##### **VUTHEJ KRISHNA REDDY VERAMA REDDY** [vuthejkrishna.vr@gmail.com](mailto:vuthejkrishna@gmail.com) |267-329-9754 | [LinkedIn](https://www.linkedin.com/in/vuthejkrishna/) | [GitHub](https://github.com/VuthejKrishna?tab=repositories)



**PROFESSIONAL SUMMARY**

Data Scientist professional with 9+ years of IT experience in displaying impactful contributions to clients in Supply Chain Management, Aviation Industry, Insurance and Food sectors, resulting in accolades and appreciation awards, combined with proven expertise in the Insurance domain as an Onsite Liaison, Professional Programmer Analyst and as a Product Developer.

**CORE COMPETENCIES**

Proficient in managing the entire Data Science lifecycle – Data Extraction & Pre-processing, Data Visualization, Statistical Analysis, Feature Engineering, AI Machine Learning Model Building & Model Evaluation

**TECHNICAL SKILLS**

**Programming Languages:** Python, R (novice), Matlab, PySpark, SQL, HTML (novice), COBOL, CICS, VSAM, CA7

**Packages/Frameworks:** Scikit-learn, Pandas, Regex, Numpy, NetworkX, OpenCV, TensorFlow, Keras, NLTK, Statsmodels, Matplotlib, Seaborn

**Machine Learning:** Regression, Classification, Clustering, Supervised & Un-supervised Learning

(Linear Regression, Logistic Regression, Support Vector Machine, Decision Tree, Random Forest, K-means, K Nearest Neighbors, Boosting & Ensemble Methods, Naïve Bayes, Time Series Forecasting, Deep Learning, Natural Language Processing Methods, PCA)

**Cloud Services:** Microsoft Azure, AWS (S3, Glue Databrew, SageMaker), IBM Cloud Services (Cloud Object Storage, Watson Assistant, Watson Discovery, Watson Knowledge Studio)

**Tools:** Jupyter, Spyder, Google Colab, Alteryx, Tableau, Orange, Gephi (Social Network Analysis), Visual Studio, PyCharm, JIRA, Postman, RapidResponse

**ACHIEVEMENTS & CERTIFICATIONS**

* PSCC-1 Consulting Certificate
* **IBM Certified** - *‘Machine Learning with Python’, ‘SQL and Relational Databases 101’*
* Received exceptional rating in annual performance review for my contributions as Data Scientist at DXC Technology in 2020, 2021 & 2022.
* Received exemplary performance award by the product owner for my contributions towards Lufthansa Technik client at DXC Technology in 2022.

**PROFESSIONAL EXPERIENCE**

**Kinaxis Inc, McKinney, TX (Remote)** **January 2023 – Present**

**Business Data Scientist**

Performed forecastability analysis on the customer data to identify key areas to focus by segregating forecast items into different demand classes. Improved the accuracy of forecasting by targeting specific set of models for each demand class followed by further tuning of the models for better performance.

* Performed data cleaning, pre-processing, exploratory data analysis (EDA), data visualization and statistical analysis to ensure the quality of the data at hand.
* Engaged with business stakeholders in better understanding the data and its assumptions in order to finalize whether the sales data captured includes lost sales (negative sales/returns) and the presence of discounts or promotional offers if any.
* Aggregated the data entries over the monthly frequency by assuming the aggregation on the 1st of each month. Whether or not to place the emphasis on the zero and negative quantities as they constitute over 40% of data.
* Performed analysis on the recency of the products as the historical data overlapped with COVID time period. Hence recency is crucial in limiting volatility in sales.
* Visualizations helped analyzing the data - 99% of overall sales volume attributed to 10% of products. Thus, shifting the focus to these 10% products yields greater profits.
* Performed forecastability analysis by segregating the products based on the amount of value / volume that is being contributed to the overall sales. Data is then split into 4 quadrants depicting smooth, intermittent, lumpy and erratic demand patterns.
* Targeted each demand pattern with a separate time series model that suited the volatility of each quadrant rather than using single best model over the entire data.
* Improved accuracy results by 12% by effectively reducing Mean Absolute Percentage Error compared to clients model.
* Inaccuracy comes with a cost which can be observed in over-forecasting and sales lost to under-forecasting. Improved accuracy resulted in optimum investment in safety stock values thereby reducing over-forecasting and lowering lost sales.

**Environment:** Python, PySpark, PostgreSQL (Databricks), RapidResponse, Jupyter Notebook, VSCode, Matplotlib, Seaborn, Scikit-learn, Pandas, Numpy, Statsmodels, JIRA

**DXC Technology, Frisco, TX (Remote)****September 2019 – December 2022**

**Data Scientist, Applied AI Centre of Excellence**

**Project Title:** Data-Driven Quotation   
**Client:** Lufthansa Airlines – Lufthansa Technik  
**April 2022 – December 2022**

Designed and developed a machine learning model focused on predicting the Turnaround Time (TAT) and Defect Estimation (DE) for aircraft maintenance in driving Artificial Intelligence powered web application to assist quotation engineers with the automated quotation process. Our success in the project has landed a multimillion-dollar deal with the client.

* Utilized SQL to extract data from the client database before performing further operations.
* Performed data quality checks, EDA and went on to feature engineering as the client was specific about using least number of features possible to attain the expected accuracy.
* Experimented with various regression and classification models before performing hyperparameter tuning to achieve optimal values for TAT and DE.
* Leveraged Azure AutoML to get the best working model along with hyperparameter set.
* Successfully built and implemented an algorithm that outperformed the Mean Absolute Error of client’s in-house model.
* Worked on Alteryx, Azure ML Services along with python to pre-process the data, to build, deploy & monitor the model.
* Prepared and delivered multiple project status demos in addition to presenting the final deliverable to the stakeholders.

**Environment:** Python, Databricks SQL, Alteryx, Databricks Machine Learning, Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, Microsoft Azure – ML Studio, AutoML, JIRA

**Project Title:** Galena **Client:** AXA Equitable   
**October 2021 – March 2022**

Worked on developing an AI-powered chatbot using IBM cloud services where-in the chatbot takes questions in natural language while it needs to maneuver using the conversational interface to deliver a list of high probable answers back to users.

* Obtained data using Databricks SQL, performed data cleaning techniques using python and visualizations using Tableau.
* Created buckets in Cloud Object Storage to set up communication with Watson Discovery and Watson Assistant.
* Set up the Watson Assistant with necessary dialogue flow along with search skills.
* Created collections in Watson Discovery which will then be used to extract keywords and labels (POS) and utilized NetworkX for graph dat analysis.
* Created workspace in Knowledge Studio to ingest documents and identify entities/relationships.
* Performed sentiment analysis using NLP to tag the content of each paragraph with positive/neutral/negative tags.
* Performed N-gram analysis for text generation (Generative AI) on top of sentiment analysis to effectively relay the results corresponding to the users query.
* Built and trained the machine-learning model with domain concepts and dictionaries.
* Performed relevancy training to improve the model efficiency.

**Environment:** Python, Databricks SQL, Tableau, Machine Learning, Natural Language Processing, Natural Language Toolkit, IBM Cloud Services (Cloud Object Storage, Watson Assistant, Watson Discovery, Watson Knowledge Studio), JIRA  
 **Project Title:** Airport Runway Traffic Optimizer  
**Client:** American Airlines  
**April 2021 – September 2021**

Worked on building an effective AI-driven machine learning model to predict the aircraft arrival time for a given airport. This was a compilation of three separate goals – ‘Predicting the Arrival Fix’, ‘Predicting the Runway’ and ‘Predicting the Transit Time’ which were later put together into a pipeline to achieve the end goal of estimating aircraft arrival time.

* Worked closely with the American Airlines Engineering and Analytics team to obtain domain knowledge and to be concurrent with the features being used for the model.
* Performed exploratory data analysis to shortlist prominent features thereby performing experiments with various machine learning models (Utilized Alteryx for data cleaning/preparation and leveraged multiple libraries from python).
* Compiled a final model to test with real-time data from Federal Aviation Authority (FAA).

**Environment:** Python, PySpark, Databricks SQL, Alteryx, Databricks Machine Learning, Microsoft Azure – ML Studio, AutoML, JIRA  
 **Project Title:** AWS Hackathon **Client:** Internal  
**January 2021 – March 2021**

Health Insurance sector witnesses extensive information being pooled in while making decisions (approvals or denials) to new policy seekers. It is crucial for the insurance providers to identify beforehand the probability of making a claim by the customer. This manual process of assessing the risk level that a customer belongs to is pretty time consuming. Tackled this issue by creating a predictive modeling approach that accurately classifies the risk level (likeliness of claiming insurance policy) using AWS.

* Created buckets in Amazon-S3 to store the data and to set up communication with SageMaker Studio and Canvas.
* Created a notebook instance in SageMaker Studio using python to perform EDA, feature engineering and to experiment with various machine learning models along with parameter tuning.
* Performed experiments in SageMaker Canvas using AutoPilot.
* Evaluated and analyzed the performance metrics of custom-built ML model to AutoPilot model.
* Deployed the best fit model using EC2 service.

**Environment:** Python, Pandas, Numpy, Scikit-learn, Matplotlib, Seaborn, Machine Learning, AWS - S3, Glue Databrew, SageMaker (SageMaker Studio for experiments & SageMaker Canvas for autopilot), EC2, JIRA  
 **Project Title:** Food Advisor  
**Client:** Aramark  
**September 2019 – December 2020**

Responsible for designing and building a food-based recommender engine to drive the cross-platform mobile application aimed at advising healthy food recommendations to users based on their profile and food ratings. Apart from contributing to the team as a Data Scientist, this project allowed me to showcase my capabilities as a Scrum Master and a bit of a Backend Developer.

* Performed data cleansing and data transformations on the structured JSON data obtained from the client’s database.
* Contributed to the design & development of logic for the recommendation engine using python.
* Implemented a hybrid-recommender model to tackle the cold start problem.
* Performed as Scrum Master which included organizing sprint planning, creation of epics and user stories, daily stand-up, review and retrospection and client demo. Proficient with Agile/Scrum concepts and workflow management tools.
* As a backend developer, implemented the functionality of logging in to the application using Facebook authentication, formulated the rules for triggering push-notifications and later implemented using Google’s ‘Firebase Cloud Messaging’.
* Responsible for preparing and delivering demos towards end of each sprint to the client and project management team.
* Involved in elaborative documentation process aimed at making the transition easier for new members into the AI team.

**Environment:** Python, Pandas, Numpy, Matplotlib, Tableau, Scikit-learn, Machine Learning, Azure Function App (Time Trigger), Google Firebase, GitHub, JIRA

**DXC Technology, Hyderabad, India** **May 2018 – August 2019****Professional Product Developer - Data Analytics**

**Project Title:** US-LIFE-TPA (Data Analysis, Python Scripting & Reporting)  
**Client:** Allstate & Prudential

* Worked on data pre-processing and applied data cleaning techniques to prepare the data for visualization.
* Performed quantitative analysis and generated data visualizations to understand the trends inside the data.
* Implemented logistic regression to understand and predict the likelihood of customer in making the claim. Model evaluation has been done using recall and F1 score per client’s requirement.
* Worked as Senior Product Developer in designing new modules for Cyberlife and enhancing the product per client specifications while assisting parallelly for production support.
* Responsible for code optimization by reducing runtime of correspondence and annual statements generation by utilizing python scripting.
* Presenting demos to management and stakeholders capturing the time/resources saved through optimized code to receive production approval.
* Providing knowledge transfer sessions to other team members on the newly designed modules.

**Environment:** Python, Pandas, Matplotlib, Sklearn, Jupyter Notebook, COBOL, VSAM, CICS, Easytrieve, Rexx, Mainframes

**DXC Technology, Jacksonville, IL September 2016 – April 2018****Professional Programmer Analyst - Data Analytics**

**Project Title:** US-LIFE-TPA (Chatbot creation, Python Scripting & Reporting)  
**Client:** Allstate

* Contributed as Liaison between developers, business analysts, project managers and stakeholders.
* Responsible for requirement gathering and building technical specifications document.
* Worked on building a conversational AI chatbot to quickly address the issues during policy underwriting and in guiding the underwriters by referring to Allstate’s Database.
* Resolving complex, high-priority and immediate response issues by interacting with clients and end users.
* Handling the deliverables of the offshore team by planning and estimating the tasks/assignments to be performed.
* Designed, built, documented and deployed the architecture for generating correspondence and annual statements.

**Environment:** Python, IBM Watson, COBOL, VSAM, CICS, Easytrieve, Rexx, Mainframes

**DXC Technology, Hyderabad, India August 2014 – August 2016****Associate Professional Product Developer - Data Analytics**

**Project Title:** US-LIFE-TPA (Mainframes Technology)  
**Client:** Allstate

* Overlooking configuration and version management activities.
* Generating monthly, quarterly and annual reports for client audits.
* Testing and processing code deployment for pilot, model and production endeavors using CA7.

**Environment:** COBOL, VSAM, CICS, Easytrieve, Rexx, Mainframes

**EDUCATION**

Drexel University, Philadelphia, PA September 2019 - June 2021

**Masters in Data Science**, GPA **3.98**/4.00  
***(Earned fully funded scholarship by DXC Technology)***

*Related Coursework*: Data Acquisition and pre-processing, Data Analysis and Interpretation, Data Mining, Machine Learning using Python, Social Network Analysis, Math/Programming Foundations, Information Visualization, Human Computer Interaction, Deep Learning using NLP and Applied Cloud Computing.

**ACADEMIC PROJECTS**

* Built end-to-end audio application for Music Genre Classification and Singer Identification.   
  (Git [Repo-1](https://github.com/VuthejKrishna/Music_Classifier_and_Singer_Identifier_EDA), [Repo-2](https://github.com/VuthejKrishna/Music_Classifier_and_Singer_Identifier_Modeling), [Repo-3](https://github.com/VuthejKrishna/Music_Classifier_and_Singer_Identifier_Application))
* Built and deployed an end-to-end model on Text-Based Search Engine using Python (Deployed in Heroku). (Git [Repo](https://github.com/VuthejKrishna/Text_Based_Search_Engine))
* Sentiment Analysis and Summarization of Online Text Reviews for products from Amazon.   
  (Git [Repo](https://github.com/VuthejKrishna/Summarization-of-Online-Text-Reviews))
* Social Distancing Video Monitoring System to detect lapses in social distancing in public areas. (Git [Repo](https://github.com/VuthejKrishna/Social_Distance_Monitoring))
* Built a Stock Market Exchange Database by Web-scraping Yahoo Finance website. (Git [Repo](https://github.com/VuthejKrishna/YahooFinance_WebScraping))
* Predicting prominent factors contributing to Bad Attrition from IBM Employee Attrition Data. (Git [Repo](https://github.com/VuthejKrishna/Employee_Attrition_Data_Analysis))
* Social Network Analysis on Airways Transport Network clusters using Gephi and Python.   
  (Git [Repo](https://github.com/VuthejKrishna/Airways_Transport_Network_Analysis))