

ARI R

Chennai, Tamil Nadu, India | +91-7530020980 | ariranalyst@gmail.com

linkedin.com/in/r-ari | ari-ds.vercel.app

SUMMARY

Driven Junior Data Scientist skilled in Machine Learning, NLP, and Predictive Analytics. Proficient in Python, SQL, Scikit-learn, Pandas, NumPy, TensorFlow, and Flask. Strong communicator experienced in collaborating with cross-functional teams to deliver data-driven solutions. Experienced in EDA, Data Preprocessing, Feature Engineering, and Model Development. Background in Email Marketing and Automation, leveraging Data Analysis, A/B Testing, and Performance Optimization to improve business outcomes. Strong foundation in Software Engineering, Statistical Analysis, and Problem Solving, with a focus on building scalable, data-driven solutions.

TECHNICAL SKILLS

- Python (Pandas, NumPy, NLTK, Scikit-learn, TensorFlow) | SQL (Joins, Subqueries, Data Querying) | Regular Expressions (Regex)
- Exploratory Data Analysis (EDA) | Data Preprocessing & Cleaning | Matplotlib, Seaborn | Excel (Formulas, Pivot Tables, Charts)
- BI Tools (Power BI & Tableau) | Supervised Learning (Linear Regression, Logistic Regression, Decision Trees, etc..)
- Unsupervised Learning (K-Means, Hierarchical Clustering) | Model Evaluation (Accuracy, F1-Score, Confusion Matrix, Precision-Recall)
- NLP | Text Preprocessing (Tokenization, Stop word Removal, Lemmatization, Stemming) | Sentiment Analysis (NLTK)
- Feature Extraction (Bag-of-Words, TF-IDF) | Readability Scoring (FOG Index) | Web Scraping (BeautifulSoup, Requests)
- Scikit-learn Pipeline Design | Data Wrangling | Feature Engineering | Predictive Modeling
- Statistical Analysis | Version Control (Git, GitHub) | PostgreSQL | MySQL | MongoDB
- Agile/Scrum Methodologies | Problem Solving | Team Collaboration | Communication | Business Insight Generation | Data Storytelling

PROFESSIONAL EXPERIENCE

FemtoSoft Technologies · Data Science · Internship

Oct 2025 - Present

- Developed predictive analytics models using supervised and unsupervised learning for client projects.
- Performed data preprocessing, Exploratory Data Analysis (EDA), and feature engineering to improve model accuracy.
- Applied model evaluation metrics (Accuracy, F1-Score, Confusion Matrix) for performance validation.
- Tools: Python, Scikit-learn, Pandas, NumPy, SQL, Flask, Web Scraping

Rubixe - AI Solutions Company · Data Science · Internship

Mar 2025 – Sep 2025

- No-Churn Telecom – Customer Churn Prediction · Internship Project
- Tech Skills: Python, Pandas, NumPy, Scikit-learn, LightGBM, XGBoost, Logistic Regression, Random Forest, Matplotlib, Seaborn, EDA, Data Preprocessing, Feature Engineering, Model Evaluation (Accuracy, F1-Score, ROC-AUC)
- Developed customer churn prediction model (4,617 records) achieving 97.18% accuracy with LightGBM.
- Conducted data preprocessing, Exploratory Data Analysis (EDA), feature engineering, and model evaluation.
- Built and compared Logistic Regression, Random Forest, XGBoost, LGBM models.
- Identified key churn drivers (international plan, call duration, voicemail) to guide retention strategy.
- Visualized insights using Matplotlib and Seaborn.
- GitHub: No-Churn-Telecom

Genxlead Solutions Pvt Ltd · Software Engineer

Apr 2022 – Oct 2024

- Applied data analysis, automation, and problem-solving to optimize business processes and improve efficiency.
- Designed and optimized email marketing campaigns, achieving a 30% higher open rate and 25% increase in CTR.
- Collaborated with cross-functional teams to develop scalable, data-informed solutions using Python and SQL.

EDUCATION

Selvam College of Technology – Bachelor of Engineering (B.E.) in Electronics and Communication Engineering

- 2017 – 2021 | CGPA: 8.09

Government Boys Higher Secondary School

- Higher Secondary (HSC), Science Stream – 2016–2017 | Percentage: 93.5%
- Secondary School (SSLC) – 2014–2015 | Percentage: 90.2%

PROJECTS

Chronic Kidney Disease Prediction (CKDP) · Self-Based Project

- Tech Skills: Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Flask, Machine Learning, Data Preprocessing, Feature Engineering, Exploratory Data Analysis (EDA), Logistic Regression, Random Forest, Model Evaluation (Accuracy, Confusion Matrix, F1-Score)
- Developed a machine learning model to predict Chronic Kidney Disease from patient medical data using Python, Pandas, NumPy, and Scikit-learn.
- Performed data cleaning, preprocessing, feature engineering, and exploratory data analysis (EDA) to identify key health indicators impacting disease prediction.
- Implemented, compared, and optimized classification algorithms (Logistic Regression, Random Forest) achieving high predictive accuracy and robust model performance.
- Built and deployed an interactive Flask web application for real-time predictions, data visualization, and reporting of patient risk indicators.

Portuguese Bank Marketing Analysis · Self-Based Project

- Tech Skills: Python, Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn, Machine Learning, Data Preprocessing, Exploratory Data Analysis (EDA), Feature Engineering, Logistic Regression, Random Forest, XGBoost, Classification, Predictive Modeling, Model Evaluation (Accuracy, Precision, Recall, F1-Score, ROC-AUC)
- Developed a predictive machine learning model to forecast customer subscription to term deposits using the Portuguese Bank Marketing dataset.
- Performed data preprocessing, feature engineering, and exploratory data analysis (EDA) to extract meaningful patterns and improve model accuracy.
- Implemented and compared classification algorithms (Logistic Regression, Random Forest, XGBoost) using evaluation metrics: Accuracy, Precision, Recall, F1-Score, ROC-AUC.
- Delivered actionable business insights to optimize marketing strategies, improve customer segmentation, and enhance campaign performance.

CERTIFICATIONS

- Certified Data Scientist – DataMites
- Python for Data Science Development
- Python for Machine Learning – Great Learning
- Advanced Google Analytics