Soumen Khatua

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

https://github.com/soumenkhatua soumenkhatua60@gmail.com

https://soumenkhatua.github.io/portfolio_soumen/

PROFILE

Data Science with 1+ years of Industry Experience with expertise in **Attribution Models** and **Machine Learning.** Having a strong **statistical** background, and proficiency in **Python, Advanced MS Excel, MySQL, and R**.

WORK EXPERIENCE

Data Analyst – Data Science, Attribution 09/2023 – present
Ipsos Research Pvt Ltd

- Conducted Placement Level **two stage modelling** Marketing Analysis using **Multiple Linear regression**, employing 1st **bounded regression** for Ad Channels analysis and **Ridge regression** for **placement-level** insights.
- Contributed to developement of **Model** Methodology to enhance **Advanced MMM Model** performance.
- Automated process workflow using **Python**, significantly reducing human intervention.
- Worked closely with Clients to drill down their Marketing strategies and **optimize** them.
- **Languages used**: **Python, SQL, Advanced MS Excel, R.**
- **Industries worked in**: Retail, Pet Care.

Data Scientist trainee 02/2023 – 05/2023
Objectsol Technologies Pvt Ltd

- 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**, and **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**.

SKILLS

Data Science
(Regression and Classification Techniques, Statistical Modeling, Principal Component Analysis)

Machine Learning
Bagging&Boosting : (Decision Trees, Random Forest, AdaBoost, Gradient Boosting, XGBoost)

Deep Learning
(ANN, Convolutional Neural Networks (CNN), Computer Vision)

Programming & Query Language
(Python3, MySQL, R)

PROJECTS

Phishing Website Classifier Using Machine Learning Algorithms

- The dataset contains **phishing** and **legitimate** websites from open-source platforms.
- **Analyze** and **preprocess** the dataset by using EDA techniques.
- **Cluster Analysis** on the training datasets.
- Run selected **machine learning** models like **Logistic Regression**, **Random Forest, Gradient Boosting, and XGBoost Classifier** on the different clusters. Choose the best model for each **cluster** by analyzing the **performance metrics**.

EDUCATION

M.Sc in Applied Statistics and Analytics 2021 – 2023
Maulana Abul Kalam Azad University

B.Sc. in Statistics Honours 2018 – 2021
Vidyasagar University

COURSES

CampusX Data Science Mentorship Program(DSMP) 2023 – 2023

Full Stack Data Science Bootcamp 2022 – 2023
iNeuron.ai