Vibhor Tyagi

✉ vtyag@illinois.edu in [LinkedIn](#) ⌖ [Github](#) ▢ [Portfolio](#)

EDUCATION

University of Illinois at Urbana-Champaign — Urbana, IL
Master of Science - Computer Science; Score: 4.0 — *Jan 2024-May 2025(Expected)*

SRM Institute of Science and Technology — Chennai, India
Bachelor of Technology - Computer Science and Engineering; Score: 87.4% — *July 2017-June 2021*

WORK EXPERIENCE

Graduate Research Assistant — March 2024 - Ongoing
University of Illinois Urbana-Champaign
- Automated data collection for sustainability research using Selenium and Scrapy, targeting companies' carbon emissions and regulatory compliance.
- Engineered and optimized ETL workflows with Apache Airflow and Apache NiFi, ensuring efficient data integration and accuracy for large datasets.
- Utilized Apache Spark and Kafka for high-performance, scalable processing and analysis of extensive datasets, focusing on the impact of environmental regulations on corporate carbon footprints.

Data Automation Engineer 2 — September 2021 - June 2023
PricewaterhouseCoopers
- Orchestrated end-to-end automation for various Fortune 500 companies using ETL pipelines, cloud dashboards, ML model deployment, and RPA bot creation resulting in major cost savings and operational efficiencies
- Led a team of 8 in designing and developing robust RPA models using Python, UiPath, Power Automate, and Power Apps, achieving a reduction of over 10,000 manual work hours annually.
- Implemented ETL pipelines with Apache Airflow, Alteryx Designer, and PowerBI, integrating with Tableau for actionable insights. Enhanced data-driven decision-making, leading to up to 18% expense reduction for clients.
- Constructed predictive and classification models using machine learning and deep learning frameworks to forecast costs and expenses, contributing to a 16% improvement in strategic financial planning and resource allocation.
- Developed impactful visualizations and optimized SQL queries, reducing database response time and contributing to an 8% increase in client engagement, as evidenced by enhanced click-through rates and user interactions.

Undergraduate Research Assistant — Jan 2021 - May 2021
SRM Institute of Science and Technology
- Spearheaded the implementation of a cutting-edge movie recommendation system leveraging K-means, collaborative filtering, and SVD, resulting in significantly enhanced personalized content delivery.
- Improved system sparsity and accuracy through innovative use of implicit and explicit feedback mechanisms, coupled with Association Rule Mining.
- Optimized key metrics including precision, recall, and F-measures, leading to substantial improvements in recommendation quality.

SELECTED PROJECTS (2/5)

- **Real-Time Data Pipeline for E-commerce Analytics**
  - Designed and implemented a robust pipeline using Apache Airflow for orchestration, Apache Kafka for data streaming, and Apache Spark for real-time processing, handling millions of transactions daily with fault tolerance and scalability.
  - Employed Cassandra for distributed storage and PostgreSQL for structured data warehousing, optimizing indexing and partitioning strategies to reduce data retrieval latency by 30% and enhance query performance.
  - Utilized Docker for containerization and Kubernetes for orchestration, ensuring consistent deployment across environments. Automated CI/CD pipelines, cutting deployment times by 25% and minimizing manual intervention.

- **Advanced Text Analytics for Sentiment and Topic Modeling**
  - Utilized BERT for sentiment analysis, achieving a 92% accuracy on benchmark datasets, significantly improving customer sentiment analysis capabilities.
  - Applied LDA to identify key themes and trends from large text corpora, enhancing content categorization and recommendation systems.
  - Deployed models on AWS SageMaker using Docker for scalable, low-latency predictions. Developed interactive dashboards with Plotly and Dash to visualize sentiment trends and topic distributions, providing actionable insights.

SKILLS SUMMARY

- **Languages / Database Management:** Python, C, C++, Java, SQL, R, MySQL, PostgreSQL, MongoDB, NoSQL, Cassandra

- **Big Data Technologies / Data Warehousing:** Apache Hadoop, Spark, Kafka, Hive,Amazon Redshift, Google BigQuery

- **Cloud Platforms:** AWS (S3,EC2, Sagemaker, Lambda)

- **ETL Tools:** Apache NiFi, Airflow, Alteryx Designer

- **DevOps / Version Control:** Docker, Terraform, Jenkins, Git, Github

- **Python Frameworks:** Numpy, Pandas, Scikit, NLTK, TensorFlow, PyTorch, Keras, Matplotlib, Seaborn, Plotly,PySpark , Shiny