APRATIM MISHRA

Champaign, IL | apratim941208@gmail.com

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

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.

PROFESSIONAL EXPERIENCE

Data Science Intern – AstraZeneca

Gaithersburg, MD | May 2023 – Aug 2023

- Developed novel protein embeddings leveraging Protein Language Models (PLMs), Graph Neural Networks (GNNs), PyTorch, and TensorFlow, enhancing predictive accuracy and precision by ~10%.
- Scaled deep learning pipelines across multiple GPUs using parameter-efficient fine-tuning techniques (LoRA) and optimization tools (DeepSpeed, Accelerate), significantly improving inference speed and memory efficiency.
- Implemented experiment tracking and model versioning with MLflow, Weights & Biases (W&B), and AWS S3.

Graduate NLP Programmer – The Cline Centre for Advanced Social Research

Champaign, IL | May 2021 – Aug 2021

- Engineered scalable NLP pipelines for entity classification and quotation extraction using spaCy, CoreNLP, and PySpark.
- Fine-tuned transformer-based language models (BERT, XLNet) for precise quote detection; enhanced model reliability via comprehensive unit tests and cross-validation. Utilized Apache Spark for efficient text processing and data handling at scale.

Energy Data Analyst – Reliable Power Alternatives Corporation

Garden City, NY | Nov 2018 – Aug 2019

- Built and deployed scalable end-to-end ML pipelines (PySpark, scikit-learn), achieving ~10% client cost reduction through optimized energy forecasting.
- Delivered insights via Flask API deployments and interactive dashboards, clearly communicating key performance metrics to stakeholders. Extracted and processed large-scale energy datasets using Hive and SQL for efficient analytics.

Research Intern – Delaware Army National Guard

Wilmington, DE | Sep 2017 – Jan 2018

- Developed advanced time-series forecasting models (ARIMA, XGBoost, LSTM) for energy consumption prediction, significantly improving forecast accuracy and reliability.
- Conducted rigorous statistical analyses (t-tests, ANOVA) to validate model performance and data integrity.

PROJECTS AND RESEARCH

- **PhD Thesis:** Quantified "hype" in biomedical literature using NLP features (author traits, topical context, collaboration diversity measures) and Large Language Models (LLMs) with prompt-based sense classification to categorize texts via probabilistic mixture models.
- **Hybrid Citation Recommendation System:** Developed a citation recommendation system integrating Graph Neural Networks (GNNs) and LLM-based embeddings, leveraging network features and novel hype scores.
- **Tweet Stance Detection:** Investigated stance detection in tweets by benchmarking neural network methods against matrix-based topic modeling, optimizing ensemble models (XGBoost, Random Forests) for improved classification accuracy.
- **Command Translation Model:** Built language-to-command translation systems using Seq-to-Seq transformers (T5) and integrated slot-filling techniques to automate parameter extraction for dynamic bash command generation.
- **Master's Thesis:** Compared univariate and multi-step forecasting methods (ARIMA, feature-based machine learning, and LSTM), evaluating their accuracy and scalability trade-offs for energy load prediction.

TECHNICAL SKILLS

- **Programming and Analytics**: Python (pandas, NumPy, scikit-learn, nltk), SQL, R
- **ML & Deep Learning**: PyTorch, TensorFlow, Transformers, LangChain, Lightning, PyG, PySpark, MLflow, Matplotlib, Seaborn
- **MLOps and Tools**: W&B, Comet, ZenML, Ray, DeepSpeed, Accelerate, Docker, Git, Linux
- **Big Data and Cloud**: Apache Spark, Hive, AWS (S3, EC2, SageMaker), GCP (BigQuery, Vertex AI), Hopsworks, Superlinked
- **Storage & Databases**: Snowflake, MongoDB

EDUCATION

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