

NITIN KUMAR

Data Scientist

Email: nktyagi423@gmail.com

LinkedIn: <https://www.linkedin.com/in/nktyagi423/>

Website: <https://nktyagi123.github.io/portfolio/>

Mobile: +91-7830260919

kyndryl

IBM



Summary

- Data Scientist with 6+ years of expertise in Generative AI (OpenAI, Gemini, Llama, Bert etc.), intelligent data processing, handling and analysis.
- Worked on transformers models and applied NLP techniques, particularly LSTM for sentiment analysis, ticket classification and summarization.
- Developed GenAI solutions: Japanese Translator, Text-to-Elasticsearch Generator, Vehicle Insurance Assistant, GenAI Ticket Resolver, and Nutritionist's Eye, leveraging Gemini Pro and GPT for innovation.
- Worked and implemented different ML/DL algorithms and Hugging Face pre-trained models for anomaly detection and text classification.
- Worked on Kibana and Power BI to create interactive and visually compelling dashboards, reports, and data visualizations.
- Experience in NLP, End to End Deployment, interpreting data and generating useful reports/dashboards.

Education

- 2024 - 2026 Master of Technology – Artificial Intelligence
- 2014 - 2018 Bachelors of Technology – Computer Science & Engineering

Certifications

- 2024 L400 Google Generative AI Badge
- 2024 Google Certified Professional Machine Learning Engineer
- 2023 AWS Certified Machine Learning - Specialty
- 2021 Microsoft Certified – Azure AI Fundamentals

Work Experience

Senior Associate II Data Scientist

Sep 2021 – PRESENT

KYNDRYL

As a Data Scientist with 6+ years of expertise in data analysis, processing, and deployment. Skilled in leveraging pre-trained models like RNNs, applying Generative AI techniques, and creating predictive dashboards using Elastic ML, Kibana, and Power BI to deliver actionable insights for informed decision-making.

Sample of Customers, Projects and Roles:

Client – Kaiser Permanente, AT&T, Anthem, iHub

- Generative AI Based Ticket resolver application: Application to the process of managing student inquiries by automating message categorization and email responses, enhancing communication and efficiency without human interaction.
- TexttoElasticQueryGeneration: To leverage the capabilities of the Gemini Pro large Language Model to facilitate the generation of Elasticsearch queries from natural language text inputs.
- Generative AI Based Vehicle Insurance Assistant: Vehicle Damage assessment and cost estimate for a Vehicle Insurance solution.
- Developed ETL workflows for storage optimization, applying ML to forecast utilization and drive cost reductions.
- Applied NLP techniques like LSTM in creating the ITCE for ticket classification.
- Utilized TensorFlow and Keras in projects involving Service Health calculations for various entities.
- Crafted interactive reports and dashboards with Kibana and Power BI, delivering actionable insights.
- Implemented anomaly detection methods based on size, time, and utilization for proactive planning.

Previous Roles:

Data Scientist

Mar 2019 – Aug 2021

IBM

Client – Storage Analytics, Kaiser Permanente

- Leveraged Elastic ML in the Storage Analytics project, proficiently handling large datasets to identify anomalies in timeseries data and enhance predictive accuracy through effective anomaly detection methodologies.

Client – GTS Analytics, Kaiser Permanente

- Engineered a log parsing system to transform unstructured raw data into structured formats, leveraging Golang and text analysis for extracting key-value pairs from log files.

Client - iPAT Analytics (IBM's Internal Project)

- Developed pricing models with Java and microservices, while leveraging Python APIs for efficient data processing tasks.

Data Engineer

Apr 2018 – Feb 2019

PhloxIT Global

- Executed data extraction and ingestion in Python, gaining hands-on experience with Apache Flume and Kafka for a Sentiment Analysis project. Proficient in data visualization techniques, showcasing skills essential for a dynamic role in data engineer.

Skills & Proficiencies

Primary: Python, AWS, SQL, ELK, Power BI, Machine Learning, Deep Learning (RNNs), Statistical Technique, NLP (Natural Language Processing), SpaCy, Time Series Modelling, Pandas, NumPy, Matplotlib, Scikit-learn, DSA, Agile methodology.

Secondary: Generative AI, Tableau, Regex, Large Language Models, Kubeflow.