

# SHIVAM TANEJA

## AI/ML FULL-STACK ENGINEER

23shivamtaneja@gmail.com  
+91 9821090076  
linkedin.com/in/shivam-taneja200

[Leetcode Profile](#) | [Kaggle Profile](#) | [Codeforces Profile](#) | [Github Profile](#)

B.E. Computer Engineering | Thapar Institute Of Engineering and Technology, Patiala | 8.88 CGPA

### EXPERIENCE

#### Software Engineer - J.P. Morgan Chase & Co., Bengaluru (Hybrid) Jan 2024 - Present

- Designed and implemented the front-end for a dashboard using React.js for performance metrics
- Built a document parser using PaddleOCR and clustering algorithms like DBSCAN and UMAP, which achieved a **76.3% precision** and **81.2% recall**.
- Developed a clustering solution using the DBSCAN algorithm for error categorization, enhancing customer support efficiency. **Cluster Cohesion: 91.4% and Cluster Separation of 90.8%**
- Ported **1.2 Petabytes** of data to AWS S3 and optimized application retrieval, achieving estimated cost savings exceeding **\$1 million** annually by reducing on-premise storage and operational costs.

#### Samsung R&D Institute of India, Bengaluru (Remote) Nov 2022 - May 2023

- Led a team of six members in conducting research focused on the Region of Interest in the context of sports.
- Extracted data through web scraping to curate a dataset intended for model training and assessment, which is now available for public use under Apache 2.0 license.
- Established a streamlined process for detecting Regions of Interest in sports, achieving an accuracy of **over 84%** by employing PaddleOCR and DistilBERT.

### PROJECTS

#### Kaggle - LLM Science Exam [Certificate Link](#)

- Utilized K-Means clustering to **reduce RAG size by 95%**, solving the problem of Out Of Memory errors.
- Employed a Retrieval Augmented Generation strategy along with ensembling, which increased Mean Average Precision by **7% to 0.902**
- Utilizing **FAISS on TF-IDF embeddings** resulted in a **20% reduction in retrieval time**.
- Received a **bronze medal** on Kaggle and a rank of **172/2662 (Top 7% Worldwide)**

#### Kaggle - The Learning Agency Lab - PII Data Detection [Certificate Link](#)

- Using **quantization** achieved a **30% speed-up** for smaller models and **200%** for larger LLMs.
- A weighted ensemble of 4 different DeBERTa models was used to get a **F-score (Beta = 5) of 0.957**
- Received a bronze medal on Kaggle and a rank of **164/2048 (Top 9% Worldwide)**

#### Sahar Wellness | [saharwellness.in](https://saharwellness.in)

- Developed a responsive slot booking application from the ground up using **MongoDB, React, and Node.js**, enabling seamless scheduling for users, therapists, and administrators.
- Designed the **complete system architecture**, including **dedicated testing** and **production server setups**, and **implemented CI/CD pipelines** to ensure a stable, automated, and scalable deployment process.
- Integrated **AWS SQS** and **Lambda** to **decouple** payment verification, significantly reducing failures **caused by users closing the payment page prematurely**, resulting in improved reliability and a smoother user experience.
- Optimized front-end performance by implementing **React Lazy** and **useContext**, reducing initial UI loading time from **3 seconds to 0.3 seconds** and significantly enhancing responsiveness across devices.

### ACHIEVEMENTS

- ICPC Preliminary Round 2023: **All India Rank: 125**
- ICPC Regionals Finals 2023: Kanpur Site
- ICPC Regionals Finals 2023: Chennai Site
- [AWS Cloud Practitioner Certified](#)
- 3 star Codechef: 1600+ Rating
- 1750+ Rating LeetCode: Top 12% worldwide
- Kaggle: Competitions, Notebook & Dataset Expert
- Toastmasters Public Speaker