Evaluation of ML models

Rana Hassan Hosny
GIZA SYSTEMS | RANA.HOSNY@GIZASYSTEMS.ONMICROSOFT.COM

Table of Contents

1. Pipeline for Preprocessing Input Features and Target Feature

Input Features:

1. **Normalization for Income**:
   - The income values are scaled between 0 and 1 to standardize the feature range.
2. **Boolean Conversion**:
   - The Own_Car and Own_Housing columns are converted from "Yes"/"No" to 1/0 for numerical processing.
3. **One-Hot Encoding for Gender**:
   - One-hot encoding is applied to the Gender column to ensure that both gender categories are represented fairly.

Target Feature:

- **Target Conversion**:
  - The target feature is mapped from "Approved" to 1 and "Denied" to 0, transforming it into a binary format suitable for model training.

---

2. Model Fairness Check

Process:

- **Gender Distribution Analysis**:
  - We analyzed the gender distribution across training, testing, and validation datasets to assess fairness.
- **Weighting Strategy**:
  - If a significant gender imbalance were detected, we would apply weights to underrepresented groups.
- Outcome:

- As gender distribution was found to be nearly equal, additional weighting was not needed.

Dataset	Percentage of Females	Percentage of Males
Training Set	49.94%	50.06%
Validation Set	49.57%	50.43%
Test Set	50.06%	49.94%

---

3. Ensuring Model Management of Variance and Bias

1. **Cross-Validation for Variance**:
    - Cross-validation was implemented to evaluate model performance consistency across data subsets, helping address variance and generalization.
2. **Accuracy for Bias**:
    - Accuracy was used to gauge overall predictive performance and detect any potential bias in the model's predictions.

Best Performing Models:

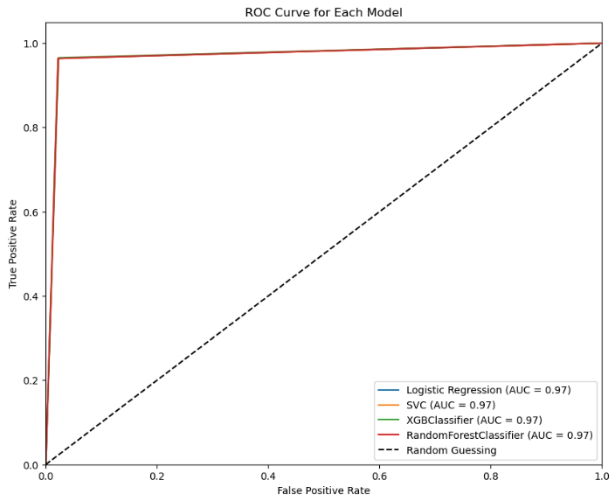
- Logistic Regression: Best Parameters - {'max_iter': 50}
- SVC: Best Parameters - {'C': 0.1, 'kernel': 'linear'}
- XGBoost Classifier: Best Parameters - {'learning_rate': 0.1, 'max_depth': 5}
- Random Forest Classifier: Best Parameters - {'max_depth': 10, 'min_samples_split': 5}

---

4. Best Model Selection Using ROC Curve and Confusion Matrix Analysis

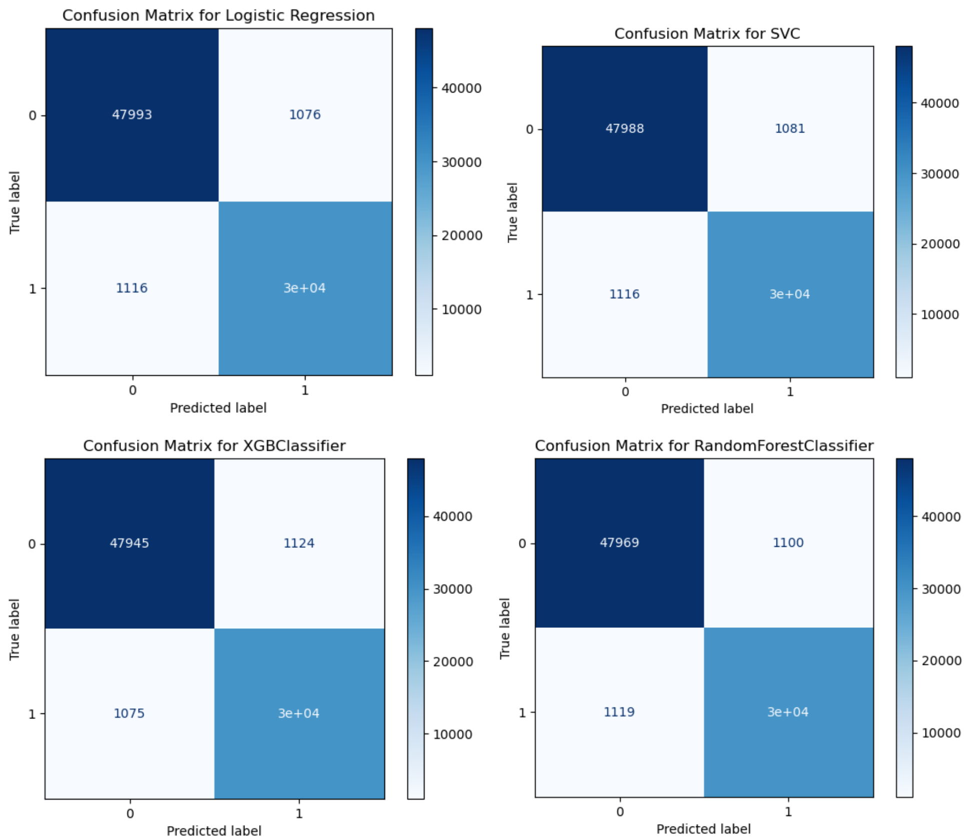
ROC Curve

- The Receiver Operating Characteristic (ROC) curve was used to compare model performance. Since all models showed nearly identical Area Under the Curve (AUC) values, further metrics were considered.



Confusion Matrix Analysis

- Given the critical goal of minimizing false positives (FP) in credit approval decisions, we prioritized models with fewer FPs. Logistic Regression showed the fewest FPs, making it the preferred model.



5. Error Analysis by Gender

To ensure fairness, we conducted an error analysis by gender. Results indicated minimal differences in error rates between genders, suggesting that gender bias is not significantly impacting the model.

Error rates per model:

Model	Female Error Rate	Male Error Rate
Logistic Regression	0.0287	0.0261
SVC	0.0288	0.0261
XGBoost	0.0287	0.0263
Random Forest	0.0290	0.0265

6. Model Interpretability: Feature Importance Scores

Feature Importances by Model

1. **Logistic Regression**:
   - Highest importance for **Income** (56.63), with negative contributions for gender features, indicating slight bias.
2. **SVC**:
   - **Income** has the largest importance (26.41), with small negative contributions from gender.
3. **XGBoost**:
   - **Income** remains significant (0.58), with a minor positive contribution from Gender_Female.
4. **Random Forest**:
   - **Income** is crucial (0.84), though interpretability is reduced due to the ensemble structure.

Feature	Logistic Regression Importance	SVC Importance	XGBClassifier Importance	RandomForestClassifier Importance
Income	56.626187	26.413798	0.584681	0.846550
Own_Housing	4.420723	2.071554	0.096817	0.031146
Own_Car	1.913403	0.903336	0.033657	0.007294
Num_Children	-0.001828	-0.002890	0.000380	0.000777
Gender_Male	-8.355581	1.467956	0.000000	0.042168
Gender_Female	-14.650782	-1.467956	0.284466	0.072065

Conclusion

- **Logistic Regression** provides the highest interpretability, followed by **SVC** and **XGBoost**, while **Random Forest** is more complex and less transparent in terms of feature impact.

Request Body Example:

POST http://127.0.0.1:8000/api/predict/

Content-Type: application/json

{

"Num_Children": [1, 2],

"Gender": ["Male", "Female"],

"Income": [40690, 41696],

"Own_Car": ["No", "Yes"],

"Own_Housing": ["Yes", "No"]

}

Response Example:

HTTP/1.1 200 OK

Content-Type: application/json

{

"predictions": [1, 0]

}

Note: In the predictions, 1 indicates that issuing the credit card is approved and 0 indicates that issuing the credit card is denied.