Shania Mitra

Indian Institute of Technology Madras

shaniamitra9@gmail.com • shania99.github.io • Shania99

in shania-mitra

Education

Indian Institute of Technology, Madras (IITM) **Jul '18 - Jul '23**

Integrated M.Tech in Data Science, B.Tech Chemical Engineering *CGPA: 9.24/10*

CGPA: 9.71/10 (Data Science) and 9.24/10 (B.Tech), Minor in Computing

- **Research Interests:** Explainable AI, Fairness in AI
- **Relevant Coursework:** Pattern Recognition and Machine Learning, Natural Language Processing, Process Optimization, Computer Architecture and Organization, Design and Analysis of Algorithms
- Department **Rank 1** out of 27 in the batch of dual degree students in Chemical Engineering 2018

Publications

- **S. Mitra** and A. Tangirala, "Causal Discovery from Natural Language Text Using Context and Dependency Information", Society of Instrumentation and Control Annual Conference 2022 **(Accepted)**
- K. Sarathi*, **S. Mitra***, D. Padmanabhan, and S. Chakraborti "Counterfactuals as Explanations for Monotonic Classifiers", XCBR Workshop, International Conference for Case Based Reasoning 2022 **(Accepted) (Shared First Author)**
- S. Chinta, **S. Mitra** and R. Rengaswamy "Machine learning based QSPR approaches to predict solvation free energy of Quinone molecules for flow battery applications", Computers and Chemical Engineering (To be submitted)

Research Experience

- **Counterfactuals as Explanations for Monotonic Classifiers** **[Paper]**

  Guide: Prof. Sutanu Chakraborti and Prof. Deepak P, IIT Madras *Jan '22 - Aug '22*

  - Developed a novel explanation algorithm to generate actionable counterfactuals incorporating domain knowledge using monotonic constraints
  - Generated counterfactuals 10% closer to the queries as compared to the baseline, with 70% more coverage

- **Subset Selection Algorithms for Prediction of LogP** **[Report]**

  MITACS Globalink Scholarship, Guide: Prof. Paul Ayers, McMaster University *May '22 - Jul '22*

  - Predicted logP on the PhysProp dataset to demonstrate the effectiveness of DiverseSelector
  - Profiled the performance of all subset selectors in DiverseSelector for different selection proportions and data size

- **Senior Thesis: Causal Discovery in Natural Language Text** **[Paper]**

  Guide: Prof. Arun Tangirala, IIT Madras *May '21 - Sep '22*

  - Developed a novel BiLSTM and Node2Vec based architecture to perform causal extraction on text
  - Built a framework that uses linguistic features and relationships between words, using dependency graphs
  - Achieved a 38% increase in F1-Score over the baseline of BiLSTM-Conditional Random Field on BECAUSE 2.0

- **Estimation of Overlap across Interdisciplinary Courses** **[Paper]**

  Guide: Prof. Sutanu Chakraborti, IIT Madras *May '21 - Sep '21*

  - Estimated extent of overlap among interdisciplinary courses across departments at IIT Madras
  - Identified overlapping concepts using phrase extraction techniques
  - Obtained lists of overlapping courses for 100+ cases; Improved average nDCG@10 by 5% over baseline

- **Caching in Deep Learning Architectures for Robustness** **[Report]**
  Guide: Prof. Pratyush Kumar, IIT Madras *May '20 - Sep '20*
  - Conceptualised a cache-like framework on existing models to accelerate inference time, maintaining similar accuracy
  - Used triplet loss to retrain a segment of existing neural network architectures and obtain similar hidden output vectors for similar inputs
  - Implemented a shallow neural network to emulate kNN on the modified hidden layer outputs
  - Reduced the inference time of DenseNet-121 by 23% maintaining top-1 accuracy at 74.32% on CIFAR100 (original accuracy: 76.58%), while increasing robustness to adversarial attacks

- **Graph Algorithms for Kidney Exchange** **[Report] [Website]**
  Guide: Prof. Meghana Nasre, IIT Madras *May '20 - Sep '21*
  - Built an Integer Linear Programming model to facilitate kidney transplant among patient-donor pairs having an incompatible but willing family donor
  - Improved over current system by incorporating new parameters to maximize the likelihood of a successful match
  - Designed a stable platform for hospitals to automate the matching process

- **Predicting Solubility of Quinone Derivatives for Flow Batteries** **[Paper]**
  Guide: Prof. Raghunathan Rengaswamy, SENAI Lab, IITM *Apr '20 - May '22*
  - Conceptualised a robust model to estimate the solvation free energy of quinone derivatives
  - Demonstrated that model learnt fundamental organic chemistry principles in an unsupervised fashion
  - Identified novel Quinone derivatives to improve battery performance, significantly improving on literature

Professional Experience

- **Goldman Sachs Inc., Model Risk Management Division** **Bangalore**
  Statistical Modelling of Revolver Bias in Loans *Jun '21 - Jul '22*
  - Performed extensive parametric and non-parametric hypothesis testing on loan data from 3 divisions to detect the presence of revolver bias across stress and non-stress periods
  - Rejected current formulation of bias and devised a new one showing improved statistical properties and interpretability
  - Built regression model for revolver model wrt. macroeconomic factors & achieved an adjusted R^2 score of 0.938
  - Received pre-placement offer

- **Bewgle Inc.** **Bangalore**
  Review Classification using BERT *May '20 - Jul '20*
  - Aimed to predict product categories and locate keywords in reviews from the Amazon Review Dataset
  - Built BERT models for each product category and tested predictions on customer reviews from E-commerce sites
  - Achieved an average validation F1-score of 0.99 over different categories
  - Received return offer

- **Antariksh Waste Ventures** **Chennai**
  Image Inferencing using Deep Learning *Dec '20 - Jan '21*
  - Predicted incorrect placement of dustbins, using deep learning, with a limited dataset of 50 images
  - Detected boundaries of bin by transfer learning on YOLO, supplemented limited data using data augmentation
  - Performed binary multi-label classification to detect occurrence of overflow and/or pileup of garbage with 96% accuracy

- **DataHive Labs** **Bangalore**
  Breakdown Detection in Time Series Sensor Data *May '19 - Jul '19*
  - Predicted sequences similar to input queries from sensor data, using an epsilon nearest neighbour graph
  - Used Gray Level Co-occurence Matrix (GLCM) and FLANN Feature Matching on spectrogram images of time sequences
  - Received return offer

Course Projects

- **Natural Language Processing CS6370**: **[Report] [Code]**.
  Built search engine on Cranfield Dataset using Vector Space Model, Latent Semantic Analysis, Explicit Semantic Analysis

- **Pattern Recognition and Machine Learning CS5691**: **[Report] [Code]**.
  Designed LightGBM model to suggest cycling tours to bikers & achieved a MAP of 0.71, ranking 19 out of 150 team
- **Mathematical Foundations for Data Science CH5091**: **[Report] [Code]**.
  Analysed data of COVID-19 patients from the Indian Council of Medical Research (ICMR) and accurately predicted hotspots and future cases for the 1st wave of COVID-19 in India
- **Machine Vision and its Applications ME6326**: **[Report]**.
  Built shell classifier by ensembling histogram, GLCM, template & FLANN feature matching method
- **Foundations of Computer System Design CS2300**: **[Report][Code]**.
  Designed a computer bottom-up from a set of logic gates
- **Introduction to Data Analytics ID4160**: **[Report] [Code]**.
  Performed hierarchical clustering on Salonpas product review dataset using TextRank, LDA and HDBSCAN
- **Big Data Laboratory CS4830**: **[Report] [Code]**.
  Predicted parking violation on the NYC Dataset using cloud functions and in real-time using Kafka streaming

Technical Skillset

- **Languages:** Python, C/C++, HTML, OpenMP, SQL, CypherQL, GraphQL, Bash Shell Scripting, HDL
- **Databases:** Oracle, MariaDB, MySQL, Neo4j
- **Libraries/Frameworks & Tools:** PyTorch, Keras, NLTK, Gensim, NetworkX, OpenCV, Git
- **Softwares:** MATLAB, SIMULINK, Marvin Sketch, ASPEN, VESTA, LaTeX

Scholastic Achievements

- Selected for MITACS Globalink Scholarship among top-250 students in India
- Recipient of the prestigious Kishore Vaigyanik Protsahan Yojana Fellowship (KVPY) by Govt. of India with All India Rank-603 out of 150,000 students *(99.6 percentile)*
- Secured All India Rank-20 in COMEDK and All India Rank-112 in KCET out of 200,000 students *(99.9 percentile)*
- Placed 1st in Wells Fargo ⟨She/Hacks⟩-2021, an all-India Machine Learning hackathon
- Selected from 1000+ candidates to receive IT training under Infosys Catch Them Young-2014

Voluntary Work

- **Coordinator, Career Development Cell, IIT Madras**: .
  Organized 100+ higher studies and placement events, coding events and skill development webinars for 500+ students
- **Student Mentor, MITr, IIT Madras**: .
  Mentored a batch of 5 freshmen through 2 semesters, in all academic and non-academic endeavors
- **Coordinator, Analytics Club, Centre for Innovation, IIT Madras**: .
  Conducted hands-on sessions to explain machine learning models and data analysis techniques to 200+ newcomers
- **Manager, Startup Services, Entrepreneurship Cell, IIT Madras**: .
  Strategized and executed 2 seasons of Internfair, successfully providing summer and winter internships to 300+ students
- **Member, National Cultural Association, Fine Arts, IIT Madras**: .
  Achieved distinction in Drawing Grade Higher Examination held by the Govt.; Member of NCA Fine Arts group in 2018