

Model Optimization and Tuning Phase Template

Date	21 JULY 2024
Team ID	739717
Project Title	Unlocking Silent Signals :Decoding Body Language With Mediapipe
Maximum Marks	10 Marks

Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining neural network models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (8 Marks):

Model	Tuned Hyperparameters
-------	-----------------------

<p>Ridge Classifier</p>	<p>#importing the library for LogisticRegression from sklearn.linear_model import LogisticRegression</p> <p>The LogisticRegression from sklearn.linear_model, sets the solver to 'lbfgs' and the maximum number of iterations to 1000, and fits the model to data XXX and yyy. This configuration is important for training a logistic regression model with a specific solver and iteration limit, which can enhance model convergence.</p> <pre>42]: from sklearn.linear_model import LogisticRegression model = LogisticRegression(solver='lbfgs', max_iter=1000) model.fit(X, y)</pre> <pre>42]: LogisticRegression LogisticRegression(max_iter=1000)</pre>
<p>Random Forest Classifier</p>	<p>The Random Forest Classifier is an ensemble learning method that constructs multiple decision trees during training and outputs the mode of their predictions, improving accuracy . random feature selection to reduce overfitting and handle large datasets</p> <pre>39]: from sklearn.pipeline import make_pipeline from sklearn.preprocessing import StandardScaler from sklearn.linear_model import LogisticRegression, RidgeClassifier from sklearn.ensemble import RandomForestClassifier, GradientBoostingClassifier 40]: pipelines = { 'lr':make_pipeline(StandardScaler(), LogisticRegression()), 'rc':make_pipeline(StandardScaler(), RidgeClassifier()), 'rf':make_pipeline(StandardScaler(), RandomForestClassifier()), 'gb':make_pipeline(StandardScaler(), GradientBoostingClassifier()), } 41]: pipelines.keys() 41]: dict_keys(['lr', 'rc', 'rf', 'gb']) 42]: from sklearn.linear_model import LogisticRegression model = LogisticRegression(solver='lbfgs', max_iter=1000) model.fit(X, y)</pre> <pre>42]: LogisticRegression LogisticRegression(max_iter=1000)</pre> <pre>43]: from sklearn.preprocessing import StandardScaler from sklearn.linear_model import LogisticRegression from sklearn.pipeline import make_pipeline</pre>

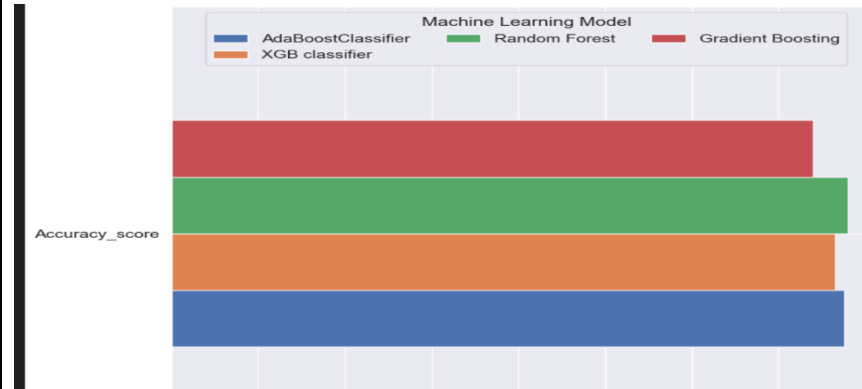
<p>XGBoost Classifier</p>	<p>These pipelines are stored in a dictionary for easy access and comparison. The LogisticRegression model is then separately instantiated and fitted, demonstrating individual model training and usage.</p> <pre> from sklearn.model_selection import GridSearchCV from xgboost import XGBClassifier rf = XGBClassifier() param_grid= {'n_estimators': [100,200,300], 'criterion':['entropy','gini'], 'max_depth' : [10,20,30], 'max_features':['auto','sqrt']} grid_search = GridSearchCV(rf, param_grid , cv = 5, n_jobs = - 1, verbose = 3) grid_search.fit(X_train,y_train) </pre> <p>✓ 46s</p> <p>Fitting 5 folds for each of 36 candidates, totalling 180 fits</p> <pre> GridSearchCV ③ ④ └─ best_estimator_ : XGBClassifier └─ XGBClassifier </pre>
-------------------------------	--

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
-------------	-----------

Random Forest

Random Forest model is chosen for its robustness in handling complex datasets and its ability to mitigate overfitting while providing high predictive accuracy.



Above all the models Random Forest model have the highest accuracy among all the models.