Murali Krishna Bezawada

FAIRBORN, OHIO | 4143346234 | muralibezawada10@gmail.com | www.linkedin.com/in/murali-krishna-bezawada-a39b46206

Data Scientist

Data Scientist with expertise in Python, PySpark and cloud platforms, specializing in forecasting and tree-based models. Designed and optimized ARIMA and demand forecasting solutions to enhance financial and supply-chain planning. Developed and tuned Random Forest and gradient boosting models, improving claims classification accuracy by 20% and reducing default rates by half. Skilled at turning imperfect data into actionable insights for enterprise decision-making.

Work Experience

Elevance Health **Mar 2025 - Present**

Data Scientist *Waukesha, WI, USA*

Elevance Health is a leading health benefits organization focused on transforming healthcare delivery through data-driven insights, value-based care models, and advanced analytics to improve quality, accessibility, and cost efficiency for millions of members.

- Developed and deployed forecasting models and predictive algorithms for patient risk stratification, fraud detection, and care optimization using Python, Scikit-learn, and Spark, enabling early identification of high-risk patients and reducing claim costs
- Improved diagnostic accuracy by 10% through deep learning models (PyTorch, TensorFlow, Keras) applied to medical image analysis.
- Enhanced fraud detection accuracy by 15% using computer vision techniques to analyze healthcare documentation and claims data.
- Reduced default/delinquency rates from 5% to 2% through advanced predictive modeling and behavioral analytics.
- Designed and implemented classification models (Logistic Regression, Gradient Boosting, Random Forest, SVM, Neural Networks) to improve healthcare service segmentation, resulting in more precise member targeting
- Built and tuned multi-layered neural networks and LLM-based architectures with Hugging Face, GPT, and other NLP models for patient risk assessment, which enhanced assessment accuracy and forecasting reliability
- Performed distributed hyperparameter tuning in PySpark to optimize model performance, reducing training time and improving overall accuracy
- Balanced imbalanced healthcare datasets using over-sampling and under-sampling techniques, improving minority-class detection rates
- Engineered scalable ETL pipelines across Amazon Redshift, Azure Cloud, and GCP to aggregate and transform large healthcare datasets, enabling faster data access for analytics teams
- Automated manual workflows with PySpark, Python, and AWS, improving operational efficiency
- Processed large-scale healthcare data in Hadoop (HDFS, MapReduce, Hive, Pig) to generate summary insights that fed downstream analytics systems
- Developed data ingestion pipelines using Azure Data Factory and Databricks for real-time analytics and reporting, supporting operational dashboards
- Conducted correlation analysis, feature engineering, and exploratory data analysis (EDA) to uncover healthcare utilization patterns, informing care-management strategies
- Identified data quality issues and collaborated with BI teams to standardize documentation and implement governance practices, enhancing data consistency across reports
- Applied customer segmentation and profiling models (KNIME, ML clustering) to design targeted engagement strategies, increasing customer base by 5% and portfolio growth by 9%.
- Increased cross-product adoption from 40% to 60% by leveraging predictive insights and behavioral modeling.
- Built robust data preprocessing pipelines to ensure training-production consistency, resulting in stable model performance after deployment
- Deployed models on AWS EC2 and AWS Lambda for scalable, production-grade healthcare analytics, allowing on-demand scaling of workloads
- Collaborated with engineering and cross-functional teams to translate AI models into enterprise-level healthcare applications.

Athenahealth **Mar 2024 - Feb 2025**

Data Scientist *Milwaukee, WI, USA*

Athenahealth is a healthcare technology company delivering cloud-based Electronic Health Records (EHR), revenue cycle management (RCM), and practice management solutions. The organization supports providers with scalable analytics and data-driven tools to enhance clinical efficiency, regulatory compliance, and financial performance.

- Developed predictive models to improve claims processing accuracy and revenue cycle performance using Logistic Regression, Random Forest, Gradient Boosting, and SVM.
- Improved claims classification accuracy by 20% through clustering and classification modeling using Scikit-learn and Spark MLlib.
- Applied tree-based models using boosting with hyperparameter tuning via Random Search to optimize prediction performance and reduce denial rates
- Built ARIMA/ETS time-series forecasting models to analyze billing trends and improve financial forecasting for healthcare providers.
- Implemented Natural Language Processing (NLP) techniques, including Word2Vec and GPT-based summarization models, to enhance EHR search functionality and automate clinical documentation summarization.

- Developed supervised ML models (Decision Trees, Neural Networks, Naïve Bayes, Random Forest) to support patient engagement analytics and workflow automation.
- Conducted feature engineering and model evaluation using precision, residual splits, and adjusted R^2 to identify optimal performing models.
- Extracted and transformed large-scale healthcare datasets using SQL and migrated legacy data into HDFS and HBase with Sqoop, enabling faster downstream analytics and reporting
- Leveraged Spark MLlib for distributed machine-learning processing in production, reducing model-training time and improving scalability
- Designed data pipelines to support scalable analytics across Oracle and SQL databases.
- Conducted exploratory data analysis (EDA), data wrangling, and feature selection using Python (pandas, NumPy, matplotlib, seaborn) and R (dplyr, caret, ggplot2), uncovering key patterns that guided model development
- Performed advanced data mining on anonymized patient datasets to identify utilization patterns and operational inefficiencies.
- Applied clustering techniques (K-Centroid/K-Means) to segment provider behavior and optimize service strategies.
- Built Tableau dashboards integrating Oracle and SQL data sources to present operational KPIs and predictive insights to executive stakeholders.
- Delivered actionable analytics to improve clinical workflow efficiency and decision-making processes.

American Express **Jan 2022 - Nov 2023**

Data Scientist *Mumbai, India*

American Express is a global financial services leader providing payment solutions, credit services, risk management, and data-driven financial products. Contributed to enterprise-scale analytics initiatives focused on customer intelligence, risk mitigation, and revenue optimization.

- Conducted large-scale customer behavior analysis using RFM segmentation and clustering techniques (K-Means, Hierarchical Clustering) to identify high-value and at-risk customers.
- Designed personalized product recommendation systems using Collaborative Filtering and Gradient Boosting Trees to increase customer engagement and acquisition.
- Improved campaign targeting effectiveness through advanced segmentation, contributing to a measurable increase in customer satisfaction by 60%.
- Developed predictive models for delinquency and default risk reduction using Logistic Regression, Gradient Boosting, and ensemble methods.
- Reduced default rates from 6% to 3% through advanced statistical modeling and optimized customer risk profiling.
- Evaluated model performance using ROC-AUC, Confusion Matrix, Cross-Validation, RMSE, and A/B testing in both controlled and real-world environments, achieving recall rates as high as 94%.
- Applied hypothesis testing, ANOVA, survival analysis, longitudinal analysis, and experimental design to assess customer lifecycle behavior.
- Designed and executed A/B testing frameworks to optimize marketing strategies and product offerings.
- Determined optimal sample sizes and experimental structures for statistically robust business decision-making.
- Addressed imbalanced datasets using SMOTE oversampling, undersampling, and cost-sensitive ensemble algorithms, improving model recall on minority classes
- Optimized custom ML algorithms using Stochastic Gradient Descent and Bayesian Optimization for hyperparameter tuning.
- Built scalable preprocessing pipelines to ensure consistency between training and production datasets.
- Aggregated and processed enterprise-level data from internal databases and web-scraped sources to build unified analytical datasets.
- Performed extensive data cleansing, missing value imputation, and outlier detection using Pandas and NumPy.
- Engineered predictive features from digital behavioral data (Adobe Analytics, Tealeaf, enterprise systems) to improve model accuracy.
- Translated complex analytical insights into executive-level dashboards and reports.
- Communicated predictive insights to cross-functional stakeholders, improving data-driven decision-making across functional teams.
- Streamlined enterprise data pipelines by identifying redundancies and improving process efficiency across analytics teams.

DHL **Jun 2019 - Dec 2020**

Data Analyst *Mumbai, India*

DHL is a global logistics and supply chain leader specializing in international shipping, e-commerce fulfillment, and end-to-end supply chain optimization. Contributed to data-driven initiatives supporting demand forecasting, pricing optimization, and operational efficiency.

- Collaborated with database engineers to design and implement ETL pipelines using PySpark for data extraction and integration from SQL Server databases, reducing data latency and speeding up analyst access to fresh data
- Optimized complex SQL queries to improve data retrieval efficiency and streamline reporting workflows.
- Automated data transformation and preprocessing using Python (Pandas, NumPy) and Alteryx to improve pipeline reliability.
- Designed and implemented forecasting models leveraging analytical skills to assess sales demand, evaluate price elasticity, and predict operational risk, delivering more accurate insights for planning
- Built predictive models to estimate campaign participation probabilities, supporting marketing and promotional strategy optimization and helping prioritize high-response customer segments
- Improved forecasting accuracy by applying feature engineering, scaling, and principal component analysis (PCA).

- Developed classification models (Logistic Regression, SVM, Random Forest, AdaBoost, Gradient Boosting) to optimize customer discount strategies across millions of records, increasing discount allocation efficiency
- Applied SMOTE resampling techniques to address class imbalance in large-scale logistics datasets, improving model recall for minority classes
- Built and trained multi-layer neural networks using PyTorch and TensorFlow, enhancing predictive modeling capabilities and raising overall accuracy
- Resolved overfitting issues through regularization techniques including batch normalization and dropout layers, improving model generalization on unseen data
- Conducted comprehensive EDA using Matplotlib and Seaborn to identify operational bottlenecks and shipment trends, informing process-improvement initiatives
- Applied feature selection techniques such as Information Value (IV) and PCA to improve model performance.
- Performed data cleansing, missing value imputation, and outlier treatment to enhance data quality.
- Developed Tableau dashboards to visualize KPIs, demand forecasts, and campaign performance metrics, enabling real-time monitoring for stakeholders
- Generated executive-ready reports summarizing predictive insights for product, sales, and marketing teams, guiding strategic decisions
- Communicated analytical findings to cross-functional stakeholders, supporting data-driven decision-making and increasing adoption of insights

Education

Concordia University Wisconsin
Masters, Computer Science *Wisconsin, USA*

Certification

- **Microsoft Certified Azure Data Scientist Associate: Microsoft**
- **AWS Technical Essentials: Amazon Web Services (AWS)**
- **Get started with Azure Stream Analytics: Microsoft**
- SAS Certified Data Scientist

Technical Skills

- **Programming Languages**: Python, R, SQL, PL/SQL, T-SQL, SQLite3, SAS, C, C++, Java, C#, HTML, JavaScript, CSS, PowerShell
- **Libraries & Frameworks**: TensorFlow, PyTorch, Scikit-learn, Pandas, NumPy, OpenCV, YOLO, Keras, SciPy, Beautiful Soup, Django, Apache Spark, JSON, PySpark, MNE, Caffe, NLP, ggplot2
- **Machine Learning & Data Analysis**: Linear Regression, Logistic Regression, SVM, KNN, Naive Bayes, CART, Random Forest, K-means Clustering, Hierarchical Clustering, TensorFlow, Caffe, Neon, Forecasting Models, Tree-based Models, Analytical Skills, Data Science
- **Cloud & Platforms**: AWS, S3, EC2, Azure, Google Cloud Platform, Oracle Cloud Platform, Docker, Hadoop, GitHub
- **Database Platforms**: Oracle, SQL Server, SharePoint, Spark, POI
- **Data Visualization & Reporting Tools**: Tableau, Power BI, R-Shiny, Matplotlib, Pentaho, Google Cloud Prediction API
- **Scripting & Development**: JavaScript, AngularJS, NodeJS, Shell Scripting, Linux, ECMAScript, ASP.NET
- **Software & Tools**: Microsoft Visio, Microsoft Power Platform, Power Apps, Power Automate, AutoCAD
- **SharePoint Technologies**: SharePoint Online/2016/2013/2010/2007, SharePoint Designer (2013/2010/2007), Nintex Designer (2010/2013), InfoPath (2013/2010/2007), MS Project (2016/2013)