Arindam Paul

(440)622-1087| arindampaul.bits@gmail.com| www.arindampaul.me|linkedin.com/in/arndmpaul/

SUMMARY	**Computer Scientist** working in **Machine Learning** with 10+ years of research experience.
INTERESTS	Machine Learning, Deep Learning, Natural Language Processing

EDUCATION

Northwestern University, Evanston, Illinois USA
Ph.D., Computer Science **Sep '19**
Master of Science, Computer Science

Birla Institute of Technology & Science, Pilani, Rajasthan India
Master of Engineering, Software Systems **May '12**
Bachelor of Engineering, Chemical Engineering

PROGRAMMING SKILLS

Python, Pandas, Scikit-Learn, XGBoost, Streamlit, Dash, Tensorflow, Google Cloud Libraries

PROFESSIONAL EXPERIENCE

Lead Data Scientist (ML Research Team), **American Family Insurance**, Greater Boston, Massachusetts
Oct'24-

- Generative AI-based Conversational Intelligence for QA :
  ◇ Designed effective prompts for Gemini Multimodal models to assess claim quality, leveraging its multimodal capabilities for comprehensive analysis.
  ◇ Collaborated with the Claims QA team to optimize auditing processes and questionnaires, streamlining workflows and improving efficiency.
- Leadership:
  ◇ Host of the AmFam Fairness & Ethics in ML Seminar.
  ◇ Collaborated closely with UW-Madison professors as part of the Amfam Data Science Institute, leading to two publications.

Senior Data Scientist (ML Research Team), **American Family Insurance**, Greater Boston, Massachusetts
Feb'24-Oct'24

- Generative AI for Audio Call Summarization & Entity Extraction :
  ◇ Contributed to Generative AI pipelines for audio call summarization and entity extraction, ensuring robust performance across formats.
  ◇ Investigated Whisper using pyannote for speaker identification and used models like Gemini-1.5-Pro, Mistral-Au-Large, and GPT-4-Turbo for summarization.
- Custom Generative AI Chatbot Development:
  ◇ Implemented a custom Generative AI chatbot leveraging technologies like streamlit, langchain, and GPT-4-Turbo/Claude-Sonnet, enhancing customer experience.

Data Scientist (ML Research Team), **American Family Insurance**, Greater Boston, Massachusetts
Oct'19-Feb'24

- User-Based Insurance (in collaboration with major US automaker):
  ◇ Developed generalized linear and additive models based on telematics features for auto insurance in collaboration with a major US automaker.
- Claims-Channeling System:
  ◇ Co-designed a claims channeling system utilizing transfer learning and performed an ablation study to enhance model accuracy for routing claims to experts.
- Financial Forecasting:
  ◇ Developed long and mid-term financial forecasts using an ensemble ESRNN+SARIMA+GBM model, deploying a financial dashboard for leadership.
- Motor Vehicle Violation:
  ◇ Built an ML decision system to predict motor vehicle violation risk using internal and external data.
- Leadership/Outreach:
  ◇ Mentored rotational associate data scientists.
  ◇ Participated as a panelist for company-wide data privacy discussions on fairness, privacy, and bias.

Data Science Intern, **Northwestern Mutual**, Milwaukee, Wisconsin **Jun - Aug '18**
- Developed distributed OCR algorithms for detecting responses from scanned questionnaires

Research Intern, **Boeing Cybersecurity**, Sunnyvale, California **Jun - Sep '13**
- Developed machine learning system using generated synthetic user profiles with different demographic and interest features for analyzing ads across profiles

RESEARCH
PROJECTS

Research Assistant, **Northwestern University**, Evanston, Illinois   **Sep '12 - Aug'19**

- Chemical Property Prediction using Molecular Fingerprints (Tensorflow, Scikit Learn)
  - ◇ Developed a multi-input input neural network architecture by merging different molecular representations (SMILES and fingerprints) for predicting chemical properties and reduced the mean absolute error by half compared to state-of-the-art architectures (CheMixNet architecture)
  - ◇ Designed Bagged Ensemble models for predicting power conversion efficiency of solar cells using chemical fingerprints, and achieved mean absolute percentage error between 1.5-2 %
  - ◇ Developed a transfer learning solution to predict solar cell properties with mean absolute percentage error below 1 % (SINet architecture)

- Developed Predictive Model for Additive Manufacturing (Tensorflow, Keras)
  - ◇ Created time series models for temporal analysis of temperature and heat flux data
  - ◇ Investigated Recurrent Neural Network models to predict point-wise temperature information for accelerating additive manufacturing simulations
  - ◇ Developed an iterative real-time predictive model using bagged decision trees

- Classification of Anonymous Posts using Recurrent Neural Networks (Keras, Scikit Learn)
  - ◇ Developed customized vector model using crowd-sourced (Urban Dictionary) & psycho-lingual (LIWC) dictionaries
  - ◇ Explored Word2vec, GloVe and FastText embedding schemes (Gensim)
  - ◇ Attained prediction accuracy of 79.8 % and 78.1 % using ensemble and LSTM models respectively

SELECT
PUBLICATIONS
(12 OF 25)

Fa Li, Qing Zhu, Kunxiaojia Yuan, Fujiang Ji, **Arindam Paul**, Peng Lee, Volker C. Radeloff, Min Chen. "**Projecting large fires in the western US with a more trustworthy machine learning method**", *Earth's Future, 2024*

Y. Mao, M. Hasan, **A. Paul**, V. Gupta, K. Choudhary, F.M. Tavazza, W. Liao, A. Choudhary, P. Acar and A. Agrawal. "**An End-to-End AI-Driven Microstructure Optimization Framework for Elastic Properties of Titanium Beyond Cubic Crystal Systems**", *Nature Partner Journal on Computational Materials, 2023*

A.Dimri, **A.Paul**, D.Girish, P.Lee, S.Afra and A.Jakubowski. "**A Multi-input Multi-label Claims Channeling System Using Insurance-Based Language Models**", *Expert Systems With Applications, 2022*

Y.Mao, Z.Yang, D.Jha, **A. Paul**, W. Liao, A. Choudhary and A. Agrawal. "**Generative Adversarial Networks and Mixture Density Networks based Inverse Modeling for Microstructural Materials Design**", *Integrating Materials and Manufacturing Innovation Journal, 2022*

K.Ness, **A. Paul**, L. Sun and Z. Zhang. "**Towards a generic physics-based machine learning model for geometry invariant thermal history prediction in additive manufacturing**", *Journal of Materials Processing Technology, 2022 (Special Issue on AI in Advanced Manufacturing)*

R.Richards, and **A. Paul**. "**An Attention-driven LSTM Network for High Throughput Virtual Screening of Organic Photovoltaic Candidate Molecules**", *Solar Energy, 2021*

Z.Yang, D. Jha, **A. Paul**, W. Liao, A. Choudhary and A. Agrawal. "**A General Framework Combining Generative Adversarial Networks and Mixture Density Networks for Inverse Modeling in Microstructural Materials Design**", *NeurIPS Workshop on Machine Learning for Engineering Modeling, Simulation and Design, 2020*

D.Jha, L.Ward, **A. Paul**, W. Liao, A. Agrawal, A. Choudhary and C. Wolverton."**ElemNet: Deep Learning the Chemistry of Materials From Only Elemental Composition**", *Nature Scientific Reports, 2018*

A. Paul, M.Mozaffar, Z. Yang, W. Liao, A. Choudhary, J.Cao and A. Agrawal."**A real-time iterative approach for temperature profile prediction in additive manufacturing processes**", *6th IEEE International Conference on Data Science and Advanced Analytics (DSAA), 2018*

A. Paul, D.Jha, R. Al-Bahrani, W. Liao, A. Choudhary and A. Agrawal."**Transfer Learning Using Ensemble Neural Nets for Organic Solar Cell Screening**", *International Joint Conference on Neural Networks, 2019*

A. Paul, D.Jha, R. Al-Bahrani, W. Liao, A. Choudhary and A. Agrawal."**CheMixNet: Mixed DNN Architectures for Predicting Chemical Properties using Multiple Molecular Representations**", *NeurIPS Workshop on Machine Learning for Molecules and Materials, 2018*

M.Mozaffar, **A. Paul**, R. Al-Bahrani, S. Wolff, A. Choudhary, A. Agrawal, K. Ehmann and J.Cao."**Data-Driven Prediction of the High-Dimensional Thermal History in Directed Energy Deposition Processes via Recurrent Neural Networks**", *Manufacturing Letters, 2018*

J.Birnholtz, N.A.R. Merola, and **A. Paul**. "**Is it Weird to Still Be a Virgin?: Anonymous, Locally Targeted Questions on Facebook Confession Boards**", *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems. ACM, 2015.*