**Extended Project Summary**

**1. Project Chosen & Why**

The project I chose for my extended work is Credit Risk Modeling.  
I selected this project because it is one of the most practical and important applications of machine learning in the finance and banking sector.  
In real life, banks need to decide whether a person is likely to repay a loan or default.  
Using data and predictive models helps them make smarter, safer decisions instead of relying only on manual judgment.

I personally enjoyed this project because it involves both financial understanding (like debt, income, and payment behavior) and technical learning (like regression, accuracy scores, and ROC curves).  
It gave me a chance to apply data analysis and machine learning concepts to a real-world problem that financial institutions face every day.

**2. Enhancements Made**

In the original version of my project, I used only one algorithm — Logistic Regression — to predict whether a person will default or not.  
While Logistic Regression worked well, I wanted to see if I could make the model more powerful and accurate.

For the extended version, I added one new model:

* **Random Forest Classifier**

**What Random Forest does:**  
It creates many small decision trees and then combines their results to make a final decision.  
Because of this, it can capture complex relationships between features such as income, debt, and payment history.  
It doesn’t assume the data is linear like Logistic Regression does — which makes it more flexible and accurate.

**Steps taken for enhancement:**

1. Trained both models — Logistic Regression and Random Forest — on the same dataset.
2. Calculated evaluation metrics like Accuracy, Precision, Recall, and ROC-AUC score for both.
3. Compared their results in a simple table and ROC Curve plot.

This enhancement was simple to implement but made the project much stronger and more practical.

**3. Results and Impact**

After adding the Random Forest model, I compared both models’ side by side.

| **Model** | **ROC-AUC Score** | **Result** |
| --- | --- | --- |
| Logistic Regression | **0.8835** | Good baseline model |
| Random Forest | **0.9982** | Excellent, near-perfect performance |

**What these results mean:**

* The ROC-AUC score tells how well a model can separate defaulters from non-defaulters.
* A score closer to 1.0 means excellent accuracy.
* My original model (Logistic Regression) had a good score of 0.88,  
  but after adding Random Forest, the score improved to 0.99 — almost perfect.

**Impact of this improvement:**

* Random Forest identified more real defaulters (better recall).
* It reduced wrong predictions, meaning fewer safe customers were marked as risky.
* This kind of improvement is extremely valuable for banks, as it can help reduce financial losses and make lending safer.

In short, adding Random Forest made the model smarter, more accurate, and more reliable.

**4. Key Takeaways**

Here’s what I learned and understood from this extended version of the project:

1. **Better Models Give Better Results:**  
   Simple models like Logistic Regression are good for understanding, but advanced models like Random Forest give better accuracy and performance.
2. **Ensemble Learning Concept:**  
   I learned how Random Forest combines multiple decision trees to improve predictions — this idea of “many models working together” is called ensemble learning.
3. **Evaluating Models:**  
   I understood how to use metrics like Confusion Matrix, Classification Report, and especially **ROC-AUC** to measure model performance.
4. **Feature Importance:**  
   Some financial features such as **Credit Utilization** and **Debt-to-Income Ratio** play a bigger role in predicting default risk.
5. **Real-World Relevance:**  
   This type of project is used in the real banking industry for **credit scoring**, **loan approval**, and **risk assessment**, so it has real practical value.
6. **Practical Learning:**  
   This project improved my understanding of Python libraries like **Pandas, scikit-learn, and Matplotlib**, and how to train, test, and visualize ML models.

**5. Conclusion**

The extended version of my Credit Risk Modeling project shows how a small improvement — adding just one new model — can make a big difference in results.  
While Logistic Regression gave good baseline predictions, Random Forest significantly improved performance, achieving a very high ROC-AUC score of 0.9982.

This experiment taught me the importance of trying multiple models and comparing them before deciding which one works best.  
Overall, this project helped me understand how machine learning can make better financial decisions, reduce risk, and support smarter lending policies in the banking sector.