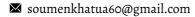
Soumen Khatua

in https://www.linkedin.com/in/soumenkhatua 6296029592

https://github.com/soumenkhatua





₽ PROFILE

As an aspiring Data Scientist Trainee, I possess expertise in Machine Learning and Deep Learning. With a strong statistical background and proficiency in Python, Excel, MySQL and Power BI, I am ready to tackle challenging data science roles, driving impactful business outcomes.

EDUCATION

M.Sc in Applied Statistics and Analytics	8.66	2021 - 2023
Maulana Abul Kalam Azad University 🛮		India, Nadia

B.Sc Statistics 6.66 2018 - 2021Vidyasagar University 🛮 India, West Bengal

SKILLS

Python MySQL Microsoft Excel Git & Github

Data Visualization Power BI Deep Learning Statistical Analysis

(ANN, CNN, Computer Vision) **Machine Learning** API **Time Series Analysis**

Natural Language Processing(NLP)

Flask, Streamlit

Data Scientist trainee 02/2023 - 05/2023

MongoDB

- Developed and implemented various machine learning models, including decision trees, random forests, boosting and neural networks.
- Used deep learning techniques such as convolutional neural networks (CNNs) for image classification.
- Preprocessed and cleaned large datasets for machine learning models using Pandas, NumPy.
- Evaluated model performance using metrics such as accuracy, precision, recall, F1 score, mean squared error, mean absolute error, and R-squared and Adjusted R-squared.
- I possess proficiency in Power BI, enabling me to create robust data visualizations and drive insightful business analytics.

Data Science Intern at iNeuron.ai

■ PROFESSIONAL EXPERIENCE

• Collect and prepare and visualize the data.

- Build the regression model and validate the model.
- Fine-tune the model and present the results.

PROJECTS

Phishing Website Classifier Using Machine Learning Algorithms

• Dataset containing **phishing** and **legitimate** websites from the open source platforms.

- **Analyze** and **preprocess** the dataset by using EDA techniques.
- **Clustering** the training datasets.
- Run selected machine learning model like Random Forest, XGBoost Classifier on the different clusters and Choose the best model for each **clusters** by analyzing the **performance metrics**.

Age Detection using Deep Learning

Developed a **real-time age detection** model using **CNN architectures**.

- Utilized **LeNet**, **AlexNet**, **VGG**, **Inception**, and **ResNet** architectures.
- Applied data **preprocessing** techniques.
- Trained the model using **TensorFlow**.
- Evaluated performance metrics **MAE**.
- Optimized model for real-time inference.

◯ COURSES

Full Stack Data Science Bootcamp

2022 - 2023iNeuron.ai Bengaluru, Karnataka

2022 - 2022

2022 - 2022

10/2022 - 02/2023

Bengalore, India