Gradient Boosting Classifier Report

# 1. Dataset Overview

The dataset consists of 43,190 rows and 12 columns. The target variable is 'risk\_flag', which indicates whether an individual is high-risk (1) or low-risk (0). The features include demographic and financial information such as income, age, experience, marital status, house ownership, car ownership, profession, city, state, current job years, and current house years.

# A blue and white graph AI-generated content may be incorrect.2. Model: GradientBoostingClassifier

The Gradient Boosting model was trained using scikit-learn's GradientBoostingClassifier with a maximum tree depth of 5. The model was trained to predict the 'risk\_flag' using the provided features.

# 3. Model Evaluation

Evaluation metrics used include confusion matrix, classification report, and accuracy score.

Confusion matrix and classification report code:

A screenshot of a computer screen

AI-generated content may be incorrect.

# 4. Feature Importance

The importance of each feature in the Gradient Boosting model was evaluated using `model.feature\_importances\_`. This helps understand which variables most influenced the predictions.

# A graph with blue squares AI-generated content may be incorrect.5. ROC Curve and AUC

Although Gradient Boosting does not expose probability scores as naturally as logistic regression, we can still evaluate the ROC curve and compute AUC using the predicted probabilities.

# A graph of a graph AI-generated content may be incorrect.6. Conclusion

The Gradient Boosting Classifier was moderately successful in classifying high-risk versus low-risk individuals. Further improvement may involve hyperparameter tuning, SMOTE for class imbalance, and model stacking or ensembling.