Abhradeep Das

Arlington, VA | (571)-259-6367 | abhradeep.das@gwu.edu | LinkedIn | GitHub

EDUCATION

The George Washington University, Washington, DC **Exp. Dec 2024**
Master of Science, Data Science
Relevant Coursework: Intro to Data Science, Intro to Data Mining, Data Warehousing

Techno India University, Kolkata, India **Jul 2023**
Bachelor of Technology, Computer Science & Engineering ; CGPA: 7.98/10
Relevant Coursework: Data Structures and Algorithms, Data Analytics, Operating Systems, Discrete Mathematics

TECHNICAL SKILLS

Programming Languages: **Python, R**, SQL, Java.
Tools: Excel, **Tableau**, Sklearn, Stats Model, Keras, Pandas, Numpy, BeautifulSoup, PyTorch, TensorFlow.
Machine Learning & Applications: Data Pre-processing, Supervised & Unsupervised Learning, Deep Learning.
Certifications: Google Data Analytics Professional Certificate (Coursera), Google's Crash Course on Python (Coursera)

PROFESSIONAL EXPERIENCE

Calcutta University | Research Intern **Aug 2022 - Oct 2022**

- Skilfully processed large volumes of textual data from emails, websites, and text messages, with a primary focus on identifying keywords to combat website payment and identification fraud, as well as email and SMS phishing threats.
- Utilized advanced natural language processing methods, such as BERT and CNN models, to predict sentence context, achieving a commendable accuracy rate of 79.6% and making substantial contributions to bolstering digital communication security.

GlobalShala (with St. Louis University) | Data Analyst Intern **Nov 2021 - Dec 2021**

- Led a 6-member team to efficiently analyze Facebook ad campaign data, identifying and discontinuing the two least effective campaigns. This strategic move led to a 15% budget reduction, optimizing advertising efficiency.
- Utilized Excel and various data visualization techniques like bar charts, line graphs, and histograms for data cleaning and analysis.

ACADEMIC PROJECTS

Credit Card Churn Reduction:

- Conducted comprehensive Exploratory Data Analysis (EDA) on credit card churn data, gaining valuable insights into customer behavior and trends.
- Implemented advanced machine learning models, including logistic regression, decision tree, and random forest. The random forest model demonstrated outstanding performance in identifying customers at risk (recall) at a rate of 83.6%, and achieving a balanced precision and recall with an F1 score of 88.7%. This optimized model holds significant potential to mitigate revenue loss and decrease customer attrition for banking institutions.

Breast Cancer Detection:

- Employed advanced data analysis techniques to construct a precise and efficient model for the early identification of breast cancer.
- Leveraged Numpy and Scikit-Learn to meticulously scrutinize a dataset comprising 100,000+ data points, facilitating the development of a robust breast cancer detection model.

Sentiment Analysis (using movie reviews):

- Data was scraped from IMDb using the request module, then cleaned, split into training/test sets, and prepared with tf-idf embedding using Scikit-Learn's vectorizer, including bigrams.
- Utilized Seaborn for movie sub-category visualization and implemented logistic regression and SGDC models, achieving 75.12% and 58.29% accuracy, respectively.

LEADERSHIP AND VOLUNTEER EXPERIENCE

GOOGLE DEVELOPER STUDENT CLUB **Mar 2020 - Jul 2022**
Actively participated & have presented in Flutter and Machine Learning sessions.