Md Shahrukh

ASPIRING DATA SCIENTIST

mdshahrukhbme@gmail.com
+91 7908920410
Kolkata, India
Linkedin

I am an aspiring Data Scientist with a solid foundation in statistics and machine learning, supported by hands-on project experience. I earned a Post Graduate Diploma in Applied Statistics from the Indian Statistical Institute in Kolkata. I recently completed internships at Upskill Campus and UniConverge Technologies Pvt. Ltd., where I developed a Crop and Weed Detection project using YOLO. Additionally, I gained experience in groundwater level analysis at ISI, Kolkata. My skill set includes proficiency in Python, R, Pandas, NumPy, and scikit-learn, and I hold certifications in AI in Healthcare from Stanford and Managing Big Data with MySQL from Duke. Furthermore, I have over 3 years of experience in the healthcare sector as a field service engineer.

CORE COMPETENCIES

Proficiencies

| Regression |Supervised models Python | R | Power BI | SQL | Advance Excel | Visualization | Team Management | Team Management | Problem Solving |

Core Skillsets

| Statistics | Machine Learning | Data Analysis | Feature Engineering | Time Series Analysis and Forecasting | Tensorflow | NLP | Large Language Models(LLMs) | Pyspark

Tools

| R Studio | Jupyter Notebook | VS Code | MS Excel | MS Power Point | Power BI | Github |

CERTIFICATIONS

- AI in Healthcare, Stanford University
- Managing Big data with MySQL, Duke University
- Drive Insights from BigQuery, Google Cloud Skills Boost
- Python and Pandas for Data Engineering, Duke University

POSITION OF RESPONSIBILITY

- Placement Representative during ISI PGDAS course (Feb 2023 – Feb 2024)
- An Active Member of the Institute of Actuaries of India (IAI)

INTERESTS

Cricket

Travelling

Data Science and Statistics

LANGUAGES

English

Hindi

Bengali

EXPERIENCE

Data Science Intern
July 2024 – Oct 2024

Upskill Campus and UniConverge Technologies Pvt. Ltd.

- I contributed to a crop and weed detection project leveraging machine learning and computer vision. I developed a system for selective pesticide spraying, which aimed to reduce waste and prevent pesticide contamination of crops, promoting sustainable agriculture.
  I designed and implemented object detection models using YOLO (You Only Look Once) and OpenCV in Python, optimizing the system's efficiency in identifying and distinguishing crops from weeds. I applied machine learning algorithms to enhance the accuracy of the detection system, ensuring that only weeds were targeted for pesticide spraying.

  I applied deep learning techniques and statistical methods to tackle complex agricultural challenges, gaining hands-on experience in real-world problem-solving. My work focused on integrating AI into agriculture to drive more sustainable practices.
- Gained hands-on experience in applying machine learning and computer vision techniques to real-world agricultural challenges.

Tools Use/Learn: Python, Pandas, Numpy, YOLO, Tensorflow, OpenCV, Team Management, Leadership, Timeline Management.

Project Intern under Dr. Debasish Sengupta
Mar 2024 – June 2024

Indian Statistical Institute

- I worked on a project named "Analyzing the Groundwater Level in West Bengal," which investigates the groundwater crisis in West Bengal, India. The main objective was to predict the impact of the upcoming monsoon season on groundwater levels. The project provided valuable insights to understand and address the groundwater crisis in West Bengal using comprehensive data analysis techniques such as machine learning and statistical analysis.
- In this project, I worked with block-wise data spanning 10 years, including pre-monsoon and post-monsoon data. I cleaned the data and applied various statistical techniques, data visualization, and data mining algorithms to extract meaningful insights from it.
- Achieved consistent pre-monsoon groundwater levels throughout the year and significantly increased post-monsoon groundwater levels over the years.

Tools Use/Learn: R, Data Visualization, Statistical Analysis, Data Exploration, Deployment etc.

Movie Recommender System
July 2024

- Developed a comprehensive movie recommendation system using machine learning techniques. The system processes and analyzes movie metadata to suggest similar films based on user input, leveraging advanced text processing and similarity algorithms.
- Engineered features by parsing genres, keywords, cast, and crew, and created consolidated tags. Applied text normalization and vectorized text data using CountVectorizer, excluding stop words. Implemented cosine similarity to measure likeness between movies and constructed a recommendation function for top similar movies. Utilized pickle for model persistence.
- Demonstrated proficiency in data preprocessing, feature extraction, and machine learning algorithms.

Tools Use/Learn: Python, Pandas, NumPy, scikit-learn, Text Vectorization, Cosine Similarity, Pickle for model persistence

EDUCATION

- **Post Graduate Diploma in Applied Statistics with Data Analytics and Official Statistics** — **Aug 2022 – Feb 2024**
  Indian Statistical Institute, Kolkata, India
  Percentage: 86.05%
- **Bachelor of Science in Mathematics** — **2018 - 2021**
  Ranchi University
  Percentage 76.96%

ACHIEVEMENTS

I was ranked 10th in the Indian Statistical Institute's Postgraduate Diploma in Applied Statistics entrance examination.