

Ethical AI Validator - Compliance Report

Report Information

Generated: 2025-08-11 10:30:52
Report ID: 20250811_103052

Model Information

model_name: XGBoost
config_name: aggressive
version: 1.3

Training Scenario and Hyperparameters

Scenario: XGB-aggressive

Hyperparameter	Value
base_score	None
booster	None
callbacks	None
colsample_bylevel	None
colsample_bynode	None
colsample_bytree	None
device	None
early_stopping_rounds	None
enable_categorical	False
eval_metric	None
feature_types	None
feature_weights	None
gamma	None
grow_policy	None
importance_type	None
interaction_constraints	None
learning_rate	0.2

max_bin	None
max_cat_threshold	None
max_cat_to_onehot	None
max_delta_step	None
max_depth	8
max_leaves	None
min_child_weight	None
missing	nan
monotone_constraints	None
multi_strategy	None
n_estimators	200
n_jobs	None
num_parallel_tree	None
objective	binary:logistic
random_state	42
reg_alpha	None
reg_lambda	None
sampling_method	None
scale_pos_weight	None
subsample	None
tree_method	None
validate_parameters	None
verbosity	None

Audit Criteria

bias_threshold	0.3
fairness_threshold	0.7

Bias Analysis Results

Protected Attribute	Group	Bias Score
gender	female	0.198

gender	male	0.256
age_group	18-25	0.198
age_group	26-35	0.198
age_group	36-50	0.551
age_group	50+	0.138
education	bachelor	0.194
education	high_school	0.562
education	master	0.198
education	phd	0.198

Fairness Assessment Results

Protected Attribute	Fairness Score
gender	0.772
age_group	0.713
education	0.740

Hyperparameter Impact Analysis

Parameter	Value	Risk	Rationale
class_weight	None	LOW	No class weighting can under-serve minority groups if data is imbalanced.

Likely Contributing Factors

- 1. Low fairness score observed for 'age_group' (score=0.713).
- 2. Highest bias in education -> high_school (bias_score=0.562).
- 3. Suspected hyperparameters: class_weight=None (LOW) - No class weighting can under-serve minority groups if data is imbalanced.

Overall Compliance Summary

Overall Status	PARTIALLY COMPLIANT
Bias Issues	Yes

Fairness Issues	No
Total Issues	1
Summary	Minor issues detected. Some compliance requirements need attention.

GDPR Compliance Assessment

Requirement	Status	Notes
Data Minimization	Compliant	Audit completed successfully
Purpose Limitation	Compliant	Audit completed successfully
Transparency	Non-Compliant	Bias detected - transparency compromised
Accountability	Compliant	Audit completed successfully
Right to Explanation	Non-Compliant	Bias/fairness issues affect explainability

AI Act Compliance Assessment

Requirement	Status	Notes
Risk Assessment	Non-Compliant	Bias/fairness risks identified
Transparency Requirements	Non-Compliant	Bias affects transparency
Human Oversight	Non-Compliant	Bias/fairness issues require oversight
Accuracy Requirements	Compliant	Audit completed successfully
Documentation	Non-Compliant	Bias/fairness issues need documentation

Recommendations

- 1: HIGH PRIORITY: Apply post-processing bias correction
- 2: Implement equalized odds post-processing
- 3: Implement comprehensive bias monitoring in production
- 4: Document all mitigation strategies implemented
- 5: Establish regular bias monitoring procedures
- 6: Provide model explanations for affected groups
- 7: Consider human oversight for high-stakes decisions