Sudha

Email - sudha90singh51@yahoo.in
sudf251@gmail.com
Mobile -+919311475272
+919555489005

Data Analyst/Statistical Analyst

Summary

Business Analyst with 2+ years of experience in interpreting and analyzing data for driving business solutions. Proficient knowledge in Statistics, Machine Learning & Predictive Modeling. Good understanding of business operations and analytics tools for effective analysis of data.

Academic Details

- **PG Diploma in Applied Statistics with specialization in industrial statistics** (2017) from IGNOU**,** Delhi, Aggregate 86%.
- **B. Tech in Computer Science** (2013) from Maharaja Agrasen Institute of Technology (GGSIPU), Delhi, Aggregate 70%.
- **Relevant Coursework:** Data Science (Edureka!), Machine Learning (Stanford University), Analytics Edge (MIT).

Skills

- Programming Languages: R, Python, SQL
- Tools: R Studio, Tableau, MS Excel, JIRA,MS SQL Server, Python (Orange3 GUI)
- Theoretical Knowledge: Machine Learning, Predictive Modelling, Descriptive and Inferential Statistics, Probability, Statistical Quality control

Work Experience

Doitcall India **Nov, 2015 – May, 2017**

Business Analyst

- Skilled in researching, collecting, organizing data by different categorical segmentation for business use.
- Analyzed business web applications of competitors and assisted in the development of online classified portal.
- Maintained all data as per management policies and procedures.
- Contributed in the continuous update of classified website and database according to current trends.

US Tech Solutions **Feb, 2015 – Aug, 2015**

Quality Analyst

- Imported, sorted, cleaned, and transformed data according to business objectives.
- Analyze data across assessments (grades, submissions), attendance and user engagement (login, reading content, watching videos etc) to look for behavioral patterns
- Ensure deliverables (Daily, Weekly & Monthly MIS Reports & Dashboard) are prepared to satisfy the project requirements cost and schedule.

Academic Projects

Testing of Hypothesis

- Performed one Sample test such as Z-test for Population mean & proportion, to-test for population mean, chi-square test for population variance and Two-Sample test such as Z-test for difference of two independent population means, proportions & variance, t-test for difference of two dependent population mean and F-tests for two independent population variances.
- Formulated Null and Alternate Hypothesis and studied Type 1 & Type 2 Error.

Quality Control Charts

- Used the appropriate Control Charts according to the data such as **X-Chart** for process mean, **R** & **S** Chart for process variability, **p** & **np** chart for Defectives, **c** & **u** charts for Defects.
- Constructed a suitable control chart and analyzed whether the process is under Statistical Quality Control or not to determine whether quality improvement project should aim to prevent specific problems or to make fundamental changes to the process.

Retail sales forecasting

- Built a regression model for Retail sales forecasting by selecting appropriate regressors using forward, backward, stepwise selection method.
- Prepared a scatter matrix plot to assess the linear relationship among the variables using the matrix approach.
- Used ANOVA table for analysis of variance (F-test) to test the significance of regression model.
- Performed hypothesis testing to test the significance at 5% level of significance and built 95% confidence interval of the regression parameter using matrix approach.
- Checked the linearity and normality assumptions by plotting residual and Probability plot for the regression analysis.

Stock Market Forecasting

- Used the best statistical-learning time series model to predict the future values for the S&P 500 Stock Index.
- Used ggplot2 and plotly for interactive data visualizations to make inferences about important components of the time-series data, such as trend, seasonality, heteroscedasticity and stationarity.
- Applied accuracy function on all the forecast methods to check the performance of the different time series model on a test data.

Credit card fraud detection

- Collected and analysed highly imbalanced dataset containing credit card transactions made by European cardholders.
- Divided the dataset into training and test datasets with 75/25 split then balanced the imbalanced training dataset by oversampling method.
- Applied different machine learning algorithm and measured the accuracy using the Area under the Precision-Recall Curve (AUPRC) then predicted fraud with about 99% accuracy using Random Forest.

Independent Projects

Text Mining and Sentimental Analysis

- Fetched feeds from Twitter into R, cleaned it and created Word-Cloud visualization of frequently tweeted words.
- Sentiment of the tweets is analyzed from sentiment scores and classified into positive, negative and neutral tweets for further analysis.

Early diseases detection

- Created training and test datasets with 75/25 split from D2Hawkeye's healthcare data.
- Built a Random Forest model to predict the occurrence of diabetes in vulnerable patients with 81.98% accuracy.
- Reduced the number of false negatives by choosing a threshold value using confusion matrices and ROC curve.

Employee Salary Prediction

- Predicted the salary of employees based on the various factors by applying Naïve Bayes classifier with 76.16% accuracy.
- Minimized the cost of the company and revenue leakage on the same time retaining and/or acquiring the best available talent in the Industry.

Loan Default Prediction

- Cleaned and transformed the Customer loan dataset containing samples of about 2 lakh unique customer details.
- Applied Feature scaling to standardize the range of independent variables and Dimension Reduction technique using Principal Component Analysis (PCA) on customer dataset to reduce it to two dimensions.
- Predicted the Loan Status (Approved/ Declined) of customers using Naïve Bayes classifier with 73.1% accuracy.

Customer Churn Analysis

- Inspected data quality issues and performed Univariate, Bivariate, Multicollinearity Analysis of variables to explore their impact on the customer churn.
- Built a logistic regression models to predict the customer propensity & churn with 81% accuracy.

Market Basket Analysis

- Performed association analysis of a set of groceries transactions to discover rules that indicate the likely occurrence of an item based on the occurrences of other items in the transaction using Apriori Algorithm
- Applied minimum support threshold to find all frequent item sets in a database and minimum confidence constraint to form rules and used arulesViz package to visualize it.