**DETAIL PROJECT REPORT (DPR)**

**Attendance Based Face Recognition System**

**Abstract**

Gradual increase in the number of students, workers and so on across schools, colleges and work stations have led to an increased need of technology in taking attendance across board. Attendance maintenance is a significant function in all institutions to monitor the performance of people.

**Problem Statement**

Jettisoning the old, archaic and sometimes unreliable ways of taking attendance across various institutions is the reason why modern and better ways of taking attendance through the use of technology is the way. The project therefore involves using face recognition system to take attendance of workers.

**Benefits of the Project**

* It is easier, faster and more reliable ways of attendance taking
* Save storage of attendance for further statistical operations

**Project Architecture**

**Q1. Tell me about your Project**

The aim of the project is to accurately identify a person and correctly mark the person present. The essence is to jettison the archaic system of taking attendance. . First, the user has to register himself/herself in the application providing the required details. The application takes multiple snaps of the user and then stores it into the database. Once the same user comes before the camera again, the application captures the image, references it against the already stored images in the database, and then marks the attendance, if the user is present in the database.

**Q2. What was the data type?**

The data used for training this model consist of images of the worker in the organization and they are then converted to tensor objects, which have a float 32 representation.

**Q3. What was the team size and distribution?**

The team consisted of:

* Product Manager,
* 1 Technical Architect,
* 1 Lead,
* 2 Dev-Ops engineers,
* 2 QA engineers,
* 2 UI developers, and
* 3 Data Scientist

**Q4. How were you maintaining the failure cases?**

Suppose the model was not able to make a correct prediction for an image. In that case, that image gets stored in the database. There will be a report triggered to the support team at the end of the day with all the failed scenarios where they can inspect the cause of failure. Once we have a sufficient number of cases, we can label and include those images while retraining the model for better model performance.

**Q5. What is the kind of automation for data processing?**

A full-fledged ETL pipeline is in place for data extraction processing and loading into the database

**Q6. Have you used any scheduler?**

Yes!

A scheduler was used to retrain the model after a combine period of twenty days.

**Q7. How are you monitoring job?**

There are logging set-ups done. We regularly monitor the logs to see for any error scenarios. For fatal errors, we had email notifications in place. Whenever a specific error code, which has been classified as a fatal error occurs, email gets triggered to the concerned parties.

**Q8. What were your roles and responsibilities in the project?**

I participated in both core and non-core data science area particularly the image validation, data preprocessing, model selection and dash boarding

**Q9. What was your day to day task?**

My day to day tasks involved completing the task assigned to me by the project manager. Some of these task include

* Attending the scrum meetings,
* Participating in design discussions and requirement gathering,
* Doing the requirement analysis,
* Data validation,
* Unit test for the models

**Q10. In which area you have contributed the most?**

I contributed the most to Data transformation and model training areas. Also, we did a lot of brainstorming for finding and selecting the best algorithms for our use cases. After that, we identified and finalized the best practices for implementation, scalable deployment of the model, and best practices for seamless deployments as well

**Q11.In which technology you are most comfortable?**

I have worked across diverse areas in the industry using different technological stack ranging from Machine Learning, Deep Learning and NLP. With my experience and exposure, I love working in Machine Learning and Deep Learning

**Q12.How you rate yourself in big data technology?**

Being someone that is hungry for growth across the tech industry, Big data is the area I am working seriously on to become proficient. I’m currently learning SPARK

**Q13. In how many projects you have already worked?**

Giving a precise figure will be difficult but I have worked in various small and large scale projects e.g Deep learning, Computer Vision, NLP projects, Chatbot building, Machine learning with regression, and classification problems.

**Q14. How would you rate yourself in distributed computation?**

I will rate myself 7

**Q15. In which part of machine learning have you already worked on?**

I have worked on both supervised and unsupervised machine learning approaches and building different models using the as per the requirement of the client/user

**Q16. What are the areas of machine learning algorithms that you already have explored?**

I have explored various machine learning algorithms like Linear Regression, Logistic Regression, L1 and L2 Regression, Polynomial Regression, Multi Linear Regression, Decision Trees, Random Forests, Extra Trees Classifier, PCA, XGBoost, CAT Boost, ADA Boost, Gradient Boosting, Light Boost, K-Means, K-Means ++, LDA, QDA, KNN, SVM, SVR, Naïve Bayes, Agglomerative clustering, DBSCAN, Hierarchical clustering and so on

**Q17. At what frequency are you retraining and updating your model?**

The model gets retrained every 20 days.

**Q18. How would you rate yourself in machine learning?**

On the scale of 10, I will rate myself 8

**Q19. Why are you leaving your current organization?**

Being in the organization was a blessing but what we do is becoming monotonous and boring. The quest to explore and seek for knowledge, skills and a fresh challenge is the reason behind my job search

**Q20. In which mode have you deployed your model?**

I have deployed the model both in cloud environments as well in the on-premise ones based on the client and project requirements

**Q21. How were you doing deployment?**

The mechanism of deployment depends on the client's requirement. For example, some clients want their models to be deployed in the cloud, and the real-time calls they take place from one cloud application to another. On the other hand, some clients want an on-premise deployment, and then they do API calls to the model. Generally, we prepare a model file first and then try to expose it through an API for predictions/classifications. The mechanism in which he API gets called depends on the client requirement

**Q22. What is your area of specialization in machine learning?**

Versatility is my strength, I don’t really have an area of machine learning that is my favorite. I have worked with a number of algorithms according to the use case I am solving

**Q23. How did you do Data enrichment?**

Data enrichment in vision problems mostly consists of image augmentation. Apart from image augmentation, we tried to train the model with images with different lighting conditions, with b/w and colored images, images from different angles, etc.

**Q24. How did you do Data validation?**

Data validation is done by looking at the images gathered. There should be ample images for the varied number of cases like change in the lighting conditions, distance from the camera, movement of the user, the angle at which camera is installed, the position at which the camera is installed, the angle at which the snap of the user has been taken, the alignment of the image, the ratio of the face and the other areas in the image etc.

**Q25. What is your future objective?**

My future objective is to continuously grow in skills, knowledge and also to contribute to technological inventions that will advance humanity in general

**Q26.What kind of challenges have you faced during the project?**

The biggest challenge that we face is in terms of obtaining a good dataset, cleaning it to be fit for feeding it to a model, and then labeling the prepared datasets. Labeling is a rigorous task and it burns a lot of hours. Then comes the task of finding the correct algorithm to be used for that business case. Then that model is optimized. If we are exposing the model as an API, then we need to work on the SLA for the API as well, so that it responds in optimum time.

**Q27. What was the data type?**

The data used for training this model consisted of the images of workers in a particular organization then are converted to tensor objects, which have a float 32 representation.

**Q28. How were you creating and maintaining the logs?**

The logs are maintained using SQL server. The logging starts with the start of the application. The start time of the application gets logged. After that, there are loggings for entry and exits to the individual methods. There are loggings for the error scenarios and exception block as well.

**Q29. How were you doing deployment?**

Generally speaking, the mechanism of deployment depends on the client's requirement. The client specifically asked for an on-premise deployment of the model and that was what we delivered

**Q30. What is your expectation?**

My expectation is to work on multiple projects across several domains in order to further develop my skill and to generally contribute to the overall success of the company, society and the entire country