# Rahul Yadav

## Data Scientist

rahulkryadav93@gmail.com | +91 7860373636 | Delhi, IN | LinkedIn

Data scientist with over 5+ years of experience in statistical analysis, data mining, and machine learning. Adept in generative artificial intelligence and prompt engineering, with a track record of generating important insights that drive business choices. Capable of leading cross-functional teams and driving growth through data visualization and predictive modelling.

### TECHNICAL SKILLS

- **Tools & Languages** Python, Jupyter, Spyder, R, R Studio, C, C++, Java

### KEY SKILLS

- Generative Artificial Intelligence
- Natural Language Processing (NLP)
- Natural Language Generation (NLG)
- Prompt Engineering
- Representational Learning
- Machine Learning
- Predictive Analytics
- Sentiment Analysis
- Deep Learning

### CERTIFICATIONS

- Google Professional Machine Learning Engineer
- Certificate in Big data and Hadoop at Studybay Technology

### KEY PROJECTS

**Domain: Social Network Analytics (Link Prediction) | Jun '17**

- *Objective*: **Link prediction** is used to predict future possible links in the network (E.g., Facebook People You May Know Feature)
- *Tech Stack*: R
- *Solution*: Designed a feature-based model to predict the probability scores.
- *Key Achievement:* Presented the work in **Indian Institute of Technology (IIT-K), Kanpur**. Published in Springer and International Journal of Web Based Communities

**Open-Source Contributions:**

- Created a library named TextVectorizer for representation learning of Text using Transformers such as BERT, AlBERT, RoBERTA and spacy.

### KEY PUBLICATION

- A Novel Similarity-based parameterized method for Link Prediction | Chaos, Solitons & Fractals — Sept'23
- WeedNet: A deep neural net for weed identification | Elsevier — Feb'22
- Network Embedding Based Link Prediction in Dynamic Networks | Elsevier — Sept'21
- Incorporating Communities' Structures in Predictions of Missing Links | Springer — May'20
- Hybrid Feature-Based Approach for Recommending Friends in Social Networking Systems | Inderscience — Feb'20
- Hybrid Approach for Predicting and Recommending Links in Social Networks | Springer — Sept'18

### PROFESSIONAL EXPERIENCE

**Sr. Data Scientist | Tata Consultancy Services**

- ***Guardrails for LLMs*** **Dec'23 - Present**
  - ***Client:*** *Centre of Excellence, TCS Research & Insight* ***| Category:*** *Gen AI*
  - ***Objective:*** *Establish and enforce comprehensive ethical guidelines for Large Language Models (LLMs) to promote responsible and accountable use of generative AI.*
  - ***Solution:*** *Developed and integrated an ethical framework encompassing aspects such as quality, toxicity, bias, etc. I collaborated with the legal team to align guidelines with regulations and implemented monitoring systems to ensure continuous adherence to ethical standards in Large Language Model (LLM) utilization. Additionally, I utilized techniques such as RAG, CoT (Chain of Thought), and Few-shot prompting to enhance the responsiveness of the LLM.*
- ***Regulation Compliance*** **Aug'23 – Nov'23**
  - ***Client:*** *Stellantis* ***| Category:*** *Gen AI*
  - ***Objective*** *Automate the consolidation of automotive regulations, corrections, and amendments, ensuring Stellantis' regulatory compliance.*

- ***Solution:*** *Implemented an automated system for Stellantis, significantly reducing processing time from weeks to 10-15 minutes, ensuring accuracy and compliance with automotive regulations. The solution encompasses data collection, preprocessing, collaboration with domain experts, and the development of models for the automatic identification and incorporation of regulatory changes. At the backend, GPT-3.5-Turbo is utilized, along with prompting techniques such as CoT (Chain of Thought) and Few-shot prompting, as well as LangChain framework.*

- ***Code Migration*** **Nov'22 – Jul'23**
  - ***Client:*** *Cigna* ***| Category:*** *Gen AI*
  - ***Objective*** *Facilitate seamless code migration, ensuring compatibility, improved maintainability, and optimal utilization of the target programming language.*
  - ***Solution:*** *Leveraging Generative AI techniques, I optimized code migration for Cigna, enhancing compatibility and maintainability in the target programming language. This involved implementing custom modularization of the source code, collaborating with Subject Matter Experts, and working with software engineers to address challenges, ensuring a more efficient migration process. The utilization of OpenAI's GPT-3 models, along with the PAL (Program-Aided Language Models) prompting technique, played a key role in achieving these improvements.*
- ***Insight Generation*** **Sept'21 – Oct'22**
  - ***Client:*** *Microsoft* ***| Category:*** *Gen AI*
  - ***Objective*** *Extract strategic insights from campaign data to inform optimal decision-making and performance optimization.*
  - ***Solution:*** *Conducted preprocessing and cleaning of campaign data, collaborated with domain experts to contextualize insights, and iteratively refined models based on feedback and evolving data patterns. Presented generated insights in a narrative format, facilitating informed decision-making and the development of performance optimization strategies. This process involved the utilization of GPT-3 models and few-shot prompting techniques.*
- ***Customer Persona*** **Jan'21 – Aug'21**
  - ***Client:*** *Verizon* ***| Category:*** *Gen AI*
  - ***Objective*** *Demonstrate the effectiveness of advanced AI in training customer support personnel by creating a chatbot that accurately emulates customer behavior.*
  - ***Solution:*** *Engaged in data analysis to discern customer behavior, fine-tuned OpenAI's Text-Davinci (GPT-3) model for chatbot development, and crafted algorithms for natural language understanding and response generation. Iteratively tested and enhanced chatbot performance, collaborating closely with customer support teams to integrate domain-specific knowledge.*

---

**Data Scientist | Tata Consultancy Services**

- ***Narrative Generation*** **Jul'20 – Jan'21**
  - ***Client:*** *Amgen* ***| Category:*** *Gen AI*
  - ***Objective:*** *Autonomously generate tailored narratives for clinical trials to enhance communication and comprehension.*
  - ***Solution:*** *Implemented a system that utilizes both a template-based and a custom probabilistic approach, leveraging GPT-3 models and few-shot prompting techniques to generate narratives specifically tailored for clinical trials, improving communication and comprehension in the context of clinical research.*
- ***Deal Intelligence*** **Dec'18 – Jun'20**
  - ***Client:*** *Centre of Excellence, TCS Research & Insight* ***| Category:*** *Predictive Analytics*
  - ***Objective:*** *Develop a system to determine the outcome (onboarded, shelved, or rejected) of deals during customer interactions.*
  - ***Solution:*** *Implemented a hybrid ML approach for the Deal Intelligence system, combining both probabilistic and deterministic techniques. This innovative methodology ensures the precise classification of deals based on customer interactions, helping the organization take appropriate action to win the deals.*

---

**System Engineer | Tata Consultancy Services**

- ***Training*** **Sep'18 – Nov'18**
  - ***Client:*** *Human Resource – TCS* ***| Category:*** *ML*
  - ***Development:*** *Designed and implemented a system for bank executives, enabling them to visualize charts and distributions of banking data for enhanced insights.*
  - ***Award & Recognition:*** *Received the LIREL Award for embodying TCS Values (Leading change, Integrity, Respect for the individual, Excellence, Learning, and Sharing). Acknowledged among 250 associates for making a significant difference in the TCS organization.*

## EDUCATION

**MCA (Gold Medallist) | University of Allahabad** **Uttar Pradesh IN | Jul '15 – Jun '18**