Neha Reddy Yenugu

Fairfax, VA
571-564-8783 | nreddy2709@gmail.com | linkedin.com/in/nehareddy | github.com/nehareddy/

Technical Skills

Languages: C, C++, Python, Java, SQL, HTML, CSS, JavaScript
Frameworks: Pandas, NumPy, MLlib, Tidyverse, Matplotlib, scikit-learn, ggplot2, Seaborn
Data Analytics tools: Tableau, R, Databricks, MySQL, NoSQL, PySpark, PowerBI, Jupyter, Eclipse, Qlik Sense, Visual Studio, MS Excel
Cloud Technologies: AWS (EC2, S3, RDS, Glue), Azure
Statistical Methods:Hypothesis Testing,Principal Analysis(PCA), Time Series, Correlation(Chi-square test, covariance), Bayes Law
Machine Learning: Linear Regression, Logistic Regression, Naive Bayes, Decision Tree, Random Forest, Support Vector Machine, K-Means Clustering, K-Nearest Neighbors(KNN), Random Forest, Gradient Boosting Trees, Natural Language Processing.

Experience

Junior Software Developer May 2022 – Jan 2023
SYNYCS Enterprise PVT LTD *Hyderabad, India*

- Elevated software application efficiency by 15% through design, development, and maintenance.
- Orchestrated cross-functional teams, resulting in a 20% bug reduction in instances through rigorous coding and software testing.
- Facilitated code and design documentation, ensuring quality and best practices adherence, slashing code-related errors by 10%.

Data Engineer May 2021 – Dec 2021
Fluxtex Solutions Inc. *Texas, USA*

- Designed, developed, and maintained 5+ ETL pipelines for data integration, ensuring accurate and reliable data through validation and cleansing processes.
- Assisted in setting up and managing 3+ relational and non-relational databases, optimizing performance through indexing and query tuning.
- Collaborated with cross-functional teams to meet data requirements, supported analytical needs, and documented 10+ data workflows and integration activities.

Intern Software Test Engineer Jun 2019 - Jun 2020
Rudra Labs PTE LTD *Hyderabad, India*

- Spearheaded the launch of "Smart - Pixie (in-house product)" while collaborating with SDEs and program members. Directed a 5-member team, resolving 17 bugs through systematic testing, resulting in a 45% increase in efficiency.
- Developed and executed test plans and cases, leveraging innovative technologies, resulting in a 30% reduction in testing time.
- Utilized UML diagrams to enhance architectural understanding and documented product details, contributing to a 15% increase in team productivity.

Projects

Anomaly Detection and Failure Prediction Unveiled | *Python, Qlik Sense, R, ML* Apr 2023 – May 2023

- Analyzed sensor data and failure reports from the MetroPT-3 dataset to identify key factors contributing to compressor failures.
- Partitioned data into 70% training and 30% test sets with the target variable "Compressor."Employed five-fold cross-validation to enhance model robustness and prevent overfitting.
- Reduced operational disruptions by 20% and enhanced passenger safety in metro train systems.

Exploring Music Through Textual Insights | *NLP, ML* Feb 2023 – May 2023

- Explored NLP methods to leverage extensive text collections for automatic music knowledge discovery, addressing training data scarcity in machine learning.
- Implemented a comparative analysis on the 79 Musical Genre Classification dataset, constructing and evaluating multiple ML models.

- Developed and evaluated models, including Support Vector Machine (SVM) and Linear SVC, with Linear SVC achieving the highest accuracy of 74%.

Sales Data ETL Pipeline | *AWS Glue, S3 bucket, RDS* Oct 2023 – Dec 2023

- The pipeline automatically extracts data from S3, transforms it to a structured format using Glue ETL jobs, and loads the cleaned data into a relational database on RDS.
- Implemented data quality checks and scheduling for regular updates, providing a centralized and reliable source for sales analysis.

Leveraging Diverse Factors for Diabetes Risk Prediction | *Python, ML* Aug 2023 - Oct 2023

- Pioneered a robust framework for diabetes prediction where the outlier rejection, filling the missing values, data standardization, feature selection, and different Machine Learning resulted in a 25% increase in prediction accuracy.
- Strategically managed data imbalance issues, with over 80% of records representing non-diabetic entries, to ensure model reliability and accuracy.
- Applied domain-specific features to investigate the relationship between affected features, gaining an accuracy of 70% and a weighted precision of 83%.

Predicting the Relationship between Socio-economic Status and Cancer | *R Studio* Mar 2023 – May 2023

- Unearthed significant predictors of socioeconomic status are found using simple linear regression and k-fold cross-validation, resulting in the identification of 8 key predictors with an average prediction accuracy of 85%.
- Employed the principal components method, reducing the error rate by 17% and elevating the R-squared value, resulting in the best model.
- Evaluated correlations between variables using correlation coefficients and assessed statistical significance using p-values [0.65 (p ¡ 0.05)].

Detection of Covid-19 from Chest X-Ray | *Python, AI* May 2022 – Jul 2022

- The application achieved an impressive accuracy rate of 92% in diagnosing COVID-19 cases.
- Implemented transfer learning to fine-tune a pre-trained CNN model, securing a 95% accuracy rate in diagnosing medical conditions.
- Collaborated with 10 healthcare professionals to validate results, ensuring clinical relevance and accuracy of predictions.

Empirical Analysis for Crime Prediction and Forecasting | *ML, Deep Learning* Feb 2022 – Mar 2022

- Collected and cleaned data from various sources, including law enforcement agencies, enhancing reliability by 30%.
- Applied machine learning algorithms (Random Forest, Gradient Boosting) achieving a 25% improvement in crime prediction accuracy.
- Employed deep learning techniques, including recurrent neural networks (RNNs), to capture temporal patterns in crime data, increasing predictive performance by 20%.
- Visualized crime trends on geographic maps, enabling a 40% reduction in response time for law enforcement agencies through intuitive decision-making.

Education

George Mason University - GPA - 3.74/4.00 Fairfax, USA
Masters in Data Analytics Engineering *Jan 2023 - Dec 2024*

Guru Nanak Institute of Technology - GPA - 7.41/10.0 Hyderbaad, India
Bachelors in Computer Science and Engineering *Jul 2018 - Aug 2022*

Certifications

Accenture North America Data Analytics and Visualization Job Simulation on Forge - Dec 2023.
Quantum Data Analytics Job Stimulation on Forge – Dec 2023.