APRATIM MISHRA

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

SUMMARY

Machine Learning professional specializing in NLP, LLMs, and GNNs, with expertise in developing scalable deep learning models, conducting statistical analysis, and applying data science solutions to product development.

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

PROFESSIONAL EXPERIENCE

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

- Utilized protein language models (PLMs), graph neural networks (GNNs), and deep learning frameworks (PyTorch, TensorFlow) to develop novel protein features, improving modeling performance (accuracy, precision) by ~ 5-10%.
- Optimized deep learning models using parameter-efficient methods (LoRA, DeepSpeed, Accelerate) for multi-GPU scalability, improving evaluation, reducing inference time and memory footprint by streamlining 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), developed robust testing pipelines (unit tests, cross-validation), and enhanced model documentation.
- 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.

PROJECTS AND RESEARCH

- Thesis: Quantified "hype" in biomedical literature using NLP (author traits, topical features) and LLMs for sense classification and categorizing texts with probabilistic mixture models.
- Designed a GNN-based citation recommendation system integrating network features with novel hype scores.
- Investigated tweet stance detection, contrasting neural networks with matrix-based topic modeling and optimizing boosting/bagging models (XgBoost/ Random Forests) via benchmarking.
- 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.
- Developed statistical models to analyze diversity in co-authorship networks, pinpointing expertise as the primary predictor of scientific impact.
- Master's Thesis: Authored a research paper comparing univariate and multi-step forecasting—ARIMA, feature-based ML, and LSTM—assessing accuracy and scalability trade-offs.

TECHNICAL SKILLS

- Programming & Analytics: Python (pandas, NumPy, scikit-learn), SQL, R
- ML Frameworks: PyTorch, TensorFlow, Transformers, LangChain, PyTorch Lightning, PyTorch Geometric, PySpark
- Tools & Platforms: AWS, Hive, Snowflake, Ray, Linux, Git, Docker, DeepSpeed, Accelerate, Weights and Biases.
- Visualization: Tableau, Power BI, Matplotlib, Seaborn

EDUCATION

- PhD in Information Sciences | University of Illinois at Urbana-Champaign | Champaign, IL | (*Aug 2019 – May 2025*)
- Master in Energy and Environmental Policy | University of Delaware | Newark, DE | (*Aug 2016 – May 2018*)
- Bachelor of Engineering in Chemical Engineering | Birla Institute of Technology and Science | Pilani, India | (*Aug 2012 – May 2016*)