Data Resarcher/ Data Scientist Data Resarcher/ Data Scientist Data Resarcher/ Data Scientist - USAA San Antonio, TX Professional qualified Data Scientist/Data Analyst with experience in Data Science and Analytics including Data Mining , Deep Learning/Machine Learning and Statistical Analysis Involved in the entire data science project life cycle and actively involved in all the phases including data cleaning, data extractionand datavisualization with large data sets of structured and unstructured data, created ER diagrams and schema. Experienced with machine learning algorithm such as logistic regression, KNN, SVM, random forest, neural network, linear regression, lasso regression and k-means Implemented Bagging and Boosting to enhance the model performance. Experience in implementing data analysis with various analytic tools, such as Anaconda Jupiter Notebook 4.X, R 3.0 (ggplot2, , dplyr, Caret) and Excel Solid ability to write and optimize diverse SQL queries, working knowledge of RDBMS like SQL Server 2008/2010/2012, NoSql databases like MongoDB 3.2 Excellent understanding Agile and Scrum development methodology Used the version control tools like Git 2.X and build tools like Apache Maven/Ant 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. Developed predictive models using Decision Tree, Naive Bayes, Logistic Regression, Random Forest, 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, Scipy, Numpy and Pandas for data analysis. 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. Work Experience Data Resarcher/ Data Scientist USAA - San Antonio, TX April 2019 to Present Description: The United Services Automobile Association (USAA) is a San Antonio, Texas-based Fortune 500 diversified financial services group of companies including a Texas Department of Insurance-regulated reciprocal inter-insurance exchange and subsidiaries offering banking, investing, and insurance to people and families who serve, or served, in the United States Armed Forces.At the end of 2017, there were 12.4 million members. Responsibilities: Test and determine whether new technology is potentially useful for USAA Extracting Opearations data from Workday and storing it in Hadoop Getting the data in Pickle file Format and Parsing it by CSV. Build a graphic database using DataStax in Python (Object-Oriented Programming) Query data/information from graphic database using Gremlin to support model validation. Work in a group to develop Machine Learning guidance for model development and validation Build analytic tools to prepare data and graphs using Power Query and GraphX. Draft the procedures for reporting. Worked on ETL tools such as Apache Airflow Build a auto DAG system which would automatically trigger the workflow whenever the Airflow will receive HTTP request. Write python scripts to parse documents. Take online courses as a means to continuously learn new subjects Did intensive research on the tools like Graph Networks, Scheduling Tools and Data Wrangling Tools to determine the best tool available in the market that would be best fit for company's need. Environment: Windows, Python 3, DataStax, GraphX, Apache Airflow, SQL. Data Scientist/ Machine

Learning Rauxa - New York, NY August 2018 to March 2019 Description: Makers of results, Rauxa applies data, technology, and content to create measurable impact at maximum speed for clients that include Gap Inc., TGI Fridays, and Verizon. The country's largest woman-owned independent advertising agency. Responsibilities: 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, 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 Tensorflow, a Deep Learning Framework 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. 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 OLAP databases/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. 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 Charter communication - St. Louis, MO May 2017 to July 2018 Description: Charter Communications, Inc. is an American telecommunications and mass media company that offers its services to consumers and businesses under the branding of Spectrum. Responsibilities: Utilized Scala, Hadoop, pySpark, Data Lake, TensorFlow, MongoDB, AWS, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. Developed sentiment analysis framework

for email conversation and Customer Satisfaction score(CSAT) correlation. 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. Sentiment Analytics engine was completely built in Python3. Email chain was considered for analysis. Emails were pulled into DB from a tool Email2DB. Each email transactions was pre-processed, sentiment calculation, subjectivity, polarity was taken and linked to the CSAT score for dashboarding. Emite was used as a tool of choice for visualization. RapidMiner and PredictionIO was used as tool of choice for predictive analysis. ITSM data was taken and build a engine for predicting CPU failure, Memory Issue, Disk Space Issue, Server failure Errors. Training data was build from either of proactive and reactive ticket data which was in turn to be used to make Classification Model for prediction. Identifying and evaluating potential vendors for technology fitments, performing proof of value exercise, conducting commercial and contract negotiations. Performed Data Profiling to learn about behavior with various features such as traffic pattern, location, Date and Time etc. Categorized comments into positive and negative clusters from different social networking sites using Sentiment Analysis and Text Analytics 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 datafrom Oracle database and used ETL for data transformation. Performed Data Cleaning, features scaling, features engineering using pandas and numpy packages in python. Exploring DAG's, their dependencies and logs using AirFlow pipelines for automation Performed data cleaning and feature selection using MLlib package in PySpark and working with deep learning frameworks such as Tensorflow and PyTorch 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 unlabeled data. Communicated the results with operations team for taking best decisions. Collected data needs and requirements by Interacting with the other departments.

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 analyst EDAC Technologies Corp - Cheshire, CT January 2016 to April 2017 Description: EDAC Technologies Corporation provides design, manufacturing, and services for tooling, fixtures, molds, jet engine components, and machine spindles in the aerospace, industrial, semiconductor, and medical device markets Responsibilities: Worked with BI team in gathering the report requirements and also Sqoop to export data into HDFS and Hive Involved in the below phases of Analytics using R, Python and Jupyter notebook. Data collection and treatment: Analysed existing internal data and external data, worked on entry errors,classification errors and defined criteria for missing values Data Mining: Used cluster analysis for identifying customer segments, Decision trees used for profitable and non-profitable customers, Market Basket Analysis used for customer purchasing behaviour and part/product association. Emite was used as a tool of choice for visualization. Assisted with data capacity planning and node forecasting. Installed, Configured and managed Flume Infrastructure Worked closely with the claims processing team to obtain patterns in filing of fraudulent claims. Worked on performing major upgrade of cluster from CDH3u6 to CDH4.4.0 Developed Map Reduce programs to extract and transform the data sets and results were exported back to RDBMS using Sqoop. Patterns were observed in fraudulent claims using text mining in R and Hive. Exported the data required information to RDBMS using Sqoop to make the data available for the claims processing team to assist in processing a claim based on the data. Developed Map Reduce programs to parse the raw data, populate staging tables and store the refined data in partitioned tables in the EDW. Created tables in Hive and loaded the structured (resulted from Map Reduce jobs) data Created Hive queries that helped market analysts spot emerging trends by comparing fresh data with EDW reference tables and historical metrics. Enabled speedy reviews and first mover advantages by using Oozie to automate data loading into the Hadoop Distributed File System and PIG to pre-process the data. Provided design recommendations and thought leadership to sponsors/stakeholders that improved review processes and resolved technical problems. Managed and reviewed Hadoop log files. Tested raw data and

executed performance scripts. Environment: HDFS, PIG, HIVE, Map Reduce, Linux, HBase, Flume, Sqoop, R, VMware, Eclipse, Cloudera, Python. Data Analyst Dorman Products Inc - Colmar, PA March 2014 to December 2015 Description: Dorman Products, Inc. supplies automotive replacement parts, automotive hardware, and brake products to the automotive aftermarket and mass merchandise markets in the United States, Canada, Mexico, Europe, the Middle East, and Australia. Responsibilities: Created and maintained Logical and Physical models for the data mart. Created partitions and indexes for the tables in the data mart. Performed data profiling and analysis applied various data cleansing rules designed data standards and architecture/designed the relational models. Maintained metadata (data definitions of table structures) and version controlling for the data model. Developed SQL scripts for creating tables , Sequences , Triggers , views and materialized views Worked on query optimization and performance tuning using SQL Profiler and performance monitoring. Developed mappings to load Fact and Dimension tables, SCD Type 1 and SCD Type 2 dimensions and Incremental loading and unit tested the mappings. Utilized Erwin's forward / reverse engineering tools and target database schema conversion process. Worked on creating enterprise wide Model EDM for products and services in Teradata Environment based on the data from PDM. Conceived, designed, developed and implemented this model from the scratch. Building, publishing customized interactive reports and dashboards, report scheduling using Tableau server Write SQL scripts to test the mappings and Developed Traceability Matrix of Business Requirements mapped to Test Scripts to ensure any Change Control in requirements leads to test case update. Responsible for development and testing of conversion programs for importing Data from text files into map Oracle Database utilizing PERL shell scripts &SQL*Loader. Involved in extensive DATA validation by writing several complex SQL queries and Involved in back-end testing and worked with data quality issues. Developed and executed load scripts using Teradata client utilities MULTILOAD, FASTLOAD and BTEQ. Exporting and importing the data between different platforms such as SAS, MS-Excel. Generated periodic reports based on the statistical analysis of the data using SQL Server Reporting Services (SSRS). Worked with the ETL team to document the Transformation Rules for Data Migration from OLTP to

Warehouse Environment for reporting purposes. Created SQL scripts to find data quality issues and to identify keys, data anomalies, and data validation issues. Formatting the data sets read into SAS by using Format statement in the data step as well as Proc Format. Applied Business Objects best practices during development with a strong focus on reusability and better performance. Developed Tableau visualizations and dashboards using Tableau Desktop. Used Graphical Entity - Relationship Diagramming to create new database design via easy to use, graphical interface. Designed different type of STAR schemas for detailed data marts and plan data marts in the OLAP environment. Environment: Erwin, MS SQL Server 2008, DB2, Oracle SQL Developer, PL/SQL, Business Objects, Erwin, MS office suite, Windows XP, TOAD, SQL*PLUS, SQL*LOADER, Teradata, Netezza, SAS, Tableau, Business Objects, SSRS, tableau, SQL Assistant, Informatica, XML.. Python Developer Dabur India Ltd - Ghaziabad, Uttar Pradesh December 2012 to February 2014 Description: Dabur is one of the India's largest Ayurvedic medicine & natural consumer products manufacturer. Dabur demerged its Pharma business in 2003 and hived it off into a separate company, Dabur Pharma Ltd. German company Fresenius SE bought a 73.27% equity stake in Dabur Pharma in June 2008 at Rs 76.50 a share. Responsibilities: Involved in the design, development and testing phases of application using AGILE methodology. Designed and maintained databases using Python and developed Python based API (RESTful Web Service) using Flask, SQLAlchemy and PostgreSQL. Designed and developed the UI of the website using HTML, XHTML, AJAX, CSS and JavaScript. Participated in requirement gathering and worked closely with the architect in designing and modeling. Worked on Restful web services which enforced a stateless client server and support JSON few changes from SOAP to RESTFUL Technology Involved in detailed analysis based on the requirement documents. Involved in writing SQL queries implementing functions, triggers, cursors, object types, sequences, indexes etc. Created and managed all of hosted or local repositories through Source Tree's simple interface of GIT client, collaborated with GIT command lines and Stash. Responsible for setting up Python REST API framework and spring frame work using Django Develope consumer based features and applications using Python, Django, HTML, behavior Driven Development (BDD) and pair based

programming. Designed and developed components using Python with Django framework. Implemented code in python to retrieve and manipulate data. Involved in development of the enterprise social network application using Python, Twisted, and Cassandra. Used Python and Django creating graphics, XML processing of documents, data exchange and business logic implementation between servers. orked closely with back-end developer to find ways to push the limits of existing Web technology. Designed and developed the UI for the website with HTML, XHTML, CSS, Java Script and AJAX Used AJAX&JSON communication for accessing RESTfulweb services data payload. Designed dynamic client-side JavaScript codes to build web forms and performed simulations for web application page. Created and implemented SQL Queries, Stored procedures, Functions, Packages and Triggers in SQL Server. Successfully implemented Auto Complete/Auto Suggest functionality using JQuery, Ajax, Web Service and JSON. Environment: Python 2.5, Java/J2EE, Django1.0, HTML,CSS Linux, Shell Scripting, Java Script, Ajax, JQuery, JSON, XML, PostgreSQL, Jenkins, ANT, Maven, Subversion, Python Data Analyst Kotak Mahindra Bank - Mumbai, Maharashtra January 2011 to November 2012 Description: Kotak Mahindra Bank is an Indian private sector bank headquartered in Mumbai, Maharashtra, India. In February 2003, Reserve Bank of India issued the licence to Kotak Mahindra Finance Ltd., the group's flagship company, to carry on banking business. Responsibilities: Analyzed data sources and requirements and business rules to perform logical and physical data modeling. Analyzed and designed best fit logical and physical data models and relational database definitions using DB2. Generated reports of data definitions. Involved in Normalization/De-normalization, Normal Form and database design methodology. Maintained existing ETL procedures, fixed bugs and restored software to production environment. Developed the code as per the client's requirements using SQL, PL/SQL and Data Warehousing concepts. Involved in Dimensional modeling (Star Schema) of the Data warehouse and used Erwin to design the business process, dimensions and measured facts. Worked with Data Warehouse Extract and load developers to design mappings for Data Capture, Staging, Cleansing, Loading, and Auditing. Developed enterprise data model management process to manage multiple data models developed by different

groups Designed and created Data Marts as part of a data warehouse. Wrote complex SQL queries for validating the data against different kinds of reports generated by Business Objects XIR2. Using Erwin modeling tool, publishing of a data dictionary, review of the model and dictionary with subject matter experts and generation of data definition language. Coordinated with DBA in implementing the Database changes and also updating Data Models with changes implemented in development, QA and Production. Worked Extensively with DBA and Reporting team for improving the Report Performance with the Use of appropriate indexes and Partitioning. Developed Data Mapping, Transformation and Cleansing rules for the Master Data Management Architecture involved OLTP, ODS and OLAP. Tuned and coded optimization using different techniques like dynamic SQL, dynamic cursors, and tuning SQL queries, writing generic procedures, functions and packages. Experienced in GUI, Relational Database Management System (RDBMS), designing of OLAP system environment as well as Report Development. Extensively used SQL, T-SQL and PL/SQL to write stored procedures, functions, packages and triggers. Analyzed of data report were prepared weekly, biweekly, monthly using MS Excel, SQL & UNIX Environment: ER Studio, Informatica Power Center 8.1/9.1, Power Connect/ Power exchange, Oracle 11g, Mainframes,DB2 MS SQL Server 2008, SQL,PL/SQL, XML, Windows NT 4.0, Tableau, Workday, SPSS, SAS, Business Objects, XML, Tableau, Unix Shell Scripting, Teradata, Netezza, Aginity Education Bachelor's Skills SQL

Name: Michelle Benitez

Email: mollyharris@example.net

Phone: 3668392785