Describe the tasks you performed after training the initial model to further improvise the project score.

After training the initial model, I undertook the following steps to further improve the project score:

1. \*\*Addressing Class Imbalance\*\*:

- The dataset exhibited a class imbalance, which could potentially bias the model towards the majority class, reducing its ability to correctly predict the minority class.

- To mitigate this, I used techniques like \*\*Random Oversampling\*\* or \*\*SMOTE (Synthetic Minority Over-sampling Technique)\*\* to generate synthetic samples for the minority class, ensuring that the model had a more balanced distribution of data to train on.

2. \*\*Feature Transformation with Quantile Transformer\*\*:

- To handle non-normal features, I applied the \*\*Quantile Transformer\*\* to transform the features into a normal distribution. This helped improve model performance, especially for models sensitive to data distribution.

- The quantile transformation made the features more normally distributed, reduced skewness, and helped the model learn better relationships between features.

3. \*\*Fine-tuning Hyperparameters\*\*:

- After the initial training, I performed \*\*hyperparameter tuning\*\* using techniques like \*\*Grid Search\*\*. This helped in identifying the optimal set of parameters for models like \*\*Random Forest\*\* to achieve better accuracy and precision.

6. \*\*Model Evaluation and Comparison\*\*:

- I evaluated multiple models such as \*\*RandomForestClassifier\*\*, and \*\*BaggingClassifier\*\*, analyzing metrics like \*\*accuracy\*\*, \*\*precision\*\*, \*\*recall\*\*, and \*\*F1-score\*\*. This allowed me to select the best performing model based on various evaluation metrics rather than focusing solely on accuracy.

- \*\*Ensemble techniques\*\* were also explored to combine different models, aiming to reduce variance and improve generalization.

These steps helped refine the model, making it more robust, balanced, and capable of better generalization, resulted in a **1% improvement on the test set** and a **5% improvement on the validation set**.