

PYTHON DEVELOPER PYTHON DEVELOPER PYTHON DEVELOPER - CAPITAL ONE 5+ years of experience in Machine Learning, Data-mining with large datasets of Structured and Unstructured data, Data Acquisition, Data Validation, Predictive modeling, Data Visualization. Experience in coding SQL/PL SQL using Procedures, Triggers, and Packages. Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python. Excellent Knowledge of Relational Database Design, Data Warehouse/OLAP concepts, and methodologies. Data Driven and highly analytical with working knowledge and statistical model approaches and methodologies (Clustering, Regression analysis, Hypothesis testing, Decision trees, Machine learning), rules and ever-evolving regulatory environment. Professional working experience in Machine Learning algorithms such as Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, K-Means Clustering and Association Rules. Experience with data visualization using tools like Ggplot, Matplotlib, Seaborn, Tableau and using Tableau software to publish and presenting dashboards, storyline on web and desktop platforms. Experienced in python data manipulation for loading and extraction as well as with python libraries such as NumPy, SciPy and Pandas for data analysis and numerical computations. Well experienced in Normalization, De-Normalization and Standardization techniques for optimal performance in relational and dimensional database environments. Experience in multiple software tools and languages to provide data-driven analytical solutions to decision makers or research teams. Familiar with predictive models using numeric and classification prediction algorithms like support vector machines and neural networks, and ensemble methods like bagging, boosting and random forest to improve the efficiency of the predictive model. Good Knowledge of NoSQL databases like Mongo DB and HBase. Develop, maintain and teach new tools and methodologies related to data science and high-performance computing. Extensive hands-on experience and high proficiency with structures, semi-structured and unstructured data, using a broad range of data science programming languages and big data tools including R, Python, Spark, SQL, Scikit Learn, Hadoop Map Reduce Expertise in Technical proficiency in Designing, Data Modeling Online Application,

Solution Lead for Architecting Data Warehouse/Business Intelligence Applications. Cluster Analysis, Principal Component Analysis (PCA), Association Rules, Recommender Systems. Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile methodologies. Adept in statistical programming languages like R and Python including Big Data technologies like Hadoop, Hive. Hands on experience with RStudio for doing data pre-processing and building machine learning algorithms on different datasets. Worked and extracted data from various database sources like Oracle, SQL Server, and DB2. Implemented machine learning algorithms on large datasets to understand hidden patterns and capture insights. Predictive Modelling Algorithms: Logistic Regression, Linear Regression, Decision Trees, K-Nearest Neighbors, Bootstrap Aggregation (Bagging), Naive Bayes Classifier, Random Forests, Boosting, Support Vector Machines. Flexible with Unix/Linux and Windows Environments, working with Operating Systems like Centos5/6, Ubuntu13/14, Cosmos. Work Experience PYTHON DEVELOPER CAPITAL ONE - McLean, VA July 2018 to Present Responsibilities: Implemented Machine Learning, Computer Vision, Deep Learning and Neural Networks algorithms using TensorFlow and designed Prediction Model using Data Mining Techniques with help of Python, and Libraries like NumPy, SciPy, Matplotlib, Pandas, scikit-learn. Used pandas, NumPy, Seaborn, SciPy, matplotlib, scikit-learn in Python for developing various machine learning algorithms. Used scikit-learn in modeling various classification, regression and clustering algorithms including support vector machines, random forests, gradient boosting, k-means. 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 Participated in all phases of data mining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis. Data Manipulation and Aggregation from a different source using Nexus, Toad, Business Objects, Powerball, and Smart View. Good knowledge of Hadoop Architecture and various components such as HDFS, Job Tracker, Task Tracker, Name Node, Data Node, Secondary Name Node, and Map Reduce concepts. Programmed by a utility in Python that used

multiple packages (SciPy, NumPy, pandas) Implemented Classification using supervised algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes. Responsible for design and development of advanced R/ Python programs to prepare to transform and harmonize data sets in preparation for modeling. 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 Participated in all phases of data mining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis. Data Manipulation and Aggregation from a different source using Nexus, Toad, Business Objects, Power BI and Smart View. 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. Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS. Interaction with Business Analyst, SMEs, and other Data Architects to understand Business needs and functionality for various project solutions Researched, evaluated, architected, and deployed new tools, frameworks, and patterns to build sustainable Big Data platforms for the clients. 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. Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS. Identifying and executing process improvements, hands-on in various technologies such as Oracle, Informatica, and Business Objects. Designed both 3NF data models for ODS, OLTP systems and dimensional data models using Star and Snowflake Schemas. Environment: Erwin r9.6, Python, SQL, Oracle 12c, SQL Server, SSRS, PL/SQL, T-SQL, Tableau, MLlib, regression, Cluster analysis, Scala NLP, Spark, Kafka, Mongo DB, logistic regression, Hadoop, PySpark, Teradata, random forest, OLAP, Azure, MariaDB, SAP CRM, HDFS, ODS, NLTK, SVM, JSON, Tableau, XML, Cassandra, MapReduce, AWS. DATA

SCIENTIST/MACHINE LEARNING RESEARCH ASSISTANT COX AUTO - Atlanta, GA July 2016 to June 2018 Responsibilities: Utilized Spark, Scala, Hadoop, HQL, VQL, oozie, pySpark, Data Lake, TensorFlow, HBase, Cassandra, Redshift, Mongo DB, Kafka, Kinesis, Spark Streaming, Edward, CUDA, MLLib, AWS, Python, a broad variety of machine learning methods including classifications, regressions, dimensionally reduction etc. 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, Mat lab. Used the version control tools like Git 2.X and build tools like Apache Maven/Ant Worked on analyzing data from Google Analytics, Ad Words, and Facebook etc. Evaluated models using Cross Validation, Log loss function, ROC curves and used AUC for feature selection and elastic technologies like Elastic Search, Kibana. Used Python scripts to update content in the database and manipulate files Skilled in using dplyr and pandas in R and Python for performing exploratory data analysis. Performed Multinomial Logistic Regression, Decision Tree, Random forest, SVM to classify package is going to deliver on time for the new route. Used Jenkins for Continuous Integration Builds and deployments (CI/CD). 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. 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. 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. CICD pipeline implementation for Java applications. CICD implementation on Azure cloud platform. Ensured that the model has low False Positive Rate and Text classification and sentiment analysis for unstructured and semi-structured data. 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. Implemented rule based expertise system from the results of exploratory analysis and information gathered from the people from different departments. 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. Communicated the results with operations team for taking best decisions. Collected data needs and requirements by Interacting with the other departments. Environment: R, Python, HDFS, ODS, OLTP, Oracle 10g, Hive, OLAP, DB2, Metadata, MS Excel, Mainframes MS Vision, Map-Reduce, Rational Rose, SQL, and MongoDB. DATA ANALYST/DATA SCIENTIST Bengaluru, Karnataka May 2015 to June 2016 Responsibilities: Worked closely with data mapping SME and QA team to understand the business rules for acceptable data quality standards. Created and provided support on various monitoring and control reports which includes, Customer verification report that accepted offer in sales engine, AMF waiver report, Credit fulfillment report, qualification and offer load volume reconciliation report and upgrade performance monitoring report. Wrote complex SQL queries to identify granularity issues and relationships between data sets and created recommended solutions based on analysis of the query results. Wrote the SQL queries on data staging tables and data warehouse tables to validate the data results. Performed data profiling on datasets with millions of rows on Teradata environment, validating key gen elements, ensuring correctness of codes and identifiers, and recommending mapping changes. Performed unit testing on transformation rules to ensure data moved correctly. Created Python scripts to take client content documents and images as input and create web pages, including home page, table of contents and links. Involved in full life cycle of Business Objects reporting Application. Worked directly with Cloud System Administrators and project managers supporting Amazon Web Services (AWS) migration. Delivered Enterprise Data Governance, Data Quality, Metadata, and ETL Informatica solution. Maintained Excel workbooks, such as development of pivot tables, exporting data from external SQL databases, producing reports and updating spreadsheet information. Used Python's Panda library in the process of analyzing the data Involved in extracting and transforming

of Data from Enterprise data warehousing. Actively involved in communication with Business Analyst, user acceptance testers and BPM for requirements gathering. Engaged in performing various ad-hoc queries. Implanted various Data transformation such as SQL, extract, Split and data validation. Analyzing and supporting day to day marketing strategy solutions. Researched and fixed data issues pointed out by QA team during regression tests. Interfaced with business users to verify business rules and communicated changes to ETL development team. Created Tableau views with complex calculations and hierarchies making it possible to analyze and obtain insights into large data sets. Created and executed SQL queries to perform Data Integrity testing on a Teradata Database to validate and test data using TOAD. Worked with data architects team to make appropriate changes to the data models. Worked on the ETL Informatica mappings and other ETL Processes (Data Warehouse). Worked with the data governance team to ensure the data quality of compliance reports for EDI transactions. Utilized Tableau server to publish and share the reports with the business users. Experienced in designing complex Drill-Down & Drill-Through Reports using Business Objects. Experienced in creating UNIX scripts for file transfer and file manipulation. Generated ad-hoc or management specific reports using Tableau and Excel. Analyzed the subscriber, provider, members and claim data to continuously scan and create authoritative master data. Prepare the data rules spreadsheet using MS Excel that will be used to update allowed values, findings, and profiling results. Performed auditing in the development phase in order to assure data quality and integrity. Provided insight and ideas in support of customer management processes and reporting. Involved in data cleansing mechanism in order to eliminate duplicate and inaccurate data. Environment: Windows 7, Linux, Tableau desktop, Tableau Server, Business Objects, R, SQL Developer, MySQL, MS-Access, MS Excel and SQL.

DATA ANALYST INTERN AMARA RAJA - Tirupati, Andhra Pradesh April 2014 to April 2015

Responsibilities: Worked with Data governance, Data quality, data lineage, Data architect to design various models and processes. Implemented end-to-end systems for Data Analytics, Data Automation and integrated with custom visualization tools using Informatica, Tableau and business objects. Designed, developed, tested, and maintained Tableau functional reports based on user

requirements. Mastered the ability to design and deploy rich Graphic visualizations using Tableau and Converted existing Business objects reports into tableau dashboards Interaction with Business Analyst, SMEs and other Data Architects to understand Business needs and functionality for various project solutions. Used Informatica Power Center for (ETL) extraction, transformation and loading data from heterogeneous source systems into target database. Created mappings using Designer and extracted data from various sources, transformed data according to the requirement. Involved in extracting the data from the Flat Files and Relational databases into staging area. Developed Informatica Mappings and Reusable Transformations to facilitate timely Loading of Data of a star schema. Developed the Informatica Mappings by usage of Aggregator, SQL overrides usage in Lookups, source filter usage in Source qualifiers, and data flow management into multiple targets using Router. Created Sessions and extracted data from various sources, transformed data according to the requirement and loading into data warehouse. Used various transformations like Filter, Expression, Sequence Generator, Update Strategy, Joiner, Router and Aggregator to create robust mappings in the Informatica Power Center Designer. Imported various heterogeneous files using Informatica Power Center 8.x Source Analyzer. Developed several reusable transformations that were used in other mappings. Prepared Technical Design documents and Test cases Involved in Unit Testing and Resolution of various Bottlenecks came across. Environment: SAS/Base, SAS/Connect, SAS/UNIX, SAS/ODS, SAS/Macros, SQL, Tableau, MS Excel, Power Point, Mainframe, DB2, Teradata, SAS Enterprise guide. Education BE (Honors in Electronics and Communication Engineering Jawaharlal Nehru Technological University

Name: Michael Smith

Email: connie39@example.com

Phone: (732)513-7478x003