

**DATA SCIENTIST PROFESSIONAL**

An insatiable intellectual curiosity and ability to mine hidden gems located within large datasets using applied statistics and visualization through healthy sense of exploration.



**PROFESSIONAL SUMMARY**

* Strong problem solving and abstract thinking skills.
* Passion to learn new technologies, techniques and apply in practical business problems to stay ahead of the Analytics industry curve.
* Liaising with the clients to understand their requirements, challenges etc. and keeping them up to date on the progress of our proposed solution.
* Ability to be flexible and change with the environment, industry and business demands.
* Proficiency in end to end coding statistical programming language R, and a database querying language SQL.
* Experienced in mathematical modeling and programming, statistical analysis, forecasting/predictive modeling, visualizations, machine learning, data mining, etc.
* Process & clean raw data structures to create data suitable to be used in a statistical analysis.



**WORK EXPERIENCE**

Working as Data Analyst for **TRANSQUBE TECHNOLOGIES PRIVATE LIMITED**, Hyderabad - 500072 Telangana, India from **Sep 2014 to till date**.



**TECHNICAL SKILLS**

* **Data science:** general analytics, dash boarding, reporting, predictive modeling.
* **Machine Learning:** classification, regression, clustering, feature engineering.
* **Mathematical statistics**: data mining, visualization, hypothesis testing and confidence intervals, linear regression, logistic regression, model selection, nonlinear regression, time series analysis, and principal component analysis
* **Software and Programming Languages:** advanced R, SQL, Microsoft PowerPoint, and Microsoft Excel. Basic knowledge of Tableau, Python and SAS.
* **Operating systems:** Familiar with Windows, Linux variants.
* **Computer savvy**.



**PROJECT PROFILE**

**Project# 3**

“**Prediction of patients at risk of developing diabetes using Machine Learning Techniques**”

[Dec, 2016 to present]

Responsibilities include doing literature review and come up with possible solutions, conducting research on a large database of Diabetes patients for finding temporal patterns.

* Tactical and practical approach for treating missing and outliers in the data.
* Extracting insights through EDA by doing statistical tests and Visualizations based on type of data.
* Understanding various complications that take place during data preparation.
* Spotting the risk factors that are more likely affecting the patients at risk of Diabetes for providing better and cost effective treatment to patients.
* Stacking of models is done to develop a robust model for classification of Readmission of diabetes patients.
* Performance of a model is evaluated by various performance measures: Classification accuracy, sensitivity and specificity.

**Project# 2**

“**Understanding and prediction of Hospital Readmission rates**”

[Feb, 2016 to Oct, 2016]

Responsibilities include conducting statistical analysis to determine key factors for planning and conducting experiments to understand readmission of patients using prescriptive and predictive analytics.

* Worked closely with domain experts to understand various challenges involving at the different levels of the project.
* Identifying the potential data sources based on the analytical problem statement and Generating synthetic features from health insurance claims data.
* Feature selection and Feature extraction regarding to both conceptual and practical considerations that optimize statistical efficiency and data quality.
* Eliminating redundant features without losing essential classificatory information.
* Applying Statistical knowledge to identify and develop various significant models based on the data available through hypothesis testing.
* Applied appropriate machine learning algorithms like **Logistic regression**, **Naive Bayes**, **SVM** and various Ensemble models like **Random Forests** to predict the patients at risk of Diabetes followed by appropriate evaluation techniques.

**Project# 1**

“**Forecasting Bed days in Hospital**”

[Aug, 2015 to Nov, 2015]

Responsibilities include,

* Exploration of time series to identify missing values and outliers in data.
* Plotting series and decomposed series in order to identify patterns such as Level, Trend, Seasonality and Cycles in the data.
* Plot the Autocorrelation function [**ACF**] and Partial Autocorrelation function [**PACF**] to identify possible models.
* Testing for stationarity using **Adf test** or **KPSS test** while building ARIMA models and doing necessary transformations.
* Transforming the series into a stationary series (differencing / seasonal differencing / taking log values etc.)
* Forecasting the finalized model along with confidence intervals.



**CERTIFICATIONS**

"Data Science, a 10-course specialization by Johns Hopkins University on Coursera". [**Verify**](https://www.coursera.org/account/accomplishments/specialization/certificate/BXSP6TJW259P)

The Data Science Specialization covers the concepts and tools for an entire data science pipeline. Successful participants learn how to use the tools of the trade, think analytically about complex problems, manage large data sets, deploy statistical principles, create visualizations, build and evaluate machine learning algorithms, publish reproducible analyses, and develop data products.



**EDUCATION**

* Masters’ Degree in Pharmacy, 2014.

