APRATIM MISHRA

901 S 1st Street, Apt 1, Champaign, IL. 61820 ◦ (302)-407-2633 ◦ apratim941208@gmail.com

SUMMARY

Ph.D. candidate in Information sciences adept in building machine learning and natural language processing (NLP) solutions. Seeking a full-time position in Data Science or Machine learning to drive data-driven innovations.

PROFESSIONAL EXPERIENCE

Data Science Intern – *AstraZeneca*
Gaithersburg, MD | *May 2023 – Aug 2023*

- Utilized protein language models (PLMs) and graph neural networks (GNNs) to identify protein features and predict biological behaviors, significantly improving modeling performance (accuracy, precision) by ~ 5-10%.
- Deployed quantized deep learning models (parameter-efficient) for multi-GPU scalability, reducing inference time and memory footprint by streamlining the model architecture.

Graduate NLP Programmer – *The Cline Centre for Advanced Social Research*
Champaign, IL | *May 2021 – Aug 2021*

- Engineered an NLP pipeline for quotation extraction and entity classification (spaCy, CoreNLP), implementing testing strategies, evaluation methods, and documentation updates.
- Fine-tuned transformer models (BERT, XLNet) for quote classification, enabling accurate *span* detection within text

Energy Data Analyst – *Reliable Power Alternatives Corporation*
Garden City, NY | *Nov 2018 – Aug 2019*

- Optimized data retrieval via SQL and built ML pipelines (PySpark, scikit-learn), delivering cost savings for clients by ~ 10%.
- Deployed Flask-based model APIs and presented performance metrics to stakeholders.

Research Intern – *Delaware Army National Guard*
Wilmington, DE | *Sep 2017 – Jan 2018*

- Analyzed energy load trends with ARIMA, XGBoost, and LSTM models, improving forecast accuracy.
- Conducted statistical tests (t-test, ANOVA) to validate data integrity for client benchmarks.

Energy Efficiency Intern – *NYC Dept. of Citywide Administrative Services*
New York, NY | *Jun 2017 – Aug 2017*

- Created interactive dashboards in Tableau to visualize cost-saving KPIs for cross-departmental use.

EDUCATION

University of Illinois at Urbana-Champaign • Champaign, IL
PhD in Information Sciences (*Aug 2019 – May 2025*)

- Analyzed *scientific "hype"* in biomedical literature using NLP-based feature engineering (e.g., author features, topical interests), modeled with language models (*sense classification*), and categorized using a probabilistic mixture model.
- Engineered citation recommendation system leveraging traditional network features incorporating *"hype"* scores.
- Investigated *stance detection* in tweets, contrasting neural methods vs. matrix-based topic modeling; designed and benchmarked tree-based models (XGBoost, Random Forest).
- Developed *language-to-command* translation models using Seq-to-Seq transformers (T5) and integrated slot-filling techniques to extract and map parameters for bash template generation.
- Employed statistical models to quantify collaboration network diversity indices (expertise, ethnicity, gender), identifying expertise as the key predictor of scientific impact.

University of Delaware • Newark, DE
Master in Energy and Environmental Policy (*May 2018*)

- Presented a final research paper comparing *univariate vs. multi-step* forecasting methods—integrating ARIMA, feature-based ML models, and deep learning sequential architectures (LSTM) to assess accuracy and scalability.

Birla Institute of Technology and Science • Pilani, India
Bachelor of Engineering in Chemical Engineering (*May 2016*)

TECHNICAL SKILLS

- Technical Skills - Programming & Analytics: Python (pandas, NumPy, scikit-learn), SQL, R
- ML Frameworks: PyTorch, TensorFlow, Transformers, PyTorch Lightning, PyTorch Geometric
- Tools & Platforms: AWS, Linux, Git, Docker, Deepspeed, Accelerate
- Visualization: Tableau, Power BI, matplotlib

LinkedIn: https://www.linkedin.com/in/apratim94/ **GitHub**: https://github.com/apratim-mishra