

Data Scientist Data Scientist Data Scientist - CUNA Mutual Group * 7+ years of Data Science experience in architecting and building comprehensive analytical solutions in Marketing, Sales and Operations functions across Technology, Banking, Manufacturing, Healthcare and Retail industries.

- * Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python. *
- Expertise in transforming business requirements into analytical models, designing algorithms, building models, developing data mining and reporting solutions that scale across a massive volume of structured and unstructured data. *
- Expert knowledge in supervised and unsupervised learning algorithms such as Ensemble Methods (Random forests), Logistic Regression, Regularized Linear Regression, SVMs, Deep Neural Networks, Extreme Gradient Boosting, Decision Trees, KMeans, Gaussian Mixture Models, Hierarchical models, and time series models (ARIMA, GARCH, VARCH etc.) *
- Expertise writing production quality code in SQL, R, Python and Spark. Hands on experience building regression and classification models and other unsupervised learning algorithms with large datasets in distributed systems and resource constrained environments. *
- Familiar with predictive models using classification algorithms like KNN, Naive base, regression and decision trees. *
- 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. *
- Worked on Text Mining and Sentimental analysis for extracting the unstructured data from various social Media platforms like Facebook, Twitter and Reddit. *
- Ability to translate analytic ideas into R/ Python production quality scripts *
- Strong background & hands-on in Data Science, Big Data, data structures, statistics, algorithms like Regression, Classification etc. *
- Strong background & hands-on of Supervised learning (Decision Trees, Random Forest, Logistic Regression, SVMs, GBM, etc) and unsupervised learning (K-Means, KNN). *
- Strong background & hands-on of Deep learning using Tensforflow, Keras, Theano, H2o. *
- Sound understanding of Deep learning using CNN, RNN, ANN, reinforcement learning, transfer learning. *
- Strong background & hands-on in Natural Language Processing and text analytics. *
- Experience and passion for solving analytical problems involving big data sets using

quantitative approaches to generate insights from data. * Identify, analyze, and interpret trends or patterns in complex data sets. Strong in Predictive and Prescriptive analytics approaches and experienced in operating tools like R and in programming using Python. * Has working experience in Data Science, Machine Learning implementation in cloud platforms like Google Cloud Platform, AWS, Azure, Bluemix. * Familiar with predictive models using different cloud based Machine learning tools like Microsoft Azure ML. * Oversee all activities related to data cleansing, data quality and data consolidation using industry standards and processes. * Perform exploratory data analysis; generate and test working hypothesis; and, prepare and analyze historical data and identify patterns * Work closely with Cross Functional teams to encourage best practices for experimental design and data analysis. * Proficient with Python 3.x including Numpy, Scikit-learn, Pandas, Matplotlib and Seaborn. * Extensive experience in RDBMS such as SQL server 2012, MySQL 5.x. * Experienced in Non-relational database such as MongoDB 3.x. * Experienced in Hadoop 2.x ecosystem and Apache Spark 2.x framework such as Hive, Pig, Pyspark. * Proficient at data visualization tools such as Tableau, Power BI, Python Matplotlib and Seaborn. * Experienced in Amazon Web Services (AWS) and Microsoft Azure, such as AWS EC2, S3, RD3, Azure HDInsight, Machine Learning Studio, Azure Data Lake. * Apply various data modeling techniques to model user behavior and identify actionable levers for retaining and growing users. * Define key metrics, conduct A/B testing, and oversee statistical measurement of new algorithms and approaches. * Expert at distilling questions, wrangling data, and driving decisions with data analytics. * Strong knowledge of relational databases and ability to write SQL code at an expert level. * Aptitude with numbers, intellectual curiosity about metrics and measuring impact. * Present models, findings and insights to senior management to catalyze business decisions. * Ability to solve complex problems by applying analytical techniques and predictive models to massive data sets, and translate business needs into mathematical abstractions for algorithms to solve. * Research and prototype models and pipelines, and work with engineers to put them into production at scale. * Raw data analysis such as assessing quality, cleansing, structuring for downstream processing. * Use mathematical, statistical, and programmatic knowledge to spec out, design, and

build first-class predictive models about customer behavior. * Design and prototype of accurate and scalable prediction algorithms. * Collaboration with engineering team to bring analytical prototypes to production. * Creative and pragmatic quantitatively minded individual with a passion for understanding location and human behavior. * Ability to identify issues quickly and rapidly determine root cause and effective resolution approach. * Very solid data analysis skills including application of analytical techniques such as statistical and machine learning. * Fundamental coding skills enabling the building of analytical pipelines and the development of prototypes for core company products and systems. * Excellent verbal and written communication skills, ability to communicate technical topics to non-technical individuals. * Ability to manage own time and work effectively with others on projects.

Work Experience Data Scientist CUNA Mutual Group - Madison, WI April 2017 to Present

This project was to support auditing team and claim department to improve accounting accuracy and reduce risk of fraudulent activities via providing machine learning and modeling solutions to identify suspicious insurance claims.

Claims severity prediction in real-time
Built classification models to predict the fraudulent claims by severity in real-time reducing the time for execution from 6 hours to 4 seconds. Implemented the models as a predictive solution for finding the fraudulent claims for the credit disability and debt protection products. Forwarded the high-risk claims for further investigation.

Text analytics for fraud prediction
Executed topic modelling for finding different topics based on the notes made corresponding to the claimant's claim. Attributed the resulting topics to classify into fraud and not fraud categories.

Risk assessment prediction
Incorporated models built in Python and R into the business processes using clustering techniques to assess the risk involved with a customer. The models built are used in assessing the premiums amount required to be paid by the customer.

Customer churn/attrition prediction
Developed models that predict whether a customer's propensity to churn leveraging the information related to insurance policies, demographics, claims, related to the customer, payment frequency, home ownership status, household tenure etc.

Responsibilities

- * Analyze Data and Performed Data Preparation by applying historical model on the data set in AZURE ML.
- * Perform Data cleaning process applied Backward - Forward filling methods on dataset for handling missing value
- *

Perform Data Transformation method for Rescaling and Normalizing Variables. * Develop a predictive model and validate KNN model for predict the feature label. * Plan, develop, and apply leading-edge analytic and quantitative tools and modeling techniques to help clients gain insights and improve decision-making. * Leverage the most appropriate algorithms and be prepared to justify your decisions. * Work closely with key stakeholders in product, finance and operations to form deep understanding of growth and marketplace dynamics, including product and pricing patterns, outlier detection, forecasting, and imputation. * Collaborate with product and engineering to integrate various sources of data. * Apply strict sampling, statistical inference, and survey techniques to derive insights from small samples of data. * Utilize Sqoop to ingest real-time data. Used analytics libraries Sci-Kit Learn, MLLIB and MLxtend. * Extensively use Python's multiple data science packages like Pandas, NumPy, matplotlib, Seaborn, SciPy, Scikit-learn and NLTK. * Performed Exploratory Data Analysis, trying to find trends and clusters. * Develop rigorous data science models to aggregate inconsistent real-time signals into strong predictors of market trends. * Automate and own the end-to-end process of modeling and data visualization. * Collaborate with Data Engineers and Software Developers to develop experiments and deploy solutions to production. * Work on data that was a combination of unstructured and structured data from multiple sources and automate the cleaning using Python scripts. * Extensively perform large data read/writes to and from csv and excel files using pandas. * Tasked with maintaining RDD's using SparkSQL. * Communicate and coordinate with other departments to collection business requirement. * Tackle highly imbalanced Fraud dataset using under sampling with ensemble methods, oversampling and cost sensitive algorithms. * Improved fraud prediction performance by using random forest and gradient boosting for feature selection with Python Scikit-learn. * Implemented machine learning model (logistic regression, XGboost) with Python Scikit-learn. * Optimize algorithm with stochastic gradient descent algorithm Fine-tuned the algorithm parameter with manual tuning and automated tuning such as Bayesian Optimization. * Write research reports describing the experiment conducted, results, and findings and also make strategic recommendations to technology, product, and senior management. Built executive Banco

Santander, S.A 2000 to Present doing business as Santander Group, is a Spanish banking group. As its name suggests, the company originated in Santander, Spain. The group has expanded since 2000 through a number of acquisitions, with operations across Europe, South America, North America and Asia. Santander has been ranked as 37th in the Forbes Global 2000 list of the world's biggest public companies. Santander is Spain's largest bank.

KEY PROJECTS

Credit History Predictive Modeling: Analyzed and predicted the customer's credit history and past bill payments based on the Credit card offers, to create a predictive model, and send offers to customers on the base of Model and past data. A system was successfully created on the past data of Credit History and Payments activity of customers; Ran model against the historical data and get predicted label if customers were eligible for credit card offer, and on that basis send them an offer. As a result, customers actually got an offer they liked and increased the number of offer acceptance, which lead to profit for the bank.

Forecasting Loan balance: Forecasted bank-wide loan balances under normal and stressed macroeconomic scenarios using R. Performed variable reduction using the stepwise, lasso, and elastic net algorithms and tuned the models for accuracy using cross validation and grid search techniques.

Top down Models (Commercial Real Estate): Automated the scraping and cleaning of data from various data sources in R and Python. Developed Bank's loss forecasting process using relevant forecasting and regression algorithms in R. The projected losses under stress conditions helped bank reserve enough funds per DFAST policies.

Loan Payment Default Prediction: Built classification models using several features related to customer demographics, macroeconomic dynamics, historic payment behavior, type and size of loan, credit scores and loan to value ratios and with accuracy of 95% accuracy the model predicted the likelihood of default under various stressed conditions.

Marketing Campaign Measurement: Built executive dashboards in Tableau that measured changes in customer behavior post campaign launch; the ROI measurements helped to strategically select the effective campaigns.

Credit Risk Scorecards: Built credit risk scorecards and marketing response models using SQL and SAS. Evangelized the complex technical analysis into easily digestible reports for top executives in the company. Developed several interactive dashboards in Tableau to visualize nearly 5 Terabytes of credit data

by designing a scalable data cube structure. Responsibilities

- * Gathered, analyzed, documented and translated application requirements into data models, supported standardization of documentation and the adoption of standards and practices related to data and applications.
- * Queried and aggregated data from Amazon Redshift to get the sample dataset.
- * Identified patterns, data quality issues, and leveraged insights by communicating with BI team.
- * In preprocessing phase, used Pandas to remove or replace all the missing data, and feature engineering to eliminate unrelated features.
- * Balanced the dataset with Over-sampling the minority label class and Under-sampling the majority label class.
- * In data exploration stage used correlation analysis and graphical techniques to get some insights about the claim data.
- * Tested classification algorithms such as Logistic Regression, Gradient Boosting and Random Forest using Pandas and Scikit-learn and evaluated the performance.
- * Implemented, tuned and tested the model on AWS EC2 with the best algorithm and parameters.
- * Set up data preprocessing pipeline to guarantee the consistency between the training data and new coming data.
- * Deployed the model on AWS Lambda, collaborated with develop team to build the business solutions.
- * Collected the feedback after deployment retrained the model to improve the performance.
- * Discovered flaws in the methodology being used to calculate weather peril zone relativities; designed and implemented a 3D algorithm based on k-means clustering and Monte Carlo methods.
- * Observed groups of customers being neglected by the pricing algorithm; used hierarchical clustering to improve customer segmentation and increase profits by 6%.
- * Designed, developed and maintained daily and monthly summary, trending and benchmark reports in Tableau Desktop.

Environment: AWS EC2, S3, Redshift, Lambda, Linux, Python (Scikit-Learn/Numpy/Pandas/Matplotlib), Machine Learning (Logistic Regression/Gradient Boosting/Random Forest), Tableau.

Data Scientist Santander - Boston, MA March 2016 to April 2017

Data Scientist Banco Santander, S.A - Austin, TX March 2014 to December 2015

Whole Foods Market Inc. is an American supermarket chain that specializes in selling organic foods products without artificial additive products for growing foods, colors, flavors, sweeteners, and hydrogenated fats. It has 473 stores in North America and the United Kingdom.

KEY PROJECTS Customer Purchase Propensity Modelling: Built machine learning based

regression models using scikit-learn python frameworks to estimate the customer propensity to purchase based on attributes such as customer verticals they operate in, revenue, historic purchases, frequency and regency behaviours. These predictions helped estimate propensities with higher accuracy improving the overall productivity of sales teams by accurately targeting the prospective clients.

Coupon Recommender System: Developed a personalized coupon recommender system using recommender algorithms (collaborative filtering, low rank matrix factorization) that recommended best offers to a user based on similar user profiles. The recommendations enabled users to engage better and helped improving the overall customer retention rates.

Outlier /Anomalous Pattern Detection: Created interactive dashboard suite that illustrated outlier characteristics across several sales-related dimensions and overall impact of outlier imputation in R (Shiny). Used iterative outlier detection and imputation algorithm using multiple density-based clustering techniques (DBSCAN, kernel density estimation).

Cross Sell and Upsell Opportunity Analysis: Implemented market basket algorithms from transactional data, which helped identify coupons used/purchased together frequently. Discovering frequent coupon sets helped unearth cross sell and up selling opportunities and led to better pricing, bundling and promotion strategies for sales and marketing teams.

Forecast Process Innovations: Forecast Sales and improved accuracy by 10-20% by implementing advanced forecasting algorithms that were effective in detecting seasonality and trends in the patterns in addition to incorporating exogenous covariates. Increased accuracy helped business plan better with respect to budgeting and sales and operations planning.

Price Elasticity Analysis: Measured the price elasticity for products that experienced price cuts and promotions using regression methods; based on the elasticity, Whole Foods made selective and cautious price cuts for certain categories.

Customer Churn Prediction: Predicted the likelihood of customer churn based on customer attributes like customer size, RFM loyalty metrics, revenue, type of industry, competitor products and growth rates etc. The models deployed in production environment helped detect churn in advance and aided sales/marketing teams plan for various retention strategies in advance like price discounts, custom licensing plans etc.

Responsibilities: * Working as project technical lead, paired with business

lead in scoping, researching, and assessing project feasibility, outcomes, and product deliverables on knowledge graph project (Hadoop ecosystem, Spark, ElasticSearch, JanusGraph, etc.). * Working with business lead, successfully launched end-to-end knowledge graph MVP in 7 months' time whereas previous attempts at knowledge graph project had languished for 3-4 years. * Using NLP and ML built and maintained a variety of cloud-based natural language web scrapers and parsers to augment existing data sources: Python (NumPy, SciPy, SpaCy, etc.), Spark, TensorFlow, Jupyter notebooks. * Mentored junior teammates, shared knowledge of NLP and information retrieval best practices, and performed code reviews on a weekly basis. * Prototyped, conducted, and reported on data science experiments (to technical and non-technical audiences) using supervised, semi-supervised and unsupervised learning techniques: anomaly detection, named entity recognition, ontology creation, etc. * Using Python and NLTK, aggregated natural language data from online documents and developed pipelines to ingest scraped data into relational databases. * Performed ad-hoc data science analyses for departments across the business using R, Python, Natural Language Toolkit (NLTK), machine learning.

Analyst/Data Scientist Banco Santander, S.A - Hyderabad, Telangana April 2012 to March 2014

Bharti AXA General Insurance Company Ltd is a joint venture between Bharti Enterprises, a leading Indian business group and AXA, a world leader in financial protection. Bharti AXAs offers insurance coverage across various categories - Motor, Health , Travel , Home, student travel and more. The objective was to load data, analyze, and provide monthly reports for the predictions on a claim's potential of a third-party recovery. Tableau and SSRS were used to build claim and recovery reports.

Responsibilities: * Assembled a Predictive Modeling module by using supervised learning for Subrogation Claim Prediction to identify which claims would be classified as having Subrogation potential. * Implemented models such as Logistic Regression and Na ve Bayes, in Python using scikit-learn, to predict the claim potential outcome. * Dimensionality Reduction techniques applied to refine the attribute lists and feature selection applied to rank selected features to generate accurate results. * Gathered requirements and business rules from business users to implement Predictive Modeling. * Designed and developed ETL packages using SSIS to create Data Warehouses from different

tables and file sources like Flat and Excel files, with different methods in SSIS such as derived columns, aggregations, Merge joins, count, conditional split and more to transform the data. *

Designed reporting solutions for different stakeholders from mock-up till deployment in different areas such as Potential Subrogation claims, Monthly Revenue from Subrogation & Transactions. *

Performed data visualization and designed dashboards with Tableau, and provided complex reports, including charts, summaries, and graphs to interpret the findings for Adjustors to view various claim information. *

Optimized queries in T-SQL by removing unnecessary columns and redundant data, normalized tables, established joins and indices; developed complex SQL queries, stored procedures, views, functions and reports that meet customer requirements. Environment: Python 3.x (Scikit-learn, Matplotlib), Jupyter, SQL Server 2012, MS SQL Server Management Studio, MS BI Suite (SSIS/SSRS), T-SQL, Visual Studio BIDS, Tableau. Jr. Python Developer Banco Santander, S.A - Hyderabad, Telangana June 2010 to April 2012 Responsibilities: *

Worked on development of customer support and complains registration system. This is a Customer feedback and complains management system. *

Design, develop, test, deploy and maintain the website. *

Coding and execution of scripts in Python/Unix/VB. *

Development of Application using Java and Python. *

Recording of Scripts (Web, Web Services HTML) using Vugen and SoapUI and script validation through co correlations, parameterizations and other methods. Scripting- web and web services. *

Data set up using SQL/ORACLE/Teradata. *

Resolving Complexity in the scripts of the website due to the complex logic and correlations. *

Script validation sometimes becomes challenging as it demanded many web based logic rather than correlation and parameterization. *

Running load/endurance tests using Vugen, ALM and controller, server monitoring, analysis using Dynatrace, UNIX putty, SQL logs and other tools and reporting the performance. Analyzing errors and exceptions using putty logs (UNIX), etc. *

Testing in citrix protocol with scripts and scenario. *

Execution of batch jobs in Control M, Perfmon and other tools. *

Scripting and validation of scripts through correlation, parameterization and web based logic testing (Smoke test, Load test, Endurance) using Controller for a duration further analysis, checking response times, CPU utilizations, memory leaks of servers and other performance characteristics of the website through

capturing Perfmon logs and creating PAPAL reports and creating test reports. * Designed and developed data management system using MySQL. * Rewrite existing Python/Django/Java module to deliver certain format of data. * Used Django Database API's to access database objects. * Wrote python scripts to parse XML documents and load the data in database. * Generated property list for every application dynamically using python. * Responsible for search engine optimization (SEO) to improve the visibility of the website. * Handled all the client side validation using JavaScript. * Creating unit test/regression test framework for working/new code. * Using Subversion version control tool to coordinate team-development. * Responsible for debugging and troubleshooting the web application. Environment: Python, Putty, SQL, Teradata, SoapUI,

ControlM, PerfMon, MySQL, Linux, HTML, XHTML, CSS, AJAX, JavaScript, Apache Web Server.

Education Bachelor of Technology in Computer Science & Engineering Sreenidhi Institute of Science & Technology - Hyderabad, Telangana Skills ALGORITHMS (10+ years), BI (10+ years), BUSINESS INTELLIGENCE (10+ years), LINUX (10+ years), LOGISTIC REGRESSION (10+ years)

Additional Information TECHNICAL SKILLS Languages Java 8, Python, R Python and R Numpy, SciPy, Pandas, Scikit-learn, Matplotlib, Seaborn, ggplot2, caret, dplyr, purrr, readxl, tidyr, Rweka, gmodels, RCurl, C50, twitter, NLP, Reshape2, rjson, plyr, Beautiful Soup, Rpy2 Algorithms Kernel

Density Estimation and Non-parametric Bayes Classifier, K-Means, Linear Regression, Neighbors (Nearest, Farthest, Range, k, Classification), Non-Negative Matrix Factorization, Dimensionality Reduction, Decision Tree, Gaussian Processes, Logistic Regression, Na ve Bayes, Random

Forest, Ridge Regression, Matrix Factorization/SVD NLP/Machine Learning/Deep Learning LDA (Latent Dirichlet Allocation), NLTK, Apache OpenNLP, Stanford NLP, Sentiment Analysis, SVMs, ANN, RNN, CNN, TensorFlow, MXNet, Caffe, H2O, Keras, PyTorch, Theano, Azure ML Cloud

Google Cloud Platform, AWS, Azure, Bluemix Web Technologies JDBC, HTML5, DHTML and XML, CSS3, Web Services, WSDL Data Modelling Tools Erwin r 9.6, 9.5, 9.1, 8.x, Rational Rose, ER/Studio, MS Visio, SAP Power designer Big Data Technologies Hadoop, Hive, HDFS,

MapReduce, Pig, Kafka Databases SQL, Hive, Impala, Pig, Spark SQL, Databases SQL-Server, My SQL, MS Access, HDFS, HBase, Teradata, Netezza, MongoDB, Cassandra. Reporting Tools

MS Office (Word/Excel/Power Point/ Visio), Tableau, Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x, Cognos7.0/6.0. ETL Tools Informatica Power Centre, SSIS. Version Control Tools SVM, GitHub BI Tools Tableau, Tableau Server, Tableau Reader, SAP Business Objects, OBIEE, QlikView, SAP Business Intelligence, Amazon Redshift, or Azure Data Warehouse Operating System Windows, Linux, Unix, Macintosh HD, Red Hat

Name: Mary Schultz

Email: knappnicholas@example.org

Phone: 001-993-808-2539x7386