SIDHARTH GUPTA

+(91)7543898755 ⋄ hello.sidharth.gupta@outlook.com
linkedin.com/in/gupta-sidharth

PROFESSIONAL SUMMARY

Analytical and meticulous data professional with a Master's degree in Statistics and a Diploma in Data Science. Proficient in Python, R, and SQL with hands-on experience in statistical methodologies, machine learning, and data analytics. Demonstrated ability to solve real-world problems, including class imbalance, regression, and classification challenges, using advanced algorithms and techniques. Passionate about leveraging AI-driven solutions to deliver actionable insights and meet business needs effectively.

SKILLS

Programming and Query Languages: Python, R, SQL
Data Science and Machine Learning: Statistics, Regression Models, Hypothesis Testing, Probability Theory, Data Preprocessing, Feature Engineering, Model Evaluation, Optimization
Libraries and Frameworks: Pandas, NumPy, Matplotlib, Seaborn, SciPy, Scikit-learn, Imbalanced-learn, Tesseract OCR, OpenPyXL, FFmpeg, OpenCV, XGBoost, LightGBM, CatBoost, PyTorch (foundational), TensorFlow (foundational)
Tools and Platforms: Jupyter Notebook, RStudio, MS Excel, MS Word, MS PowerPoint, LaTeX, Kaggle
Languages: English (Fluent), Hindi (Fluent), Maithili (Native)

PROJECTS

ML Project: Classification Model for Bank Telemarketing Campaign Success Prediction

- Implemented a comprehensive preprocessing pipeline to streamline model-based imputations, handling of categorical features, stratification, and standardization.
- Addressed class imbalance using techniques like SMOTE and Borderline-SMOTE, achieving a macro F1 score above 0.75.
- Trained and fine-tuned diverse classifiers, including Random Forest, Logistic Regression, XGBoost, AdaBoost, LightGBM, CatBoost, and neural networks, to evaluate and compare their performance.
- Generated actionable insights using advanced visualizations, heatmaps, and detailed performance analytics.

Data Processing Pipeline: Automated Video-to-Excel Data Extraction and Cleaning

- Designed a pipeline for processing videos, extracting frames, and performing OCR to gather structured data from video content.
- Utilized FFmpeg for frame extraction and Tesseract OCR for text recognition, enhancing accuracy in noisy environments through advanced preprocessing techniques like grayscale conversion and adaptive thresholding.
- Streamlined batch processing for enhanced scalability and efficiency by using Bash scripting, enabling seamless handling of multiple video sources while significantly reducing processing time.
- Developed Python scripts for automating data cleaning, standardizing formats, and saving results to Excel. Included detailed logging to track execution and ensure data accuracy.

Master's Thesis: Regression Models Using Composite Weibull-WTP Model

- Engineered an innovative composite statistical model by integrating the Weibull and tempered Pareto distributions, addressing limitations in modeling heavy-tailed data.
- Integrated the model into a regression framework, enabling robust predictions and analysis of real-world datasets such as financial risk and survival data.
- Conducted rigorous evaluations, showcasing improved adaptability and performance over traditional methodologies.
- Authored a detailed academic report and presented findings using LaTeX, effectively communicating complex statistical concepts.

CERTIFICATIONS

Diploma in Data Science, IIT Madras (Pursuing)
Coursework includes Mathematics for Data Science, Statistics for Data Science, Machine Learning Foundations, Business Data Management, Machine Learning Techniques, Business Analytics, and Tools in Data Science. The program also includes hands-on projects and skill enhancement courses focused on machine learning and data management.

EDUCATION

M.Sc. in Statistics Dec 2022 – May 2024
Central University of Rajasthan *CGPA: 7.1*

B.Sc.(Hons) in Statistics Aug 2018 – Oct 2021
Patna Science College *Percentage: 74%*

ADDITIONAL INFORMATION

Hobbies: Badminton, Chess, FPS gaming, and staying updated with the latest tech news.