## Loan Default Prediction Project — Summary Report

**🎯 Project Goal**

The objective of this project is to **predict the likelihood of loan default** using **LendingClub’s peer-to-peer lending dataset**.  
By applying **Exploratory Data Analysis (EDA)** and multiple **Machine Learning algorithms**, the project aims to identify key risk factors that influence borrower default and assist financial institutions in minimizing lending risks.

**Models Tried**

* **Decision Tree Classifier**
* **Random Forest Classifier**
* **Decision tree Classifer**
* **XGBoost Classifier** ✅ *(Best Performer)*
* **Support Vector Classifier (SVC)**
* **Naive Bayes Classifier**
* **Artificial Neural Network (ANN)**

After performance comparison, **XGBoost Classifier** achieved the **best accuracy and generalization**.

**Final Model Performance**

* **Training Accuracy:** **89.9 %**
* **Testing Accuracy:** **88.9 %**
* **Precision / Recall / F1-Score:** Balanced across classes
* **ROC-AUC:** ≈ 0.91

✅ *The XGBoost model demonstrated excellent predictive power with minimal overfitting.*

**Key Insights from EDA**

1. **Interest Rate (int\_rate)** — Higher rates strongly correlate with higher default risk.
2. **Loan Grade & Subgrade (grade, sub\_grade)** — Lower grades (F, G) show the most defaults.
3. **Debt-to-Income Ratio (dti)** — Borrowers with high DTI are more likely to default.
4. **Annual Income (annual\_inc)** — Lower income groups default more frequently.
5. **Employment Length (emp\_length)** — Shorter employment durations link to instability.
6. **Loan Purpose (purpose)** — “Small business” and “debt consolidation” loans have higher risk.
7. **Revolving Utilization (revol\_util)** — High utilization (> 80 %) indicates credit stress.
8. **Verification Status (verification\_status)** — Verified incomes correlate with better repayment.

**Conclusion**

EDA and modeling highlight that **credit quality, income stability, and debt burden** are the strongest determinants of loan performance.  
The **XGBoost Classifier**, with nearly **89 % accuracy**, provides the most reliable predictions and can be effectively used in **risk analytics** and **credit scoring** systems for peer-to-peer lending platforms.