**Introduction**  
  
  
Hello, my name is Prachi, and I am a Data Scientist with over four years of experience in Python, SQL, Machine Learning, and NLP. **Throughout my career, I have worked extensively on AI-driven solutions, specializing in Natural Language Processing (NLP), deep learning, and large language models (LLMs).**

I have hands-on experience in developing end-to-end pipelines, leveraging cloud technologies like AWS Textract, S3, Lambda, and SageMaker to build scalable AI applications. My expertise includes data preprocessing, feature engineering, model development, and deployment using FastAPI.

Some of my key projects include developing AI-powered chatbots for education and e-commerce, automating document processing workflows using OCR and NER models, and building predictive models for logistics optimization. My work has been focused on solving real-world business challenges using advanced AI and ML techniques.

I am passionate about continuous learning and staying updated with the latest advancements in AI to drive impactful, data-driven solutions. I am excited about the opportunity to contribute my skills and expertise to your team.

**SNP project N23-**----

I recently worked on a document classification and Entity extraction system powered by Large Language Models (LLMs) on AWS. The goal of this project was to automate the classification of various documents regards to **OIL and GAS** and extracting entities with their values by leveraging LLMs for semantic understanding.

The system was fully cloud-native, using AWS services for scalability, automation, and cost efficiency.

The entire workflow follows from uploading the documents over s3, we take documents from s3 and perform ocr extraction using AWS textract service. The first page text is sent to the model deployed over Amazon ECS , where model performs binary document classification or document classifcation between OIL and GAS and OTHER category, if confidence score comes above the threshold we proceed for further classification and if it falls down , we trigger human review there using label studio and correct the classifcation. Futher entire text of document is sent to model for entity extraction , if confidence score of entities is above the threshold , we proceed further to deliver the results to users else here we triggered human review using label studio to correct the results.

A confidence score is a numerical value that represents how certain a machine learning model is about its prediction. It is usually expressed as a percentage or a probability between 0 and 1.

**Choose a threshold** where precision and recall are balanced, based on business needs.

* Example:
* If precision > recall is more important (e.g., legal documents), set a higher threshold (e.g., 0.90).
* If recall > precision is more important (e.g., medical reports), set a lower threshold (e.g., 0.75).That is ensuring false negatives to be less

For our case :  
 🔹 If misclassification or False Positive is costly (e.g., contracts), set Threshold = 90%.

We monitored performance post-deployment using AWS CloudWatch & SageMaker Model Monitor.

**Questions has been asked** ---

**Que1.. How did you handled with the people of US?**

### **## Effective Communication**

"Working with a US client, I quickly adapted to their communication style by ensuring clarity, conciseness, and responsiveness. I used structured emails and detailed meeting agendas to keep conversations productive. Understanding time zone differences, I scheduled meetings during overlapping hours and maintained asynchronous communication through Slack, Jira, or Trello. For updates, I prepared concise reports so they could review them at their convenience. This approach reduced back-and-forth discussions and improved efficiency."

💡 Example:  
"For instance, when we needed urgent approvals, I provided detailed yet to-the-point emails with action items so the client could make quick decisions without confusion. This ensured smooth workflows despite time differences."

**Que2.. How did you handle the timings working with the US people.**  
We starts our day around 4PM and communicate with them through scheduled meetings ducirn this overlapping hours.

1️⃣ Identifying Overlapping Work Hours  
"Since the Pacific Time Zone is 12.5 to 13.5 hours behind IST, I identified the best overlapping hours. Typically, 8:00 AM - 12:00 PM PST aligns with 8:30 PM - 12:30 AM IST, so I scheduled critical meetings and discussions during this period."

💡 Example:  
"For a project review, I scheduled calls at 9:00 AM PST (9:30 PM IST), ensuring both teams could discuss progress without inconvenience."