Artha Pillai

arthapillai2000@gmail.com | +1 617-504-3230 | www.arthapillai.com | Weymouth, MA | www.linkedin.com/in/artha-pillai

EDUCATION

Northeastern University | Boston, MA   **September 2022 - August 2024**

University of Mumbai | Mumbai, India   **August 2019 - May 2022**

Maharashtra State Board of Technical Education | Mumbai, India   **August 2016 - June 2019**

SKILLS

Languages: Python, R, SQL (MySQL, PostgreSQL, MongoDB, Oracle), C, C++, Java, HTML, JavaScript, PHP, Unix/Linux
Libraries: Scikit-learn, NumPy, Pandas, Matplotlib, Plotly, Seaborn
Frameworks: Streamlit, PyCharm, Spark, Hadoop, ETL
Software: RStudio, Tableau, Power BI, QlikSense, Jupyter Notebook, Git, Databricks, Stata, Google Analytics, Looker
Cloud Platforms: AWS-Amazon Web Services (EC2, Athena, S3), Microsoft Azure, Google Cloud Platform, Snowflake
Certifications: Career Essentials in Data Analysis by Microsoft and LinkedIn (bit.ly/MsLiDataAnalysisCert)
Analytical: Data Mining, Data Visualization, Data Manipulation, Big Data Analytics, Predictive Analytics, Statistical Analysis

WORK EXPERIENCE

Bright Mind Enrichment and Schooling | Remote, USA   **November 2024 - Present**
Data Analyst

- Led implementation of data collection workflows by designing SQL queries, automating Python pipelines, and integrating R models to process datasets of 10,000+ records, reducing data retrieval time by 40% and enhancing data accuracy
- Spearheaded the creation of comprehensive KPI dashboards and the generation of over 15 visual reports using Tableau and Python-based metrics analysis, enhancing campaign performance tracking and improving forecasting accuracy by 33%
- Managed in-depth data manipulation, cleansing, and analysis of thousands of data points, identifying actionable patterns, trends, and key insights that drove a 15% improvement in strategic decision-making and resource allocation

World Justice Project | Washington, D.C.   **September 2023 - March 2024**
Data Analyst

- Created and published comprehensive USA and North Macedonia reports featuring 25+ data analysis charts using R and HTML, enhancing global awareness, and supporting policymaking. Link: RuleofLawUSA, RuleofLawNorthMacedonia
- Executed complex SQL queries to extract data from various databases, automating the generation of 27 Excel reports from AWS data using Python pipelines for summarizing news data and sentiment analysis, cutting report generation time by 98%
- Revolutionized Rule of Law report creation, collaborating with stakeholders to build an innovative HTML webpage using Flask, cutting development time by 93%, simplifying code by 85%, and automating reports for greater efficiency
- Translated complex business requirements into technical specifications by deploying a Naive Bayes model to classify over 10,000 rule of law news articles, achieving 84% accuracy and streamlining the news data analysis process by 67%

AINE AI | Mumbai, India   **August 2021 - January 2022**
Data Scientist

- Analyzed telecom customer data using statistical techniques to identify patterns and trends across 25+ variables, resulting in a 90% accuracy rate in predicting high-risk customer churn to inform market strategy and revenue optimization
- Established data quality management protocols for customer data pipelines, implementing metadata standards and data validation checks that improved data integrity by 87% and reduced error rates from 12% to under 5%
- Developed interactive dashboards using Business Intelligence tools like Tableau and Power BI to visualize engagement metrics, enabling senior management to optimize retention strategies, resulting in a 20% increase in customer retention

PROJECTS

S&P 500 Stock Price Prediction Model

- Engineered a predictive model using Random Forest and XGBoost to forecast S&P 500 index closing prices with a 68% accuracy rate during weekly options expirations, aiding financial analysis and decision-making for TrustNet market analysts
- Optimized key features like moving averages and Bollinger Bands, boosting model performance by 25%, and delivered market trend insights through visualizations, enhancing stakeholders' data-driven strategy adoption by 30%

Credit Risk Analysis

- Built a risk assessment model using XGBoost, LightGBM, and CatBoost, performing complex analysis on data from 45,000 clients across 43 features to identify key risk factors, achieving an 85% accuracy rate in predicting loan defaults
- Enhanced underwriting strategies by implementing predictive analytics and statistical models, improving model interpretability by 30% and reducing default rates by 20%, optimizing decision-making in credit risk management

Stroke Prediction

- Conducted detailed data analysis on 20+ healthcare process metrics using GBM and Logistic Regression models that led to optimization recommendations, improving stroke detection efficiency by 85% and enhancing data-driven decision-making
- Reduced false negatives in stroke detection by 65% through effective data preprocessing techniques like ordinal encoding, one-hot encoding, and SMOTE up sampling, enhancing the accuracy of stroke prediction in the imbalanced dataset