**VENKAT BACHU | Data Scientist**

+1 (510) 764-8921| Charlotte, NC | [venkat@myjobtech.com](mailto:venkat@myjobtech.com)

**PROFESSIONAL SUMMARY**

* Data Scientist with 4+ years of experience specializing in data science, machine learning, and data mining, contributing to advancing diverse projects and initiatives.
* Expertise in a wide range of machine learning algorithms, including Linear Regression, Logistic Regression, Decision Trees, Neural Networks, NLP, and proficient in utilizing Deep Learning frameworks such as Keras, PyTorch, ensuring the implementation of robust and scalable models tailored to specific project requirements.
* Skilled in statistical modelling and proficient management of databases across multiple platforms, including SQL Server, MySQL, PostgreSQL, and MongoDB, as well as retrieval and manipulation for enhanced decision-making processes and business outcomes.

**SKILLS**

**Languages and Database Technologies** – Python, Java, C, C++, JavaScript, SQL, Shell Scripting, SQL Server, MySQL, PostgreSQL, MongoDB

**Machine Learning** – Linear Regression, Logistic Regression, K-Mean & hierarchical Clustering, Decision Trees, Neural Networks, Random Forest, Navies Bayes, SVM, NLP, Keras, PyTorch, NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Spark MLib

**Statistics** **–** Descriptive and Inferential Analysis, A/B and Hypothesis Testing, Experimental design, Statistical Modeling

**Cloud and DevOps** – Microsoft Azure, AWS, Redshift, GCP, Docker, Kubernetes, GitLab, Jenkins

**Data Engineering Tools** – Apache NiFi, Google BigQuery, Tableau, Power BI

**Others** – SDLC, Waterfall, Agile, Visual Studio Code, PyCharm, Jupyter Notebook, Git, MS Excel, Jira

**PROFESSIONAL EXPERIENCE**

**Data Scientist | Metlife, USA | 01/2023 – Current**

* Employed ensemble methods like Random Forest to augment predictive accuracy by amalgamating diverse models, mitigating overfitting concerns, and fortifying model resilience, thereby enhancing the reliability of predictions in real-world scenarios.
* Leveraged Tableau to develop and deploy machine learning algorithms to forecast pivotal metrics, empowering stakeholders with actionable insights for strategic decision-making and resource allocation.
* Executed comprehensive descriptive and inferential analyses, including A/B testing and hypothesis testing, and meticulously crafted experimental designs to support data-driven decision-making processes, ensuring robustness and reliability in drawing actionable insights from complex datasets.
* Architected and implemented efficient database schemas and structures within SQL Server and MySQL, optimizing data storage and manipulation processes, fostering seamless data management, and ensuring the integrity and reliability of stored information.

**Graduate Intern | University of North Carolina at Charlotte, Charlotte, NC, US | 02/2022 – 12/2023**

* Designed and developed a Python application using Selenium to gather and extract Yelp reviews and engineered a sentiment analysis process, with NLP, Linear Regression, K-Mean Clustering, analyzing 2000+ reviews, resulting in identification of top-rated establishments.
* Computed and catalogued extracted data into a CSV file, facilitating easy access & utilization for business evaluation.

**Data Scientist | Hexaware Technologies, India | 01/2020 – 11/2021**

* Leveraged advanced clustering methodologies, such as Density-Based Spatial Clustering of Applications with Noise (DBSCAN), K-means, and Hierarchical clustering, to segment data effectively, unveil hidden patterns, and enrich data comprehension.
* Demonstrated proficiency in data visualization using R programming language and ggplot2 package, producing insightful visual representations that facilitated data-driven decision-making processes.
* Enhanced the efficiency of linear regression algorithms through optimization techniques, resulting in a notable reduction in processing time and bolstering the performance of real-time applications.
* Spearheaded developing and implementing Tableau-based dashboards tailored to analyze call center metrics, offering actionable insights into team performance and streamlining operational processes.
* Provided effective leadership to a sales team by defining strategic objectives, conducting regular performance evaluations, and fostering an environment conducive to high team performance and productivity.
* Successfully deployed K-Nearest Neighbors-based recommendation systems, which contributed to heightened user engagement in content delivery platforms and optimized content dissemination strategies.

**Data Analyst | Streebo Inc, India | 01/2019 – 12/2019**

* Crafted efficient Python scripts tailored for handling extensive datasets and extracting data from various file formats, ensuring smooth data processing and extraction workflows.
* Utilized a wide array of Python's data science libraries including but not limited to Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-learn, and NLTK, demonstrating versatility in implementing diverse data analysis techniques.
* Employed Tableau Dashboard to visualize data sourced from MySQL and MS Excel, enabling comprehensive monitoring of email performance metrics with intuitive visualizations.
* Played a pivotal role in all stages of data mining, encompassing data collection, thorough data cleaning, model development, rigorous validation, and insightful visualization to derive meaningful conclusions.

**EDUCATION**

**Master of Science in Computer Science | 01/2022 – 05/2023**

University of North Carolina at Charlotte, Charlotte, NC

**Bachelor of Technology in Electronics and Communication Engineering | 06/2015 – 05/2019**

Jawaharlal Nehru Technological University, Hyderabad, India