SAMEER MAHAJAN

Portfolio | sam7736@psu.edu | +1 949-501-2750 | ResearchGate | GitHub | LinkedIn

EDUCATION

MS, Data Analytics, Penn State University — May 2023
Courses: Applied Statistics, Data Visualization, Data Mining, Machine Learning, Deep Learning — **GPA: 4.0/4.0**

BE, Computer Engineering, University of Mumbai — May 2021
Courses: Data Structures, Advanced Algorithms, Operating Systems, Machine Learning, Big Data Analytics, NLP — **GPA: 8.9/10**

PROFESSIONAL EXPERIENCE

Research Assistant, Penn State University — Oct 2021 – Present

- Conducting research on data privacy attacks and security threats in federated learning ecosystem to find defense mechanism for mitigation of the same
- Collaborating with Dr. Youakim Badr to implement robust federated learning environment against data poisoning attacks
- **Technologies**: Python, Deep Learning, Federated Learning, Unsupervised ML, TensorFlow, VS Code, Jupyter Notebooks

Data Analyst Intern, Bhangale Hydraulics Pvt. Ltd, India — Jan 2020 – July 2020

- Collected raw data from sales and implemented data preprocessing pipeline to push clean data to a database
- Presented cleaned data using Tableau and provided stakeholders with dashboards that positively helped in decision making
- Developed predictive models for price estimation with analysis to reduce transportation costs that helped increase sales by 15%
- Technologies: Python, NumPy, Pandas, Matplotlib, Regression modeling, Descriptive and Inferential Statistics, MySQL, Tableau

Full Stack Developer Intern, Bhangale Hydraulics Pvt. Ltd, India — Aug 2019 – Oct 2019

- Developed and managed a robust front end for the Website, showcasing catalogs, contact information, and achievements for an effective web presence.
- **Technologies**: Python, Django, React, HTML, CSS, REST API, MySQL, Postman, VS Code

PROJECT EXPERIENCE

Analyzing US Air Pollution Data with Unsupervised Techniques, Penn State University — Oct 2021 - Dec 2021

- Scraped over 1.75 million rows of US Air pollution data of 16 years and used data preprocessing to achieve cleaned data
- Assessed if there is a regional effect on air pollutant values using unsupervised Machine Learning to cluster according to US cities
- Developed SARIMAX models to predict seasonal variations of air pollutants and predicted pollutant values for upcoming years
- **Technologies**: Python, Unsupervised ML, Kmodes, Kmeans, PCA, Dimensionality Reduction, SARIMAX, Regression, Plotly, Folium

Happiness Analyzer: Statistics and Machine Learning, Penn State University — Aug 2021 - Oct 2021

- Preprocessed data and developed statistical methods to prove a hypothesis that Covid-19 had no effect on world happiness score
- Identified linear relations among dependent and independent variables using Multivariate Regression
- Constructed visualizations using Tableau, built dynamic dashboards to explain impact of variables that govern happiness score
- **Technologies**: Python, Statistics, Hypothesis Testing, Regression, Tableau, t-tesbt, z-test, Jupyter Notebooks

Early Detection of Alzheimer's Disease using Deep Learning, Independent Project — July 2021 - Aug 2021

- Implemented Deep Learning models to detect and classify 4 stages of Alzheimer's Disease with 95.67% accuracy
- **Technologies**: Python, TensorFlow, Keras, InceptionV3, Xception, ResNet152V2, DenseNet CNNs, Transfer Learning, Google Colab

ShareConn, a social media website, University of Mumbai — Nov 2020 – Mar 2021

- Lead a team of 4 to develop a social media website with functionalities like sharing images, videos, comments and liking posts
- The users had the ability to create subconns which act as community groups, under which all members of that subconn can post
- **Technologies**: Python, Django, React, HTML, CSS, Bootstrap, REST, Postman, VS Code

Movie Recommender System with Sentiment Analysis, Independent Project — Nov 2020

- Headed a team to implement a movie recommending web-app and performed sentiment analysis on movie reviews
- The web-app provides every movie cast, directors, reviews, and accompanying sentiment analysis as good, bad, or neutral
- **Technologies**: Python, Django, NumPy, Pandas, NLTK, ReactJS, HTML, CSS, VS Code

SKILLS

- **Python**: Django, Flask, NumPy, Pandas, Scikit-learn, BeautifulSoup, Matplotlib, NLTK, TensorFlow & Keras, PyTorch, OpenCV
- **Modeling**: Regression, Classification, Decision Tree, KNN, Naïve Bayes, SVM, Time Series Analysis, Convolutional Neural Networks, Recurrent neural Networks, Kmeans, Kmodes, PCA, SARIMAX, ARIMA
- **Programming Languages & Tools**: R, C, SQL, Github, Anaconda, Postman, Tableau, KNIME, Minitab,

RESEARCH PUBLICATIONS

- Robust Federated Machine Learning and Cybersecurity leveraging Causal Inference against Privacy Attacks — 2022 – Present
- Deep Learning for Internet of Things Applications, Springer AI IOT 2020 Internet of Things — 2022
- Causal Inference and its Applications in Healthcare and Finance, JETIR — 2021
- Exploration-Exploitation problem in Policy-Based Deep Reinforcement Learning, IJEAT — 2021