A close-up of a logo

Description automatically generated

**BHASKAR REDDY**

**Data Scientist**

**Email: surya.t@saxonglobal.com**

**Phone: +1(469)8286236**

**PROFESSIONAL SUMMARY**

* Professional qualified **Data Scientist/Data Engineer** with over **7+ years of experience** in **Data science and Analytics including Machine learning, Data Mining and Statistical Analysis.**
* Involved in the entire **data science** project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modeling, and data visualization with large data sets of structured and unstructured data.
* Experienced with **machine learning** algorithms such as **NLP**, **logistic regression, random forest**, **XGboost, KNN, SVM, neural network, linear regression, lasso regression and k-means.**
* Implemented Bagging and Boosting to enhance the model performance.
* Strong skills in **statistical methodologies such as A/B test, experiment design, hypothesis test, ANOVA**
* Extensively worked on **Python 3.5/2.7 (NumPy, Pandas, Matplotlib, NLTK and Scikit-learn)**
* Experience in implementing data analysis with various analytic tools, such as Anaconda 4.0JupiterNotebook 4.X, R 3.0 (ggplot2, Caret, dplyr) and Excel.
* Solid ability to **write and optimize diverse SQL queries**, working knowledge of RDBMS like **SQL**Server2008, No**SQL** databases like MongoDB3.2
* Proficient in **Python** and its libraries such as **NumPy,** Pandas, Scikit-learn, **NLTK**, Matplotlib and Seaborn.
* Expert in preprocessing data in Pandas using visualization, data cleaning and engineering methods such as looking for **Correlations**, **Imputations**, **Scaling** and **Handling Categories.**
* Experience in building various **machine learning** models using algorithms such as **Linear Regression**, **Logistic Regression**, **Gradient Descent**, **Support Vector Machines (SVM)**, **KNN**, **Decision Tree**, **Ensembles** such as **Random Forrest**, **Gradient Boosting Trees.**
* Experience of building **machine learning** solutions using **PySpark** for large sets of data.
* Extract Transform and Load data from Sources Systems to Azure Data Storage services using a combination of Azure Data Factory, T-**SQL**, Spark **SQL,** and U-**SQL** Azure Data Lake Analytics. Data Ingestion to one or more Azure Services - (**Azure Data Lake, Azure Storage, Azure SQL, Azure DW**) and processing the data in **Azure Data bricks.**
* Experience in using source control websites like **GitHub**, **Gitlab** and **Bit bucket.**
* Experience in using cloud services like **Amazon Web Services (AWS)** such as **EC2** to work with different virtual machines.
* Experience in tuning algorithms using methods such as **Grid Search**, **Randomized Search**, **K-Fold Cross Validation**.
* Experience in dimensionality reduction techniques such as **PCA，LDA,** and **SVD.**
* Experience in **Unsupervised** learning working on social network datasets using **K-means Clustering** and **Dimension Reduction methods**.
* Knowledge and hands-on experience with **Text Mining** and **Natural Language Processing.** Used **Python libraries like Pandas, NumPy, NLTK, and RE** to handle unstructured data.
* Developed enterprise applications using **Scala.**
* Extensive hands-on experience with **Python and R for predictive and statistical analytics.**
* Proficient in creating **customized and interactive Tableau dashboards** using multiple data sources - Oracle, **SQL** Server, Netezza, SAS datasets, CSV files & TDEs.
* Excellent understanding of concepts like **Star-schema, Snowflake schema** using fact and dimension tables and relational databases MS **SQL** Server, My**SQL**, Postgre**SQL**, and client/server applications.
* Identified **Business Data and requirements** and modified it to logical data models.
* Skilled in **accessing and transforming massive datasets** through filtering, grouping, aggregation, and statistical calculation.
* Good experience in **dashboard design using large data sets.**
* Experienced in creating different **visualizations using Bars, Lines, Pies, Maps, Scatter plots****, Gantt charts, Bubbles, Histograms,** Bullets, Heat maps and Highlight tables.
* Expert in developing **user objects (filters, prompts, metrics, Sets and Groups, Calculated Fields)** to develop reports.
* Created **ad-hoc reports** based on business needs.
* Worked on **creating interactive, real-time dashboards** using drill down and customization.
* Implemented **design patterns** in Scala for the application.
* Skilled in **performing Data Manipulation and Data Preparation** with methods including describing Data contents, computing descriptive statistics of Data, regex, split and combine, remap, merge, subset, reindex, and reshape.
* Experience in using various packages in **Python and R like pandas, NumPy, Seaborn, SciPy, matplotlib, SciKit-learn, Beautiful Soup, nltk, ggplot2, caret, dplyr, Rweka, twitter, and gmodels**,
* Generating Data Visualization using Python and R creating dashboards using tools like **Tableau**
* Hands on experience with **Big Data tools** like **Spark (Spark SQL, PySpark MILib)**
* Hands on experience in implementing LDA, Naive Bayes and skilled in Random Forests, Decision Trees, XGboost, light GBM, Linear and Logistic Regression, SVM, Clustering, Principal Component Analysis.
* Experienced in gathering necessary Data for analysis from different sources, prepared Data for Data exploration using Data wrangling.
* Good industry knowledge, analytical & problem-solving skills and ability to work well within a team as well as an individual.
* Highly creative, innovative, committed, intellectually curious, business savvy with good communication and interpersonal skills.
* Excellent performance in building, publishing customized interactive reports and dashboards with customized parameters and user-filters using **Tableau.**
* Deep knowledge of **SQL** languages for writing **Queries**, **Stored Procedures**, **User-Defined Functions** etc.
* Experience in integrating data from different sources by building **ETL** process in **SSIS**.
* Quick learner in any new business industries or software environments to deliver the best solutions adapted to new requirements and challenges.
* Driven by delivering the best results and taking the ownership of work while being an effective communicator with teammates and a constant learner in this changing technology.
* Experienced the full software life cycle in SDLC, Agile and Scrum methodologies.
* Experienced in Python to manipulate data for data loading and extraction and worked with python libraries like Matplotlib and Pandas for data analysis.
* Experienced in Big Data with **Hadoop, HDFS, MapReduce, and Spark.**
* Experienced in Data Integration Validation and Data Quality controls for ETL process and Data Warehousing with reporting in **Tableau.**
* Worked on various versions of **Tableau (2019.x, 2018.x, 10.x, 9.x)** to create interactive visualizations for ad hoc reporting.
* **Predictive analytics** models and **forecasting data** sets in **Tableau**.

**CERTIFICATIONS:**

1. Business Analytics in R

<https://drive.google.com/file/d/15VMwIALIvRtVeOGewWjwspyp_Vkx4FUT/view?usp=share_link>

1. Data Analyst using SQL

<https://drive.google.com/file/d/1i89b5A0g_1MBMcASSRr2hZYL9rWjFffr/view?usp=share_link>

1. Visualization using Power BI

<https://drive.google.com/file/d/1S-DnJRa5b4xGdcMvc79jM_uPTqKifEMV/view?usp=sharing>

1. Data Engineer on AWS

**PROFESSIONAL EXPERIENCE**

**Blue Cross Blue Shield, TX January 2023 to Present**

**Data Scientist**

**Responsibilities:**

* Developed and applied methods to identify, collect, process, and analyze large volumes of data to build and enhance products, processes, and systems.
* Conducted data mining and retrieval, and applied statistical and mathematical analyses to identify trends, solve analytical problems, optimize performance, and gather intelligence.
* Visualized information using a range of tools (e.g., Python, RStudio), developed scripts and algorithms, created explanatory analysis and predictive models, and performed comparative analyses to address complex problems.
* Develop and drive activities like data mapping framework between various data items with regulations such as **CCPA**, GDPR, PIPA, PDPA and PIPEDA.
* Responsible for reporting of findings that will use gathered metrics to infer and draw logical conclusions from past and future behavior.
* Performed Multinomial Logistic Regression, Random Forest, Decision Tree, SVM to classify package is going to deliver on time for the new route.
* Used Principal Component Analysis & Factor Analysis in feature engineering to analyze high dimensional data. GIT commands code checks in, checks out branch, merging, push, pull, stash, and commit using GIT bash. Used web-based repositories such as **GitLab** and **GitHub**. Did code review by reviewing pull request in **GitHub.**
* Performed data analysis by using **SQL** to retrieve the data from Hadoop cluster.
* Used R **machine learning** library to build and evaluate different models.
* Responsible for reviewing the code from team members and merging the code into source repository **GitLab.**
* Implemented rule-based expertise system from the results of exploratory analysis and information gathered from the people from different departments.
* Collected data needs and requirements by Interacting with the other departments.
* Created various types of data visualizations using Python and **Tableau** schemas.
* Communicated the results with operations team for taking best decisions.
* Utilized Python to develop and execute Big Data Analytics and **Machine learning** applications, prototyped **machine learning** use cases under SciKit-Learn.
* Led technical implementation of advanced analytics projects and defined the time-series approaches.
* Worked with various Python Integrated Development Environments such as NetBeans, PyCharm, PyScripter and Eclipse.
* Set up Python **REST API** framework using Django.
* Develop new and effective analytics algorithms and write the key pieces of mission-critical source code.
* Implemented advanced **machine learning** algorithms including NLP, regression trees, kernel PCA, among other methods in Scala, Hadoop 1.x, R and in other tools and languages as needed.

**Giant Eagle, Hyderabad April 2020 to August 2021**

**Data Scientist/Data Engineer**

**Responsibilities:**

* Explored availability of rural broadband and its effect on property prices: Created an R Shiny app to validate, compare and map federally collected data to our aggregated database.
* Applied random forests and boosting trees **text classification** algorithms and **neural networks** trained using Googles AI's BERT system to detect product innovation.
* Develop and drive activities like data mapping framework between various data items with regulations such as **CCPA**, GDPR, PIPA, PDPA and PIPEDA.
* Conducted a **predictive modeling** workshop for end users.
* Performed **statistical analysis** to determine peak and off-peak time periods for ratemaking purposes.
* Conducted analysis of data for the purpose of designing costs. Identified root causes of problems and facilitated the implementation of cost-effective solutions with all levels of management.
* Application of various **machine learning** algorithms and statistical modeling like decision trees, regression models, **clustering**, **SVM** to identify Volume using Scikit-learn package.
* Worked on different data formats such as **JSON, XML** and performed **machine learning** algorithms in Python.
* Responsible for reviewing the code from team members and merging the code into source repository **GitLab.**
* Performed K-means clustering, Regression and Decision Trees.
* Developed penalized regression models based on data provided by the client.
* Work independently or collaboratively throughout the complete analytics project lifecycle including data extraction/preparation, design and implementation of scalable **machine learning** analysis and solutions, and documentation of results.
* Partner with technical and non-technical resources across the business to leverage their support and integrate our efforts.
* Pro-actively analyze data to uncover insights that increase business value and impact.
* Support various business partners on a wide range of analytics projects from ad-hoc requests to large-scale cross-functional engagements.
* Prepared Data Visualization reports for the management using **R** and **PowerBI**.
* Approach analytical problems with an appropriate blend of statistical/mathematical rigor with practical business intuition.
* Hold a point-of-view on the strengths and limitations of statistical models and analyses in various business contexts and can evaluate and effectively communicate the uncertainty in the results.
* Approach analysis in multiple ways to evaluate approaches and compare results.

**ICICI Bank, Hyderabad India November 2017 to March 2020**

**Data Scientist**

**Responsibilities:**

* Used R machine learning library to build and evaluate different models.
* Communicated the results with the operations team to take best decisions.
* Utilized Python to develop and execute Big Data Analytics and Machine learning applications, prototyped machine learning use cases under SciKit-Learn.
* Prepared Data Visualization reports for the management using R and PowerBI.
* Approach analytical problems with an appropriate blend of statistical/mathematical rigor with practical business intuition.
* Worked with several R packages including knitr, dplyr, SparkR, Causal Infer, Space-Time.
* Coded R functions to interface with Caffe Deep Learning Framework.
* Used Pandas, NumPy, Seaborn, SciPy, Matplotlib, Sci-kit-learn, and NLTK in Python for developing various Machine learning algorithms.
* Forecasted default consumers in future using XGBOOST, and Linear Regression
* Segmented consumers on basis of delinquency and defaults in loans.
* Quantitative Time Series Forecasting using R Packages Tidy quant, Tidy verse, Parsnip, Model time, Timetk.
* Experience in tuning algorithms using methods such as Grid Search, Randomized Search, K-Fold Cross Validation.

**ICICI Bank, Hyderabad, India July 2015 to October 2017**

**Data Scientist**

**Key Contributions:**

* Identified and predicted product launch articles using random forest and naive-bayes algorithms on our database acquired by web scraping non-traditional data sources.
* Understood STEM education and workforce pathways by exploring massive open online courses (MOOCs) data and understand college going trends among high school students in Virginia by implementing a penalized regression model on household survey data.
* Supported data process flow for Quantitative Risk Management Function (QRM) function – establish / document / enhance process to load and reconcile detailed positions each month, in a timely fashion for monthly analysis / reporting.
* Provided thorough documentation of data flow, processes, and policies, to provide sufficient governance, ensure business continuity and full compliance with audit/regulatory requirements.
* Worked on installation of Docker using Docker toolbox. Created custom Docker container images, tagging and pushing the images.
* Dockerized applications by creating Docker images from Docker file.
* Involved in reviewing business requirements and analyzing data sources form Excel/Oracle **SQL** Server
* Experienced in creating the groups, users, workbooks and, also permission sets for the **tableau** server logons and checks.
* Worked with several R packages including knitr, dplyr, SparkR, Causal Infer, Space-Time.
* Coded R functions to interface with Caffe Deep Learning Framework.
* Used Pandas, NumPy, Seaborn, SciPy, Matplotlib, Sci-kit-learn, and NLTK in Python for developing various **Machine learning** algorithms.
* Installed and used Caffe Deep Learning Framework.
* Hands on with **Git Lab/ GitHub** for code check-ins/checkouts and branching etc.
* Used pandas, NumPy, Seaborn, matplotlib, Scikit-learn, SciPy, TensorFlow LSTM
* Data Manipulation and Aggregation from a different source using Nexus, Business Objects, Toad, Power BI and Smart View.
* Designed, developed, and tested various **Tableau**/**Tableau** visualizations for dashboard and ad-hoc reporting solutions.
* Preparing Dashboards using the calculations, parameters in the **Tableau** and creating **SQL** datasets for **Tableau** and Ad-hoc reports.
* Expert on creating report schedules on **Tableau** server.
* Developed story telling boards in desktop and published into the server which allowed users to understand the data with the filters when they needed certain information.
* Perform data/systems analysis to determine best BI solution (reports, dashboards, scorecards, data cubes) using **Tableau**.
* Extensive experience in developing Stored Procedures, Functions, Views and Triggers, Complex **SQL** queries using **SQL** Server, T**SQL** and Oracle PL/**SQL**.
* Proficient in developing Entity- relationship diagrams, star/snowflake schema designs and expert in modelling the transactional databases and data warehouse.

**Education**

|  |
| --- |
| Master’s in engineering management and information systems from Indiana tech, Fort Wayne, IN May 2023, Gpa-3.55/4. |
| Bachelors from Guru Nanak Institute of Technology (JNTUH), May 2016, GPA- 6.1/10 |
| Projects on Banking and Retail collaborated with IIT Kanpur |

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Programming Languages** | SQL, Python, R, Scala. |
| **Big data tools** | Spark, Hadoop, Kafka |
| **Machine learning** | PyTorch, TensorFlow, sci-kit-learn |
| **RDBMS** | **Snowflake data warehouse**, SQL Server, Azure Data  Factory- Data Bricks, AWS-Data Bricks |
| **Scripting Language** | JSON, HTML, XML, Python. |
| **Operation** | Window 2008/2010/2011/XP |
| **Tools** | **MATILLION**, Informatica, Airflow, Hive, Spark, Kafa, Scala, ADF, Azure Synapse, Terraform, Data Bricks, NLP, AWS **stack S3**, Airflow, SQL Developer, GitHub, CSV, JSON, Parquet, Looker, Tableau, BI, Talend. |
| **Methodologies** | Agile, Waterfall. |
| **Cloud** | Azure, Aws, Hadoop Distribution System. |