WANSHIKA PATRO

+1-716-390-2392 | wanshika85@gmail.com | LinkedIn | Github | Portfolio

SUMMARY

Graduate Data Science professional with strong experience designing and deploying machine learning systems for text, image, audio, and time-series data. Skilled in Generative AI and Large Language Models (LLMs), including Transformer-based architectures and prompt engineering. Proficient in the end-to-end ML pipeline: data collection, cleaning, feature engineering, model training, evaluation, and deployment. Skilled at collaborating across engineering and research teams to deliver measurable, high-impact digital experiences at scale

EDUCATION

- **University at Buffalo, SUNY** | Buffalo, NY
  Master of Science in Data Science | *Jun 2025*
  Relevant Coursework: Deep Learning, Natural Language Processing, Reinforcement Learning, Cloud Computing, Scalable Data Pipelines, Big Data Analytics, Recommender Systems
- **KIIT University** | Bhubaneswar, India
  Bachelor of Technology in Computer Science | *Aug 2021*
  Undergraduate Projects: Credit Card Fraud Detection, Object Recognition using Neural Networks

TECHNICAL SKILLS

- **Programming:** Python, Java, C++, SQL, R
- **ML Frameworks:** SparkML, PyTorch, TensorFlow, Scikit-learn, XGBoost, Hugging Face
- **Techniques:** Generative AI, Large Language Models (LLMs), Transformers, Prompt Engineering
- **Data Platforms:** Snowflake, PostgreSQL, MySQL, BigQuery
- **Systems:** Apache Spark, PySpark, Docker, Multi-GPU Training
- **APIs & DevOps:** REST APIs, Flask, Streamlit, GitHub Actions, CI/CD
- **Visualization:** Tableau, Power BI, Seaborn, Matplotlib

PROFESSIONAL EXPERIENCE

Application Development Analyst
ACCENTURE - Pune, India| *Jul 2021 - Jun 2022*

- Automated ServiceNow workflows, reducing manual ticket resolution time by 35% and improving SLA adherence
- Analyzed 100,000+ incident and change logs using Python and SQL, increasing triage accuracy by 40%
- Developed real-time dashboards to monitor system health and ticket KPIs, driving 25% faster decision-making
- Collaborated with engineering teams to integrate data pipelines across 3+ modules, enhancing consistency and speed
- Contributed to REST API automation and configuration deployments using GitHub Actions, cutting manual errors by 20%

PROJECTS

Generative AI and NLP Systems

Toxic Comment Detection | *Hugging Face, Transformers, Flask* | Jan 2025 – May 2025

- Fine-tuned BERT on 160,000+ text samples with prompt-switching and variant evaluation, cutting manual moderation efforts by 30% through a real-time Streamlit UI.
- Deployed a real-time Streamlit UI with confidence scoring and model toggles, enabling 50% faster review turnaround.

Spelling Error Clustering | *BERT, Autoencoder, LLMs* | Jan 2024 – May 2024

- Clustered 50,000+ spelling errors with BERT embeddings and unsupervised methods, boosting downstream NLP token correction by 15%, improving LLM pipeline input quality.
- Integrated clusters into an LLM preprocessing pipeline, increasing adaptive correction accuracy by 18% on test sets.

Scalable ML Systems and Forecasting

Revenue Forecasting | *XGBoost, Streamlit, Tableau* | Feb 2025 – Present

- Modeled 145,000+ time-series records using XGBoost with window-based features, identifying anomalies with 95% precision and visualizing results in Streamlit dashboards
- Visualized demand anomalies via Streamlit dashboards, reducing forecast reporting time by 40% for stakeholders.

Crime Trend Analysis | *PySpark, MLlib, Pandas* | Aug 2024 – Dec 2024

- Processed 1,000,000+ crime records via PySpark for spatiotemporal regression and clustering, achieving 83% accuracy in location-based prediction and supporting proactive planning.
- Delivered heatmaps and proactive alert simulations, enabling 20% improvement in deployment efficiency for field teams.

Stock Price Prediction | *Python, Docker, SHAP, Streamlit* | Jan 2024 – May 2024

- Engineered sequential indicators (MACD, RSI, SMA), trained RF/SVM models, containerized with Docker, and benchmarked stream vs. batch latency with 20% faster inference.
- Benchmarked stream vs. batch inference, achieving 20% faster response and reducing prediction error rates by 12%.

Recommendation Systems

Music & Entertainment Recommender | *Python, TensorFlow, Scikit-learn, Flask* | Mar 2025 - Present

- Built a hybrid system combining collaborative filtering and sequence models, serving recommendations via Flask APIs, and driving 25% increase in user engagement through A/B testing.
- Served real-time recommendations via Flask APIs with sub-second latency, improving recommendation response times by 35%.

CERTIFICATION & LEADERSHIP

- **Supervised Machine Learning –** Stanford/DeepLearning.AI *(In Progress)*
- **Career Skills in Data Analytics –** LinkedIn Learning
- **Meta Kaggle Hackathon (2025):** Top 20% out of 1,200+ participants; designed trading signal classifiers across public code metadata
- **Digital Strategy Lead –** MUN Leadership Team (2018–2020)
- **Volunteer Analyst –** Animal Rescue NGO: Led real-time analytics for donation optimization and campaign conversion