Amitabha Dey

Greensboro, NC 27405
amitabhadey.github.io

a_dey@uncg.edu
linkedin/amitabhadey
github/amitabhadey

Education

- **University of North Carolina at Greensboro** — Greensboro, North Carolina
  Masters in Computer Science (CGPA: 3.67) — *August 2021 - May 2023*
  - Relevant Coursework: Algorithm Analysis and Design, Big Data and Machine Learning, Data Science, Advanced Database Systems, Software Engineering, Data Visualization, Introduction to Probability.

Work Experience

- **University of North Carolina at Greensboro** — Greensboro, North Carolina
  Lecturer, Department of Computer Science — *August 2023 - Present*
  - CSC 105: Data, Computing, and Quantitative Reasoning: Fall 2023
  - CSC 250: Foundations of Computer Science I: Fall 2023
  - CSC 330: Advanced Data Structures: Spring 2024
  - CSC 362: System Programming: Spring 2024, Fall 2024

- **DevResonance Ltd.** — Dhaka, Bangladesh
  Data Scientist — *January 2018 - May 2020*
  - Utilized Python to implement a CNN model on 1TB of unstructured data, performed PCA and other dimensionality reduction techniques to reduce process time by 20% and improved classification accuracy by 15% by optimizing loss function. Increased customer retention rate by 13% as a result of these improvements.
  - Generated dynamic and interactive 3D visualizations with both linear and non-linear trendlines by integrating Plotly and Streamlit to allow clients to evaluate the impacts of interventions and monitor progress. Won grants of over $50,000 from Bill & Melinda Gates Foundation, UNICEF, WHO, Save the Children, etc.

- **Redgreen Corporation** — Dhaka, Bangladesh
  Data Science Intern — *July 2017 - September 2017*
  - Developed and implemented predictive regression models to project future sales by constructing feature space, performing data-preprocessing steps, and doing PCA resulting in an 87% accuracy, 12% better than previous years.
  - Built models to predict the possibility of faulty products and identify the manufacturers responsible. Cutting these manufacturers reduced the number of faulty components in the next quarter by 35% and increased MRR by $5K/mo. Developed a marketing analytics metrics dashboard to monitor sales conversion rate from Facebook Ads.

Technical Skills

- **Frameworks and Libraries:** TensorFlow, PyTorch, Hugging Face Transformers, spaCy, AllenNLP, Gensim
- **Pre-trained Language Models:** GPT-4, BERT, RoBERTa, T5, XLNET, ALBERT, DistilBERT
- **Techniques:** Transformer Architecture, Attention Mechanism, Transfer Learning, Seq2Seq, Beam Search, RLHF, Knowledge Distillation
- **Technologies:** Google Cloud Natural Language, AWS Comprehend, Azure Text Analytics, ElasticSearch, Word2Vec, GloVe, FastText, Docker, UTM
- **Statistics:** Probability, Hypothesis Testing, Regression, Time Series, Bayesian Statistics, K-means Clustering
- **Visualization:** Matplotlib, Plotly, Seaborn, GGplot2, Geoplotlib, Tableau, Qlik, D3.js, Microsoft Excel & Visio, Google Sheets
- **Language:** Python, R, SQL, PHP, Java, C, C++, JavaScript, Hadoop, Bootstrap, Ruby on Rails, MATLAB

Selected Publications

- Dey, Amitabha, et al. "Fake news pattern recognition using linguistic analysis." 2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision& Pattern Recognition (icIVPR). IEEE, 2018.
- Dey, Amitabha, and Shan Suthaharan. "LDEB–Label Digitization with Emotion Binarization and Machine Learning for Emotion Recognition in Conversational Dialogues." arXiv preprint arXiv:2306.02193 (2023).

Selected Research

- **Transformers for Multimodal Information Retrieval in Large-Scale Document Collections:** Utilized Hugging Face Transformers for fine-tuning BERT and RoBERTa on text data. Integrated CLIP from OpenAI to handle image data, enabling cross-modal embedding. Implemented a dual-encoder architecture to process text and images and cross-attention layers to fuse multimodal information.
- **Knowledge Distillation for Efficient Transformer-Based Summarization Models:** Implemented T5 and BART for abstractive summarization. Transferred knowledge from large models to DistilBERT by utilizing knowledge distillation. Used a mix of cross-entropy loss on the student's outputs and distillation loss on the intermediate layers to optimize DistilBERT's performance.
- **Robust Named Entity Recognition Using Adversarial Training:** Fine-tuned BERT using adversarial training techniques and leveraging the TextAttack library to generate adversarial examples and augment training data. Used gradient-based perturbation methods to craft adversarial examples that challenge model's robustness.
- **Unsupervised Machine Translation with Multilingual Pre-training:** Used mBART for multilingual pre-training and incorporated back-translation and denoising autoencoding. Fine-tuned the pre-trained model on unsupervised translation tasks using a combination of back-translation and dual learning methods.
- **Contextualized Sentiment Analysis in Financial Texts Using Transformers:** Utilized FinBERT on financial sentiment dataset, FiQA. Implemented a multi-task learning approach where the model is jointly trained on sentiment classification and domain-specific tasks like volatility prediction or risk assessment. Used contextualized embeddings to capture nuances in financial language, and employed attention mechanisms to focus on key phrases and entities.

Selected Projects

- **Loan Defaulter Prediction:** Performed exploratory data analysis, feature engineering, and five-fold cross-validation. Applied Regularization - Ridge, Lasso, Elastic Net - on a Linear model to predict loan defaulting probability of A/C holders. Optimized loss function using Stochastic Gradient Descent. Improved classification accuracy by 10% compared to previous models.
- **Image Scraping WebApp:** Developed WebApp using Streamlit. Performed Canny Edge Detection, Convex Hull Contour Detection and Adaptive Thresholding with OpenCV. Deployed to Heroku. Developed optimized web crawlers using BeautifulSoup and Selenium - bypassed CAPTCHA & credential authentication. Exports JSON file.
- **Voice Controlled Jarvis:** Created a voice-controlled program using gTTS and speech recognition; the app can search and play songs on YouTube, search images, give weather updates, report time and date, report breaking news; Created Python libraries for separate components and functions.
- **Sentiment Analysis on Financial News Articles:** Conducted sentiment analysis on financial news articles to predict stock movements. Preprocessed text data using tokenization and TF-IDF. Implemented a Long Short-Term Memory (LSTM) model with attention mechanism, optimizing with Adam optimizer. Achieved a 15% accuracy improvement over traditional models.
- **Named Entity Recognition for Legal Documents:** Developed a Named Entity Recognition (NER) system to extract entities from legal texts using a fine-tuned BERT model. Added a CRF layer for label dependency capture. Improved entity extraction accuracy by 12% over baseline models.

Selected Honors and Awards

- **UNCG Outstanding Graduate Student Award (2023):** Awarded the most prestigious award by the Department of Computer Science for the academic year 2022-23 in recognition of scholarly accomplishment and contribution to the department.
- **UNCG Merit Scholarship (2021):** Awarded $16,000 for 14 months and In-state and Out-of-state full tuition waiver by the Department of Computer Science and the Graduate School.
- **The Daily Star Award (2010):** Awarded the National Daily Star Award for Edexcel IGCSE Students in 2010 for academic results - Further Mathematics (A*), Mathematics (A*), Chemistry (A*), Physics (A*), Economics (A), English (A), Bengali (A).