MOLVENO CONSULTING

Santosh Hale
Mobile: +91-8801371269
hallaysantoshaman@gmail.com

Professional Summary

- 5+ years of experience in the analytics consulting environment and statistical model building.
- Hands-on experience on building Information Retrieval (IR) system for query document search text data of clinical data, FDA Regulatory data, marketing data.
- Expertise in NLP (natural language processing) modeling and language modeling using TF-IDF algorithm python numpy, pandas, scipy and sklearn.
- Writing read, write operations and dumping text documents in MongoDB using python.
- Applying statistical modeling like (Poisson distribution, bayesian model) for predicting patient recruitment in clinical trial study.
- Extensive focus in Big Data Analytics and Optimization of Data Science.
- Expertise problem-solving and decision-making skills in a Startup Environment patiently.
- Proficient in working and writing analytical queries on databases like: SQL, MySQL, MSSQL, NoSQL, MongoDB, CouchDB, Cassandra and HBase.
- Applied Natural Language Processing (NLP) for Sentiment Analysis on Twitter using Twitter API.
- Performed various analysis and model building using Bayesian Modelling, Monte Carlo Method, Genetic Algorithm, Simulated Annealing, LASSO and Ridge Regressions and Random Forest.
- Creating visualization using R and Tableau.
- Excellent knowledge and worked on Hadoop Like Map Reduce, SPARK, SCALA, Pig, Hive, Sqoop, Flume, kafka, storm, Samza, Avro, Oozie, Zookeeper, Ambari, Impala, Mahout and Chukwa.
- Excellent insights in the field of data science technically and analytical abilities.

Areas include:

Regression Analysis	Data Visualization	Statistical Modeling
Big Data	Predictive Analytics	Clustering Analysis
Time Series Analysis	Text Mining	Machine Learning

Education

- Bachelor of Technology in Electronics & Communications Engineering
- Vignan Institute Of Technology And Science| Hyderabad, India | May 2012
- Diploma In Electronics & Communications Engineering
- Q.Q Govt. Polytechnic | Hyderabad, India | April 2009

Work Experience

- Working as Data Scientist with Makrocare in Hyderabad since Dec 2017
- Worked as Assistant Manager with Vimbri Media Pvt ltd in Hyderabad from Sep 2016 – Sep 2017
- Worked as Process Analyst with Solugenix India Pvt Ltd in Hyderabad from Mar 2014 – May 2016
- Worked as Data Analyst with HP in Bangalore from Feb 2013 – Feb 2014

MOLVENO CONSULTING

Makrocare | Hyderabad | Dec 2017 – Present
Data Scientist

- Data mining in clinical trial supply chain manage to determine the parameters for generating the scenarios for demand supply in clinical trial study to reduce the time and cost of the drug manufacturing.
- Working on statistical algorithms to forecast the subject recruitment in clinical trial study.
- Implementing time series algorithms like moving averages, exponential smoothing, holt-winter, Croston and other methods in the IRT AI product.
- Applying time series algorithms on the data recorded from the IRT system at clinical trial study.
- Working on demand supply of drug at clinical study in clinical trial supply chain management in the IRT AI product.
- Worked on the FDA Regulatory data, clinical trials, marketing text data and built Information Retrieval (IR) query-document search system on that data using TF-IDF algorithm. (python, numpy, pandas, scipy, sklearn).
- Applied Natural Language Processing (NLP) modeling techniques on text data crawled from client server.
- Designed Information Retrieval (IR) System for query-document search in the product using TF-IDF algorithm in python.
- Worked on clustering algorithm like K-means, classification algorithms like bayes, svm, decision tree on the text data using python (sklearn library).

Vimbri Media Pvt Ltd | Hyderabad | Sep 2016 – Sep 2017

Assistant Manager

- Designing and implementing the database architecture for big data capture and storage using Apache Solr and MySQL for transforming the unstructured data into structure by analyzing and analytical queries.
- Data mining using state-of-the-art methods.
- Extending company's data with third party sources of information when needed.
- Enhancing data collection procedures to include information that is relevant for building analytic systems by team.
- Maintained MySQL scripts to create and populate tables in data warehouse for daily reporting across departments.
- Hands-on experience on model building for predicting imports for next Month or quarter needed using regression techniques like Linear Regression, Neural Network and SVM validation by RMSE value <=0.45.
- Analyze and contrive prediction model for exports for next Month using statistical methodology Regression analysis using Excel, R and Python.
- Collect data from different sources and database for data Visualization and reporting using Excel, R, ggplot2, plotly, Tableau and Microsoft Power BI.
- Agile Initiatives and solve complex business and technical problems, prioritize work and make decisions under challenging environments.

MOLVENO CONSULTING

Solugenix India Pvt. Ltd. |Hyderabad | Mar 2014 – May 2016

Process Analyst

- Acquire data from primary or secondary data sources and maintain databases.
- Identify, analyze, and interpret trends or patterns in complex data sets.
- Processing, cleansing, and verifying the integrity of data used for analysis.
- Data extractions, Data purges, Data fixes and transform Weekly data of machine processing downtime collected to predict next downtime of machine processing to reduce cost, time, energy and manpower.
- Monthly data of machine processing failure collected to analyze and predict next failure to reduce the overall productivity of machine using statistical testing, regression and predictive modeling using Excel, R, Minitab, Tableau and Python.
- Ad-hoc analysis and presenting results in a clear manner.
- Perform exploratory data analysis using visualizations with initial insights.
- Data visualizations using Excel, R, ggplot2 and Tableau.

HP | Bangalore | Feb 2013 – Feb 2014

Data Analyst

- Fetching the data needed from database like SQL server.
- Data Pre-processing analysis using Excel and R.
- Extracting data insights using Excel and R.

Skills

R, Python (numpy, pandas, scipy, sklearn) , SAS, Tableau, Minitab, SQL, MySQL, MSSQL, Microsoft Office Suite, Microsoft Power BI, ggplot, R-shiny, ggplot2, Hadoop, Spark MLlib, Pig, Hive, HBase, Sqoop, Flume, kafka, Impala.

Certifications

- Completed Certificate Program in Engineering Excellence (Business Analytics & Optimization) from Carnegie Mellon University (CMU), US (June 2012 – Nov 2012).
- Completed Certificate Program in Engineering Excellence (Big Data Analytics & Optimization) from INSOFE, India (Feb 2016 – Jul 2016).

Analytics Projects

Prior Authorization

An insurance company serves their customers by bearing the expenses of the patients as per their health policies. Over 10,000 drug records and 20 attributes were given to analyze the data and to predict if prescribed medication requires Prior Authorization or not. Used decision tree(C50), k-nearest neighbours, Naive Baye's, Neural Network, Random Forest and Ensemble methods models to predict. Reduced 30% service time for pharmacy and insurance.

MOLVENO CONSULTING

INTUCEO

INTUCEO is an analytics product that is being developed which in its core has five patent-pending analytic and predictive engines. My work focused on rule extraction in high dimensional spaces. Variants of Random subspace method, Simulated Annealing and Beam search were used to generate rules. Used R and Python.

Electricity Price Forecasting

As complex non-linear relationships are difficult to model with conventional methods. AI methods like neural-net and data-mining methods such as Random Forest, Support Vector Machines were used to forecast the hourly prices. Several causal variables were derived and an ensemble of these methods along with the traditional Time Series Analysis was used and successfully forecasted hourly prices for the next 7 days with a MAPE (mean absolute percentage error) of 4.48%. Used R for the analysis.

Restaurant Revenue Prediction

A TFI company wants to invest on development of new restaurant sites and save time and capital. From 1995 to till date over 100000 records with 43 variables are provided. My work focused on rule extraction high dimension spaces. Variant of Random Forest, Conditional Random Forest, Support Vector Machine and Neural Network were used to generate rules and Saved 32% time and capital. Using R, Excel, Tableau and Python.

Allstate

Allstate, a personal insurer in the United States, is continually seeking fresh ideas to improve their claims service for the over 16 million households they protect. Develop methods of predicting the cost, and hence severity of claims is the main objective by creating algorithm, which accurately predicts claims severity based on mean absolute error (MAE) with 0.39 error using algorithms like Linear Regression, Neural Network (nnet), Support Vector Machine (SVM), Random Forest (RF), Conditional Random Forest, eXtreme Gradient Boosting (XGB) and Ensemble Method.