pandas), SQL, PL/SQL, Pig Latin, HiveQL, Shell Scripting. Cloud Computing Tools Amazon AWS, Azure. Databases Microsoft SQL Server 2008 [] MySQL 4.x/5.x, Oracle 10g, 11g, 12c, DB2, Teradata, Netezza NO SQL Databases HBase, Cassandra, MongoDB, MariaDB Build Tools Maven, ANT, Toad, SQL Loader, RTC, RSA, Control-M, Oozie, Hue, SOAP UI Development Tools Microsoft SQL Studio, Eclipse, NetBeans, IntelliJ Development Methodologies Agile/Scrum, Waterfall, UML, Design Patterns Version Control Tools and Testing API Git, SVN, GitHub, SVN and JUNIT ETL Tools Informatica Power Centre, SSIS Reporting Tools MS Office (Word/Excel/PowerPoint/ Visio/Outlook), Crystal Reports XI, SSRS, Cognos7.0/6.0. Operating Systems All versions of UNIX, Windows, LINUX, Macintosh HD, Sun Solaris

Name: Walter Brown

Email: tim08@example.org

Phone: (792)397-1324