

## **1. Complete Understanding of the Job**

- Managed the full lifecycle for the **SpiceReclaim** booking forecast model—from initial development and optimization to cloud deployment.
- Handled complex scheduling and orchestration for high-compute models, ensuring reliable performance in production using services like EC2, lambda, S3, EventBridge.
- Improved the accuracy of **Payload Prediction** models and successfully transitioned them to production environments from On-prem.
- Engineered a production-grade routing API to identify all possible source-to-destination paths under complex operational constraints, using Advance graph-based solution.
- Designed, deployed, and maintained a custom **Vector Database** to optimize costs and improve response quality for a CX (Customer Experience) email generator.
- Built scalable, secure frontends and integrated them with serverless AWS backends for both CX Chatbots and Crew Pairing generators.

## **2. Initiative and Drive to Meet Job Targets**

- Spearheaded the migration of multiple model deployments from on-premise to AWS, utilizing **EC2, Lambda, EventBridge, and S3** for automated scheduling.
- Proactively introduced **MLflow and DVC** to existing projects to ensure better experiment tracking, data versioning, and model reproducibility.
- Established **CI/CD pipelines** across all projects to streamline code versioning and ensure seamless, low-risk rollouts using Gitlab runners and AWS.
- Developed a standardized framework for LLM-based projects while working in CX chatbot, significantly reducing the development time for future AI initiatives.

## **3. Team Player**

- Partnered closely with my own team to refine models like Booking Forecast and Payload Prediction, contributing code and architectural suggestions.
- Worked with the **Kargo360** team to integrate routing APIs, incorporating feedback to meet specific operational requirements.
- Worked with non-technical team like CX, converting high-level business problems into viable technical solutions.

- Collaborated with Product and Testing teams to ensure rigorous data validation and transparent sharing of results.
- Constantly In touch with Data Mart team, explaining them requirement and solving Query related to any of our projects.

#### **4. Planning, Organizing, and Setting Priorities**

- Successfully balanced multiple high-priority projects (e.g., SpiceReclaim and CX Chatbot) simultaneously, delivering all milestones ahead of schedule.
- Managed legacy model improvements and feature additions while concurrently migrating older architectures to the AWS for payload and booking forecast model.
- Maintained high uptime for production environments, for project like SpiceReclaim and CX Chatbot, providing rapid hotfixes for live projects without slowing down the development of new features.

#### **5. Work Under Pressure**

- Investigated and resolved critical bugs in the Booking Forecast model during a high-stakes "time crunch" just days before the **SpiceReclaim** launch.
- Delivered a fully functional CX Chatbot—complete with a custom Vector DB and secure frontend—with a tight 5-month window while working and supporting parallel project.
- Provided support during the CX Chatbot testing phase, resolving production issues and solved bugs over a single weekend to meet the final lunch deadline.

#### **6. Working Beyond Job Confines**

- Though focused on Data/Backend, I took the initiative to build secure **Streamlit** frontends and integrated them using **AWS** Infra like lambda, ALB and Step-Fucntion.
- Self-taught and implemented **FastAPI** security protocols, including authentication and input validation, to ensure API integrity for the Cargo360 and CX teams.
- Took ownership of the production pipeline by creating **GitLab CI/CD** workflows and managing multi-branch (Dev/Prod/Feature) environments.
- Leveraged serverless AWS architectures to minimize operational costs while maintaining high scalability and security standards, for project like crew-pairing and CX Chatbot.
- learning and using tools like Docker, MIFlow, DVC to Gitlab-action to provide better solution for real problems.