**Venkata Bogolu**

**Phone**- (559)468-6639 || **Email Id**: bogoluvenky1@gmail.com

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Professional Summary:**

* Above 5+ years of professional IT experience in**Web Applications**, **Machine Learning, Deep Learning, Reinforcement Learning, Statistics, Transforming Data with large datasets of Structured and Unstructured data, Data Engineering, Data Validation, Predictive modeling,**and**Data Visualization.**
* **Extensive use of Natural Language Processing(NLP) for Text Analytics,developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations usingPython2.7\3.4.3, R 3.2.2, SAS, SQL, PL/SQL**and **Tableau9.4.**
* Experienceworking with **Flask** and **Django framework** of **Python**for building **Web Applications**.
* Hands on experience in implementing **LDA**, **Naïve Bayes**and skilled in **Random Forests, Decision Trees**, **Linear** and **Logistic Regression**, **SVM**, **Clustering**,**Bagging**, **Boosting,****Neural networks**, **Principle Component Analysis** and good knowledge on Recommender Systems.
* Proficient in Statistical Modeling and Machine Learning techniques **(Linear, Logistics, Decision Trees, Random Forest, SVM, K-Nearest Neighbors, Bayesian, XG Boost) in Forecasting/ Predictive Analytics, Segmentation methodologies, Regression based models, Hypothesis testing, Factor analysis/** **PCA,**and**Ensembles.**
* Expertise in transforming business requirements into **Analytical models**, **designing algorithms**, **building models**, developing datamining and reporting solutions that scales across massive volume of **structured** and **unstructured data.**
* Strong experience in **Software Development Life Cycle (SDLC)** including RequirementsAnalysis, Design Specification and Testing as per Cycle in both Lean Six Sigma and Agile methodologies.
* Adept in statisticalprogramminglanguages like **R3.2.5, MATLAB**and**Python2.7.3/3.4.3** including BigData technologies like **Hadoop**,**HDFS**, **Hive, HBase,** and **Pig**.
* Experienced over translating data driven insights combined with domain knowledge into business stories and dashboards using traditional presentation tools as well as visualization tools like **Tableau 9.4**, **R3.2.2programing** and **Python 3.4.3**.
* Expert in working on all activities related to the development, implementation, administration and support of **ETL** processes for large-scale Data Warehouses using **Data Transformation Services (DTS)** and **(SQL Server Integration Services)SSIS** with **MS SQL 2014/2012/2008R2**.
* Solid understanding of **RDBMS** database concepts including Normalization and hands on experience creating database objects such as tables, **views**, **stored procedures**, **triggers, row-level audit tables**, **cursors**, **indexes**, user-defined data-types and functions.
* Experience with **Data Analytics**, **Data Reporting, Ad-hoc Reporting, Graphs, Scales, Power PivotTables, Power View,** and **OLAP reporting using SQL Server Reporting Services(SSRS).**
* Good understanding of Microsoft SQL Management Studio and data load / export utilities like **BTEQ**, **Fast Load**, **Multi Load, Fast Export** and **Performance Tuning** for enhancing extraction speed of data.

|  |  |
| --- | --- |
| Machine Learning Algorithms | Multivariate regression analysis, SVM, Cluster Analysis, NLP, Parametric and Non-Parametric Tests and Analysis of variance technique, SVM, Polynomial regression, Cluster Analysis, Decision tress, Random Forest, Apriori, MBA, Factor Analysis, Principal Component Analysis, K Nearest Neighbors, and Reinforcement Leaning. |
| ProgrammingLanguages | Python2.7/3.4.3, R3.2.2, SQL, PL/SQL, SAS,C, C++, and Java. |
| Databases | SQL Server 2014/2012/2008R2, Oracle 11g/10g, MS Access, DB2, and Teradata. |
| Tools | HQL, Pig, Apache Spark, PySpark, Spark, Scala, Microsoft Visual Studio, Minitab, Arena, SPSS, Microsoft SQL Integration Services, Microsoft SQL Analyzing Services, Microsoft SQL Reporting Services, and Oracle Database Integrator. |
| Environments and Libraries | Jupyter, R Studio, Anaconda, Spyder, Python Console, Pycharm and Tensor Flow, Keras, Theano, Scikit- learn, ggplot 2, Numpy, Pandas, Scipy, caret and Matplotlib. |
| Operating system | Windows, Linux, and Unix. |
| Data Warehousing Tools | SQL Server Business Intelligence Development – SSIS, SSAS, SSRS, Informatica, Alteryx and SAP Business Objects and Business Intelligence. |
| BI and Power BI Tools | Tableau, Kibana, Tableau server, Tableau Reader, SAP Business Objects, OBIEE, QlikView, SAP Business Intelligence and Power Pivot, Power Query, Power Map, and Power View. |

**Technical Skills:**

**Work Experience:**

**Active Health Management, Alpharetta, GA** **September 2016– Present**

Active Health Management is a global leader in the creation of innovative prescription medicines for all health, mental health, and anesthesia – products that contribute to the health of people and their quality of life.Its objective is to be a reliable partner of health care professionals and patients by providing products and services for the worldwide improvement of human health and quality of life.

The project mostly consists of data processing activities and analyzing data to build patient and treatment models, capture patient and treatment behaviors and aggregate rules for customer treatment classification there by creating customer treatment recommendations.

**Designation: Data Scientist**

* Experience working in **Data Requirement analysis** for transforming data according to business requirements.
* Worked thoroughly with data compliance teams such as Data Analysts and Data Engineers to gathered require **raw data** and define source fields in **Hadoop**.
* Applied **Forward Elimination** and **Backward Elimination** for data sets to identify most statically significant variables for Data analysis such**PBDIT**, **PBIT** and **PBT** ratios as statically significant in Profitability ratios.
* Utilized **Label Encoders** in **Python** to create dummy variables for geographic locations to identify their impact on pre-acquisition and post acquisitions by using **2 sample paired t test**.
* Worked with ETL **SQLServer Integration Services (SSIS)** for data investigation and mapping to extract data and applied **fast parsing** and enhanced efficiency by 17%.
* Developed Data Science content involving **Data Manipulation**and**Visualization, Web Scraping, Machine Learning, Python programming, SQL, GIT** and **ETL for Data Extraction.**
* Developed Analytical systems, data structures, gather and manipulate data, using statistical techniques.
* Designing suite of **Interactive dashboards**, which provided an opportunity to scale and measure the statistics of the HR dept. which was not possible earlier and schedule and publish reports.
* Provided and created data presentation to reduce biases and telling true story of people by pulling millions of rows of data using **SQL** and performed **Exploratory Data Analysis**.
* Applied breadth of knowledge in programming (Python, R), **Descriptive**, **Inferential**, and **ExperimentalDesign statistics**, **advanced mathematics**, and database functionality (SQL, Hadoop).
* Migrated data from Heterogeneous Data Sources and legacy system (DB2, Access, Excel) to centralized SQL Server databases using **SQL Server Integration Services (SSIS).**
* Applied **Descriptive statistics** and **Inferential Statistics** on varies data attributes using **SPSS** to draw insights of data regardingproviding products and services for patients.
* Utilized data reduction techniques such as **Factor analysis** to identify most correlated values to underlying factors of the data and categorized the variable according to factors.
* Applied**Wilcoxon sign test to patient**and**treatment data** for pre-acquisition and post-acquisition for different sectors to find the statistical significance in **R programming.**
* **Performance Tuning**: Analyze the requirements and fine tune the stored procedures/queries to improve the performance of the application.
* Designed and built **scalable infrastructure** to support real- time analytics with the help of **Scala**.
* Used **PySpark** Streaming to divide **streaming data** into batches for batch processing and **Spark-SQL**/Streaming for faster testing.
* Used **Kafta** to perform real time processing on patient’s lab data, patient’s record and testing reports to get immediate results.
* Developed various **Tableau9.4 Data Models by** extractingand using the data from various sources **files, DB2, Excel, Flat Files** and **Bigdata**.
* Handled importing data from various data sources, performed transformations using **Hive**, **Map Reduce,** and **loaded data into HDFS**.
* Utilized **Amazon Web ServicesS3**, **EC2**, **EMR** and **RDS**, **Redshift** to setup storage and data analysis tools like Python (Scikit Learn, Pandas, Numpy, and Scipy).
* Interaction with **Business Analyst, SMEs** and **other Data Architects to understand Business needs** and **functionality for various project solutions.**
* Identifying and executing process improvements, hands-on in various technologies such as Oracle, Informatica, and BusinessObjects.
* Designed both **3NF data models for ODS, OLTP** systems and dimensionaldatamodels using **Star** and **SnowflakeSchemas.**

**Environment:**Python,Jupyter, Tensor Flow, Keras,R Programming,SPSS, SQL Server 2014, SSRS, SSIS, SSAS, PySpark, Spark, Scala, AWS, EMR, RDS, RedShift, Microsoft office, SQL Server Management Studio, Business Intelligence Development Studio, MS Access, Informatica, SAP Business Objects and Business Intelligence.

**MISO, Carmel, IN June 2015-August 2016**

MISO is a cost-effective delivery of electric power consumption in Carmel, IN that specializes in power service. The company to find new ways to bringing new customers.

The project involves data extraction and applying data integrity and analytical techniques for story telling from the data. The major key performance indicators such as power reliabilityand improved interconnected transmission data etc. using supervised and unsupervised machine learning techniques.

**Designation: Data Scientist**

* Involved in gathering, analyzing and translating business requirements into analytic approaches**.**
* Developed scripts in **Python**(**Pandas** and**Numpy**) for **data ingestion**, **analyzing** and **data cleaning**.
* And utilized **SAS** for developing **Pareto Chart** for identifying highly impacting categories in modules to find the work force distribution and created various data visualization charts.
* Performed**univariate**, **bivariate** and **multiple analysis** of approx. 4890 tuples using **bar charts**, **box plots** and **histograms**.
* Participated in features engineering such as **feature creating, feature scaling and One-Hot encoding with Scikit-learn.**
* Converted raw data to processed data by **merging**, finding **outliers**, **errors**, trends, missing values and distributions in the data.
* Generated detailed report after validating the graphs using **R**, and adjusting the variables to fit the model.
* Worked on **Clustering** and **factor analysis** for classification of data using machine learning algorithms.
* Developed **Descriptive statistics** and **inferential statistics** for **Logistics optimization,Value** throughput data to at 95% confidence interval.
* Imported data by using **Power Query** in MS Excel from **API’s** and **Web API’s** then created relationship between **data tables** by using **Power Pivot**.
* Used **Power Map** and **Power View** to represent data very effectively to explain and understand **technical** and **non-technical users**.
* Written **MapReduce** code to process and parsing the data from various sources and storing parsed data into **HBase** and **Hive** using **HBase** - Hive Integration.
* Created **SQL** tables with referential integrity and developed advanced queries using stored procedures and functions using **SQL** server management studio.
* Used **Pandas**, **NumPy**, **seaborn**, **SciPy**, **Matplotlib**, **Scikit-learn**, **NLTK** in Python for developing various machine learning algorithms and utilized machine learning algorithms such as **linear regression, multivariate regression, naive Bayes, Random Forests, K-means,**and**KNN for data analysis**.
* Used packages like **dplyr**, **tidyr** and **ggplot2** in **R** Studio for data visualization and generated scatter plot and high low graph to identify relation between different variables.
* Worked on **Business forecasting, segmentation analysis** and **Data mining**and prepared management reports defining the problem; documenting the analysis and recommending courses of action to determine the best outcomes.
* Experience with **risk analysis, root cause analysis, cluster analysis, correlation** and optimization and**K-means algorithm** for clustering data into groups.
* Coordinated with data scientists and senior technical staff to identify client's needs.

**Environment**: SQL Server 2012,Python, Jupyter, R 3.1.2, MATLAB, SSRS, SSIS, SSAS, MongoDB, HBase, HDFS, Hive, Pig, SAS, Power Query, Power Pivot, Power Map, Power View, Microsoft office, SQL Server Management Studio, Business Intelligence Development Studio, MS Access.

**Educational Testing Service, Princeton, NJ January 2014- May 2015**

Educational Testing Service is a leading testing service organization. It provided and maintained all testing information of test takers.

The project consisted of data preprocessing activities from OLTP server using ETL and analyzing data to build testing report and test models, capture test taker behaviors and aggregate rules for test takers by their recommendations.

**Designation:Data Analyst**

* Developed complex **SQL** queries using group by, join, where clause to answer tester questions.
* Communicated and coordinated with other departments to collect **Business RequirementAnalysis**and developed data flow mapping to load OLAP server for analysis of policies details.
* Worked on **missing value imputation**, **outlier’s identification** using **Random Forest** and **Box Plots**.
* Tackled highly imbalanced dataset using under sampling with **ensemble methods**, **oversampling** and **cost sensitive algorithms**.
* Improved prediction performance by using **random forest** and **gradient boosting** for feature selection with **PythonScikit-learn**.
* Utilized **Parametric** and **Non-Parametric** test in **SPSS** to draw insight from data for making business decisions.
* Implemented machine learning models(logistic regression, XGboost) with Python.
* Validated and selected models using **k-fold cross validation**, **confusion matrices** and worked on optimizing models for high recall rate.
* Involved with **data profiling** for multiple sources and answered complex business questions by providing data to business users.
* Participated in **Agile planning process** and **daily scrums**, providing details to create stories.
* Experience with routine DBA activities like **Query Optimization, Performance Tuning** and **Effective SQL Server configuration** for better performance and cost reduction.
* Developed **Tabular Reports, Sub Reports, Matrix Reports, Drill down Reports** and **Charts using SQL Server Reporting Services (SSRS).**
* Designed rich data visualizations with **Tableau 9.4** and dynamic dashboards for business analysis.

**Environment**: SQL Server 2012, R programming, Python, MATLAB, SSRS, SSIS, SSAS,SPSS, Minitab, Microsoft office, SQL Server Management Studio, Business Intelligence Development Studio, MS Access.

.

**Decision Craft, Bangalore, India September 2012- December 2013**

Decision Craft is a provider of information technology (IT) services. The Company delivers a range of IT services through globally integrated onsite and offshore delivery locations. It offers its services to customers through industry-focused practices, including insurance, manufacturing, financial services, healthcare and telecommunications, and through technology-focused practices.

The project consisted of extracting data from various sources and transforming the data according to the business requirements and loading data to target destination. The SQL Server Integration Services(SSIS) is used to create mapping for automating data pipelines from OLTP to OLAP. The Data cleansing and Data integrity checks are performed on data and customized business report are developed and deployed on to SQL Server Reporting Services (SSRS) server for making Business Decisions.

**Designation: Data Analyst**

* Gathered requirements and documented requirements with **Use cases** in Requisite Pro and created different traceability views by **MS Visio**.
* Worked with Data Compliance teams to identify the most suitable **source of record** and **outlined** the **data**.
* Functioned with project team representatives to ensure that logical and physical **ER/Studio data models** were established in line with business standards and guidelines.
* Deeply analyzed the clients’ data by using **SQL Server Analytic Services (SSAS)**.
* Completed data **cleaning process** to discover like taking care of missing values by utilizing strategies, like, supplanting by **mean**, **forward/backward fill**, evacuating whole **rows/columns/values**, **expelling outliers** and **errors**, **normalizing**, and **scaling information** in data set.
* Implemented **metadata** repository, maintained **data quality**, **data cleanup procedures**, **transformations**, data standards, data governance program, scripts, **stored procedures**, **triggers** and **executed test plans**.
* Worked with team by Extracting Mainframe Flat Files **(Fixed or CSV)** onto **UNIX Server** and then converting them into **Teradata** Tables using **SQL Server Integration Services (SSIS)**.
* Developed **visualizations** and **dashboards** using SQL Server Reporting Services (SSRS) to present analysis outcomes in terms of **patterns**, **anomalies**, and **predictions** use of **bar charts**, **Line Plots**, **Scatter Plots**, **3D plots**, and **histograms** and connect with **Hive** for generating daily reports.
* **Documented** the **complete process flow** to describe **program development**, **logic**, **testing**, **implementation**, **application integration**, and **coding** by using **SQL Server Reporting Services (SSRS)**.
* Created customized **SQL Queries** using **SQL Server 2008/2008R2 Enterprise** to pull specified data for analysis and report building in conjunction with **Crystal Reports**.
* Designed & developed various Ad hoc reports for different teams in Business (**Teradata** and **MS ACCESS, MS EXCEL)**.

**Environment:** SSRS, SSIS, SSAS, SQL Server 2008/2008 R2 Enterprise,MS Visio, MS Excel, MS Project, Teradata, Crystal Reports, ER Studio, Crystal reports, and Business Objects.

**Education:**

Bachelors of Technology – Jawaharlal Nehru technological university,India.