

# KOUSHIK KHAN

+91 (750) 192-3251  Bangalore, India

 koushik.khan.ds@gmail.com  linkedin.com/in/koushikkhan/  koushikkhan.github.io

## SUMMARY

---

Full Stack Data Scientist with nearly a decade of experience in consulting and developing data-driven products. Skilled at delivering impactful solutions that extract actionable insights from data across diverse domains. Proficient in all phases of the Data Science lifecycle, including data sourcing, ETL, analytics, ML/DL model development, and visualization. Adept at adapting to varied business needs to drive value and empower stakeholders with data-informed decision-making.

## SKILLS

---

<b>Machine Learning &amp; AI</b>	Predictive Modeling, Statistical Analysis, Forecasting, Supply Chain Analytics, Recommender Systems, Natural Language Processing (NLP), Information Retrieval
<b>Programming Languages</b>	R, Python
<b>Database</b>	SQL, NoSQL (MongoDB)
<b>Tools &amp; Frameworks</b>	Pandas, Polars, PySpark, Scikit-Learn, PyTorch, TensorFlow, Linux, Docker, Kubernetes, CI/CD, Power BI
<b>Cloud Platforms</b>	IBM Cloud, Microsoft Azure, AWS

## WORK EXPERIENCE

---

### Lead Data Scientist

October 2024 - Till Date

Tiger Analytics India Consulting Pvt. Ltd.

*Bangalore, India*

- Currently leading a high-performing team to design, develop, and maintain robust data science frameworks and tools that serve as the backbone for critical customer projects. These frameworks standardize workflows, optimize resource usage, and ensure scalability, enabling faster model development and deployment. By incorporating best practices in machine learning engineering, automation, and CI/CD pipelines, the team has significantly reduced manual overhead and streamlined end-to-end processes. As a result, delivery timelines for customer-facing solutions have been accelerated by 60%, enhancing productivity, consistency, and overall business impact.
- Developing a Retrieval-Augmented Generation (RAG)-based application for a banking client to automate the extraction of numerical summaries from complex internal records. The application converts natural language queries into accurate SQL statements to retrieve information from the backend database, ensuring precision in data extraction. It also provides visual summaries through dynamically generated charts, enhancing the interpretability of insights. Leveraging advanced natural language processing (NLP) techniques, the solution seamlessly integrates with the client's systems, enabling stakeholders to access critical data efficiently. This automation reduces manual effort, minimizes errors, and accelerates data-driven decision-making while aligning with the client's operational goals.

**Methods & tech stacks used: Deep Learning, Python, LangChain, LLM, RAG, Linux, Docker**

### Principal Data Scientist

August 2021 - September 2024

S&P Global, Market Intelligence

*Bangalore, India*

- Led a cross-functional team in the design and development of robust DevOps pipelines and advanced predictive models for maritime supply chain and emission products. The DevOps pipelines automated the end-to-end workflows, enabling continuous integration and continuous delivery (CI/CD) for model deployment, testing, and monitoring. This ensured consistent, scalable, and efficient operations while reducing the time-to-market for new features. The predictive models leveraged historical data, machine learning, and statistical techniques to forecast supply chain dynamics and emissions, providing actionable insights that supported strategic decision-making. These innovations enhanced product performance, accuracy, and client satisfaction, driving operational efficiency and reducing the environmental footprint for key stakeholders in the maritime industry.

- Developed advanced time series models using deep learning techniques to forecast fuel oil consumption for container ships, leveraging historical operational data and environmental factors. The predictive models accurately estimated fuel usage patterns, enabling more precise forecasting of  $CO_2$  emissions associated with maritime transport. This solution helped the company optimize fuel consumption, reduce carbon emissions, and comply with environmental regulations, contributing to sustainability efforts. The models were integrated into the broader maritime analytics platform, providing real-time insights and supporting strategic decision-making for clients seeking to lower their environmental impact.
- Engineered a Retrieval-Augmented Generation (RAG)-based application to automatically detect potentially risky bill-of-lading documents by identifying dual-use goods, which could have military and commercial applications. By applying advanced natural language processing (NLP) and machine learning techniques, the application efficiently flagged high-risk documents, reducing the likelihood of non-compliance and trade financing risks. This proactive detection mechanism streamlined the document review process, cutting down manual effort and enhancing decision-making speed. As a result, the application contributed to a 13% improvement in the efficiency of financial decision-making, allowing for faster approvals and more secure trade transactions.
- Transformed a significant portion of the ETL codebase to be cloud-ready, enabling seamless scalability and more efficient processing of large datasets for supply chain analytics. By developing Python packages that implemented core ETL functionalities, I led the transition to a cloud-based infrastructure, which reduced manual intervention by approximately 40%. This shift not only enhanced operational efficiency but also contributed to a 15% increase in revenue through the acquisition of new clients for the supply chain analytics product. The transformation process involved introducing best practices in software development, modularizing the codebase for improved maintainability, and standardizing workflows across the team, ensuring faster deployment cycles and greater reliability for clients.

**Methods & tech stacks used: Python, Deep Learning, LLM, RAG, PowerBI, DevOps, AWS**

## **Data Scientist**

October 2016 - July 2021

IBM India Pvt. Ltd.

*Bangalore, India*

- Developed an advanced framework leveraging natural language processing (NLP) and deep learning to automate the analysis of tech support tickets and identify root causes of recurring issues. By utilizing Long Short-Term Memory (LSTM) networks for text classification and later fine-tuning a BERT model, the solution achieved high accuracy in understanding and categorizing support queries. This innovation not only streamlined the support process but also provided valuable insights into common issues, enabling quicker resolution times and more proactive customer service. The solution was successfully deployed across multiple industries, including banking and insurance, leading to the acquisition of a deal valued at \$200K and securing funding for future development. This initiative enhanced client satisfaction, reduced operational costs, and opened up new business opportunities.
- Achieved an average 10% sales growth across multiple markets by applying market-mix modeling (MMX) to sales and advertisement data, providing actionable insights for optimizing marketing strategies. Specifically, I worked with datasets covering beverage items in the Indonesian and Pakistani markets for one of our leading retail customers. By analyzing the impact of various marketing and sales factors, the MMX models enabled the client to allocate resources more effectively and drive higher sales. Additionally, I developed robust frameworks that automated the generation of detailed reports and visualizations, integrating the MMX models to present data-driven insights in an easily digestible format. This not only enhanced reporting efficiency but also provided real-time actionable insights to the client, further improving decision-making processes.
- Built an advanced contextual search engine for a FAQ dataset using Latent Semantic Indexing (LSI), resulting in a 20% boost in customer satisfaction. The project involved implementing the LSI model on top of the existing question-answer datasets, significantly improving the search engine's ability to understand context and retrieve more relevant results. Unlike traditional keyword-based indexing, the LSI model captured semantic relationships between words and phrases, enabling users to receive more accurate answers to their queries. This enhancement not only improved the efficiency of the search process but also contributed to a more user-friendly experience, leading to higher customer engagement and satisfaction. The success of this solution led to \$50,000 in funding and opened up additional business opportunities with the client, further strengthening the partnership and driving continued growth.

- Served as a Machine Learning Engineer for one of IBM's largest automobile clients, focusing on building a cutting-edge automated vehicle diagnostic tool. Refined and fine-tuned a BERT-based model using transcribed call logs to accurately identify and classify vehicle issues. The solution reduced false warranty claims by 10%, significantly cutting operational costs. Additionally, played a key role in productionizing the model, ensuring seamless integration with existing systems and delivering a scalable, robust solution to support customer service operations.

**Methods & tech stacks used:** Deep learning, PyTorch, TensorFlow, Docker, Kubernetes, REST API

### Data Analyst

October 2015 - August 2016

BCausE Enterprise Pvt. Ltd.

Noida, India

- Developed automation tools to scrape customer reviews periodically for an in-house product, feeding the data into downstream sentiment classification tasks. This automation significantly reduced the time required for manual data collection and ensured a steady stream of real-time feedback for sentiment analysis, enabling the company to better understand customer perceptions. Additionally, I crafted interactive dashboards that simplified monthly report generation, providing clear insights into sales numbers and consumer activities. By automating the report generation process, the solution reduced manual effort by 15%, improving the efficiency of the team and enabling faster decision-making for business operations.

**Methods & tech stacks used:** Python, MongoDB, R Shiny

### EDUCATION

---

- **M.Sc, Statistics:** Visva Bharati University, West Bengal, India 2013 - 2015
- **B.Sc, Statistics:** Visva Bharati University, West Bengal, India 2010 - 2013

### CERTIFICATIONS

---

- Microsoft Certified: Azure Fundamentals [↗](#)
- Microsoft Certified: Azure Data Scientist Associate [↗](#)
- Microsoft Certified: Azure AI Fundamentals

### EXTRA-CURRICULAR ACTIVITIES

---

- Serving as an Industry Expert at Intellipaat [↗](#), one of India's leading EdTech platforms, where I teach students foundational and advanced concepts in Statistics, Machine Learning, Deep Learning, and their real-world applications. I design and deliver comprehensive learning modules, focusing on practical, hands-on experience that bridges theoretical knowledge with industry requirements. My role involves mentoring students, providing expert insights into the latest trends and methodologies, and guiding them on how to apply these skills in real-world scenarios. Through my instruction, I aim to equip students with the necessary tools and knowledge to succeed in the rapidly evolving fields of data science and artificial intelligence.