

Data Scientist Data Scientist Data Scientist - CVS Health Scottsdale, AZ Around 9+ years of IT experience includes in Data Science (Machine Learning, Deep Learning, NLP/ Text Mining), Data/Business Analytics, Data Visualization, Data Operations, and BI A deep understanding of Statistical Modelling, Multivariate Analysis, Bigdata analytics and Standard Procedures Highly efficient in Dimensionality Reduction methods such as PCA (Principal component Analysis), Factor Analysis etc. Implemented bootstrapping methods such as Random Forests (Classification), K-Means Clustering, KNN, Na ve Bayes, SVM, Decision Tree, BFS, Linear and Logistic Regression Methods Experience in text understanding, classification, pattern recognition, recommendation systems, targeting systems and ranking systems using Python Strong skills in statistical methodologies such as A/B test, Experiment design, Hypothesis test, ANOVA, CrossTabs, T tests and Correlation Techniques Worked with applications like R, SPSS and Python to develop predictive models Experience with Natural Language Processing (NLP) Extensively worked on Python 3.5/2.7 (Numpy, Pandas, Matplotlib, NLTK and Scikit-learn) Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0 Jupyter Notebook 4.X, R 3.0(ggplot2) and Excel Worked on Tableau to create dashboards and visualizations Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python, Power BI and Tableau Hands on experience in business understanding, data understanding, and preparation of large databases Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing datamining and reporting solutions that scales across massive volume of structured and unstructured data Documenting new data to help source to target mapping. Also updating the documentation for existing data assisting with data profiling to maintain data sanitation, validation Identifies what data is available and relevant, including internal and external data sources, leveraging new data collection processes such geo-location information Good understanding on data preparation techniques Experienced working on large volume of Data using BASE SAS programming Proficiency in application of statistical prediction modeling, machine learning classification techniques and econometric

forecasting techniques Proficiency in various type of optimization, Market Mix modeling, Segmentation, Time Series, Price Promo models etc. Experience in the application of Neural Network, Support Vector Machines (SVM), and Random Forest Creative thinking and propose innovative ways to look at problems by using data mining approaches on the set of information available Identifies/creates the appropriate algorithm to discover patterns, validate their findings using an experimental and iterative approach Applies advanced statistical and predictive modeling techniques to build, maintain, and improve on multiple real-time decision systems Closely works with product managers, Service development managers, and product development team in productizing the algorithms developed Experience in designing star schema, Snowflake schema for Data Warehouse, ODS architecture Experience in designing stunning visualizations using Tableau and Power BI software and publishing and presenting dashboards, Storyline on web and desktop platforms Experience in working with relational databases (Teradata, Oracle) with advanced SQL programming skills In-depth knowledge of statistical procedures that are applied in Supervised / Unsupervised problems Proficiency in SAS (Base SAS, Enterprise Guide, Enterprise Miner) Familiar with graphical models and deep learning models, including deep learning frameworks such as TensorFlow Experienced in working with advanced analytical teams to design, build, validate and refresh data models that enable the next generation of sophisticated solutions for global clients Excellent communication skills (verbal and written) to communicate with clients and team, prepare deliver effective presentations. Strong experience in Software Development Life Cycle (SDLC) including Requirements Analysis, Design Specification and Testing as per Cycle in both Waterfall and Agile methodologies. Strong experience in interacting with stakeholders/customers, gathering requirements through interviews, workshops, and existing system documentation or procedures, defining business processes, identifying and analyzing risks using appropriate templates and analysis tools. Mapping and tracing data from system to system in order to establish data hierarchy and lineage. Used Data Lineage and reverse engineering as a way to track back errors in data till the data source. Authorized to work in the US for any employer

Work Experience Data Scientist CVS Health - Scottsdale, AZ August 2018 to Present CVS Health is an American

pharmacy and healthcare company with nearly 10,000 stores in its network. Its health-focused business includes pharmacy services, retail, in-store health clinics, and its own Digital Innovation Lab aimed at creating smart devices and apps to improve healthcare. Involved in multiple projects with the primary objective of providing better services to the customers. Performed behavioral analysis, personalization and customer targeting.

Data Scientist Sherwin Williams - Minneapolis, MN
January 2017 to July 2018 Minneapolis, MN Jan 2017 - Jul 2018 Data Scientist The Sherwin-Williams Company is an American Fortune 500 company. It is a global leader in the manufacture, development, distribution, and sale of paints, coatings and related products to professional, industrial, commercial, and retail customers. The primary objective of the project is to perform predictive analytics using the production and client data to estimate the production quantities.

Responsibilities: Used the Classification machine learning algorithms Na ve Bayes, Linear regression, Logistic regression, SVM, Neural Networks and used Clustering Algorithm K Means. Analyzed business requirements and developed the applications, models, used appropriate algorithms for arriving at the required insights. Established partnerships with product and engineering teams and work closely with other teams. Work collaboratively with senior management to develop strategy and approach to defining business challenges to be answered by data science. Used unsupervised (K-means, DBSCAN) and supervised learning techniques (Regression, Classification) for feature engineering and did Principal Component Analysis for dimensionality reduction of features. Used Spark-Streaming APIs to perform necessary transformations and actions on the fly for building the common learner data model which gets the data from Kafka in near real time. Worked on data cleaning, data preparation and feature engineering with Python 3.X. Used ML toolkit - H2O for model fitting and to discover patterns in data. Used H2O for data visualization and for different machine learning algorithms. Used TensorFlow library for development and evaluation of Deep Learning Models. Performed text classification task using NLTK package and implemented various natural language processing techniques. Used Tableau for data visualization to create reports, dashboards for insights and business process improvement. Extensively used Python's multiple data science packages like

Pandas, NumPy, Matplotlib, SciPy, Scikit-learn and NLTK. Used TensorFlow Python API's to perform TensorFlow graphs. Worked on Spark Python modules for machine learning and predictive analytics in Spark on AWS. Worked on end to end pipe line in Spark. Explored and analyzed the customer specific features by using Spark SQL. Performed data imputation using Scikit-learn package in Python. Created the dashboards and reports in tableau for visualizing the data in required format. Collaborated with team members and translated functional requirements to technical requirements for development. Conducted Code Review for the Fit Gap done by the team members. Created Hive scripts to create external, internal data tables on Hive. Worked on creating datasets to load data into HIVE. Environment: Spark, Apache Spark, Hive, Machine learning, Python, Numpy, NLTK, Pandas, Scipy, SQL, Tableau, HDFS, Tableau, DynamoDB, Mongo DB, SQL Server, and ETL. Data Scientist TracFone - Miami, FL March 2015 to December 2016 TracFone Wireless, Inc. is a prepaid mobile virtual network operator in the United States, Puerto Rico, and the US Virgin Islands. The objective of the project is to analyze the Customer data, work on Churn and build recommendation engines and automated customer scoring systems.

Responsibilities: Enhancing data collection procedures to include information that is relevant for building analytic systems Processing, cleansing, and verifying the integrity of data used for analysis Doing ad-hoc analysis and presenting results in a clear manner Constant tracking of model performance Excellent understanding of machine learning techniques and algorithms, such as Logistic Regression, SVM, Random Forests, Deep Learning etc. Worked with Data governance, Data quality, Data lineage, Data architect to design various models. Independently coded new programs and designed Tables to load and test the program effectively for the given POC's. Extending company's data with third party sources of information when needed. Designed data models and data flow diagrams using Erwin and MS Visio. Developed Implemented & maintained the Conceptual, Logical & Physical Data Models using Erwin for Forward/Reverse Engineered Databases. Experience with common data science toolkits, such as R, Python, Spark, etc. Good applied statistics skills, such as statistical sampling, testing, regression, etc. Build analytic models using a variety of techniques such as logistic regression, risk scorecards and pattern recognition

technologies. Analyze and understand large amounts of data to determine suitability for use in models and then work to segment the data, create variables, build models and test those models. Work with technical and development teams to deploy models. Build Model Performance Reports and Modeling Technical Documentation to support each of the models for the product line. Developed new reports in SAS using SAS ODS, PROC REPORT, PROC TABULATE, and PROC SQL. Imported Data from relational database into SAS files per detailed specifications. Responsible for the development and maintenance of SAS Information Maps for Analytics and Business Forecasting Team. Performed Exploratory Data Analysis and Data Visualizations using R, and Tableau. Perform a proper EDA, Univariate and bi-variate analysis to understand the intrinsic effect/combined. Established Data architecture strategy, best practices, standards, and roadmaps. Lead the development and presentation of a data analytics data-hub prototype with the help of the other members of the emerging solutions team. Experience in SAS programming for auditing data, developing data, performing data validation QA and improve efficiency of SAS programs. Involved in analysis of Business requirement, Design and Development of High level and Low-level designs, Unit and Integration testing. Worked with several R packages including knitr, dplyr, SparkR, CausalInfer, spacetime. Interacted with the other departments to understand and identify data needs and requirements. Environment: UNIX, Python 3.5.2, MLLib, SAS, regression, logistic regression, Hadoop, NoSQL, Teradata, OLTP, Random forest, OLAP, HDFS, ODS Data Science Analyst Kindred Health - Louisville, KY April 2013 to February 2015 Kindred Healthcare Incorporated is a healthcare services company that operates hospitals, nursing centers, and contract rehabilitation services across the United States. The project involved in predictive analysis and trend analysis using the patient historical data. Responsibilities: Extracted, Transformed and loaded data from given source to analysis. Hands-on implementation of R, Python, Hadoop, Tableau and SAS to extract and import data. Had a pleasure working experience on Spark (spark streaming, spark SQL), Scala and Kafka. Also converting Hive/SQL queries into Spark transformations using Spark RDDs and Scala. Used Kafka to load data in to HDFS and move data into NoSQL databases. Worked on Spark SQL and Data frames for faster execution of

Hive queries using Spark Sql Context. Gained extreme knowledge on Map Reduce using Python, Hive queries using Oozie workflows. Performed Data cleaning process applied Backward - Forward filling methods on dataset for handling missing values. Design built and deployed a set of python modelling APIs for customer analytics, which integrate multiple machine learning techniques for various user behavior prediction and support multiple marketing segmentation programs.

Segmented the customers based on demographics using K-means Clustering. Explored different regression and ensemble models in machine learning to perform forecasting. Used classification techniques including Random Forest and Logistic Regression to quantify the likelihood of each user referring. Performed Boosting method on predicted model for the improve efficiency of the model.

Environment: R/R studio, Informatica, SQL/PLSQL, Oracle 10g, MS-Office, Teradata. Data Analyst SunTrust Bank - Atlanta, GA October 2011 to March 2013 Worked for Credit Card Management Team and involved in calculating the Credit Card Performance by dividing the data into different DataMart's and performing analysis on the data and credit risk calculation for the Customers and lending money (Mortgaging). Responsibilities: The team of Data analysts focused on providing analytics insights and decision support tools for executives for accurate decision making. Applied highly advanced data access routines to extract data from source systems for monitoring operations compliance to banking Laws, Rules and Regulations using Visual Basic Apps (VBA), SQL Server SSIS, SAS and SQL. Identified, measured and recommended improvement strategies for KPIs across all business areas. Assisted in defining, implementing, and utilizing business metrics calculations and methodologies. Designed and provided complex excel reports including summaries, charts, and graphs to interpret findings to team and stakeholders. Assisted the team for standardization of reports using SAS macros and SQL. Responsible for creation of Credit data related warehouse to help with Risk Assessment for Commercial loans. Performed competitor and customer analysis, risk and pricing analysis and forecasted results for credit card holders on demographical basis. Created macros and used existing macros to develop SAS programs for data analysis. Created and manipulated various management reports in MS Excel for sales metrics using VLOOKUP and Pivot tables. Developed transformation logic for BI tools

(Informatica) for data transformation into various layers in Data warehouse. Utilized SQL to develop stored procedures, views to create result sets to meet varying reporting requirements. Used advanced excel formulas (lookup functions, pivot table, If Statements etc.) for analyzing data. Identified process improvements that significantly reduce workloads or improve quality. Worked for BI Analytics team to conduct A/B testing, data extraction and exploratory analysis. Generated dashboards and presented the analysis to researchers explaining insights on the data.

Environment: Excel 2010, R, Informatica Power Center 9.0, MS SQL Server 200. Data Analyst/Python Developer Teleparadigm Networks Private Limited May 2009 to September 2011

Teleparadigm Networks delivers software and IT services to clientele across the globe under service models. Have a wide client base in Healthcare and Telecom Industries. Responsibilities: Brought in and implemented updated analytical methods such as regression modelling, classification tree, statistical tests and data visualization techniques with Python. Analyzed customer Help data, contact volumes, and other operational data in MySQL to provide insights that enable improvements to Help content and customer experience. Maintained and updated existing automated solutions. Improved data collection and distribution processes by using pandas and Numpy packages in Python while enhancing reporting capabilities to provide clear line of sight into key performance trends and metrics. Analysed historical demand, filter out outliers/exceptions, identify the most appropriate statistical forecasting algorithm, develop base plan, understand variance, propose improvement opportunities, and incorporate demand signal into forecast and executed data visualization by using plotly package in Python. Interacted with QA to develop test plans from high-level design documentation. Environment: MySQL, Statistical modelling, Python libraries, pandas and Numpy packages

Education Bachelor's in Feature Engineering and Selection Stanford

NER Tagger Skills APACHE HADOOP HDFS (5 years), HADOOP DISTRIBUTED FILE SYSTEM (5 years), HDFS (5 years), Python (7 years), SQL (6 years)

Name: Joan Reyes

Email: lknight@example.org

Phone: 001-556-343-5258x7860