MEDHASWETA SEN
WASHINGTON, DC | 202-956-8451 | msen22@gwu.edu
linkedin.com/in/Medhasweta | github.com/Medhasweta

PROFESSIONAL SUMMARY

- Enthusiastic and forward-thinking data enthusiast, poised to delve into the theoretical underpinnings of machine learning and data science with a humble approach to re-examining the fundamentals that drive the field.
- Committed to exploring innovative perspectives in data science, with an aim to contribute to a deeper, more nuanced understanding of algorithmic processes and their real-world implications.

EDUCATION

The George Washington University, Washington DC Anticipated May 2024
Master of Science in Data Science
Relevant Coursework: Machine Learning, Deep Learning, NLP for Data Science, Time Series Analysis and Modelling
Awards: Global Leaders Fellowship of 16000 USD

University of Calcutta, Kolkata, India August 2022
Bachelor of Science in Statistics
Relevant Coursework: Multivariate Analysis, Statistical Inference and Hypothesis Testing, Linear Models, Linear Algebra, Real Analysis
Awards: Rama Chowdhury Award for Academic Excellence, Ranked University Second

TECHNICAL SKILLS

Programming Languages: Python, Java, R, C/C++, Bash/ Shell, MySQL, GraphQL, MongoDB, Neo4j
Tools: Tableau, Power BI; Anaconda, Git, RStudio, PyCharm; Hugging Face, Scikit-Learn, NLTK, SpaCy, TensorFlow, Keras, PyTorch, LangChain, Dash, Plotly, AWS, Azure, Google Cloud Platform (GCP), Docker, LaTeX, StreamLit.
Machine Learning & Applications: Data Pre-processing, Exploratory Data Analysis, Data Modelling, Feature Engineering, Model Evaluation, Model Optimization, Machine Learning Algorithms, Computer Vision, Natural Language Processing, Automatic Speech Recognition, Anomaly Detection
Certifications: Machine Learning Pipelines with Azure ML Studio; Custom Attribution Modelling with Google Analytics; Building Custom Regional Reports with Google Analytics with Coursera Project Network, Data Analyst in R with Dataquest

TECHNICAL PROJECTS

Explainable AI: A No-Code XAI Platform with Custom UI | Python December 2023
Pioneered a cutting-edge, no-code platform for object detection, incorporating hyperparameter tuning capabilities. This project skilfully integrated GradCAM, LSTM, MLP, and other XAI tools into an intuitive, custom UI. Focused on enhancing user comprehension of complex AI models, This offers an accessible, insightful experience in optimizing and understanding object detection models through advanced XAI methodologies and interactive hyperparameter adjustments

AI-Driven Scholarly Assistant with Interactive UI | Python December 2023
Spearheaded the creation of the AI-Driven Scholarly Assistant, an innovative platform with an interactive UI, integrating advanced NLP models like BERTopic and Longformer for efficient research summarization and Q&A. This project transformed user engagement with scholarly content, leveraging large language models to enhance comprehension and interaction within the academic community

Bengali Speech Transcription and Classification | Python December 2023
Enhanced Wav2Vec2 and Whisper models for Bengali speech recognition and sentiment analysis, incorporating custom classification heads into the models. This adaptation, coupled with addressing the scarcity of diverse Bengali audio data through a developed custom dataset, significantly improved the models' ability to process and classify Bengali speech accurately, showcasing the impact of tailored model modifications and dataset integration in specialized audio processing tasks

Movie Recommendation Engine | Python November 2023
Engineered a sophisticated Python-based recommendation engine leveraging Pandas for intricate data integration and pre-processing, utilized NumPy for efficient array manipulations, and applied natural language processing (NLP) techniques for feature extraction. Implemented a cosine similarity algorithm with scikit-learn for nuanced movie matching, and encapsulated the functionality within a user-friendly Streamlit web interface

NMT with Attention | Python October 2023
Implemented a Neural Machine Translation (NMT) model using a sequence-to-sequence architecture with Bahdanau's Attention mechanism, enabling effective translation across languages by focusing on specific parts of input sequences, thus enhancing the model's applicability in real-time translation, document conversion, and development of translation tools for low-resource languages

Neural Text Synthesis: WaveNet & GPT Implementation Project | Python September 2023
Independently implemented advanced neural network architectures, including WaveNet and GPT, following Andrej Karpathy's tutorial. This involved building and training models from scratch in PyTorch, encompassing concepts such as character-level text generation, vocabulary encoding, sequential data processing, batch normalization, embedding layers, and hierarchical network design. Demonstrated proficiency in deep learning principles, data pre-processing, and algorithm optimization

Assessing the Effect of FDI Flows, as Percent GDP, on Unemployment Rates Stratified by Education Levels | R August 2023
Investigated nuanced interplay between FDI and unemployment rates stratified by education levels (2005-2021, 40 countries) using regression analyses; discovered FDI positively impacts unemployment for higher-educated individuals, guiding informed policymaking to prioritize FDI promoting educational advancement. This project is under consideration for publication

Neural Network Optimization Techniques | Python May 2023
Evaluated impact of training features, layers, activation functions, and learning rates on performance of several optimization techniques including SGD, CG and Levenberg Marquardt for backpropagation training using only self-developed NumPy codes

Energy Consumption Time Series Forecasting | Python May 2023
Utilized custom-coded SARIMA for order determination and parameter estimation alongside ARIMA, Holt-Winters, and LSTM for energy consumption forecasting in Tétouan, Morocco. While LSTM achieved RMSE of 0.47, SARIMA was deemed superior due to its whiter residual error, underscoring its precision in energy prediction

Tic-Tac-Toe | Python March 2023
Developed a Python-based AI model to play Tic-Tac-Toe using reinforcement learning, specifically Q-learning, enabling the AI to learn and adapt optimal strategies through gameplay; this process not only enhanced my understanding of AI in dynamic environments but also significantly improved my skills in algorithm optimization and data structure utilization, showcasing the potential of reinforcement learning in game AI and its broader applications

RELEVANT WORK EXPERIENCE

George Washington University Biostatistics Center- Student Technical Support Specialist-III June 2023 – Present

- Optimizing existing code to accelerate algorithm convergence, leveraging refined heuristics, faster optimization techniques, and lower-level languages, resulting in significant performance improvements
- Devising innovative visualization techniques for profile plots of maximum likelihood function, enabling detailed assessment of 'Optim' package optimizer convergence patterns
- Conducting exhaustive time profile analysis, identifying, and rectifying time-intensive code areas through custom functions, significantly boosting overall efficiency
- Upholding code robustness and accuracy through comprehensive unit tests, ensuring seamless performance between original and optimized versions
- Seamlessly integrating advanced functionalities from R packages such as Rfast and 'OptimX', further enriching package's capabilities, currently in penultimate stage, preparing for publication to CRAN

Estelle and Melvin Gelman Library – Student Data Consultant-IV November 2022 - Present

- Applying knowledge in Python, R, and general statistics to support diverse range of research projects, including those from public health, applied economics, anthropology, machine learning, epidemiology, and public policy
- Collaborating with research teams from various disciplines, resulting in co-authored papers with medicine-focused team at George Washington University Hospital and Center for the Advanced Study of Human Paleobiology
- Handling sensitive data with care, prioritizing data privacy and confidentiality to maintain trust of researchers and participants
- Contributing to GW Libraries and Academic Innovation workshops, collaborating with colleagues to promote productive academic environment
- Aiding doctoral researchers, sharing insights and expertise, while continually learning from vast array of academic projects
- Providing tailored tutoring support as and when required to GW students, faculty, and staff, adapting to both in-person and virtual environments to facilitate their academic growth and understanding

KYC Hub (Hyperion Technologies Ltd) - Data Engineering Intern July - December 2021

- Collaborated in development of data pipeline, utilizing Python and SQL, which halved data processing duration. This initiative enhanced operational efficiency, improved response time to clients
- Unearthed hidden correlations amongst entities for clients using Natural Language Processing within financial crime datasets, contributing to prevention of potential financial misconduct
- Utilized machine learning to predict customer churn, resulting in noteworthy 10% increase in retention, thereby honing melding data science skills with business insights
- Presented machine learning and data science insights to key clients, stakeholders, and senior officials, enhancing transparency and collaboration resulting in recognition with commendation for outstanding performance as intern

Exposys Data Labs – Data Science Intern January – June 2021

- Contributed to state-of-the-art machine learning algorithm to pinpoint fraud in financial sector, elevating prediction accuracy by 15%, emphasizing dedication to addressing financial services challenges, was seamlessly integrated into company's operational framework, fortifying customer trust and retention
- Architected sophisticated data pipeline capturing insights from social media platforms resulting in more effective decoding of customer sentiments that highlighted market trends, steering firm's marketing endeavors for maximized impact
- Leveraged Neural Network techniques, particularly long short-term memory, to forecast stock prices from Calcutta Stock Exchange's daily trade data resulting in minimizing losses for novice trader clientele

Silver Neuron Consulting LLP- Data Science Intern April – December 2020

- Partnered with key players in Coal Mining Industry to optimize their data collection methodologies leading to streamlined data maintenance systems, ensuring accurate and efficient data storage, ultimately aiding in better decision-making processes
- Delved deep into transport data to identify inefficiencies by leveraging regression and trend analysis and highlighting key areas of potential losses thereby helping clients reduce their product transportation losses significantly, leading to tangible financial savings
- Played integral role across diverse verticals, from debugging intricate Data Structures to fostering academic collaborations by assisting in bridging digital skill divide, connecting students with industry veterans, and contributing to enriching workshop experiences

RESEARCH WORK

- Diana Jaber, Tina Vaziri, Jennifer Beckerman, Medhasweta Sen, Yuan Rao, Pavani Chalasani, Sharad Goyal. *Breast Cancer Clinicopathological Characteristics and Outcomes in Patients Living with HIV/AIDS: A Systematic Review and Meta-Analysis*. American Society of Clinical Oncology Annual Meeting. Poster presentation, June 2023

  Annual Meeting Research Award, 2023 American Society of Clinical Oncology Annual Meeting. Granted for abstract: *Breast Cancer Clinicopathological Characteristics and Outcomes in Patients Living with HIV/AIDS: A Systematic Review and Meta-Analysis*