IDfy Interview Experience - Dhruv Ojha

Online Assessment

The assessment consisted of two main parts:

- **Multiple Choice Questions (5 questions):** These were theoretical questions covering:
    - Inverse of a symmetric matrix
    - Order of kurtosis values
    - The effect of the p-value on hypotheses
    - Lasso and Ridge regression
    - Identifying SQL queries that would produce errors
- **Coding Task:** This involved working in a Jupyter Notebook to build a complete machine learning pipeline for a given dataset. The requirements included:
    - **Preprocessing:** Dropping null values, standardizing and normalizing data, encoding, and feature engineering for time-series data.
    - **Pipeline and Model:** Creating a pipeline with all preprocessing steps and selecting a suitable model.
    - **Task 1:** Creating and visualizing a dataframe that highlighted the top 20 features of the model.
    - **Task 2:** Generating a .csv file containing two columns: the 'id' from the test data and the corresponding model predictions.

Interview Process

Round 1: Technical Interview (Resume-focused)

Duration: 20-25 minutes

This round was primarily based on your resume.

- The main focus was on the IPD project I had built using OCR and NLP, covering its implementation and model performance.
- Scenario-based questions were presented, such as:
    - How to detect face-swapping in a digital KYC scenario.
    - Explaining the face verification process used in dating apps.

Round 2: Technical Interview (ML-focused)

Duration: 1.5 hours (For me, for others it was aorund 30-40 mins)

- The interview started with the KYC scenario (like the previous round, but focused on making a product out of it). This included discussing the input process, pipeline design, required preprocessing steps, and potential model failure points.

- Questions were asked about which layers of a model to freeze during fine-tuning and why, and how to improve the quality of training data.
- I was asked to identify appropriate metrics to measure model performance, explain how they work, and how they would influence model selection.
- A deep dive into the concepts of Transformers, including:
    - Key-Query-Value (KQV) pairs and KV caching.
    - The mechanics of encoder and decoder blocks.
    - Comparison with RNNs and LSTMs, specifically how Transformers better handle long-term dependencies and why they are more computationally efficient.
- The discussion covered why I might recommend CNNs and how to adapt them to focus on global feature extraction instead of just local features.
- I was asked about my experience in training models. This led to a discussion on evaluating OCR models and other relevant metrics.
- A problem was presented with continuous, categorical, and plain text features, and the candidate had to describe the necessary preprocessing steps before applying logistic regression.

Round 3: Behavioral Interview

Duration: 15-20 minutes

This final round focused on behavioral and reflective questions.

- I was asked for feedback on the interview process and the PPT.
- Based on a passing comment, the interviewer asked me how I would design an AI-driven interview system differently.
- I had also mentioned that my previous interview was fairly long and I was asked for my opinion on whether longer interviews are better.
- A common question asked to all people was, "Who are you outside of your resume?" I responded by discussing my hobbies and how those activities have contributed to my personal growth.