Soumen Khatua

✉ soumenkhatua60@gmail.com   in https://www.linkedin.com/in/soumenkhatua

○ https://github.com/soumenkhatua   📞 6296029592

PROFILE

As a **Junior Data Scientist** with experience in **Machine Learning** and **OpenCV**, I have developed skills in computer vision techniques and tools, including object detection, image segmentation, and facial recognition. Highly skilled aspiring data scientist with a strong background in Machine Learning and Statistics. Very much proficient in **Python, MySQL, Power BI** and a variety of tools and libraries for **data analysis** and **visualization**. A team player and a highly motivated individual willing to learn new technologies and methods. Seeking a challenging data science role that allows me to apply my skills and expertise to drive business value.

SKILLS

Python

Deep Learning
(ANN, CNN, Computer Vision)

MySQL

Git & Github

Power BI

Statistical Analysis

MongoDB

Machine Learning
Supervised & Unsupervised Learning

Data Visualization

Object-oriented programming (OOP)

API
Flask, Streamlit

Microsoft Excel

Cloud
Heroku

PROFESSIONAL EXPERIENCE

Junior Data Scientist   2023 – present
- Developed a real-time age detection system using **computer vision techniques** and **deep learning** algorithms.
- Preprocessed and cleaned large datasets of facial images to ensure accuracy and completeness.
- Trained and evaluated deep learning models using tools such as **TensorFlow** with **Keras** and **OpenCV** to accurately predict the age of individuals in real-time.
- Collaborated with cross-functional teams to integrate the age detection system into a larger software application.
- Communicated findings and insights through data visualization and presentations .

Data Science Intern at iNeuron.ai 🔗   10/2022 – 02/2023
- Define the problem and the goal of the project.   Bengalore, India
- Collect and prepare the data.
- Explore and visualize the data.
- Build the regression model.
- Validate the model.
- Fine-tune the model.
- Present the results.

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 – 2021
Vidyasagar University 🔗   India, West Bengal

PROJECTS

Phishing Website Classifier Using Machine Learning Algorithms 🔗 2022 – 2022
- 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 **SVM, XGBoost Classifier** on the different cluster of training datasets and Choose the best model for each cluster by analyzing the **performance metrics**.

Scania truck aps-fault detection 🔗 2022 – 2022
- **Description:** My purpose of this project is to predict the fault of Scania truck due to air pressure system or not. For this I used various types of ml-based classification model.
- **Model Used:** Random Forest classifier, Xgboost classifier, decision tree, K-nearest neighbors classifier, gradient boosting, adaboost classifier, logistic regression.

COURSES

Full Stack Data Science Bootcamp 2022 – 2023
iNeuron.ai Bengaluru, Karnataka