

Data Analyst with 2 Years of Experience

PROFESSIONAL SUMMARY:

A Data Analytics professional with 2.8 Years of Experience in Data Analysis, Predictive Modeling using R. Experienced in Machine Learning techniques. Proficiency in extracting quality information from Quantitative and Qualitative Data and in presenting the result to the Technical and Non/Technical end users.

EDUCATION

- B.Tech in Computer Science Engineering college

ANALYTICAL & BI TOOLS:

- R
- TABLEAU
- PUTTY
- WINSXP
- MS EXCEL

WORK EXPERIENCE:

- Liaising with SME to gather requirements. Data collection, Processing, Cleaning, Analyzing, Validating and Reporting the Data by using RStudio on windows Platform.
- Analyzing Raw Data and doing EDA to extract relevant and critical information.
- Identifying and Extracting Time Series components from the Time Series Data.
- Importing Data from different sources in different formats such as CSV, XLSX, DAT and Database files.
- Precise knowledge in Model Building, Pruning and Validation parameters.
- Developing Diagnostic Tables and Graphs.
- Clear understanding of both Classification, Regression Algorithms.
- Experience in working with various versions in Tableau such as Tableau Desktop, Tableau Public.
- Able to work with Tableau to customize and Develop Interactive Dashboards.

- Good Knowledge of Mathematical concepts like Matrices, Descriptive and Inferential Statistics.
- Data Handling, Data Mining, Data Exploration, Merging, Data aggregation, Data Summarization using Base R and Dplyr Libraries.
- Data Visualizations using ggplot2, qplot packages.
- Expertise in usage of Excel features Data Analysis and Pivot tables.
- R studio
- Analyzing large Data and Extracting insights.
- Statistical Analysis such as Descriptive Statistics on the Data.
- Inferential Hypothesis tests.
- Data cleaning and Imputations.
- Extracting Classification tables and Visualizations.
- Using Supervised Learning Algorithms like Linear Regression, Logistic Regression and Random Forest.
- ROC curve evaluation.
- Validation Parameters of Regression Analysis.SQL Server Management Studio
- Integrating R studio with SSMS using ODBC connection.
- Retrieving required Tables in to R.TABLEAU
- Detail Graphical Analysis of the Features in the Data.
- Creating and Customizing Interactive Dashboards.MS EXCEL
- Basic functions in Excel.
- V- Lookup, H- Lookup, Pivot tables and Data Analysis.

STATISTICAL TECHNIQUES

Inferential Statistics: T test, Z test, Paired T test, Anova, Chi Square

Regression Analysis: Linear Regression, Logistic Regression

Tree based Classification: Decision Tree (CHAID & CART), Random Forest

Segmentation Technique: Cluster Analysis (K-Means Clustering)

Time Series Analysis: ARIMA MODEL

Dimensionality Reduction: Factor Analysis

Association Rules Mining

Market Basket Analysis

STATISTICAL PACKAGES

Dplyr, Lubridate, Rpart, Random Forest, MICE, Information Value, R part Plot, Caret...etc.

CERTIFICATIONS

- Predictive Modeling with R.
- Data Analysis & Visualization.

PROJECTS INVOLVED

Project 1: A Leading Banking Client in India wanted to reduce the Defaulter Risk on Loan processing Data by developing an Analytical solution and Identifying different Features in the Data that are contributing to the Loan eligibility.

Libraries used: R part, R part plot, RODBC, Dplyr, Caret.

- Data contains 32 features like Applicant Income, Property Area and Credit Rating, etc.
- Manipulated the Data using Dplyr Functions in R to another Subset from the Master Data as per the requirement.
- Make a detailed understanding of the Data and reduce the variables from 32 to 20 using Factor Analysis.
- Divide the Data into Train and Test Subsets.
- Approached with Decision Tree Model with CART technique and Pruned the Tree to achieve a Testing Accuracy of 78.5%
- Performed all Validation procedures.

Project 2: A leading FMCG client in India wanted to build an Analytical solution to identify precise customers collaborating for their sales and target them with relevant offers.

Libraries used: Dplyr, Lubridate, base packages of R.

- Queried Datasets using SSMS by establishing an ODBC connection through Rstudio.
- Data contains 356000 observations with 12 features such as Spending Score, Annual Income, and Age etc.
- Performed Descriptive and Statistical inferences on the data.
- Approached with non hierarchical K- means clustering for proper segmentation.
- Find out the optimal clusters by using the Cluster Dendrogram.
- By using the averages obtained grouped the relevant customers together.