**Rahul Middha**

Toronto, ON, CA | (647) 383 - 9553 | rahulmiddha@gmail.com | www.linkedin.com/in/rahulmiddha

**SUMMARY**

Results-oriented ML/Data Engineer with over 6 years of experience developing machine learning models and scalable data pipelines. Proven expertise in leveraging cloud services for optimized data processing and creating impactful dashboards. Known for collaborating with cross-functional teams, mentoring peers, and driving data strategy. Dedicated to solving real-world problems using ML, focusing on search relevance, recommendations, fraud detection, and time-series forecasting. Experienced in microservice-based architecture, scalable AI applications, and secure software development.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| * **Programming**: Python, NumPy, Pandas, Scikit-learn, SQL, Java * **ETL Tools:** Informatica, dbt, Airflow * **Visualization:** Tableau, Power BI * **ML Tools:** TensorFlow, PyTorch, Pinecone, GPT-4 | * **Cloud Platforms:** Google Cloud Platform, Azure, AWS, Snowflake * **Other Tools & Technologies:** Databricks, Azure Synapse, Docker, Kubernetes, SonarQube, JUnit, Mockito |

**EDUCATION**

**MASTER OF SCIENCE – COMPUTER SCIENCE 2017**

The University of Texas at Arlington, USA

**BACHELOR OF ENGINEERING – INFORMATION SCIENCE 2014**

Visvesvaraya Technological University, India

**PROFESSIONAL EXPERIENCE**

[](https://cmicglobal.com/) **CMiC** TORONTO, CANADA

*Data Engineer L3*  *July 2023 – June 2024*

Tech Stack: Python, Scikit, NumPy, Pandas, TensorFlow, Oracle, AWS

* Designed, developed, and deployed machine learning models focusing on recommendations and time-series forecasting, achieving a 25% increase in model accuracy using TensorFlow and Scikit.
* Developed LLM-powered chatbots that extracted key information from construction documents, reducing query response time by 40%.
* Used Pinecone as a Vector Database to store Hugging Face embeddings which improved accuracy and faster responses.
* Applied clustering techniques to identify similar projects, enhancing resource allocation and planning efficiency.
* Led the creation of scalable ML services and microservice architecture for generative AI solutions using GPT-4, reducing deployment time by 35%.
* Engineered ML models for project cost forecasting, improving financial accuracy.
* Integrated AI tools with existing systems, boosting operational efficiency by 25%.
* Applied ML Ops for reliable deployment and mentored junior engineers.
* Conducted secure code reviews and threat modeling for machine learning applications, reducing security vulnerabilities by 15%.

[Logo for Affine](https://affine.ai/)**AFFINE** BENGALURU, INDIA

*Consultant Data Engineer*  *January 2022 – July 2023*

Tech Stack: GCP, AWS Glue, BigQuery, RedShift, Python, SQL, Power BI

**Social Media Datawarehouse Project**

* Designed and implemented real-time data pipelines using AWS services, accelerating processing by 30% and reducing latency by 25%.
* Created ETL scripts in Databricks for complex data transformations, enhancing retrieval speed.
* Led ML Solution Pods to develop production-ready AI applications, focusing on fraud detection and anomaly detection, achieving a 40% reduction in false positives.
* Managed data profiling and data QA, resulting errors by 25% and improving data quality.
* Identified new opportunities and insights from the data to improve models and projected ROI of proposed modifications.

**Cosmos DB Data Views and Pipeline Creation Project**

* Developed Cosmos DB data views for popular metrics, improving data accessibility and accuracy.
* Reduced processing time by 30% through efficient data pipelines using Azure Data Factory, Azure Databricks, and Synapse Analytics.
* Increased stakeholder satisfaction by 25% by creating informative and visually appealing dashboards in Power BI.
* Improved team skills and productivity by 40% through mentoring and training sessions.
* Collaborated closely with engineering and ML platform teams to ensure adoption of best practices in building and deploying scalable ML services.

[Logo for Globalsoft, Inc.](https://www.globalss.com/)**GLOBALSOFT INC** BENGALURU, INDIA

*Software Engineer*  *August 2018 – January 2022*

Tech Stack: AWS, RedShift, Snowflake, Informatica, Python, SQL

* Developed ETL pipelines in Informatica, ensuring data integrity and security, reducing data discrepancies by 30%.
* Automated batch ETL processes in Airflow, reducing manual workload by 40% monthly and increasing operational efficiency.
* Utilized full stack GCP development for scalable AI solutions which improved daily operations.
* Modeled and shifted custom SQL queries into dbt for creating materialized views.
* Implemented large-scale data warehousing programs on Snowflake Cloud and GCP Big Query, reducing storage costs by 20%.
* Enhanced application deployment efficiency by 80% using Docker.
* Designed and developed complex aggregate, join, lookup transformation rules using Informatica PowerCenter tools.

**PROFESSIONAL CERTIFICATES**

**[Logo for Google Cloud](https://google.accredible.com/4af88ecd-6279-41b9-9ad1-6b78df2e30bd) Associate Cloud Engineer - Google Cloud**

**Issued Jun 2023 – Expires Jun 2026**

**Credential ID** [**75891934**](https://google.accredible.com/4af88ecd-6279-41b9-9ad1-6b78df2e30bd)