Data Scientist Data Scientist Data Scientist - Palo Alto Networks Dallas TX Dallas, TX Above 8+ years of experience in large Unstructured data, Datasets of Structured, Data Visualization, Data Acquisition, Predictive modeling, Data Validation. Develop, maintain and teach new tools and methodologies related to data science and high - performance computing. Data Scientist with proven expertise in Data Analysis, Machine Learning, and Modeling. Experience in Machine Learning algorithms such as Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, K-Means Clustering and Association Rules. Experience in applying predictive modeling and machine learning algorithms for analytical reports. Experience using technology to work efficiently with datasets such as scripting, data cleaning tools, statistical software packages. Developed predictive models using Decision Tree, Random Forest, Na ve Bayes, Logistic Regression, Cluster Analysis, and Neural Networks. Very Strong in Python, statistical analysis, tools, and modeling. Experienced in Machine Learning and Statistical Analysis with Python Scikit-Learn. Excellent Knowledge in Relational Data Warehouse/OLAP concepts, Database Design and methodologies. Strong SQL programming skills, with experience in working with functions, packages and triggers. Expertise in transforming business requirements into designing algorithms, analytical models, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data. Skilled in performing data parsing, data manipulation, data architecture, data ingestion and data preparation with methods including describe data contents, compute descriptive statistics of data, regex, split and combine, merge, Remap, subset, reindex, melt and reshape. Worked with NoSQLDatabase including Hbase, Cassandra and MongoDB. Experienced in Big Data with Hadoop, MapReduce, HDFS and Spark. Experienced in Data Integration Validation and Data Quality controls for ETL process and Data Warehousing using MS Visual Studio, SSAS, SSISand 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 andPython 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. Experience with Big Data technologies like Hadoop and Spark would be a plus. Worked and extracted data from various database sources like SQL Server, Oracle and DB2. Experience working at Pricing and/or Revenue Management would be a plus. Familiarity with agile principles (e.g. Scrum), facilitating workshops and prototyping. Hands on experience with R-Studio for doing data pre-processing and building machine learning algorithms on different datasets. Good Knowledge in NoSQL databases like HBase and Mongo DB. Time Series Analysis -ARIMA, Neural Networks, Sentiment Analysis, Forecasting and Text Mining. Cluster Analysis, Principal Component Analysis, Association Rules, Recommender Systems. Inferential Statistics, Hypothesis Testing, Descriptive, and Sampling. Work Experience Data Scientist Palo Alto Networks Dallas TX - Dallas, TX August 2018 to Present Description:Palo Alto Networks, Inc. an American multinational cyber security company. Its core products are a platform that includes advanced firewalls and cloud-based offerings that extend those firewalls to cover other aspects of security. Responsibilities: Extracted data from HDFS and prepared data for exploratory analysis using data munging Built models using Statistical techniques like Bayesian HMM and Machine Learning classification models like XGBoost, SVM, and Random Forest. Participated in all phases of data mining, data cleaning, data collection, developing models, validation, and visualization and performed Gap analysis. A highly immersive Data Science program involving Data Manipulation& Visualization, Web Scraping, Machine Learning, Python programming, SQL, GIT, MongoDB, Hadoop. Setup storage and data analysis tools in AWS cloud computing infrastructure. Installed and used Caffe Deep Learning Framework Worked on different data formats such as JSON, XML and performed machine learning algorithms in Python. Worked as Data Architects and IT Architects to understand the movement of data and its storage and ER Studio 9.7 Used pandas, numpy, seaborn, matplotlib, scikit-learn, scipy, NLTK in Python for developing various machine learning algorithms. Data Manipulation and Aggregation from different source using Nexus, Business Objects, Toad, Power BI and Smart View. Implemented Agile Methodology for building an internal application. Focus on integration overlap and Informatica newer commitment to MDM with the acquisition of Identity Systems. Coded

proprietary packages to analyze and visualize SPCfile data to identify bad spectra and samples to reduce unnecessary procedures and costs. Programmed a utility in Python that used multiple packages (numpy, scipy, pandas) Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, Naive Bayes, KNN. As Architect delivered various complex OLAPdatabases/cubes, scorecards, dashboards and reports. Updated Python scripts to match training data with our database stored in AWS Cloud Search, so that we would be able to assign each document a response label for further classification. Used Teradata utilities such as Fast Export, MLOAD for handling various tasks data migration/ETL from OLTP Source Systems to OLAP Target Systems Data transformation from various resources, data organization, features extraction from raw and stored. Validated the machine learning classifiers using ROC Curves and Lift Charts.

Environment: Unix, Python 3.5.2, MLLib, SAS, regression, logistic regression, Hadoop 2.7.4, NoSQL, Teradata, OLTP, random forest, OLAP, HDFS, ODS, NLTK, SVM, JSON, XML and MapReduce. Data Scientist Cisco - Dallas, TX May 2017 to July 2018 Description:Cisco Systems, Inc. is an American multinational technology conglomerate that develops, manufactures and sells networking hardware, telecommunications equipment and other high-technology services and products. Cisco helps seize the opportunities of tomorrow by proving that amazing things can happen when you connect the unconnected. An integral part of our DNA is creating long-lasting customer partnerships, working together to identify our customers' needs and provide solutions that fuel their success. Responsibilities: Utilized Spark, Scala, Hadoop, HQL, VQL, oozie, pySpark, Data Lake, Tensor Flow, HBase, Cassandra, Redshift, MongoDB, Kafka, Kinesis, Spark Streaming, Edward, CUDA, MLLib, AWS, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. Utilized the engine to increase user lifetime by 45% and triple user conversations for target categories. Application of various 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. Worked onanalyzing data from Google Analytics, AdWords, Facebook etc. Evaluated

models using Cross Validation, Log loss function, ROC curves and used AUC for feature selection and elastic technologies like ElasticSearch, Kibana. Performed Multinomial Logistic Regression, Decision Tree, Random forest, SVM to classify package is going to deliver on time for the new route.

Performed data analysis by using Hive to retrieve the data from Hadoop cluster, Sql to retrieve data from Oracle database and used ETL for data transformation. Performed Data Cleaning, features scaling, features engineering using pandas and numpy packages in python. Performed data cleaning and feature selection using MLlib package in PySpark and working with deep learning frameworks such as Caffe, Neon. Developed Spark/Scala,R Python for regular expression (regex) project in the Hadoop/Hive environment with Linux/Windows for big data resources. Used clustering technique K-Means to identify outliers and to classify un-labeled data. Tracking operations using sensors until certain criteria is met using Air Flow technology. Responsible for different Data mapping activities from Source systems to Teradata using utilities like TPump, FEXP,BTEQ, MLOAD, FLOAD etc Addressed over fitting by implementing of the algorithm regularization methods like L1 and L2. Used Principal Component Analysis in feature engineering to analyze high dimensional data. Used MLlib, Spark's Machine learning library to build and evaluate different models. Created and designed reports that will use gathered metrics to infer and draw logical conclusions of past and future behavior. Developed Map Reduce pipeline for feature extraction using Hive and Pig. 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. Environment: Python 2.x, CDH5, HDFS, Hadoop 2.3, Hive, Impala, AWS, Linux, Spark, Tableau Desktop, SQL Server 2014, Microsoft Excel, Matlab, Spark SQL, Pyspark. Data Scientist Cardlytics, Inc - Atlanta, GA January 2016 to April 2017 Description: Cardlytics uses purchase-based intelligence to make marketing more relevant and measurable. It help marketers identify, reach and influence likely buyers at scale, as well as measure the true sales impact of marketing campaigns. Responsibilities: Used SSIS to create ETL packages to Validate, Extract, Transform and Load data into Data Warehouse and Data Mart. Maintained and developed complex SQL queries, stored procedures, views, functions and reports that meet customer

requirements using Microsoft SQL Server 2008 R2.   Created Views and Table-valued Functions, Common Table Expression (CTE), joins, complex sub queries to provide the reporting solutions. Optimized the performance of queries with modification in T-SQL queries, removed the unnecessary columns and redundant data, normalized tables, established joins and created index.   Created SSIS packages using Pivot Transformation, Fuzzy Lookup, Derived Columns, Condition Split, Aggregate, Execute SQL Task, Data Flow Task and Execute Package Task.   Migrated data from SAS environment to SQL Server 2008 via SQL Integration Services (SSIS).   Developed and implemented several types of Financial Reports (Income Statement, Profit& Loss Statement, EBIT, ROIC Reports) by using SSRS.   Developed parameterized dynamic performance Reports (Gross Margin, Revenue base on geographic regions, Profitability based on web sales and smartphone app sales) and ran the reports every month and distributed them to respective departments through mailing server subscriptions and SharePoint server.   Designed and developed new reports and maintained existing reports using Microsoft SQL Reporting Services (SSRS) and Microsoft Excel to support the firm's strategy and management.   Created sub-reports, drill down reports, summary reports, parameterized reports, and ad-hoc reports using SSRS.   Used SAS/SQL to pull data out from databases and aggregate to provide detailed reporting based on the user requirements. Used SAS for pre-processing data, SQL queries, data analysis, generating reports, graphics, and statistical analyses.   Provided statistical research analyses and data modeling support for mortgage product.   Perform analyses such as regression analysis, logistic regression, discriminant analysis, cluster analysis using SAS programming.   Environment: SQL Server 2008 R2, DB2, Oracle, SQL Server Management Studio, SAS/ BASE, SAS/SQL, SAS/Enterprise Guide, MS BI Suite (SSIS/SSRS), T-SQL, SharePoint 2010, Visual Studio 2010, Agile/SCRUM. Data Analyst Beacon Healthcare Communications - Bedminster, NJ March 2014 to December 2015 Description: At Beacon, we're all Engagement Architects - people with significant industry experience engaging the 3 primary healthcare customers - consumers, providers, and payers.And because we also engage with one another, we're able to provide a more efficient integration of thinking right from the start. One that looks at all the stakeholders, regardless of what you hire us for.   Responsibilities:

Used SSIS to create ETL packages to Validate, Extract, Transform and Load data into Data Warehouse and Data Mart. Collaborating with business and technology teams. Data Analysis-Data collection, data transformation and data loading the data using different ETL systems like SSIS and Informatica. Performed source to target mapping as part of data migration from JD Edwards system to Agile PDM system. Data Migration testing and implementation activities using SSIS and SSRS tools of Microsoft SQL Server 2008. Involved in construction of data flow diagrams and documentation of the processes. Interacted with end users for requirements study and analysis by JAD (Joint Application Development). Participated in system and use case modeling like activity and use case diagrams. Analyzed user requirements & worked with data modelers to identify entities and relationship for data modeling. Actively participated in the design of data model like conceptual, logical models using Erwin. Used Exception handling application block for checking errors/exceptions across the website. Developed Report Component, so that it retrieves the data by executing Stored Procedures throw Data Access component. Environment: Windows, Oracle, MS Excel, SSIS, Informatica, GAP Analysis, ERWIN Data Analyst CybermateInfotek Limited - Hyderabad, Telangana December 2012 to February 2014 Description: CybermateInfotek Ltd. (CIL) is a Software solutions and IT services company, was founded in May 1994 at Hyderabad, INDIA and is a. CIL is an offshore software development company executing projects on Web & Web related technologies (Both Microsoft & Open Source's). Responsibilities: Involved in Data mapping specifications to create and execute detailed system test plans. The data mapping specifies what data will be extracted from an internal data warehouse, transformed and sent to an external entity. Worked closely with stakeholders to understand, define, document business questions needed. Review system/application requirements (functional specifications), test results and metrics for quality and completeness. Designed and Developed Oracle PL/SQL Procedures and UNIX Shell Scripts for Data Import/Export and Data Conversions. Analyzed the source data coming from different sources (SQL Server, Oracle and also from flat files like Access and Excel) and working with business users and developers to develop the Model. Have Used Informatica Data Quality (IDQ) and Informatica Power Center as ETL tools to extract the data from

various sources systems and transform them into one common format and load them into target database for the analysis purpose from Data Warehouse.   Accomplished data analysis, statistical reports and graphs based on the business requirement using SAS/Base, SAS/Macro and SAS/Graph, SAS/SQL, SAS/Access, SAS/ODS and SAS/Connect.   Worked on Predictive Modeling using SAS/SQL.   Executed SQL queries to validate actual test results and match expected results as per financial rules.   Responsible for maintaining the integrity of the SQL database and reporting any issues to the database architect.   Design and model the reporting data warehouse considering current and future reporting requirement   Involved in the daily maintenance of the database that involved monitoring the daily run of the scripts as well as troubleshooting in the event of any errors in the entire process.   Involved with statistical domain experts to understand the data and worked with data management team on data quality assurance.   Environment: SQL Server, Oracle PL/SQL, Informatics Data Quality (IDQ), Informatics PowerCenter (Designer, Workflow Manager Workflow Monitor), UNIX, SAS/Base, SAS/Macro and SAS/Graph, SAS/SQL, SAS/Access, SAS/ODS and SAS/Connect. Data Analyst Rsoft India Pvt.ltd - Bengaluru, Karnataka January 2011 to November 2012 Description: RSoft is a leading software product development company in India. We develop Mobile & Cloud based Customer Relationship Management (CRM) Software Solution for all size of businesses. Such as Small, Medium, Large Size of businesses, B2B and B2C enterprises to increase leads and sales opportunities.   Responsibilities: Understanding the requirements and develop various packages in SSIS.   Gathered requirements from JAD/JAR sections with developers and business clients.   Designed the business requirement collection approach based on the project scope and SDLC methodology.   Designs and develops the logical and physical data models to support the Data Marts and the Data Warehouse   Create SQL queries for product components to update FACETS backend tables and create product prefixes.   Involved in formatting data stores and generate UML diagrams of logical and physical data.   Developed project plans and manage project scope.   Identified/documented data sources and transformation rules required to populate and maintain data warehouse content.   Write and execute positive and negative test cases to ensure the data originating from the data warehouse (Oracle dB) is accurate

through to the SQL dB in the applications. Work and triage Facets configuration issues and route work back for correct processing. Document step by step Facets configuration steps for the Quality Assurance team. Assisted in building a Business Analysis Process Model using Rational Rose and Visio. Created ad-hoc and custom reports using Microsoft Access, Cognos BI and Crystal reports. Performed extensive Requirement analysis and developed use cases and workflows. Designed and developed Use Cases, Activity Diagrams, Sequence Diagrams, and OOD Played a key role in the planning, User Accepted Testing, and Implementation of system enhancements and conversions. Updated provider tables, diagnosis tables, fee schedules, and service is contained in Facetsback-end tables. Environment: MS SQL Server 2008R2, BIDS 2008 (SSIS), JAD/JAR, SSIS, Cognos BI, Use Cases, Activity Diagrams, Sequence Diagrams. Education Bachelor's Skills Db2, Microsoft sql server, Microsoft sql server 2008, Sql server, Sql server 2008, Mysql, Oracle, Sql, Cassandra, Hdfs, Impala, Mapreduce, Oozie, Sqoop, Hbase, Kafka, Flume, Hadoop, Mongodb, Splunk Additional Information TECHNICAL SKILLS BigData/Hadoop Technologies Hadoop, HDFS, YARN, MapReduce, Hive, Pig, Impala, Sqoop, Flume, Spark, Kafka, Storm, Drill, Zookeeper and Oozie Languages HTML5,DHTML, WSDL, CSS3, C, C++, XML,R/R Studio, SAS Enterprise Guide, SAS, R (Caret, Weka, ggplot), Perl, MATLAB, Mathematica, FORTRAN, DTD, Schemas, Json, Ajax, Java, Scala, Python (NumPy, SciPy, Pandas, Gensim, Keras), Java Script, Shell Scripting NO SQL Databases Cassandra, HBase, MongoDB, MariaDB Business Intelligence Tools Tableau server, Tableau Reader, Tableau, Splunk, SAP Business Objects, OBIEE, SAP Business Intelligence, QlikView, Amazon Redshift, or Azure Data Warehouse Development Tools Microsoft SQL Studio, IntelliJ, Eclipse, NetBeans. Development Methodologies Agile/Scrum, UML, Design Patterns, Waterfall Build Tools Jenkins, Toad, SQL Loader, Maven, ANT, RTC, RSA, Control-M, Oziee, Hue, SOAP UI Reporting Tools MS Office (Word/Excel/Power Point/ Visio/Outlook), Crystal reports XI, SSRS, cognos 7.0/6.0. Databases Microsoft SQL Server 2008,2010/2012, MySQL 4.x/5.x, Oracle 11g, 12c, DB2, Teradata, Netezza Operating Systems All versions of Windows, UNIX, LINUX, Macintosh HD, Sun Solaris

Name: Dr. John Owens

Email: ysmith@example.org

Phone: 001-911-953-9838x6649