

ARTHA PILLAI

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

EDUCATION

Northeastern University | Boston, MA

September 2022 - August 2024

Master of Professional Studies in Analytics / GPA: 3.6

University of Mumbai | Mumbai, India

August 2019 - May 2022

Bachelor of Engineering in Computer Engineering / GPA: 3.8

Maharashtra State Board of Technical Education | Mumbai, India

August 2016 - June 2019

Diploma in Computer Engineering / GPA: 3.5

SKILLS

Analytical: Data Mining, Data Visualization, Data Manipulation, Data Warehousing, Predictive Analytics, Statistical Analysis

Languages: Python, R, SQL (MySQL, PostgreSQL, MongoDB, Oracle), C, C++, Java, HTML, JavaScript, PHP, Unix/Linux

Software: RStudio, Tableau, Power BI, QlikSense, Microsoft Office (Excel, Word, PowerPoint, Access), Google Apps (Docs, Sheets, Slides), Jupyter Notebook, Git, Databricks, Stata, Google Analytics

Libraries: Scikit-learn, NumPy, Pandas, Matplotlib, Plotly, Seaborn

Framework: Streamlit, PyCharm, Spark, Hadoop, ETL

Cloud Platforms: AWS-Amazon Web Services (EC2, Athena, Sagemaker), Microsoft Azure, Google Cloud Platform, Snowflake

WORK EXPERIENCE AND INTERSHIPS

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](#)
- Automated the generation of 27 Excel reports from AWS data, utilizing a Python script for summarizing news data and performing sentiment analysis, resulting in a 98% reduction in report generation time
- Revolutionized the creation of Rule of Law reports, by building an innovative dynamic HTML webpage using Flask, which reduced development time by 93%, simplified code by 85%, and automated report generation, thus enhancing efficiency
- Deployed a Naive Bayes model to classify over 10,000 news articles related to the rule of law, achieving an accuracy rate of 84% and streamlining the news analysis process by 67%

AINE AI | Mumbai, India

August 2021 - January 2022

Data Scientist

- Designed Machine Learning Model that predicted customer churn for high-risk customers with 90% accuracy by extracting insights from telecom data sets and identifying trends in customer satisfaction
- Improved work efficiency of the company by 84% by implementing Machine Learning algorithms to forecast customer churn using customer usage patterns, call center interactions, and demographic information
- Developed interactive visualizations of data by employing Business Intelligence tools namely Tableau and Power BI to communicate insights to senior management resulting in a 20% increase in engagement with the data analytics team

SIES Graduate School of Technology | Mumbai, India

June 2021 - July 2021

Data Scientist

- Adopted Agile/SCRUM principles in analyzing Quantum Computing data, employing techniques such as Neural Networks
- Incorporated Natural Language Processing techniques, such as lexical ambiguity resolution, to extract insights from outcomes

PROJECTS & PUBLICATIONS

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%

Stroke Prediction

- Achieved a high recall rate of 85% in stroke prediction by implementing data mining techniques, including feature engineering and Machine Learning models for instance GBM and Logistic Regression
- 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

Clinical Healthcare Prediction Website

- Led a team to build a website using Django, SQLite, and Python to diagnose illnesses from symptoms with 95% accuracy through Data Mining and application of Machine Learning models such as Naive Bayes, Decision Tree, and Random Forest
- Publication: Artha Pradeep Kumar Pillai, Amir Basha Shaikh, et al., (2022) Clinical Healthcare System using Data Mining. *IJREAM*. Link: https://www.ijream.org/IJREAM_V08I02.html