# Model Optimization and Tuning Phase Template

| | |
|---|---|
| Date | 05 June 2024 |
| Team ID | 737568 |
| Project Title | AutoForesight : A Predictive Model for Streamlining Car Loan Repayment Planning |
| Maximum Marks | 10 Marks |

## Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining machine learning models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

## Hyperparameter Tuning Documentation (6 Marks):

| Model | Tuned Hyperparameters | Optimal Values |
|---|---|---|
| Random Forest | `(n_estimators, random_state)` | `(n_estimators=200, random_state=42)` |
| Decision Tree Classifier | `(criterion, random_state)` | `(criterion='entropy', random_state=0)` |
| K-Nearneast Neighbour | n_neighbors | n_neighbors=15 |

## Performance Metrics Comparison Report (2 Marks):

| Model | Baseline Metric | Optimized Metric |
| --- | --- | --- |
| Random Forest | Training Accuracy = 0.998252 | Test Accuracy= 0.99825 |
| K-Nearneast Neighbour | Training Accuracy =0.873415 | Test Accuracy=0.855944 |
| Naïve Bayes Classifier | Training Accuracy = 0.639234 | Test Accuracy== 0.544299 |

**Final Model Selection Justification (2 Marks):**

| Final Model | Reasoning |
| --- | --- |
| Random Forest | Random Forest Regressor emerged as a contender due to its ability to mitigate overfitting and handle noisy data. It is an ensemble method that combines multiple decision trees, providing robust predictive accuracy. Random forests are less sensitive to outliers and noise compared to individual decision trees. |
| KNN(K-Nearest Neighbors) | KNN was assessed for its simplicity and intuitive approach. It makes predictions based on the average of the k-nearest neighbors in the feature space, without making strong assumptions about the underlying data distribution. KNN is suitable for capturing complex, non-linear relationships, especially in smaller datasets. |

| | |
|---|---|
| Naïve Bayes Classifier | The Naive Bayes classifier is chosen due to its simplicity, efficiency, and effectiveness in handling large datasets with high-dimensional features, which are common in financial data. Its probabilistic nature provides clear insights into the likelihood of loan repayment default, making it easier to interpret and communicate results to stakeholders. Despite its assumption of feature independence, the classifier performs robustly in real-world applications, ensuring reliable and timely predictions for loan repayment planning. |
| Decision Tree Classifier | The Decision Tree classifier is selected for its intuitive and interpretable structure, making it easy for stakeholders to understand the decision-making process. It efficiently handles both numerical and categorical data, accommodating the diverse features present in vehicle loan data. Additionally, it can capture complex interactions between features without requiring feature independence, providing more nuanced predictions. Its ability to visualize decision paths aids in identifying key factors influencing loan repayment, enhancing the overall decision-making process. |