Lokesh Babu

Quincy, MA - Email me on Indeed: indeed.com/r/Lokesh-Babu/e07a9bb0e50e74ee

* Experience in designing, developing, testing and maintaining the applications by using SAS, R and Python in windows and UNIX environments.
* Prepared scripts to ensure proper data access, manipulation and reporting functions with R/Python programming language.
* Experience in creating user friendly automated SAS/R projects.
* Expertise on optimization by writing SQL queries using SAS/SQL.
* Trained and certified from Indian Institute of Technology (IIT), Kharaghpur on advanced Statistical Modeling for Data Analysis.
* Experience in building the POC's, RFP's and Pilot projects.
* Excellent program management skills to ensure comprehensive accountability on customer expectation management.
* Experience in leading entire gamut of analytics projects including defining scope, estimating effort & timeline, developing project road-map; driving timely project delivery, tracking project progress, highlighting risks and driving mitigation.
* Fluent in translating high-level business goals into technical specs and communicating low-level technical concerns about implementation back to business leaders.
* Experience in leading a team (Onsite and Offshore) and delivering projects with working on multidisciplinary teams and also capable of completing independent tasks.
* Rich experience in relationship management activities in coordination with Business, Development, Testing and Production support teams to achieve qualitative deliverables on time.
* Developing technical specifications which include approach, delivery stages, methodology and required tools for all business queries.
* Developing and maintaining documentation to describe program stages, logic, coding, validation and maintenance.
* Experience in working with the projects in agile methodology.
* Designing Templates/Dashboards according to the objective of the study.
* Experienced in delivering the projects with Quality Assurance (QA) standards.
* Experience on Manthan ARC Merchandise Analytics for Retail Analytics.

WORK EXPERIENCE

Lead Data Scientist

State Street Corporation - April 2017 to Present

Project: Access Management Mining (Role & Application) - The objective of this study is to define rules for the users with respective to their division, unit and group. So that state street can control the license access to their internal and external applications.

Responsibilities:
* Consider the latest one year data from different applications and users of State Street.
* Executed Explorative Data Analysis (EDA) to understand the behaviour of the data.
* Identified few interesting insights from data.
* Applied association rule mining algorithm for identifying the rules.
* Now, started implementing these rules unit to unit.

Environment: Python, R Studio, MS-OFFICE and Windows 2010.

R / Data Scientist

Fidelity Investments - April 2016 to March 2017

Project: Employer Recommendation Engine - The objective of the recommendation engine is to identify the highly recommending insurance products for the employers based on their selection. This includes the attributes like age of the company, # employees (part time + full time), employee's salary bucket, employee age bucket, industry of the employer and state of the employer .etc. Frontend of this application was developed in java which connects to R studio server pro for extracting the relevant insurance products based on the decision tree model.

Responsibilities:

* Gathering the survey data from different sources to R environment.
* Installed required R packages from CRAN to carry out the analysis and it include CARET, PARTYKIT, SQLDF and RESHAPE2.
* Created new columns based on the existing columns.
* Divided the data into train and test samples.
* Executed Explorative Data Analysis (EDA) to understand the behaviour of the data.
* Calculated Weight of Evidence (WoE) for each of the employer industry in association with insurance plan.
* Implemented ctree algorithm of Decision trees and Random Forest.
* Identified the accuracy of each of the target variable. Identified high recommended products based on high accuracy.
* Validated the accuracy of the model in comparison with the results of test data.

Environment: R Studio Server Pro, MS-OFFICE and Windows 2010.

SAS / Data Scientist

Cisco - April 2015 to March 2016

Project: Target Price Forecasting - The objective is to forecasting the price of the Manufacturing Part Number (MPN) for the next 4 quarters based on the historical data. There are more than 1.6M unique MPN's. So we divide the data into 2 parts based on their price (with change and without change in price) over quarter to quarter. Then implemented separated forecasting techniques for these two categories of data.

Responsibilities:

* Extracting the data from SAP HANA to R environment.
* Created new columns based on the existing columns.
* Divided the data into train and test samples.
* Installed required R packages from CRAN to carry out the analysis and it include READR, PLYR, DPLYR, TIDYR, TIMESERIES, FORECAST, METRICS, STRINGR, DATA.TABLE, GGFORTIFY and RESHAPE.
* Executed Explorative Data Analysis (EDA) to understand the behaviour of the data.
* Performed missing value treatment for the previous/later quarters missing data.
* Implemented different forecasting methods like ETS, AUTO ARIMA, STL, MEANF, NAÏVE, SNAIVE and RWF.
* Identifying best model (out of 7) for each of the MPN based on MAPE value.
* Based on this best method (out of 7) for each MPN, forecasted for the next four quarters.

Environment: R Studio, SAP HANA, MS-OFFICE and Windows 2010.

Data scientist

Visa - August 2012 to March 2015

Project#1: X - Border Propensity Models - The objective of this model is to identify whether a customer will participate or not given an opportunity to participate in X - border travel on a given time. Build the individual models with respective to their type transactions like Face-to-Face (F2F), Electronic (ECI) and CASH. This model help the business to target right customers based on their transaction type.

Responsibilities:
* Extracting the relevant data from SQL server to the SAS environment.
* Created a final dataset from customer dimension and transaction tables.
* Developed macro programs to clean, validate and update data to perform data analysis.
* Modified existing datasets using set, merge, sort, formats, functions and conditional statements.
* Created multiple new variables from the existing data based on their transactional data.
* Divided the data into train and test samples.
* Implemented logistic regression algorithm and measured the model accuracy.
* Validated model accuracy with test dataset and scored the new customers.
* Automated these models and delivered to the business with complete documentation.

Environment: SAS Enterprise Guide, SAS Enterprise Miner, MS-OFFICE and Windows 2007.

Project#2: Travel Segmentation - The objective of this study is to group the customer based on their travel attributes. Business wants to know key conditions out of data for segmentation. So that management can use these conditions for their future strategies.

Responsibilities:
* Extracting the relevant data from SQL server to the SAS environment.
* Created a final dataset from customer dimension and transaction tables.
* Developed macro programs to clean, validate and update data to perform data analysis.
* Modified existing datasets using set, merge, sort, formats, functions and conditional statements.
* Divided the time window into observation and performance window.
* Created multiple new variables from the existing data based on their transactional data.
* With the help of factor analysis reduced these columns into required factors.
* Divided the data into train and test samples.
* Implemented Decision trees instead of Cluster analysis to identify conditions.
* Validated these conditions on the test data.
* Presented the segmentation conditions to the business.

Environment: SAS Enterprise Guide, SAS Enterprise Miner, MS-OFFICE and Windows 2007.

Project#3: Monthly Dashboard - Developed a dashboard that will help the business to see the usage of the cards (Credit & Debit) on monthly basis w.r.to their card categories like Signature, Platinum, Gold and Silver etc. This dashboard also provides the activation status of newly issued cards. With this dashboard business can easily track the history over latest one year.

Responsibilities:
* Automated the below entire process after multiple tests. So those end users simply change the month end date and it generates complete report.
* Created data extraction SAS codes from different sources.
* Summarised the data at card level and then calculated new variables.
* Writing the data validation SAS codes with the help of Univariate, Frequency procedures.
* Performed detailed profiling at category of card level.

* Submit all developed codes and test results to QA team.
* Validated the dashboard with different samples and developed a complete documentation.

Environment: SAS Enterprise Guide, SAS Enterprise Miner, MS-OFFICE and Windows 2007.

Data scientist

Deutsche Bank - August 2010 to July 2012

Project: DBArtos (Aligned Risk Technology Reporting System) - Supports the DBArtos application, which is the central credit risk data management, calculation and reporting system which provides all Deutsche Bank credit risk reports for internal management and external/regulatory purpose. The system is based on an Oracle data management layer, SAS calculation and a Java client weblogic J2EE/ Micro strategy reporting layer. Data preparation and calculations for the standard risk costs are implemented using SAS ETL. Transaction data is loaded into Teradata database.

Responsibilities:
* Preparation of monthly production environment.
* Testing and Bug fixing of SAS programs in Production.
* Analysis of the Enhancement & Maintenance requests.
* Prepared Design documentation.
* Extensively used ETL methodology for supporting Data Extraction, Transformations and Loading processing to design ETL solutions.
* Developing SAS programs that supports for STE (Stress Test Environment) environment.
* Prepared of Unit & System Test cases.
* Working major enhancements on UNIX shell scripts for SAS batch programs.
* Also provided L2 and L3 support.

Environment: SAS/BASE, SAS/MACROS, SAS/CONNECT, ORACLE, and UNIX.

Pilot project

Sobeys - April 2008 to August 2010

Apr'08 - Aug'10) Project#1: PCA (Post Campaign Analysis) Template - The objective of this study is to analyze the campaign effectiveness by measuring the behavior of customers in 3P's. Measuring the performance of the customer behavior before promotion, during promotion and after promotion. Finally, measuring the ROI (Return on Investment) of the campaign. This campaign template consists of both Target and Mass campaign.

Responsibilities:
* Business provided their requirements in a business brief document.
* Developed a technical brief based on the business brief. This contains detailed steps and stages of developing and delivering the project including timelines.
* After sign-off from the client on technical brief then will start developing the SAS codes.
* Writing the data extraction SAS codes from different sources.
* Creating/summarizing data and then calculating new variables.
* Summarising the data at customer level by joining the datasets of customer transaction, dimension and from 3rd party sources.
* Separately calculated the KPIs for Target and Mass campaigns at pre-promo-post periods with respective to their transactions, spend and visits.
* Also measured the KPIs at MoM (Month on Month), QoQ (Quarter on Quarter) and YoY (Year on Year) with respect to pre-promo-post.
* Measured the ROI based on the differences pre-promo-post KPIs.

* Extensively used the SAS procedures like IMPORT, EXPORT, SORT, FREQ, MEANS, FORMAT, APPEND, UNIVARIATE, DATASETS and REPORT.
* Also responsible Data audit, QA of SAS code/projects and sense check of results.
* Demonstrate presentations and make recommendations to the senior management.

Environment: SAS Enterprise Guide, SAS/MACROS, SAS/ACCESS, SAS/STAT, SAS/SQL, ORACLE, Teradata, MS-OFFICE and Windows XP professional.

Project#2: Free Product Promotion - The objective of this study is to promote a product for free on spend of certain amount (say $60) in a single transaction. This free product will change from week to week and informed to customers well in an advance via mails/flyers. So that the customers can purchase other products and avail this product for free. Also targeting the customers who can spend near to this amount (say $60) will increase spend up to $60 to avail this product for free? With this strategy client gained very good ROI and planned better inventory.

Responsibilities:
* Business provided their requirements in a business brief document.
* Developed a technical brief based on the business brief. This contains detailed steps and stages of developing and delivering the project including timelines.
* After sign-off from the client on technical brief then will start developing the SAS codes.
* Writing the data extraction SAS codes from different sources.
* Creating/summarizing data and then calculating new variables.
* Summarising the data at customer level by joining the datasets of customer transaction, dimension and from 3rd party sources.
* Calculated the relevant KPIs that supports to identify free product related transactions or customers.
* Extensively used the SAS procedures like IMPORT, EXPORT, SORT, FREQ, MEANS, FORMAT, APPEND, UNIVARIATE, DATASETS and REPORT.
* Generated reports in MS-Excel.
* Also recommended possible use cases to the business.

Environment: SAS Enterprise Guide, SAS/MACROS, SAS/ACCESS, SAS/STAT, SAS/SQL, ORACLE, Teradata, MS-OFFICE, Windows XP professional and Excel VBA.

Project#3: RFM Segmentation - The objective of this model is to segment the customers based on their purchasing patterns. Segment the customers based on their key patterns of Recency, Frequency and Monitory. This segmentation model will help the managers to identify most profitable customers for business and used these insights for better marketing decisions.

Responsibilities:
* Business provided their requirements in a business brief document.
* Developed a technical brief based on the business brief. This contains detailed steps and stages of developing and delivering the project including timelines.
* Prepared sample data for analysis (Normalized, coded, Integrity checked) - Pilot project.
* Submitted the sample results to business along with the technical brief.
* After sign-off from the client on technical brief then will start developing the SAS codes.
* Writing the data extraction SAS codes from different sources.
* Creating/summarizing data and then calculating new variables.
* Summarising the data at customer level by joining the datasets of customer transaction, dimension and from 3rd party sources.

* Writing the data validation SAS codes with the help of Univariate, Frequency procedures.
* Dividing the data into train and test data sets.
* Standardised the data with the help of PROC STANDARD.
* Implemented cluster analysis (PROC CLUSTER and PROC FASTCLUS) iteratively.
* Validated and compared the results with test data results.
* Performed detailed profiling on each cluster and then assigned names to each cluster.
* Submit all developed codes and test results to QA team.
* Prepare presentation to business on final results along with the solutions.

Environment: SAS Enterprise Guide, SAS/MACROS, SAS/ACCESS, SAS/STAT, SAS/SQL, ORACLE, MS-OFFICE and Windows XP professional.

Data scientist

ICFAI University - April 2007 to March 2008

Project: Students Satisfaction Survey - The objective of this study is to measure the satisfaction level of the student with respect to their curriculum, faculty, facilities, campus and recruitment. It is continues process (on quarterly) across all branches of ICFAI. Also we will measure the satisfaction across the regions. So based on these results management will take necessary actions to improve the process.

Responsibilities:
* Assisting in designing and development of the questionnaire.
* Designing Templates in accordance with the objective of the study.
* Editing, Coding and Tabulation of the collected raw data.
* Creating the SAS data sets from the raw data with the help of INFILE statement.
* Modified existing datasets using set, merge, formats, functions and conditional statements.
* Used options like DROP, KEEP and RENAME to work with some variables.
* Sorted, filtered and joined tables based on the requirement.
* Extensively used the SAS procedures like IMPORT, EXPORT, SORT, FREQ, MEANS, FORMAT, APPEND, UNIVARIATE, DATASETS and REPORT.
* Latter developed macro programs to clean validate and update data to perform data analysis and to generate reports.
* Calculated composite scores for each parameter. Based on this score management will take necessary actions.
* Also handled the studies related to Feasibility, Evaluation and Tracking in primary market research industry.

Environment: SAS/BASE, SAS/MACROS, SAS/STAT, SAS/SQL, MS-OFFICE and Windows XP.

(Customer Service Quality Assessment) Model

ICICI Bank - June 2006 to March 2007

The objective of this model is to assess the quality of service providing by the bank branches to their customer. This includes the service areas like ATM service, cheque service, staff support, services on new/old products, other offerings and facilities in a branch. In this competitive world bankers want to gain maximum attention from the customers. So based on these results management will take necessary actions to improve their service quality.

Responsibilities:
* Assisting in designing and development of the questionnaire.
* Designing Templates in accordance with the objective of the study.
* Creating the SAS data sets from the raw data with the help of INFILE statement.

* Modified existing datasets using set, merge, formats, functions and conditional statements.
* Sorted, filtered and joined tables based on the requirement.
* Extensively used the SAS procedures like IMPORT, EXPORT, SORT, FREQ, MEANS, FORMAT, APPEND, UNIVARIATE and CLUSTER.
* Implemented Need-Gap analysis technique to know the lag in service.
* Calculated composite scores for each parameter.
* Latter developed an end to end automated process to calculate the scores at each parameter.
* Also implemented segmentation among the branches with the help of cluster analysis (PROC CLUSTER).

Environment: SAS/BASE, SAS/MACROS, SAS/STAT, SAS/SQL, MS-OFFICE and Windows XP.

EDUCATION

Masters in Computer Science & Statistics in Computer Science & Statistics
Andhra University

Bachelors in Computer Science in Statistics & Mathematics
Andhra University

SKILLS

Data Analysis (5 years), MARKETING ANALYSIS (5 years), SAS (8 years), Serial Attached SCSI (8 years), SQL (6 years), Data Scientist (10+ years), R (5 years), Python (5 years)

ADDITIONAL INFORMATION

CORE COMPETENCIES

* Domains: Market Research (MR), Retail (FMCG/CPG), Supply Chain Management (SCM), Manufacturing, e-Commerce, Telecommunications, Banking & Finance (Cards & STE) and Automobile.
* SAS Modules: SAS/Base, SAS/Macros, SAS/SQL, SAS/Stat, SAS/Enterprise Guide (EG) and SAS/ Enterprise Miner (EM).
* R Modules: R Programming, R Studio and R Studio Server Pro.
* Python: Pandas, Jupyter, Numpy, IPython, Anaconda and Scipy
* Tableau Modules: Tableau Desktop, Tableau Server, Tableau Reader and Tableau Public.
* Other Data Visualization Tools: SAS/Base, SAS/Graph, R Shinny and MS-Excel.
* Big Data Tools: Hive and MangoDB.
* Other Statistical Software's: IBM SPSS, Mini Tab, Alteryx, Matlab, Strata and Stat Transfer.
* Basic Statistics: Exploratory Data Analysis (EDA), Qualitative Analysis (Rating scales) and Sampling Techniques.
* Supervised Machine Learning Algorithms: Multiple Regression, Logistic Regression, Decision Trees, Random Forest and Support Vector Machine.
* Unsupervised Machine Learning Algorithms: Cluster Analysis, Variable Clustering, Factor Analysis and Principal Component Analysis.
* Text Mining Algorithms: Sentiment Analysis, Topic Modeling, Term Frequency, Named Entry Recognition (NER) and Event Extraction.
* Other Algorithms: Time Series Forecasting Methods, Market Basket Analysis (MBA), Campaign Analytics (Target & Mass Campaign).

* Modeling Catalog: Segmentation Models, Propensity Models, Classification Models and Market Mix Models (MMM).