**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

 $\bf Birla \ Institute \ of \ Technology \ \& \ Science,$  Pilani, Rajasthan India

Master of Engineering, Software Systems

May '12

Bachelor of Engineering, Chemical Engineering

Programming Skills Experienced: Python, Keras/Tensorflow, Scikit-Learn, Dash/Plotly, Matplotlib/Seaborn, Pandas, XG-Boost, Selenium

 $Familiar: \ SQL, Shapley, NLTK, Hugging Face, Gensim, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spark, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spacy, Py Torch/Fast AI, Spacy, Cat Boost/Light GBM, Py Spacy, Py Torch/Fast AI, S$ 

HTML/CSS

Professional Experience Machine Learning Scientist, American Family Insurance, Madison, Wisconsin

Oct'19-

- Developing ML-based safe driving behavior algorithms in collaboration with major US automaker
- Created insurance-based language models for predicting claims fraud
- Developed SARIMAX, LSTM and ES-RNN forecasting financial metrics for next calendar year
- Developed an ML decision system for predicting motor vehicle violation risk

Data Science Intern, Northwestern Mutual, Milwaukee, Wisconsin

Jun - Aug '18

- Developed distributed OCR algorithms for detecting responses from scanned questionnaires
- Designed a noise reduction algorithm to de-noise scanned and photocopied questionnaires

Research Intern, Boeing Cybersecurity, Sunnyvale, California

Jun - Sep '13

- Generated synthetic user profiles with different demographic and interest features for analyzing ads across profiles
- Developed a machine learning model for predicting user demographics and interests from ads

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
- Ensemble Learning-based Guided Optimization for Aircraft Design (MATLAB, Python)
  - Created intelligent sampling algorithms to explore the constrained search space for candidate microstructures
  - Developed Feature Ranking-based Technique for Search Space Reduction of Constrained Non-Convex Optimization
  - Achieved 100x more solutions compared to state-of-the-art methods that can accelerate the designto-experiment life-cycle
- Classification of Anonymous Posts using Recurrent Neural Networks (Keras, Scikit Learn)
  - ♦ Developed customized vector model using crowd-sourced (Urban Dictionary) & psycho-lingual (LIWC) dictionaries
  - $\diamond$  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 (15 OF 22) 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)

Z.Yang, Y.Mao, 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", Science Advances (under review)

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", Expert Systems With Applications (under review)

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

A. Paul, W. Liao, A. Choudhary and A. Agrawal. "Harnessing Psycho-lingual and Crowd-Sourced Dictionaries for Predicting Taboos in Written Emotional Disclosure in Anonymous Confession Boards", Journal of Health Informatics Research, 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, A. Furmanchuk, W. Liao, A. Choudhary and A. Agrawal. "Property Prediction of Organic Donor Molecules for Photovoltaic Applications using Extremely Randomized Trees", Journal of Molecular Informatics, 2019

A. Paul, P. Acar, W. Liao, A. Choudhary, V.Sundararaghavan and A. Agrawal. "Microstructure Optimization with Constrained Design Objectives using Machine Learning-Based Feedback-Aware Data-Generation", *Journal of Computational Materials Science*, 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.

R. Liu, D. Palsetia, A. Paul, R. Al-Bahrani, D. Jha, W. Liao, A. Agrawal and A. Choudhary. "Pinter-Net: A Thematic Label Curation Tool for Large Image Datasets", Proceedings of the Workshop on Open Science in Big Data at IEEE Bigdata Conference, 2016.

FELLOWSHIPS

McCormick Dean's Commendation Fellowship
Predictive Science and Engineering Design Fellowship
Segal Design Fellowship
Walter P. Murphy Fellowship
'18 Spring
'16-'17
Segal Design Fellowship
'14-'15
'12-'13

SELECTED TEACHIN AND LEADERSHIP

Selected Teaching Teaching Assistant & Guest Lecturer, Northwestern University

Jan'14- Jun'19

- ♦ Prepared and delivered weekly lectures for 20-50 students
- Courses: Social Media Mining, Data Structures , Introduction to Programming (Python)

President/Vice-President/Treasurer, Northwestern Toastmasters

Sep '15 -May '18

- Lead the Northwestern chapter of Toastmasters with over 30 graduate students, post doctoral fellows from 10 different departments
- ♦ Co-wrote proposal to The Graduate school and obtained 3000 USD to fund programming

## Co-Facilitator, Northwestern Dialogue Group

Oct '16 - Sep '17

- $\diamond \ \ \text{Facilitated dialogue in safe spaces for cultural exchange across international and domestic students}$
- $\diamond\,$  Organized social events to enhance group cohesion