**SRI VENKATESH KONDAVEETI**



**DATA SCIENTIST**



 srivenkatesh.ds@gmail.com

 +1 248-457-5423

 [linkedin.com/in/venkatesh-k-6672b31b2/](http://www.linkedin.com/in/venkatesh-k-6672b31b2/)

**PROFESSIONAL SUMMARY**

* A Productive Data Scientist with the excellence of more than 7 years’ experience, technical expertise, and strong communication skills in the IT industry
* Profound knowledge in developing Machine Learning models (Supervised and Unsupervised) using Python and R Programming languages
* Performed Requirements Elicitation, Data Integration, Data Pre-processing, Data Cleaning, and Data Transformation
* Experienced in working with Data Mining, Statistical Modeling, Exploratory Data Analysis (EDA), Feature Selection, Data Modeling, Dimensionality Reduction, Predictive Modeling, and Data Visualization
* Hands-on experience in developing predictive models using Linear Regression, Logistic Regression, K-Nearest Neighbors, Naïve Bayes, Decision Trees, Random Forests, Support Vector Machine (SVM), and skilled in K-means Clustering and Recommender systems
* Solid knowledge on applying regularization methods like Lasso, Ridge, and Elastic Net to overcome the problem of overfitting
* Good experience in using cloud services from Amazon Web Services (AWS) using EC2, S3, EMR, and QuickSight
* Hands-on experience in building Chatbots with Amazon Lex, AWS Lambda, and Facebook Developer Tools
* Performed advanced data analytics, reporting, and integration with the help of drag and drop tools on platforms like KNIME and Alteryx
* Experience in Object-Oriented Programming (OOP) Concepts of Python, and IDE’s like Python IDLE and Spyder
* Performed various functions like data manipulation, cleaning, regression, classification, and visualization using Dplyr, Tidyr, Caret, and Ggplot2 in RStudio
* In-depth knowledge on Hadoop (HDFS and MapReduce), Apache Spark (PySpark), and Big Data tools like Hive, Pig, and HBase
* Good experience in working with Shell Scripting, Azure Databricks, Cloudera, and Operating Systems like Windows and Linux
* Developed a website using JQuery, JavaScript, and HTML/CSS for the front-end, PHP for the backend, and API of the JSON format. Hands-on experience with building responsive websites using Bootstrap framework and Media Queries
* Implemented Demographic Targeting for Viacom’s social channels on Facebook to assure that content reaches specific demographic (age group and gender) considered valuable using the KNIME platform, cloud services from Microsoft Azure, and Visualization using Tableau
* Strong background in working with various programs offered by Microsoft Office like MS Word, Excel, PowerPoint, and Outlook

**TECHNICAL SKILLS**

|  |  |
| --- | --- |
| **Languages**  | Python, R, JAVA/ C/ C++, JavaScript, PHP |
| **Big Data Skills** | Apache Hadoop (HDFS, MapReduce, Pig, Hive, HBase, Flume, Sqoop), Apache Spark (PySpark, MLlib, Scala) |
| **Analytical Skills**  | Data Preparation, Data Modeling, Data Mining, Web Scraping (Beautiful Soup), Exploratory Data Analysis, Natural Language Processing, Time Series Analysis, Machine Learning (Supervised and Unsupervised), Opportunity Sizing, Advanced Statistics, Technical Documentation |
| **Databases** | MySQL, PostgreSQL, Oracle, MySQL Workbench, MySQL Server |
| **Data Visualization** | Tableau, MS Power BI, Seaborn, Matplotlib, Plotly, Ggplot2, rgl, Advanced MS Excel |
| **Web Technologies** | HTML, CSS, BOOTSTRAP, JQuery, XML, CGI |
| **Cloud Services** | Amazon EC2, S3, Lex, SageMaker and Lambda, Microsoft Azure, Azure Databricks |
| **Other Skills** | MS Office 365 (PowerPoint, Word, Excel) |
| **Operating Systems** | Windows, Linux (Ubuntu), VMware, MacOS  |
| **Platforms / IDE** | KNIME, Alteryx, Android Studio, Python IDLE, TensorFlow, Jupyter Notebook, Spyder, Cloudera, RStudio |

**EXPERIENCE**

1. **Novelis Inc, Atlanta, Georgia Jan 2020 - Present**

**Role: Data Scientist**

* Conducted research based on business requirements of stakeholders and designed sample methodologies
* Read data from various sources like CSV files, Excel files, and organized and maintained data in Relational Database systems using MySQL
* Used MySQL Connector/Python to extract data from the database
* Operated on Amazon Web Services (AWS) environment for storing large volumes of data using S3 buckets
* Implemented statistical analysis, hypothesis testing, and Exploratory Data Analysis on the gathered data for better comprehension about the business trends and patterns in existing customer behaviors
* Collaborated with a team of business analysts to perform analysis and predictions on complex data sets using Python
* Used libraries like NumPy, Pandas, SeaBorn, SciPy, TensorFlow and Scikit-Learn
* Performed k-Fold and Stratified k-Fold Cross Validation to overcome biased models
* Trained and tested several Machine Learning models like Random Forests, Decision Trees, Multivariate regression, Naïve Bayes, Neural Networks, and SVM Algorithms using Jupyter Notebook on the AWS SageMaker Platform
* Implemented dimensionality reduction techniques using Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA)
* Evaluated the performance using metrics like Root Mean Square Error (RMSE), Precision and Recall, and Area Under Curve (AUC) score
* Included Hyperparameter tuning using Grid Search and Random Search to optimize and increase the performance of the models
* Implemented Time Series Analysis to forecast the future behavior of business metrics like budget, sales, and profit
* Built Chatbot for client’s business website using Amazon Lex and Lambda
* Documented reports and visualized data-driven interactive dashboards using Tableau

**Environment:** Python, MySQL Workbench, MySQL Connector/Python, Jupyter Notebook, AWS S3, SageMaker, Lex, and Lambda, Tableau, Microsoft Word, and Excel

1. **Kroger, Charlotte, North Carolina Jan 2019 – December 2019**

**Role: Spark Developer**

* Captured business and functional requirements from Stakeholders
* Performed Data Ingestion from a wide range of data sources and live data streams using Azure Data Factory and built a high-throughput and fault-tolerant processing application for transactional data using Spark Streaming
* Involved in preparing large volumes of structured and unstructured data by cleaning, pre-processing, conducting ETL and large-scale analysis with Big Data tool like Apache Spark in Python (PySpark)
* Implemented PySpark on a cloud-based engineering tool - Azure Databricks and stored large amounts of data in Azure Data Lake, transformed, cleaned, and explored the data
* Created SQLContext and extracted data from RDBMS to work with structured data and also read data from Parquet files
* Worked with Resilient Distributed Datasets (RDD), DataFrames, and repartitioning
* Performed various high-level functions like map, reduce and join and implemented various User Defined Functions (UDF)
* Worked on a NoSQL database like HBase using Spark HBase connector to access HBase tables as external sources and performed operations (transformations and actions) on the Databricks platform
* Processed batch and streaming transactional data, and implemented scalable machine learning models using Apache Spark MLlib and evaluated the performance using various metrics to discover bottlenecks and make appropriate changes for improvement
* Implemented Retail Store and Price-Based Clustering using K-Means Algorithm and Agglomerative Hierarchical Clustering
* Performed Market Basket Analysis using Apriori Algorithm and Association rules to identify associations between frequent items purchased together, and Collaborative Filtering to make recommendations using TF-IDF Vectorization and Cosine Similarity
* Implemented Sentiment Analysis and Natural Language Processing on Twitter data and call logs of customers to better understand the opinions of the users, categorize them based on polarity and assign each group of users to different teams for seeking feedback and improving their experience
* Used ML Pipelines to automate machine learning workflows that enable a sequence of data transformations and correlating together
* Documented reports and presented the key findings to Stakeholders using Tableau

**Environment:** Apache Spark,PySpark, Python, MLlib and ML Pipelines, HBase, Azure Databricks, and Tableau

1. **Mitratech India LLP, Telangana, India March 2017 – December 2018**

**Role: Data Analyst**

* Identified issues that need to be addressed with effective decision making and problem-solving
* Maintained a draft of client requirements and gathered functional and technical resources based on them
* Integrated large volumes of data when needed (from several data sources and in different varieties) and stored in HDFS
* Implemented Qualitative and Quantitative data analyses to develop analytic solutions for client’s business questions
* Extracted relevant data, and performed hypotheses testing and financial analyses on it
* Analyzed large volumes of transactional data for supporting invoice activity and recognizing revenue and executed queries in Hive using DDL, DQL, and DML commands on the Cloudera Platform
* Performed EDA using Ggplot2, and rgl libraries of R language and data mining to look beyond data and numbers, and understand trends and insights
* Worked alongside a Financial Analyst to ensure accurate invoicing and improved solutions based on the end-user feedback
* Developed a Minimum Viable Product (MVP) that contains features just enough to satisfy early customers and gather feedback for future product development
* Applied Feature Engineering Techniques like handling missing values, outliers, One-Hot Encoding, Scaling and Splitting data (Training and Testing data sets) using Python
* Used advanced predictive models to analyze the company’s data, uncover patterns and insights for identifying opportunities that could improve business decisions and investment recommendations
* Evaluated performance, chose an optimal set of hyperparameters and tuned models to enhance performance
* Performed Cohort and Trend Analysis to categorize customers into different cohorts such as cohort for attriters and another for high transactors
* Generated weekly and monthly reports regarding progress and performance
* Developed compelling visuals using Microsoft Power BI that emphasized discovered insights and recommendations to the stakeholders

**Environment:** Python, R, Hadoop, HDFS, Hive, Cloudera, Ubuntu OS, Microsoft Word, Excel, and Power BI

1. **Ekincare, Hyderabad, Telangana, India July 2015 – February 2017**

**Role: Data Analyst**

* Identified gaps and opportunities in the existing business for improvement
* Communicated between technical and business personnel for ensuring mutual agreement with process and procedure of the applications
* Connected PostgreSQL with Tableau for Exploratory Data Analysis
* Worked with Dimensional Modeling for Data Warehousing in PostgreSQL and designed star and snowflake schemas
* Improved existing internal processes by transitioning older database technologies to new ones
* Partnered with a team of analysts and engineers and provided recommendations that were consumer-ready to the product managers
* Performed analysis on data consisting of health details of employees to find out insights related to their health to ensure that a happy workplace is maintained
* Developed predictive models or forecasting health conditions of employees based on various factors like Age, past health conditions, gender, etc., using Random Forests (with Bagging or Bootstrap Aggregating), Extra Trees Classifier, Gradient Boosting Classifier, and linear kernel SVM in Python
* Managed data quality and accuracy of the analytics performed
* Implemented optimized models on the company employees’ data
* Developed a User Interface (UI) for online marketplace using Bootstrap, Media Queries, and PHP
* Prepared well-documented reports and made recommendations to clients based on all the summarized insights and trends

**Environment:** Python, PostgreSQL, Bootstrap, PHP, Media Queries, and Tableau

1. **ModulusPI Pvt Ltd, Telangana, India May 2013 – June 2015**

**Role: Python Developer**

* Gathered business requirements while working with Stakeholders and documented it
* Implemented Web Scraping to parse HTML and XML documents, and extract data using the Beautiful Soup package
* Parsed JSON documents, stored it into SQL databases, and made changes using Python Scripts and SQL queries
* Implemented Object-Oriented Analysis and Design (OOAD) approach and used libraries like Matplotlib and Plotly for EDA
* Conducted Regular Expressions operations for pattern matching
* Associated with a team of Java Developers to optimize and convert existing source codes in Java to Python
* Designed scalable, reusable, and efficient code to develop applications with low-latency and high-availability for offering solutions
* Developed test cases for backend code and performance standards for evaluating work
* Analyzed logs and performed predictions using machine learning models
* Performed opportunity sizing to help key stakeholders to look at possible outcomes from different perspectives and made recommendations
* Built a secured web application for customers and Administrators using Django framework and Bootstrap, and PostgreSQL API calls
* Contributed towards documenting technical reports and training materials, and visually illustrated solid data-driven analytics to Business Stakeholders leading to actionable insights using AWS QuickSight

**Environment:** Python, PostgreSQL, Web Scraping, HTML/CSS, Bootstrap, AWS QuickSight, and Django framework

**EDUCATION**

* **Northeastern University | Boston, MA**

Master of Professional Studies in Data Analytics

 GPA: 3.89/4

**CERTIFICATIONS AND ACHIEVEMENTS**

* DataCamp certified Supervised Learning with Scikit-Learn
* DataCamp certified Unsupervised Learning with Python
* DataCamp certified Market Basket Analysis in Python
* DataCamp certified Big Data Fundamentals with PySpark
* LinkedIn issued skill badges for Python, MySQL, Hadoop, C, and HTML/CSS

**PUBLICATIONS**

* Published an International Research Paper on Sentiment Analysis of unstructured data from Twitter and Wikipedia in the International Journal of Emerging Technologies and Innovative Research (JETIR)
* Cleaned and transformed data, and generated a Polarity of either -1, 0, or +1 using Natural Language Processing (NLP), and Text Mining with TextBlob in Python IDLE
* Worked with Machine Learning algorithms like Naïve Bayes Classifier, Decision Trees, and Random Forests, and developed a User Interface (UI) through which users can make requests
* Used Common Gateway Interface (CGI) through which an HTML web page is generated based on the requests made