Data Scientist Data Scientist Data Scientist - Zoetis Inc Tampa, FL   Data Scientist with 12 Plus years of experience in, Statistical Modeling, Machine Learning, Data Mining with Structured and Unstructured Data.   Performed Data Acquisition, Data Validation, Predictive Modeling and Data Visualization.   Expertise in Python (2.x/3.x) programming with multiple packages including NumPy, Pandas, Matplotlib, SciPy, Seaborn andScikit-learn.   Hands on experience in implementing LDA, Na ve Bayes and skilled in Random Forests, Decision Trees, Linear and Logistic Regression, SVM, Clustering, Neural Networks, Principle Component Analysis and good knowledge on Recommender Systems.   Implementation experiences in Machine Learning and deep learning, including Regression, Classification,Natural Language Processing (NLP) using packages like NLTK, Spacy. Experience in tuning algorithms using methods such as Grid Search, Randomized Search, K-Fold Cross Validation and Error Analysis.   Also worked with several boosting methodologies like ADA Boost, Gradient Boosting andXGBoost.   Validated the machine learning classifiers using Accuracy, AUC, ROCCurves and Lift Charts.   Worked with various text analytics or Word Embedding libraries like Word2Vec, Count Vectorizer, GloVe, LDA etc.   Solid knowledge and experience in Deep Learning techniques including Feedforward Neural Network, Convolutional Neural Network (CNN), Recursive Neural Network (RNN).   Worked with numerous data visualization tools in python like Matplotlib, Seaborn, ggplot, pygal.   Worked and extracted data from various database sources like Oracle, SQL Server,and MongoDB.   Highly skilled in using Hadoop, HBase, Spark, and Hive for basic analysis and extraction of data in the infrastructure to provide data summarization.   Handled importing data from various data sources, performed transformations using Hive, MapReduce, and loaded data into HDFS.   Experience on working with different operating systems like UNIX, LINUX, and Windows.   Experience working with MS Word, MS Excel, MS PowerPoint, MS SharePoint, and MS Project. Work Experience Data Scientist Zoetis Inc April 2018 to Present Zoetis Inc. is the world's largest producer of medicine and vaccinations for pets and livestock. Zoetis delivers quality medicines, vaccines and diagnostic products, which are complemented by genetic tests, bio devices and a range of services. The project is to collect data from different sources and create a master data set and doing predictions on sales and profits. Measures to be taken for improving the sales by

applying machine learning strategies and statistical analyses to support animal health projects and products. Responsibilities: Developing analytical databases from different sources and create a master data set. Responsible for data identification, collection, exploration, cleaning for modeling. Performed Time Series Analysis on animal medicine and vaccine product sales data in order to extract meaningful statistics and other characteristics of the data to predict future values based on previously observed values. Determined customer satisfaction and helped enhance customer experience using NLP. Manage large data sets from a wide variety of sources and apply analytics and statistical analyses to support animal health projects and products. Analysis of biological and spatial data to develop insights into precision animal management and precision medicine. Implementing analytics algorithms in Python. Performed training Natural Language models and reinforcement learning engines to optimize intelligent agents that automate task execution. Performed k-Means clustering in order to understand customer itemized bought products and segment the customers based on the customer products for animal medicine and vaccines behavior information for customized product offering, customized and priority service, to improve existing profitable relationships and to avoid customer churn, etc using Python. Performed Text analytics on unstructured email data using Natural language processing tool kit (NLTK). Worked with text to vector representation methods including Counter Vectorizer, Tf-idf for topic modelling. Applying Clustering algorithms to group the data on their similar behavior patterns. Work with data analytics team to develop time series and optimization. Experienced in data scraping. Used PySpark Machine learning library to build and evaluate different models. Created various Proof of Concepts (PoC) and gap analysis and gathered necessary data for analysis from different sources, prepared data for data exploration using data munging. Experienced in Agile Methodology. Used Tableau to generate reports with internal records, secondary sources of data, JSON, CSV and more.Which helped the support team for better marketing. Environment: Python 3.7.0, PySpark, NLTK, SQL Server , Microsoft Excel, SQL, AWS, Tablau, Sqoop, ETL, agile. Data Scientist Mars Solutions Group, WI January 2016 to March 2018 The Client is the largest Healthcare Provider and offers health care products, insurance services, Data Analytics, Payment Integrity, and The project was to

build predictive models for customer value analysis by applying machine learning methods, principal component analysis, and regression on large dataset. Responsibilities: Worked on MachineLearning, Datamining with largedatasets of Structured and Unstructureddata, DataAcquisition, DataValidation, Predictivemodeling, DataVisualization. Performed Multinomial Logistic Regression, Random forest machine learning algorithms. Used AWS to manage the data in cloud. Good knowledge on Hadoop components such as HDFS, Job Tracker, TaskTracker, Name Node, Data Node, and MapReduce concepts. Maintained updated Log filesusing Python. Alteryx was used for Data Preparation. Used machine learning algorithms like Logistic Regression ,Knn, Decision Trees, Random Forest to make the data to fit for the desired output. Interact and brain-storm with multifunctional teams to explore the opportunities of using data to improve business and health-care outcomes. Developed machine-learning models and translate complex ideas and results into actionable management insights and solutions to achieve the expected business/health-care goals. Design, coding, unit testing of ETL package source marts and subject marts using Informatica ETL processes for Oracledatabase. Worked in Agile Methodology. Generated reports with internal records, secondary sources of data, JSON, CSV and more. Developed various QlikView Data Models by extracting and using the data from various sources files, DB2, Excel, Flat Files and Bigdata. Provided schedules, status reports, and issue resolutions to the Project team, Business Users, and Project Managers. Environment: Python 3.x, Linux, Spark, SQL Server 2012, Microsoft Excel, , Spark SQL, AWS, QlikView ,Sqoop, ETL, agile. Data Analyst / Data Scientist CMS Energy - Jackson, MI September 2013 to December 2015 CMS Energy is an energy company that is focused principally on utility operations. I was responsible for building a new data science department with the help of other departments and I was able to learn how the business is operated and helped the company to grow and stay ahead of the competition. By using machine learning we improvised the predictive algorithm for pricing strategy. And we creating alerts that would notify customers of potential issues that their system has solely based on the data available to me. Responsibilities: Worked on Data Manipulation & Visualization, Machine Learning, Python, and SQL. Transformed the business requirementsinto analytical

models, designing algorithms, building models, developing data mining and reporting solutions that scales across massive volume of structured and unstructured data.   Worked on customer segmentation using an unsupervised learning technique - clustering.   Implemented Classification using Supervised learninglike Logistic Regression, Decision trees, KNN, Naive Bayes.   Built models using Statistical techniques and Machine Learning classification models like XG Boost,SVM, and Random Forest.   Improved model's accuracy by using Gradient Boosting technique like Light GBM and gained around 82% accuracy with Random Forest and 77% with Logistic Regression.   Used Jupyter notebook for spark to make data manipulations.   Developed ETL processes for data conversions and construction of data warehouse using INFORMATICA.   Handled importing data from various data sources, performed transformations using Hive, Map Reduce, and loaded data into HDFS.   Interacted with the other departments to understand and identify data needs and requirements and work with other members of the IT organization to deliver data visualization and reporting solutions to address those needs.   Environment: Tableau 10.05, AWS, GIT, Python 3.5.2 , Anaconda-Navigator , Hadoop, NoSQL, random forest, MongoDB, HDFS, , NLTK, XML, MapReduce, Informatica. Data Analyst Karvy Financial Services Limited July 2011 to August 2013 Karvy Financial Services Limited is a company which has been playing a very proactive role in the economic growth of India by providing loans to Micro & Small Business segments and individuals like credit for the requirements of different sectors of economy. Industries, exports, trading, agriculture, infrastructure and the individual segments. We worked on various projects which handles customer analytics, Credit Risk analysis and assessing risks associated with loans like identify and prevent fraudulent loans, identify and prevent fraud detection for transactions. Responsibilities: Compiled data from various sources public and private databases to perform complex analysis anddata manipulation for actionable results.   Applied concepts of probability, distribution, and statistical inference on the given dataset to unearthinteresting findings using comparison, T-test, F-test, R-squared, P-value etc.   Applied linear regression, multiple regression, ordinary least square method, mean-variance, the theory of large numbers, logistic regression, dummy variable, residuals, Poisson distribution, Naive Bayes, fitting function etc to data with help of

Scikit, SciPy, NumPy and Pandas module ofPython.   Applied Principal Component Analysis (PCA) based unsupervised technique to determine unusual VPN log-on time.   Performed Clustering with historical, demographic and behavioral data as features to implement the personalized marketing to the customers.   Also created classification model using Logistic Regression, Random Forests to classify dependent variable into two classes which are risky and okay.   Used F-Score, Precision, recall evaluating model performance.   Built user behavior models for finding activity patterns and evaluating risk scores for everytransaction using historic data to train the supervised learning models such as Decision trees, Random Forests and SVM.   Real time analysis of customer financial profile and providing recommendation for financial products best suited.   Performed Sentimental analysis (NLP) on the email feedback of the customers to determine the emotional tone behind the series of words and gain the express of the attitudes and emotions by Long-Short Term Memory (LSTM) cells in Recurrent Neural Networks(RNN).   Forecasted demand for loans and interest rates using Time Series analysis like ARIMAX, VARMAX and Holt-Winters.   Obtained better predictive performance of 81% accuracy using ensemble methods like Bootstrap aggregation (Bagging) and Boosting (Light GBM, Gradient).   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.   Performed financial data ingestion to the Spark distribution environment, using Kafka.   Developed visualizations and dashboards using ggplot, Tableau. Prepared and presented data quality report to stakeholders to give understanding of data. Environment: Tableau 10.05, GIT, Python 3.5.2 , Anaconda-Navigator, Hadoop, Spark, Kafka, NoSQL, Random forest, MongoDB, HDFS, , NLP. Python Developer / Data Analyst Symbiosys Technologies - Visakhapatnam, ANDHRA PRADESH, IN May 2009 to June 2011 Genius Brands International is our client and we performed exploratory data analysis on corporate purchase orders, contracts and projects data using sampling and statistical methods. Identified strata, improved precision and accuracy. Works with other team members, including DBA's, Other ETL developers, Technical Architects, QA, and Business Analysts & Project Managers.   Responsibilities: Participated in requirement gathering and analysis phase of the project in documenting the business

requirements by conducting workshops/meetings with various business users.   Used Python to place data into JSON files for testing Django Websites.   Updated and manipulated content and files by using python scripts. Worked on Python Open stack API's.   Used Python scripts to update content in the database and manipulate files.Generated Python Django Forms to record data of online users.   Implemented end-to-end systems for Data Analytics, Data Automation and customized visualization tools using Python, R, Hadoop and MongoDB.   Used pandas, NumPy, seaborn, SciPy, matplotlib, scikit-learn in Python for developing various machine learning algorithms.   Worked on csv, json, excel different types of files for the data cleaning and data analysis.   Used R for statistical operations on the data and ggplot2 for the visualizing the data. Application of various ML algorithms and statistical modeling like decision trees, regression models, random forest , SVM, clustering to identify Volume using scikit-learn package in python.   Performed Classification using Supervised algorithms like Logistic Regression, Decision trees, KNN, Naive Bayes.   Performed data profiling to merge the data from multiple data sources.   Extracted data from HDFS (Hadoop Distributed File System) and prepared data for exploratory analysis using data munging.   Performed time series analysis using Tableau.   Knowledge of other relational database platforms such as Oracle, DB2, NoSQL.   Managed offshore projects and coordinated work for 24 hour productivity cycle.   Environment: Python 2.7, Django 1.4, R, Oracle, GitHub, SQL Server, HDFS, Hive. ETL Developer Sutherland Global Services - Hyderabad, Telangana April 2007 to April 2009 Sutherland builds processes for the digital age by combining the speed and insight of design thinking with the scale and accuracy of data analytics. Sutherland has customers across industries like financial services to Healthcare. My role is to assist Analytics department for the data extraction and cleaning as a data preprocessing steps to build models. Responsibilities: Involved with Business Analysts team in requirements gathering and in preparing functional specifications and changing them into technical specifications.   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.   Managed full SDLC processes involving requirements management, workflow analysis, source data analysis, data

mapping, metadata management, data quality, testing strategy and maintenance of the model. Developed ETLs to pull data from various sources and transform it for reporting applications using PL/SQL.   Designed SSIS packages to extract, transform and load existing data into SQL Server, used lots of components of SSIS, such as Pivot Transformation, Fuzzy Lookup, Merge, Merge Join, Data Conversion, Row Count, Sort, Derived Columns, Conditional Split, Execute SQL Task, Data Flow Task and Execute Package Task.   Created SSIS Packages that involved dealing with different source formats (flat files, Excel, XML) and different destination formats.   Debugged and troubleshot the ETL packages by using a breakpoint, analyzing the process, catching error information by SQL command in SSIS   Developed SQL queries in SQL Server management studio, Toad and generated complex reports forth end users.   Automated and scheduled recurring reporting processes using UNIX shell scripting and Teradatautilities such as MLOAD, BTEQ, and Fast Load   Experience with Perl.   Performed data analysis and data profiling using complex SQL on various sources systems includingOracle and Teradata.   Environment: ETL Tools, SDLC, GitHub, SQL Server, PL/SQL, Excel, XML, SQL.

Name: Robert Mason

Email: harrisontara@example.com

Phone: 001-218-666-8786x327