**Abhishek kumar sinha**

25 Walbridge st., Boston, MA | +1(857)-961-6419 | [sinha.ab@northeastern.edu](mailto:sinha.ab@northeastern.edu) | [LinkedIn](https://www.linkedin.com/in/abhishek-k-sinha) | [GitHub](https://github.com/Abhisheksinha1830) | [Portfolio](https://abhisheksinha1830.github.io/)

**EDUCATION**

|  |  |
| --- | --- |
| **Northeastern University** | Boston, USA |
| Master of Science (MS) in Data Analytics Engineering [GPA: 3.9] | Expected May 2025 |



Relevant Courses: Foundation of Data Analytics, Data Management for Analytics, Data Mining, Algorithms

**National Institute of Technology Warangal** Warangal, India

Bachelor of Technology (B. Tech) in Mechanical Engineering [GPA: 3.2] May 2021

Relevant Courses: Problem solving and computer programming, Integral calculus and Numerical method, Statistics

**SKILLS**

* *Statistical and Analytical tools***:** Python, SQL, MongoDB, R
* *Machine Learning*: Random Forest, KNN, SVM, Decision Tree, Linear and Logistic Regression, K-means clustering
* *Visualization tools*: Tableau, PowerBI, Microsoft Excel
* *Statistics*: ARIMA, BSTS, LSTM, Hypothesis testing, Financial Forecasting, Covariance, and correlation
* *IDE & Tools*: GitHub, Jupyter Notebook, Google Colab, R-Studio, AWS, Bayesian model
* *Soft Skills*: Communication, Presentation, Problem solving, Time management, Initiative, Teamwork, Planning



**PROFESSIONAL EXPERIENCE**

|  |  |  |
| --- | --- | --- |
|  |  |  |
| **Northeastern University** | Boston, USA | |
| Financial Analyst | Dec 2023 - Present | |



* Conducted detailed data analysis using Excel, Tableau and Python, providing valuable insights for informed decision-making in the finance department and managed financial data entry, ensuring accuracy and compliance.

**Ganit Inc** Chennai, India

Senior Data Analyst Aug 2021 - Jul 2023

* Led a team of 4 data analyst and developed a financial forecasting machine learning model tailored for a prominent global CPG client which helped their sales team to set target accurately by 11%
* Created an advance machine learning-powered demand forecasting model, operating at the utmost granularity (store-sku) level for over 20 thousand combinations and boosted the accuracy by 15% and reduces SKU wastage
* Streamlined financial reporting processes, reducing reporting time by 65% and increasing accuracy.
* Designed and developed a Timeseries Forecasting engine, incorporating various machine learning models such as ARIMA, BSTS, LSTM, Random Forest, and XG Boost
* Conducted a robust analysis of critical key performance indicators (KPIs) such as Return on Investment(ROI), Retail Price Index(RPI), cannibalization, and competitor performance using statistical models in R



* Implemented substantial modifications to a rule-based model dedicated to detecting fraudulent transactions for a prominent Indian Banking, Financial Services, and Insurance (BFSI) company contributing to a notable 60% reduction in team effort
* Leveraged cutting-edge advanced analytics techniques, including machine learning and deep learning algorithms, resulting in an impressive 13% enhancement in inventory forecast accuracy
* Developed 8 Tableau Dashboards for EDA and project forecasted results, enhancing the efficiency of decision-making

**ACADEMIC PROJECTS**

|  |  |  |
| --- | --- | --- |
|  |  |  |

**Taxi Service Analysis [**[**Link**](https://github.com/Abhisheksinha1830/Data-Engineering-Project-Taxi-data)**]** Apr 2023 – Jun 2023

*Technology stack: Microsoft Excel, SQL, Python, Data Pipeline tool -Mage, GCP, Tableau*

* Designed a normalized database containing 7 entities to store and mapped EER/UML relationship diagram
* Query data using complex queries, including aggregation function, window functions, subqueries to provide the analysis on Taxi data and visualized the data using python and Tableau

**Stroke risk prediction [**[**Link**](https://github.com/Abhisheksinha1830/Stroke-Prediction)**]** Aug 2023 – Sep 2023

*Technology stack: Pandas, NumPy, scikit-learn, XG Boost, SVM, Logistic Regression, Random Forest, Decision Tree*

* Orchestrated the collection of historical health datasets for stroke prediction, ensuring data quality and relevance through meticulous preprocessing.
* Conducted in-depth Exploratory Data Analysis (EDA) to unveil patterns in health data, employing feature engineering techniques like RFE that significantly enhanced the accuracy and performance of predictive models.
* Built and compared multiple machine learning algorithms, with SVM emerging as the top-performing model (95% testing set accuracy), and optimized models through hyperparameter tuning.

**International Employee Immigration Management Data [**[**Link**](https://github.com/Abhisheksinha1830/International-Employee-Immigration-Data-Management)**]** Sep 2023 – Nov 2023

*Technology stack: Pandas, NumPy, SQL, Excel, Tableau, Python*

* Developed extensive Entity-Relationship (EER) and Unified Modeling Language (UML) diagrams, translating the conceptual model into a relational model with MySQL
* Leveraged Python and MongoDB for seamless integration with the database, unlocking analytics capabilities instrumental in monitoring and managing employees effectively.

**Cricket Worldcup23 prediction [**[**Link**](https://github.com/Abhisheksinha1830/cricket-worldcup-23-prediction)**]** Sep 2023 – Oct 2023

*Technology stack: Pandas, NumPy, scikit-learn, Logistic Regression, Random Forest, Gradient Boosting, Python*

* Conducted Exploratory Data Analysis (EDA) to uncover patterns and relationships within cricket data, laying the groundwork for informed decisions in machine learning approaches and feature selection
* Implemented feature engineering techniques to enhance model accuracy, leading to the successful application of machine learning algorithms, with Random Forest emerging as the best-performing model at 76% accuracy

**EEG Classification Model [**[**Link**](https://github.com/Abhisheksinha1830/EEG-Classification-model)**]** Sep 2023 – Nov 2023

*Technology stack: Pandas, NumPy, scikit-learn, CNNs, RNNs, Tableau, Excel, Python*

* Conducted comprehensive data preprocessing on EEG signals, addressing missing values and reducing noise.
* Extracted relevant time and frequency-domain features, training deep learning models (CNNs and RNNs) with measures to prevent overfitting.
* Optimized model performance, evaluating on validation and test sets, and produced insightful visualizations for informed decision-making.

**Customer Segmentation using RFM Analysis [**[**Link**](https://github.com/Abhisheksinha1830/Customer-Segmentation-using-RFM-Analysis)**]** Sep 2023 – Nov 2023

*Technology stack: Pandas, NumPy, scikit-learn, K-Means, Random Forest, Tableau, Excel, Python*

* Conducted comprehensive data preprocessing, ensuring dataset readiness by addressing missing values and cleaning.
* Employed RFM metrics for customer segmentation, generating actionable insights for targeted marketing.
* Utilized K-Means clustering and compelling visualizations to effectively group customers, delivering tailored recommendations that enhanced retention and revenue based on in-depth analysis of customer behavior, satisfaction, product trends, and various business metrics.

**Cleaning, Analyzing and Forecasting Crime Data[**[**Link**](https://github.com/Abhisheksinha1830/Crime-Rate-analysis)**]** Sep 2023 – Nov 2023

*Technology stack: Pandas, NumPy, scikit-learn, ARIMA, BSTS, Tableau, Excel, Python*

* Conducted crucial data preprocessing, EDA, and statistical analysis on a real-world crime dataset from 2020 to present.
* Utilized data visualization to identify crime patterns, seasonal trends, and regional variations, along with examining the correlation between economic factors and crime rates in Los Angeles
* Employed predictive modeling techniques, including ARIMA and BSTS, to forecast future crime trends, contributing to data-driven decision-making foBr crime prevention and intervention.

**Analysis of University Query Database [**[**Link**](https://public.tableau.com/app/profile/abhishek.sinha2708/viz/AnalysisoftheUniversityqueryDatabase/Story1)**]** Nov 2023 – Dec 2023

*Technology stack: Pandas, NumPy, Tableau, Excel, Python*

* Analyzed a university query database to optimize performance, identify structural improvements, and enhance data management efficiency.
* Conducted data cleaning and wrangling using Python, delivering the final insights through a Tableau dashboard for informed decision-making.