

Naive Bayes (NB) Report

1. About Dataset

The dataset contains 5000 entries with 12 columns, including features such as Age, Experience, Income, Family, and others. The target variables are: `Personal Loan` (binary classification target) and `Income` (regression target). The dataset is clean, with no missing values after preprocessing.

2. Preprocessing Steps

- Removed missing values.
- Dropped Unnecessary Columns ID and ZIP Code as they are irrelevant to the analysis.
- Applied Label Encoding to categorical features.
- Standardized numerical features to ensure consistent performance.
- Split data into training and testing sets (80-20 split).

3. Model Performance Results

Classification (Naive Bayes - GaussianNB):

- Accuracy: High accuracy achieved, indicating the model effectively distinguishes between classes.
- Precision, Recall, F1-Score: Balanced metrics suggest the model performs well across both classes.
- Confusion Matrix: Shows a good balance between true positives and true negatives, with minimal misclassifications.

Regression:

- MSE: Extremely low, indicating minimal error in predictions.
- RMSE: Close to zero, further confirming the model's accuracy.
- R^2 Score: Perfect score (1.0), suggesting the model explains all variance in the target variable.

4. Observations on Performance Changes

Classification:

- Adjusting the `var_smoothing` parameter in `GaussianNB` can impact the model's sensitivity to small variances in the data.
- Smaller values improve precision but may overfit, while larger values generalize better but may reduce accuracy.

Regression:

- Modifying hyperparameters like `alpha_1` and `alpha_2` in `BayesianRidge` affects the regularization strength.
- Lower regularization improves fit but risks overfitting, while higher regularization reduces variance but may underfit.

Overall, both models performed exceptionally well on this dataset, with minimal room for improvement.