ML_Model

March 10, 2025

1 Enriching Vessel Location Data

1.1 Loading the Data :

[3]:
```
import numpy as np
import pandas as pd
import geopandas as gpd
import matplotlib.pyplot as plt
import matplotlib.image as mpimg
from shapely.geometry import Point
from scipy.spatial import cKDTree
from pathlib import Path
import requests
import zipfile
import io

from mpl_toolkits.basemap import Basemap
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler, RobustScaler, OneHotEncoder,␣
↪PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from sklearn.impute import SimpleImputer
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline, make_pipeline
from sklearn.metrics import mean_squared_error, mean_absolute_error,␣
↪accuracy_score, r2_score
from scipy.spatial import cKDTree
import xgboost as xgb

import geodatasets
import psycopg2
```

[4]:
```
# Connect to PostgreSQL database
conn = psycopg2.connect(
    dbname="ais_project",
    user="postgres",
    password="120705imad",
```

```python
    host="localhost"
)

# Query the raw AIS data
query = "SELECT * FROM raw_ais_data"
df = pd.read_sql(query, conn)

# Close the connection
conn.close()

# Check the first few rows
df.head()
```

```
/var/folders/xk/_4gq0t990fg3nxd8jrlkz2k00000gn/T/ipykernel_5605/1358359146.py:11
: UserWarning: pandas only supports SQLAlchemy connectable (engine/connection)
or database string URI or sqlite3 DBAPI2 connection. Other DBAPI2 objects are
not tested. Please consider using SQLAlchemy.
  df = pd.read_sql(query, conn)
```

[4]:

```
   id  vessel_id   latitude  longitude               timestamp  \
0   1  219598000  55.770832  20.851690 2022-07-30 23:28:58.646
1   2  376128000  59.180200  19.612433 2021-12-08 07:18:31.289
2   3  273274000  59.886713  30.190985 2018-10-18 01:20:47.556
3   4  231319000  56.052607  17.597902 2019-12-05 03:55:51.243
4   5  538007963  58.876633  21.255277 2023-07-03 00:43:11.378

                                            raw_json
0  {'mmsi': 219598000, 'type': 'Feature', 'geomet…
1  {'mmsi': 376128000, 'type': 'Feature', 'geomet…
2  {'mmsi': 273274000, 'type': 'Feature', 'geomet…
3  {'mmsi': 231319000, 'type': 'Feature', 'geomet…
4  {'mmsi': 538007963, 'type': 'Feature', 'geomet…
```

previewing data :

[6]:

```python
df['raw_json'].iloc[0]
```

[6]:


```
{'mmsi': 219598000,
 'type': 'Feature',
 'geometry': {'type': 'Point', 'coordinates': [20.85169, 55.770832]},
 'properties': {'cog': 346.5,
  'rot': 4,
  'sog': 0.1,
  'mmsi': 219598000,
  'raim': True,
  'posAcc': True,
  'heading': 79,
  'navStat': 1,
```

```
'timestamp': 59,
'timestampExternal': 1659212938646}}
```

Based on the preview and the official data source documentationn here's a breakdown of the structure:

Feature Object (Individual Vessel) - `type`: Always `"Feature"` (geospatial object with geometry + properties). - `mmsi`: **Maritime Mobile Service Identity** (9-digit unique vessel identifier). - `geometry`: - `type`: Always `"Point"` (coordinates represent a single position). - `coordinates`: [**Longitude, Latitude**] in WGS84 decimal degrees (EPSG:4326).

Properties Object (Vessel Metadata) - `sog`: **Speed Over Ground** (0.1 = 0.1 knots in this example). - `cog`: **Course Over Ground** (346.5° true north, 0-359.9°). - `navStat`: **Navigational Status** (1 = "Underway using engine" per AIS standards). - `rot`: **Rate of Turn** (±127 = degrees/minute, 4 = slow turn in this example). - `posAcc`: **Position Accuracy** (`true` = high-accuracy GPS/DGPS, `false` = >10m error). - `raim`: **Receiver Autonomous Integrity Monitoring** (`true`/`false` for GPS reliability checks). - `heading`: **True Heading** (79° in this example, 0-359°). - `timestamp`: **Internal timestamp** (59 = seconds from last minute/UTC second). - `timestampExternal`: **External timestamp** (1659212938646 = Unix epoch milliseconds = Thu Jul 30 2022 12:28:58 UTC).

Key Notes:

- Coordinates follow GeoJSON `[lon, lat]` order (not lat/lon).
- Timestamps may need conversion for human readability (use `datetime` libraries).
- MMSI prefixes indicate vessel origin (219 = Denmark in this example).

1.2 2- Data Preprocessing

[8]:
```
# Extract properties from raw_json
df['sog'] = df['raw_json'].apply(lambda x: x['properties']['sog'])
df['cog'] = df['raw_json'].apply(lambda x: x['properties']['cog'])
df['navStat'] = df['raw_json'].apply(lambda x: x['properties'].get('navStat'))
df['rot'] = df['raw_json'].apply(lambda x: x['properties'].get('rot'))
df['posAcc'] = df['raw_json'].apply(lambda x: x['properties'].get('posAcc'))
df['raim'] = df['raw_json'].apply(lambda x: x['properties'].get('raim'))
df['heading'] = df['raw_json'].apply(lambda x: x['properties'].get('heading'))
df['timestampExternal'] = df['raw_json'].apply(lambda x: x['properties'].
 ↪get('timestampExternal'))

# Drop unnecessary columns
df = df.drop(columns=['id', 'raw_json'])

# Convert timestamp to datetime
df['timestamp'] = pd.to_datetime(df['timestamp'])

# Extract vessel origin from MMSI
mid_country = pd.read_csv('mid_to_country.csv').set_index('MID')
df['mid'] = df['vessel_id'].astype(str).str[:3].astype(int)
```

```
df['country'] = df['mid'].map(mid_country['Country'])

# Convert boolean features to numerical
df['posAcc'] = np.where(df['posAcc'], 1, 0)
df['raim'] = np.where(df['raim'], 1, 0)

# Sort by timestamp
df = df.sort_values('timestamp')
```

1.3 3- Feature Engineering

[10]:
```
# Calculate time differences in seconds
df['time_diff'] = df['timestamp'].diff().dt.total_seconds().fillna(0)

# Corrected positional delta calculations
df['delta_lat'] = (df['sog'] * 0.51444 * np.cos(np.radians(df['cog'])) *␣
 ↪df['time_diff']) / 111_111
df['delta_lon'] = (df['sog'] * 0.51444 * np.sin(np.radians(df['cog'])) *␣
 ↪df['time_diff']) / (111_111 * np.cos(np.radians(df['latitude'])))

# Set deltas to 0 when time_diff <= 0
df.loc[df['time_diff'] <= 0, ['delta_lat', 'delta_lon']] = 0

# Fill NaN values
df['delta_lat'] = df['delta_lat'].fillna(0)
df['delta_lon'] = df['delta_lon'].fillna(0)
```

1.4 4- Geospatial features

[12]:
```
import requests
import zipfile
from io import BytesIO
from pathlib import Path
from shapely.geometry import Point

# Download Natural Earth dataset
url = "https://naciscdn.org/naturalearth/10m/cultural/ne_10m_admin_0_countries.
 ↪zip"
extract_path = Path("natural_earth_data")
if not (extract_path / "ne_10m_admin_0_countries.shp").exists():
    response = requests.get(url)
    with zipfile.ZipFile(BytesIO(response.content)) as z:
        z.extractall(extract_path)

# Load land geometry (Baltic Sea region)
world = gpd.read_file(extract_path / "ne_10m_admin_0_countries.shp")
```

```
land = world[world['NAME'].isin(['Sweden', 'Finland', 'Denmark', 'Germany',␣
↪'Poland', 'Lithuania', 'Latvia', 'Estonia', 'Russia'])].geometry

# Create geometry and check if points are over water
points = gpd.GeoDataFrame(df, geometry=gpd.points_from_xy(df['longitude'],␣
↪df['latitude']))
df['is_water'] = ~points.geometry.apply(lambda pt: land.contains(pt).any())
```

1.5 5- Preparing Features & Targets

[31]:
```
from sklearn.model_selection import train_test_split

# Feature set
features = ['latitude', 'longitude', 'sog', 'cog', 'navStat', 'rot', 'posAcc',␣
↪'raim', 'heading', 'time_diff', 'delta_lat', 'delta_lon', 'is_water']
X = df[features]
y = df[['latitude', 'longitude']].shift(-1)  # Next position

# Remove rows with NaN targets
mask = ~y.isna().any(axis=1)
X = X.loc[mask]
y = y.loc[mask]

# Split data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,␣
↪random_state=19)
```

1.6 6 - Building & Testing Models

[33]:
```
from sklearn.compose import ColumnTransformer
from sklearn.pipeline import Pipeline
from sklearn.preprocessing import OneHotEncoder, RobustScaler,␣
↪PolynomialFeatures
from sklearn.impute import SimpleImputer
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor
from xgboost import XGBRegressor

# Preprocessing pipeline
preprocessor = ColumnTransformer(
    transformers=[
        ('cat', OneHotEncoder(drop='first', handle_unknown='ignore'),␣
↪['navStat']),
        ('num', Pipeline([
            ('imputer', SimpleImputer(strategy='median')),
            ('scaler', RobustScaler())
```

```
        ]), ['latitude', 'longitude', 'sog', 'cog', 'rot', 'heading',␣
↪'time_diff', 'delta_lat', 'delta_lon', 'is_water']),
        ('binary', 'passthrough', ['posAcc', 'raim'])
    ])

# Define model pipelines
models = {
    "Linear Regression": Pipeline([
        ('preprocessor', preprocessor),
        ('regressor', LinearRegression(n_jobs=-1))
    ]),
    "Polynomial Regression": Pipeline([
        ('preprocessor', preprocessor),
        ('poly', PolynomialFeatures(degree=2)),
        ('regressor', LinearRegression(n_jobs=-1))
    ]),
    "Random Forest": Pipeline([
        ('preprocessor', preprocessor),
        ('regressor', RandomForestRegressor(n_jobs=-1))
    ]),
    "XGBoost": Pipeline([
        ('preprocessor', preprocessor),
        ('regressor', XGBRegressor(n_jobs=-1))
    ])
}

# Train models
for name, model in models.items():
    model.fit(X_train, y_train)
    print(f"{name} trained.")
```

```
Linear Regression trained.
Polynomial Regression trained.
Random Forest trained.
XGBoost trained.
```

1.7 7 - Predictions & Post-Processing

[35]:
```
from scipy.spatial import cKDTree

# Function to check if a point is over water
def is_over_water(lon, lat, land_geometry):
    point = Point(lon, lat)
    return not any(land_geometry.contains(point))

# Function to snap predictions to water
```

```
def snap_to_water(pred, actual, land_geometry, max_attempts=5,␣
↪buffer_distance=0.1):
    pred_adjusted = pred.copy()
    water_points = actual[~np.array([not is_over_water(lon, lat, land_geometry)␣
↪for lon, lat in actual])]
    tree = cKDTree(water_points)
    land_mask = np.array([not is_over_water(lon, lat, land_geometry) for lon,␣
↪lat in pred_adjusted])

    if any(land_mask):
        distances, indices = tree.query(pred_adjusted[land_mask], k=1)
        pred_adjusted[land_mask] = water_points[indices]

    return pred_adjusted

# Make predictions and snap
predictions = {}
for name, model in models.items():
    y_pred = model.predict(X_test)
    predictions[name] = snap_to_water(y_pred, y_test.values, land)
```

1.8 8 - Evaluation

[37]:
```
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score

# Evaluation function
def evaluate_predictions(y_true, y_pred, name):
    mse = mean_squared_error(y_true, y_pred)
    lat_mse = mean_squared_error(y_true['latitude'], y_pred[:, 0])
    lon_mse = mean_squared_error(y_true['longitude'], y_pred[:, 1])
    mae = mean_absolute_error(y_true, y_pred)
    r2 = r2_score(y_true, y_pred)
    return [name, mse, lat_mse, lon_mse, mae, r2]

# Evaluate all models
results = []
for name, y_pred in predictions.items():
    results.append(evaluate_predictions(y_test, y_pred, name))

# Display results
df_results = pd.DataFrame(results, columns=["Model", "Overall MSE", "Lat MSE",␣
↪"Lon MSE", "MAE", "R^2"])
print(df_results.round(4))
```

```
                   Model  Overall MSE  Lat MSE  Lon MSE     MAE     R^2
0      Linear Regression       1.4630   0.4469   2.4791  0.4539  0.7449
1  Polynomial Regression       1.4488   0.4310   2.4666  0.4472  0.7497
```

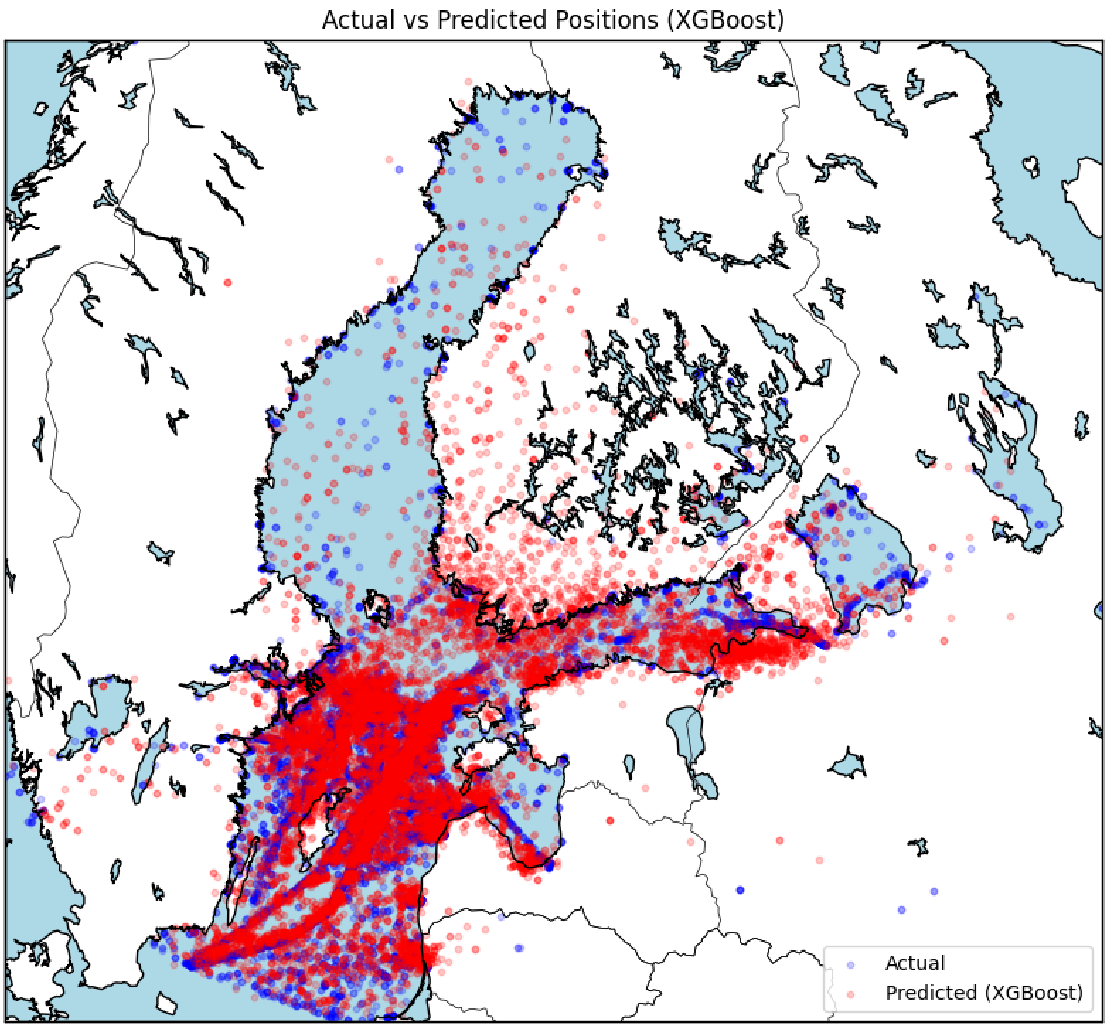
```
2         Random Forest       1.8168   0.5227   3.1109  0.5486  0.6898
3               XGBoost       1.4958   0.4444   2.5473  0.4505  0.7417
```

1.9 9 - Visualisation

[39]:
```python
import matplotlib.pyplot as plt
from mpl_toolkits.basemap import Basemap

fig, ax = plt.subplots(figsize=(12, 9))
m = Basemap(ax=ax, projection='merc',
            llcrnrlat=y_test['latitude'].min(), llcrnrlon=y_test['longitude'].
 ↪min(),
            urcrnrlat=y_test['latitude'].max(), urcrnrlon=y_test['longitude'].
 ↪max(),
            resolution='i')
m.drawcoastlines()
m.drawcountries()
m.fillcontinents(color='white', lake_color='lightblue')
m.drawmapboundary(fill_color='lightblue')

# Plot actual and predicted (e.g., XGBoost)
x_act, y_act = m(y_test['longitude'].values, y_test['latitude'].values)
x_pred, y_pred = m(predictions["XGBoost"][:, 1], predictions["XGBoost"][:, 0])
ax.scatter(x_act, y_act, alpha=0.2, s=10, color='blue', label='Actual')
ax.scatter(x_pred, y_pred, alpha=0.2, s=10, color='red', label='Predicted␣
 ↪(XGBoost)')
ax.legend()
plt.title("Actual vs Predicted Positions (XGBoost)")
plt.show()
```



1.10 10 - Tuning Hyperparameters

[41]:
```
import optuna

# Create a study and optimize the objective function with early stopping
def objective(trial):
    # Define hyperparameters
    params = {
        'n_estimators': trial.suggest_int('n_estimators', 50, 500),
        'max_depth': trial.suggest_int('max_depth', 3, 10),
        'learning_rate': trial.suggest_float('learning_rate', 0.01, 0.3),
        'subsample': trial.suggest_float('subsample', 0.5, 1.0),
        'colsample_bytree': trial.suggest_float('colsample_bytree', 0.5, 1.0),
        'gamma': trial.suggest_float('gamma', 0.0, 1.0),
        'reg_alpha': trial.suggest_float('reg_alpha', 0.0, 1.0),
```

```
        'reg_lambda': trial.suggest_float('reg_lambda', 0.0, 1.0),
        'n_jobs': -1
    }

    model = XGBRegressor(**params)
    model.fit(
        X_train,
        y_train,
        eval_set=[(X_test, y_test)],
        verbose=0
    )

    y_pred = model.predict(X_test)
    mse = mean_squared_error(y_test, y_pred)
    return mse

study = optuna.create_study(direction='minimize')
study.optimize(objective, n_trials=100)

# Get best hyperparameters
best_params = study.best_params
best_params['n_jobs'] = -1
print(best_params)

# Train XGBoost with best hyperparameters
xgb_model = XGBRegressor(**best_params)
xgb_model.fit(X_train, y_train)

# Make predictions and snap
y_pred = xgb_model.predict(X_test)
y_pred = snap_to_water(y_pred, y_test.values, land)

# Evaluate
results = evaluate_predictions(y_test, y_pred, "XGBoost (Optimized)")
new_row = pd.DataFrame([results], columns=df_results.columns)
df_results = pd.concat([df_results, new_row], ignore_index=True)
print(df_results.round(4))

# Plot actual and predicted (e.g., XGBoost)
fig, ax = plt.subplots(figsize=(12, 9))
m = Basemap(ax=ax, projection='merc',
            llcrnrlat=y_test['latitude'].min(), llcrnrlon=y_test['longitude'].
 ↪min(),
            urcrnrlat=y_test['latitude'].max(), urcrnrlon=y_test['longitude'].
 ↪max(),
            resolution='i')
m.drawcoastlines()
```

```
m.drawcountries()
m.fillcontinents(color='white', lake_color='lightblue')
m.drawmapboundary(fill_color='lightblue')

x_act, y_act = m(y_test['longitude'].values, y_test['latitude'].values)
x_pred, y_pred = m(y_pred[:, 1], y_pred[:, 0])
ax.scatter(x_act, y_act, alpha=0.2, s=10, color='blue', label='Actual')
ax.scatter(x_pred, y_pred, alpha=0.2, s=10, color='red', label='Predicted␣
  ↪(XGBoost)')
ax.legend()
plt.title("Actual vs Predicted Positions (XGBoost Optimized)")
plt.show()
```

```
[I 2025-03-10 18:56:23,301] A new study created in memory with name: no-name-e240e6ed-f449-4c4e-bf55-034a17e1885a
[I 2025-03-10 18:56:30,632] Trial 0 finished with value: 1.587729317667184 and parameters: {'n_estimators': 403, 'max_depth': 9, 'learning_rate': 0.03749283149466337, 'subsample': 0.5090998065053078, 'colsample_bytree': 0.6592197019356771, 'gamma': 0.10603731635650404, 'reg_alpha': 0.009776572831123542, 'reg_lambda': 0.058401679839618414}. Best is trial 0 with value: 1.587729317667184.
[I 2025-03-10 18:56:33,982] Trial 1 finished with value: 1.7385731717199084 and parameters: {'n_estimators': 171, 'max_depth': 10, 'learning_rate': 0.2813719779701284, 'subsample': 0.751557902153396, 'colsample_bytree': 0.9441610566327703, 'gamma': 0.980641758271928, 'reg_alpha': 0.088865052863357, 'reg_lambda': 0.1624712554980211}. Best is trial 0 with value: 1.587729317667184.
[I 2025-03-10 18:56:37,621] Trial 2 finished with value: 1.449892243153011 and parameters: {'n_estimators': 436, 'max_depth': 5, 'learning_rate': 0.0543385217533687, 'subsample': 0.7457671382295932, 'colsample_bytree': 0.6983011185932624, 'gamma': 0.14615682835613908, 'reg_alpha': 0.6737632182904713, 'reg_lambda': 0.09534171583920537}. Best is trial 2 with value: 1.449892243153011.
[I 2025-03-10 18:56:41,134] Trial 3 finished with value: 1.4818929132430188 and parameters: {'n_estimators': 419, 'max_depth': 4, 'learning_rate': 0.2095547437192582, 'subsample': 0.6794867110269865, 'colsample_bytree': 0.6773424110403728, 'gamma': 0.07671918170639602, 'reg_alpha': 0.6365425178685319, 'reg_lambda': 0.9450916737577268}. Best is trial 2 with value: 1.449892243153011.
[I 2025-03-10 18:56:46,112] Trial 4 finished with value: 1.7036181805777229 and parameters: {'n_estimators': 341, 'max_depth': 8, 'learning_rate': 0.2416984594666007, 'subsample': 0.5559612153433127, 'colsample_bytree': 0.552704193732677, 'gamma': 0.719287007300811, 'reg_alpha': 0.16495905104366537, 'reg_lambda': 0.8868321523498389}. Best is trial 2 with value: 1.449892243153011.
[I 2025-03-10 18:56:49,428] Trial 5 finished with value: 1.532152839254263 and parameters: {'n_estimators': 337, 'max_depth': 5, 'learning_rate': 0.22888997297008593, 'subsample': 0.6617246772599521, 'colsample_bytree':
```

```
0.850664162821635, 'gamma': 0.5423708434634187, 'reg_alpha':
0.16229169682106726, 'reg_lambda': 0.3246439734681881}. Best is trial 2 with
value: 1.449892243153011.
[I 2025-03-10 18:56:52,360] Trial 6 finished with value: 1.5653620094368448 and
parameters: {'n_estimators': 271, 'max_depth': 6, 'learning_rate':
0.2626190778175208, 'subsample': 0.8308017339191174, 'colsample_bytree':
0.5111274859628281, 'gamma': 0.9260231017249815, 'reg_alpha':
0.20398342848974027, 'reg_lambda': 0.08176830344769381}. Best is trial 2 with
value: 1.449892243153011.
[I 2025-03-10 18:56:55,619] Trial 7 finished with value: 1.5293021627414367 and
parameters: {'n_estimators': 339, 'max_depth': 5, 'learning_rate':
0.22688245104999072, 'subsample': 0.6653041130595141, 'colsample_bytree':
0.727013610814656, 'gamma': 0.3170868372579172, 'reg_alpha': 0.5052332239419541,
'reg_lambda': 0.3004525186109226}. Best is trial 2 with value:
1.449892243153011.
[I 2025-03-10 18:56:57,205] Trial 8 finished with value: 1.6384543085083412 and
parameters: {'n_estimators': 75, 'max_depth': 10, 'learning_rate':
0.17668814303491928, 'subsample': 0.5130935592979895, 'colsample_bytree':
0.8449074104214446, 'gamma': 0.6621546513805605, 'reg_alpha':
0.7469772976941094, 'reg_lambda': 0.033643498881472444}. Best is trial 2 with
value: 1.449892243153011.
[I 2025-03-10 18:56:59,723] Trial 9 finished with value: 1.5665395617015514 and
parameters: {'n_estimators': 234, 'max_depth': 6, 'learning_rate':
0.25826156829181307, 'subsample': 0.6360508114761396, 'colsample_bytree':
0.8913445512284879, 'gamma': 0.22575204858536824, 'reg_alpha':
0.12193501434451459, 'reg_lambda': 0.8537182140792465}. Best is trial 2 with
value: 1.449892243153011.
[I 2025-03-10 18:57:03,883] Trial 10 finished with value: 1.4200583847566084 and
parameters: {'n_estimators': 432, 'max_depth': 3, 'learning_rate':
0.06500492958403169, 'subsample': 0.9937408354503906, 'colsample_bytree':
0.5814992780663066, 'gamma': 0.3803654061735343, 'reg_alpha':
0.9948863792913043, 'reg_lambda': 0.6133120180062811}. Best is trial 10 with
value: 1.4200583847566084.
[I 2025-03-10 18:57:07,588] Trial 11 finished with value: 1.4187925573150422 and
parameters: {'n_estimators': 464, 'max_depth': 3, 'learning_rate':
0.062018354170525515, 'subsample': 0.9462816474256146, 'colsample_bytree':
0.5918128461676065, 'gamma': 0.36005686931824205, 'reg_alpha':
0.9869964464392068, 'reg_lambda': 0.6171917520161199}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:57:11,271] Trial 12 finished with value: 1.4213695295282665 and
parameters: {'n_estimators': 468, 'max_depth': 3, 'learning_rate':
0.0969969916269307, 'subsample': 0.9755517759931284, 'colsample_bytree':
0.5928816770868346, 'gamma': 0.35169635694653906, 'reg_alpha':
0.9967144302396107, 'reg_lambda': 0.6186260594979486}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:57:14,980] Trial 13 finished with value: 1.4239933968926413 and
parameters: {'n_estimators': 498, 'max_depth': 3, 'learning_rate':
0.10383099534212997, 'subsample': 0.9988403341689189, 'colsample_bytree':
```

0.6025143492400523, 'gamma': 0.42650360644670476, 'reg_alpha': 0.9583418751917582, 'reg_lambda': 0.6426744206028095}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:17,936] Trial 14 finished with value: 1.4216702019986014 and parameters: {'n_estimators': 370, 'max_depth': 3, 'learning_rate': 0.10176965376660425, 'subsample': 0.905484574884099, 'colsample_bytree': 0.7801167299539224, 'gamma': 0.5424107638169944, 'reg_alpha': 0.8465248824050465, 'reg_lambda': 0.6986983112979781}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:23,091] Trial 15 finished with value: 1.4192324966867604 and parameters: {'n_estimators': 500, 'max_depth': 4, 'learning_rate': 0.01590734589821368, 'subsample': 0.9063300043913174, 'colsample_bytree': 0.6220491594538633, 'gamma': 0.31792253049651387, 'reg_alpha': 0.8407381647277161, 'reg_lambda': 0.5002464435675367}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:30,393] Trial 16 finished with value: 1.443466216367986 and parameters: {'n_estimators': 497, 'max_depth': 7, 'learning_rate': 0.014499765324716205, 'subsample': 0.8956095278358345, 'colsample_bytree': 0.6323978732078409, 'gamma': 0.014464874920069537, 'reg_alpha': 0.8346280516429314, 'reg_lambda': 0.4448919120930204}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:32,027] Trial 17 finished with value: 1.4336017101873757 and parameters: {'n_estimators': 189, 'max_depth': 4, 'learning_rate': 0.14510557741349475, 'subsample': 0.9057246267988364, 'colsample_bytree': 0.7714498396330317, 'gamma': 0.24100082336978912, 'reg_alpha': 0.34783694390092135, 'reg_lambda': 0.4731238350445909}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:35,374] Trial 18 finished with value: 1.4288993245313393 and parameters: {'n_estimators': 289, 'max_depth': 4, 'learning_rate': 0.03194643048683521, 'subsample': 0.8268198679673342, 'colsample_bytree': 0.5235959938622222, 'gamma': 0.6848021097527559, 'reg_alpha': 0.5187604122503275, 'reg_lambda': 0.7230275192508071}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:36,683] Trial 19 finished with value: 1.553771574611303 and parameters: {'n_estimators': 54, 'max_depth': 7, 'learning_rate': 0.07303596688235464, 'subsample': 0.8389719167231222, 'colsample_bytree': 0.6246179943558585, 'gamma': 0.5006850550140122, 'reg_alpha': 0.8642477266781697, 'reg_lambda': 0.3851970603876117}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:41,694] Trial 20 finished with value: 1.4485663588276703 and parameters: {'n_estimators': 384, 'max_depth': 4, 'learning_rate': 0.13248764105773655, 'subsample': 0.9399793000708845, 'colsample_bytree': 0.7206463581056857, 'gamma': 0.2215344561953303, 'reg_alpha': 0.742599907460288, 'reg_lambda': 0.5392469372687742}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:57:45,934] Trial 21 finished with value: 1.4729441014442815 and parameters: {'n_estimators': 450, 'max_depth': 3, 'learning_rate': 0.01159271563655352, 'subsample': 0.958868258752677, 'colsample_bytree':

```
0.5640881958212871, 'gamma': 0.3930795739135511, 'reg_alpha':
0.9268653202347781, 'reg_lambda': 0.7847061679177136}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:57:50,188] Trial 22 finished with value: 1.419070469442879 and
parameters: {'n_estimators': 461, 'max_depth': 3, 'learning_rate':
0.06840707699632294, 'subsample': 0.9418130596195935, 'colsample_bytree':
0.5760873006875226, 'gamma': 0.2988247311673774, 'reg_alpha':
0.9967776546757845, 'reg_lambda': 0.5478500766636227}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:57:56,568] Trial 23 finished with value: 1.440367113234934 and
parameters: {'n_estimators': 479, 'max_depth': 4, 'learning_rate':
0.07964593882233142, 'subsample': 0.871127306699117, 'colsample_bytree':
0.6481608393230807, 'gamma': 0.2877767218322742, 'reg_alpha':
0.8893689684432273, 'reg_lambda': 0.54422594385492}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:58:02,347] Trial 24 finished with value: 1.4378730360162209 and
parameters: {'n_estimators': 449, 'max_depth': 5, 'learning_rate':
0.040720469754640116, 'subsample': 0.7838298355775264, 'colsample_bytree':
0.5397620301523369, 'gamma': 0.4541640761202492, 'reg_alpha':
0.6048562988109059, 'reg_lambda': 0.3912931377555774}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:58:06,361] Trial 25 finished with value: 1.4250295973350122 and
parameters: {'n_estimators': 401, 'max_depth': 3, 'learning_rate':
0.12129661410577156, 'subsample': 0.9366117178725117, 'colsample_bytree':
0.6106854988454808, 'gamma': 0.6001181152618881, 'reg_alpha': 0.775156702777966,
'reg_lambda': 0.2322084225055963}. Best is trial 11 with value:
1.4187925573150422.
[I 2025-03-10 18:58:11,463] Trial 26 finished with value: 1.4698470241748625 and
parameters: {'n_estimators': 466, 'max_depth': 4, 'learning_rate':
0.16350142910766746, 'subsample': 0.8637004711339569, 'colsample_bytree':
0.6842809159086162, 'gamma': 0.17013481665988692, 'reg_alpha':
0.38879244237490657, 'reg_lambda': 0.5616509084560726}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:58:15,754] Trial 27 finished with value: 1.4595014646074753 and
parameters: {'n_estimators': 371, 'max_depth': 6, 'learning_rate':
0.05396088988644869, 'subsample': 0.9349955628898187, 'colsample_bytree':
0.5004861625320051, 'gamma': 0.2887165848952214, 'reg_alpha': 0.788394073845166,
'reg_lambda': 0.7698864210479859}. Best is trial 11 with value:
1.4187925573150422.
[I 2025-03-10 18:58:20,061] Trial 28 finished with value: 1.4232419915848402 and
parameters: {'n_estimators': 499, 'max_depth': 3, 'learning_rate':
0.08719962293650486, 'subsample': 0.7898507955159138, 'colsample_bytree':
0.5679126807566953, 'gamma': 0.45111895469000135, 'reg_alpha':
0.9255244465554979, 'reg_lambda': 0.46230275724422776}. Best is trial 11 with
value: 1.4187925573150422.
[I 2025-03-10 18:58:26,832] Trial 29 finished with value: 1.4884943686547851 and
parameters: {'n_estimators': 410, 'max_depth': 8, 'learning_rate':
0.028503143760410153, 'subsample': 0.893891345166901, 'colsample_bytree':
```

0.6291296084694222, 'gamma': 0.8099735926784212, 'reg_alpha': 0.6749356616380762, 'reg_lambda': 0.6742470679672653}. Best is trial 11 with value: 1.4187925573150422.
[I 2025-03-10 18:58:28,443] Trial 30 finished with value: 1.416804169941111 and parameters: {'n_estimators': 132, 'max_depth': 5, 'learning_rate': 0.04715065676259471, 'subsample': 0.95266253661963, 'colsample_bytree': 0.997241492432885, 'gamma': 0.12748918840473364, 'reg_alpha': 0.9161363342566323, 'reg_lambda': 0.3925320272150198}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:29,657] Trial 31 finished with value: 1.4423083987428433 and parameters: {'n_estimators': 100, 'max_depth': 5, 'learning_rate': 0.04621765212972521, 'subsample': 0.9596529483135676, 'colsample_bytree': 0.9422988426284771, 'gamma': 0.11032850012001741, 'reg_alpha': 0.9144777166146003, 'reg_lambda': 0.3804910745737512}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:31,068] Trial 32 finished with value: 2.2532981718256115 and parameters: {'n_estimators': 136, 'max_depth': 4, 'learning_rate': 0.022049041879440748, 'subsample': 0.9284206153239001, 'colsample_bytree': 0.9960864059800265, 'gamma': 0.0008548970847709891, 'reg_alpha': 0.816636910257617, 'reg_lambda': 0.5164502781065384}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:33,155] Trial 33 finished with value: 1.4312236397713376 and parameters: {'n_estimators': 200, 'max_depth': 5, 'learning_rate': 0.06215160790999194, 'subsample': 0.8693894251591329, 'colsample_bytree': 0.8113693404584127, 'gamma': 0.17382087586759054, 'reg_alpha': 0.9956923495864705, 'reg_lambda': 0.21747131470273634}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:34,398] Trial 34 finished with value: 1.433421607637575 and parameters: {'n_estimators': 127, 'max_depth': 4, 'learning_rate': 0.04383470191770425, 'subsample': 0.9686195303448067, 'colsample_bytree': 0.6656470141679728, 'gamma': 0.11388405489170125, 'reg_alpha': 0.8973967124985491, 'reg_lambda': 0.5826987672352085}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:37,035] Trial 35 finished with value: 1.4250064893129006 and parameters: {'n_estimators': 305, 'max_depth': 3, 'learning_rate': 0.1161044078665202, 'subsample': 0.7498347954124545, 'colsample_bytree': 0.9903798114751909, 'gamma': 0.2597695338205181, 'reg_alpha': 0.7258207305057518, 'reg_lambda': 0.4351947688828698}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:40,170] Trial 36 finished with value: 1.5163629538001286 and parameters: {'n_estimators': 248, 'max_depth': 5, 'learning_rate': 0.29467056545132375, 'subsample': 0.9227882962364536, 'colsample_bytree': 0.7003332869648846, 'gamma': 0.33961706673528774, 'reg_alpha': 0.9480359339188547, 'reg_lambda': 0.2973752149252372}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:58:44,120] Trial 37 finished with value: 1.4373814160876262 and parameters: {'n_estimators': 431, 'max_depth': 4, 'learning_rate': 0.08422250924675403, 'subsample': 0.7227586309074359, 'colsample_bytree':

```
0.5383588646737221, 'gamma': 0.07731668496455113, 'reg_alpha':
0.8594144901255935, 'reg_lambda': 0.5041582849668311}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:58:45,897] Trial 38 finished with value: 1.5093921341241645 and
parameters: {'n_estimators': 154, 'max_depth': 6, 'learning_rate':
0.1963335008162963, 'subsample': 0.7921352122614517, 'colsample_bytree':
0.9162611267980398, 'gamma': 0.1839121658602647, 'reg_alpha':
0.6840968873661522, 'reg_lambda': 0.3535507215526565}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:58:49,020] Trial 39 finished with value: 1.4351850423663957 and
parameters: {'n_estimators': 317, 'max_depth': 5, 'learning_rate':
0.05666398370439843, 'subsample': 0.9982437975047077, 'colsample_bytree':
0.6567294376134001, 'gamma': 0.3086887857090708, 'reg_alpha':
0.5690396006707789, 'reg_lambda': 0.1628872687392029}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:58:52,782] Trial 40 finished with value: 1.4172176670419567 and
parameters: {'n_estimators': 411, 'max_depth': 4, 'learning_rate':
0.025395261219582424, 'subsample': 0.8450248920418847, 'colsample_bytree':
0.7481064597920731, 'gamma': 0.38898328596660137, 'reg_alpha':
0.390754412668192, 'reg_lambda': 0.741668065393176}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:58:58,280] Trial 41 finished with value: 1.419614709999457 and
parameters: {'n_estimators': 470, 'max_depth': 4, 'learning_rate':
0.033021784106193976, 'subsample': 0.844780993493965, 'colsample_bytree':
0.7500447154865268, 'gamma': 0.40199608481020255, 'reg_alpha':
0.3940613607278562, 'reg_lambda': 0.7667808217661011}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:01,737] Trial 42 finished with value: 1.4178563289653037 and
parameters: {'n_estimators': 397, 'max_depth': 3, 'learning_rate':
0.023969023751130678, 'subsample': 0.88352864482073, 'colsample_bytree':
0.8484283046289331, 'gamma': 0.5177734423488732, 'reg_alpha':
0.4466607473024824, 'reg_lambda': 0.9561326631949073}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:05,089] Trial 43 finished with value: 1.4176886536776956 and
parameters: {'n_estimators': 386, 'max_depth': 3, 'learning_rate':
0.048592308123988744, 'subsample': 0.8112940748719025, 'colsample_bytree':
0.8676764389223536, 'gamma': 0.5066009159669322, 'reg_alpha':
0.3017811181441911, 'reg_lambda': 0.9867334655235037}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:08,069] Trial 44 finished with value: 1.4182109518263721 and
parameters: {'n_estimators': 352, 'max_depth': 3, 'learning_rate':
0.04708237148527389, 'subsample': 0.819062125748836, 'colsample_bytree':
0.8600395414848212, 'gamma': 0.5947801522218288, 'reg_alpha':
0.23342658945201805, 'reg_lambda': 0.992693821747652}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:11,675] Trial 45 finished with value: 1.4176927793517908 and
parameters: {'n_estimators': 347, 'max_depth': 3, 'learning_rate':
0.0468729482089745, 'subsample': 0.8120742321011974, 'colsample_bytree':
```

```
0.8698374830363317, 'gamma': 0.6009855829205117, 'reg_alpha':
0.28134902160064323, 'reg_lambda': 0.9988940473505653}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:15,038] Trial 46 finished with value: 1.4199867822000947 and
parameters: {'n_estimators': 389, 'max_depth': 3, 'learning_rate':
0.026624413597411495, 'subsample': 0.7224780558634502, 'colsample_bytree':
0.8191951488131336, 'gamma': 0.7643690083907143, 'reg_alpha':
0.27716827039074365, 'reg_lambda': 0.9024985372779438}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:17,914] Trial 47 finished with value: 1.418194346916029 and
parameters: {'n_estimators': 330, 'max_depth': 3, 'learning_rate':
0.038000802256822536, 'subsample': 0.8039438645925338, 'colsample_bytree':
0.8884057286138829, 'gamma': 0.6025392287283692, 'reg_alpha':
0.006923884954332382, 'reg_lambda': 0.9860582161320336}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:21,943] Trial 48 finished with value: 1.5043625970394223 and
parameters: {'n_estimators': 358, 'max_depth': 5, 'learning_rate':
0.01149373145552695, 'subsample': 0.7665429442865586, 'colsample_bytree':
0.9648919876188401, 'gamma': 0.519205438239098, 'reg_alpha': 0.4416472431371844,
'reg_lambda': 0.863508133281546}. Best is trial 30 with value:
1.416804169941111.
[I 2025-03-10 18:59:25,662] Trial 49 finished with value: 1.4422994287349051 and
parameters: {'n_estimators': 417, 'max_depth': 4, 'learning_rate':
0.09338705939656429, 'subsample': 0.8563711200648736, 'colsample_bytree':
0.8789331804021814, 'gamma': 0.55604807145472, 'reg_alpha': 0.2914390095286323,
'reg_lambda': 0.9219921047116033}. Best is trial 30 with value:
1.416804169941111.
[I 2025-03-10 18:59:31,523] Trial 50 finished with value: 1.5881186776588898 and
parameters: {'n_estimators': 262, 'max_depth': 10, 'learning_rate':
0.05279444736592567, 'subsample': 0.6094169074441, 'colsample_bytree':
0.820656027168063, 'gamma': 0.6373558728600762, 'reg_alpha':
0.32262956277611315, 'reg_lambda': 0.950001584483975}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:34,254] Trial 51 finished with value: 1.4182959012980485 and
parameters: {'n_estimators': 319, 'max_depth': 3, 'learning_rate':
0.03626994706876035, 'subsample': 0.802555306692285, 'colsample_bytree':
0.9060375213916031, 'gamma': 0.7131478584219156, 'reg_alpha':
0.0408990081665191, 'reg_lambda': 0.9963731174965229}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:36,943] Trial 52 finished with value: 1.4173859697369067 and
parameters: {'n_estimators': 334, 'max_depth': 3, 'learning_rate':
0.07289629165971781, 'subsample': 0.8151738389300952, 'colsample_bytree':
0.9375139062298115, 'gamma': 0.47805309851975863, 'reg_alpha':
0.44642457483837317, 'reg_lambda': 0.8398815792594719}. Best is trial 30 with
value: 1.416804169941111.
[I 2025-03-10 18:59:39,283] Trial 53 finished with value: 1.4174926963004288 and
parameters: {'n_estimators': 285, 'max_depth': 3, 'learning_rate':
0.07693068281144112, 'subsample': 0.8856258026526737, 'colsample_bytree':
```

0.9369147951949381, 'gamma': 0.47199313955060873, 'reg_alpha': 0.4657072609575614, 'reg_lambda': 0.8159651779628643}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:59:44,415] Trial 54 finished with value: 1.5862656664997818 and parameters: {'n_estimators': 287, 'max_depth': 9, 'learning_rate': 0.07453416203918337, 'subsample': 0.8225382725790645, 'colsample_bytree': 0.9517316592713758, 'gamma': 0.4933133961759595, 'reg_alpha': 0.4568058864639784, 'reg_lambda': 0.8211739260542754}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:59:46,414] Trial 55 finished with value: 1.434836871417247 and parameters: {'n_estimators': 218, 'max_depth': 4, 'learning_rate': 0.10807133517449874, 'subsample': 0.7677670368818227, 'colsample_bytree': 0.9295734025962898, 'gamma': 0.46946152650624295, 'reg_alpha': 0.5471005963098338, 'reg_lambda': 0.8273087931549301}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:59:49,317] Trial 56 finished with value: 1.4209126977664033 and parameters: {'n_estimators': 350, 'max_depth': 3, 'learning_rate': 0.06530813415446235, 'subsample': 0.7267349255166169, 'colsample_bytree': 0.9711800235313093, 'gamma': 0.5709541268242643, 'reg_alpha': 0.17709739179931006, 'reg_lambda': 0.8713351513822779}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:59:51,742] Trial 57 finished with value: 1.4194252973188568 and parameters: {'n_estimators': 298, 'max_depth': 3, 'learning_rate': 0.07879090732800227, 'subsample': 0.8460912879164862, 'colsample_bytree': 0.8701761155533818, 'gamma': 0.42850101667626883, 'reg_alpha': 0.38856039783157004, 'reg_lambda': 0.8058059039704752}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:59:55,005] Trial 58 finished with value: 1.4417939696433766 and parameters: {'n_estimators': 367, 'max_depth': 4, 'learning_rate': 0.09447513053176784, 'subsample': 0.8097241323561036, 'colsample_bytree': 0.9041723496344628, 'gamma': 0.9996021357090799, 'reg_alpha': 0.23212677598788564, 'reg_lambda': 0.9217559450259905}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 18:59:57,817] Trial 59 finished with value: 1.4259111721052202 and parameters: {'n_estimators': 271, 'max_depth': 3, 'learning_rate': 0.13762994249942115, 'subsample': 0.7018090331826596, 'colsample_bytree': 0.9813547470585237, 'gamma': 0.6534003390833927, 'reg_alpha': 0.4908053490373881, 'reg_lambda': 0.7281628631572763}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 19:00:01,622] Trial 60 finished with value: 1.425177116158601 and parameters: {'n_estimators': 333, 'max_depth': 4, 'learning_rate': 0.05591843616071493, 'subsample': 0.8330386829054409, 'colsample_bytree': 0.7866544244640203, 'gamma': 0.8285014028857488, 'reg_alpha': 0.35894965632719655, 'reg_lambda': 0.8468209994195022}. Best is trial 30 with value: 1.416804169941111.
[I 2025-03-10 19:00:05,588] Trial 61 finished with value: 1.4212967402826138 and parameters: {'n_estimators': 383, 'max_depth': 3, 'learning_rate': 0.021134113004322146, 'subsample': 0.8858312244382827, 'colsample_bytree':

```
0.833583432550744, 'gamma': 0.5302536318178913, 'reg_alpha': 0.4323533679067597,
'reg_lambda': 0.9465486467529688}. Best is trial 30 with value:
1.416804169941111.
[I 2025-03-10 19:00:08,883] Trial 62 finished with value: 1.4162192828643438 and
parameters: {'n_estimators': 397, 'max_depth': 3, 'learning_rate':
0.025581815367903552, 'subsample': 0.8804860665540301, 'colsample_bytree':
0.9310084255140005, 'gamma': 0.4968520966952436, 'reg_alpha':
0.47552760674234534, 'reg_lambda': 0.9594835300850942}. Best is trial 62 with
value: 1.4162192828643438.
[I 2025-03-10 19:00:12,283] Trial 63 finished with value: 1.420563968157728 and
parameters: {'n_estimators': 424, 'max_depth': 3, 'learning_rate':
0.06936266657602716, 'subsample': 0.7741361023208285, 'colsample_bytree':
0.9330056179498555, 'gamma': 0.47044815506647697, 'reg_alpha':
0.4865448408694583, 'reg_lambda': 0.8990514865927794}. Best is trial 62 with
value: 1.4162192828643438.
[I 2025-03-10 19:00:16,380] Trial 64 finished with value: 1.4253881175292114 and
parameters: {'n_estimators': 372, 'max_depth': 4, 'learning_rate':
0.047233810412133675, 'subsample': 0.8536265380433549, 'colsample_bytree':
0.9593078624605426, 'gamma': 0.3784518076737964, 'reg_alpha':
0.30929044478125833, 'reg_lambda': 0.7437315048274553}. Best is trial 62 with
value: 1.4162192828643438.
[I 2025-03-10 19:00:20,006] Trial 65 finished with value: 1.415810356970245 and
parameters: {'n_estimators': 444, 'max_depth': 3, 'learning_rate':
0.03605478326060174, 'subsample': 0.9125571343562086, 'colsample_bytree':
0.9240203774344042, 'gamma': 0.4213669711544677, 'reg_alpha':
0.26315885046040866, 'reg_lambda': 0.9659679753194428}. Best is trial 65 with
value: 1.415810356970245.
[I 2025-03-10 19:00:23,908] Trial 66 finished with value: 1.4672699848238415 and
parameters: {'n_estimators': 442, 'max_depth': 3, 'learning_rate':
0.010191707140610452, 'subsample': 0.9152473626262846, 'colsample_bytree':
0.9276041426924442, 'gamma': 0.42418278930286585, 'reg_alpha':
0.11140312034054259, 'reg_lambda': 0.6741056898388058}. Best is trial 65 with
value: 1.415810356970245.
[I 2025-03-10 19:00:27,804] Trial 67 finished with value: 1.41984131430555 and
parameters: {'n_estimators': 413, 'max_depth': 4, 'learning_rate':
0.033012204557614035, 'subsample': 0.8754694551265438, 'colsample_bytree':
0.973474461974572, 'gamma': 0.48452715622952575, 'reg_alpha': 0.598390728072957,
'reg_lambda': 0.8871788575070199}. Best is trial 65 with value:
1.415810356970245.
[I 2025-03-10 19:00:33,646] Trial 68 finished with value: 1.4459075418081373 and
parameters: {'n_estimators': 401, 'max_depth': 7, 'learning_rate':
0.02003560772824201, 'subsample': 0.9833388254693874, 'colsample_bytree':
0.9417286833545017, 'gamma': 0.43667557347726943, 'reg_alpha':
0.41085395567805727, 'reg_lambda': 0.927704970270963}. Best is trial 65 with
value: 1.415810356970245.
[I 2025-03-10 19:00:37,358] Trial 69 finished with value: 1.4164521356469815 and
parameters: {'n_estimators': 433, 'max_depth': 3, 'learning_rate':
0.04123604649432521, 'subsample': 0.9033952854450643, 'colsample_bytree':
```

0.9122024716881633, 'gamma': 0.3527330455212929, 'reg_alpha': 0.24354228780768267, 'reg_lambda': 0.7986560876698298}. Best is trial 65 with value: 1.415810356970245.
[I 2025-03-10 19:00:41,549] Trial 70 finished with value: 1.4313197163484088 and parameters: {'n_estimators': 448, 'max_depth': 4, 'learning_rate': 0.06146444482554579, 'subsample': 0.9082404254516804, 'colsample_bytree': 0.9989362991455872, 'gamma': 0.40048145257217793, 'reg_alpha': 0.5260683640369445, 'reg_lambda': 0.8453853729303933}. Best is trial 65 with value: 1.415810356970245.
[I 2025-03-10 19:00:45,249] Trial 71 finished with value: 1.4152957952269598 and parameters: {'n_estimators': 432, 'max_depth': 3, 'learning_rate': 0.03821265131885728, 'subsample': 0.9510624507419833, 'colsample_bytree': 0.9068339501068202, 'gamma': 0.38683110367329177, 'reg_alpha': 0.34480120951271076, 'reg_lambda': 0.8074297227631062}. Best is trial 71 with value: 1.4152957952269598.
[I 2025-03-10 19:00:49,489] Trial 72 finished with value: 1.4157209605413674 and parameters: {'n_estimators': 431, 'max_depth': 3, 'learning_rate': 0.03458105999245172, 'subsample': 0.961062361217172, 'colsample_bytree': 0.9136663874853623, 'gamma': 0.3542409360965112, 'reg_alpha': 0.3500349285381573, 'reg_lambda': 0.7918807612237814}. Best is trial 71 with value: 1.4152957952269598.
[I 2025-03-10 19:00:53,203] Trial 73 finished with value: 1.4154100279772368 and parameters: {'n_estimators': 433, 'max_depth': 3, 'learning_rate': 0.040203686844967335, 'subsample': 0.9522507422674038, 'colsample_bytree': 0.9167752003702643, 'gamma': 0.34122595569903547, 'reg_alpha': 0.34105681690979495, 'reg_lambda': 0.7714401117837694}. Best is trial 71 with value: 1.4152957952269598.
[I 2025-03-10 19:00:57,612] Trial 74 finished with value: 1.4218930395544558 and parameters: {'n_estimators': 484, 'max_depth': 4, 'learning_rate': 0.037538803611264424, 'subsample': 0.9534975349768743, 'colsample_bytree': 0.9149515219534521, 'gamma': 0.3385213832624304, 'reg_alpha': 0.3460504716081887, 'reg_lambda': 0.7019313973866125}. Best is trial 71 with value: 1.4152957952269598.
[I 2025-03-10 19:01:01,557] Trial 75 finished with value: 1.416710161385152 and parameters: {'n_estimators': 432, 'max_depth': 3, 'learning_rate': 0.02726136157097507, 'subsample': 0.975053604826694, 'colsample_bytree': 0.896475751051997, 'gamma': 0.359176688079206, 'reg_alpha': 0.24687859036607124, 'reg_lambda': 0.7877274647497233}. Best is trial 71 with value: 1.4152957952269598.
[I 2025-03-10 19:01:05,412] Trial 76 finished with value: 1.4195315970159563 and parameters: {'n_estimators': 433, 'max_depth': 3, 'learning_rate': 0.01768732699871116, 'subsample': 0.9780300432344106, 'colsample_bytree': 0.8969390928591353, 'gamma': 0.2584074945279886, 'reg_alpha': 0.24339620007218637, 'reg_lambda': 0.7845218840488665}. Best is trial 71 with value: 1.4152957952269598.
[I 2025-03-10 19:01:08,804] Trial 77 finished with value: 1.4453620914331908 and parameters: {'n_estimators': 454, 'max_depth': 3, 'learning_rate': 0.24766057437175545, 'subsample': 0.965015635753232, 'colsample_bytree':

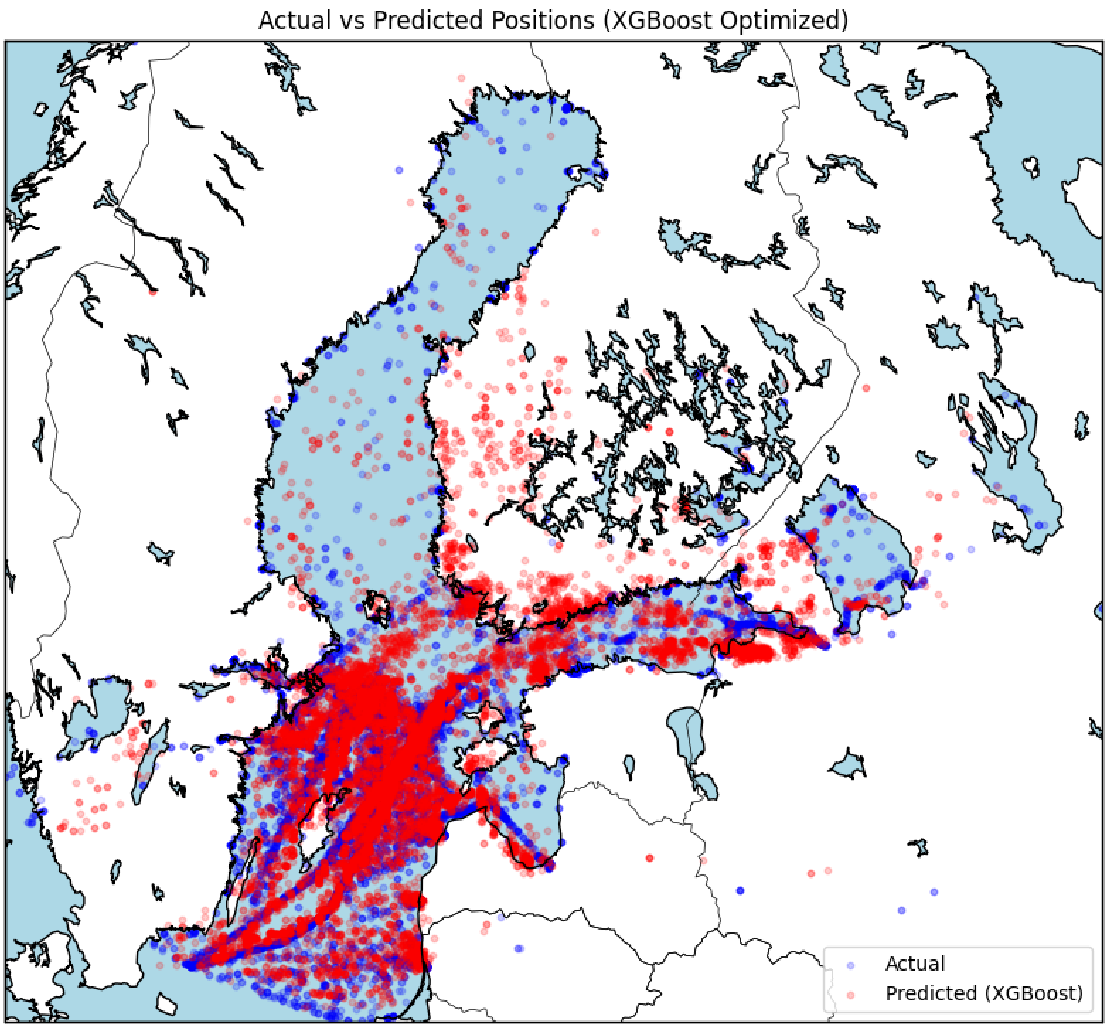
```
0.8883801395126456, 'gamma': 0.3619563839502307, 'reg_alpha':
0.16049036442827005, 'reg_lambda': 0.6321448657083372}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:12,687] Trial 78 finished with value: 1.4162256934191633 and
parameters: {'n_estimators': 484, 'max_depth': 3, 'learning_rate':
0.037986350561014286, 'subsample': 0.9481429663796013, 'colsample_bytree':
0.9165078531488381, 'gamma': 0.3212019963382283, 'reg_alpha':
0.25890604136781786, 'reg_lambda': 0.6603644737591596}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:16,513] Trial 79 finished with value: 1.4167518819434202 and
parameters: {'n_estimators': 475, 'max_depth': 3, 'learning_rate':
0.029392881573165493, 'subsample': 0.9267136238969285, 'colsample_bytree':
0.9106864812459531, 'gamma': 0.20505108169396513, 'reg_alpha':
0.25732549034098834, 'reg_lambda': 0.6784816149094335}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:21,157] Trial 80 finished with value: 1.4157058866656655 and
parameters: {'n_estimators': 482, 'max_depth': 3, 'learning_rate':
0.038655990812416156, 'subsample': 0.9870825843846468, 'colsample_bytree':
0.8830387569331319, 'gamma': 0.33305496611455526, 'reg_alpha':
0.19933482936763094, 'reg_lambda': 0.7927928480592973}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:26,544] Trial 81 finished with value: 1.41684765330338 and
parameters: {'n_estimators': 490, 'max_depth': 3, 'learning_rate':
0.03677030798726833, 'subsample': 0.9890364923192043, 'colsample_bytree':
0.9536290687724928, 'gamma': 0.3240308873138851, 'reg_alpha':
0.1781083844045914, 'reg_lambda': 0.793260528264404}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:31,776] Trial 82 finished with value: 1.4178561844822202 and
parameters: {'n_estimators': 460, 'max_depth': 3, 'learning_rate':
0.056752882501050234, 'subsample': 0.9472446572642901, 'colsample_bytree':
0.9198700948173488, 'gamma': 0.27155497486034846, 'reg_alpha':
0.19356694242067723, 'reg_lambda': 0.7032561390543438}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:36,347] Trial 83 finished with value: 1.4153438969316692 and
parameters: {'n_estimators': 437, 'max_depth': 3, 'learning_rate':
0.04124189704712837, 'subsample': 0.9683265275826703, 'colsample_bytree':
0.8858671679284708, 'gamma': 0.3571208446833711, 'reg_alpha':
0.22132695817766865, 'reg_lambda': 0.75940553378867}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:41,563] Trial 84 finished with value: 1.4166333454164752 and
parameters: {'n_estimators': 481, 'max_depth': 3, 'learning_rate':
0.03975918888438106, 'subsample': 0.9175881745853098, 'colsample_bytree':
0.921001237859811, 'gamma': 0.3070085802776488, 'reg_alpha': 0.207104233969354,
'reg_lambda': 0.7513197199438598}. Best is trial 71 with value:
1.4152957952269598.
[I 2025-03-10 19:01:47,383] Trial 85 finished with value: 1.4190090210650674 and
parameters: {'n_estimators': 443, 'max_depth': 3, 'learning_rate':
0.017537784785014707, 'subsample': 0.9400366183721526, 'colsample_bytree':
```

0.8873879655498133, 'gamma': 0.23716491265176565, 'reg_alpha':
0.14220020076166187, 'reg_lambda': 0.7631061228969692}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:53,641] Trial 86 finished with value: 1.4167956250207925 and
parameters: {'n_estimators': 425, 'max_depth': 3, 'learning_rate':
0.04074290392888927, 'subsample': 0.8980278631117504, 'colsample_bytree':
0.8565166721489432, 'gamma': 0.4108888907256645, 'reg_alpha':
0.34324961482267374, 'reg_lambda': 0.6659981497554266}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:01:59,637] Trial 87 finished with value: 1.4386820587111389 and
parameters: {'n_estimators': 462, 'max_depth': 3, 'learning_rate':
0.18619536483049573, 'subsample': 0.9611322027705567, 'colsample_bytree':
0.9102135049920498, 'gamma': 0.33049838148002175, 'reg_alpha':
0.20905646055676197, 'reg_lambda': 0.7202311759823751}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:04,767] Trial 88 finished with value: 1.4189616222224197 and
parameters: {'n_estimators': 471, 'max_depth': 3, 'learning_rate':
0.06245377884137869, 'subsample': 0.9987331654035465, 'colsample_bytree':
0.8374811440612213, 'gamma': 0.2814315660116591, 'reg_alpha':
0.3237091514483844, 'reg_lambda': 0.8802349275366039}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:09,535] Trial 89 finished with value: 1.4260228149003815 and
parameters: {'n_estimators': 440, 'max_depth': 4, 'learning_rate':
0.05268289994135855, 'subsample': 0.9321073709477531, 'colsample_bytree':
0.8789691967873278, 'gamma': 0.37220058396527517, 'reg_alpha':
0.07030648880744961, 'reg_lambda': 0.6531395625374756}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:16,287] Trial 90 finished with value: 1.5078277030615868 and
parameters: {'n_estimators': 455, 'max_depth': 8, 'learning_rate':
0.029421481380746652, 'subsample': 0.946146362227311, 'colsample_bytree':
0.946397572802991, 'gamma': 0.4457623222607391, 'reg_alpha':
0.37451118409038703, 'reg_lambda': 0.5892087811052946}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:21,005] Trial 91 finished with value: 1.4176918692395892 and
parameters: {'n_estimators': 484, 'max_depth': 3, 'learning_rate':
0.04259193033450521, 'subsample': 0.9246805708398058, 'colsample_bytree':
0.9239625795332934, 'gamma': 0.310903600708618, 'reg_alpha':
0.21179008152707132, 'reg_lambda': 0.7641558668870903}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:25,611] Trial 92 finished with value: 1.4465340871278949 and
parameters: {'n_estimators': 490, 'max_depth': 3, 'learning_rate':
0.21588124075289203, 'subsample': 0.9150848809927936, 'colsample_bytree':
0.8976500762213461, 'gamma': 0.3391033437854704, 'reg_alpha': 0.269455051702883,
'reg_lambda': 0.7337314291367992}. Best is trial 71 with value:
1.4152957952269598.
[I 2025-03-10 19:02:29,296] Trial 93 finished with value: 1.415555678040877 and
parameters: {'n_estimators': 477, 'max_depth': 3, 'learning_rate':
0.041269791568822524, 'subsample': 0.971372615231651, 'colsample_bytree':

```
0.8748039213976695, 'gamma': 0.30833689296888156, 'reg_alpha':
0.2199226440943713, 'reg_lambda': 0.8157924017703618}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:32,454] Trial 94 finished with value: 1.4162291738825228 and
parameters: {'n_estimators': 418, 'max_depth': 3, 'learning_rate':
0.03437028749981561, 'subsample': 0.9701802976258358, 'colsample_bytree':
0.8815869317959902, 'gamma': 0.3866802656733357, 'reg_alpha':
0.12622133149432688, 'reg_lambda': 0.8046307312347846}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:36,234] Trial 95 finished with value: 1.415938054110251 and
parameters: {'n_estimators': 500, 'max_depth': 3, 'learning_rate':
0.032539100288144565, 'subsample': 0.9710709448922903, 'colsample_bytree':
0.8727177127798302, 'gamma': 0.3895347205283539, 'reg_alpha':
0.08740027671806716, 'reg_lambda': 0.8609315814619523}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:40,248] Trial 96 finished with value: 1.4337283306482826 and
parameters: {'n_estimators': 499, 'max_depth': 3, 'learning_rate':
0.16238783959416414, 'subsample': 0.9876396939038714, 'colsample_bytree':
0.8637175770611647, 'gamma': 0.2824688296866523, 'reg_alpha':
0.07780787115284851, 'reg_lambda': 0.9631492344334575}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:43,819] Trial 97 finished with value: 1.4165767948105479 and
parameters: {'n_estimators': 472, 'max_depth': 3, 'learning_rate':
0.023438231200814084, 'subsample': 0.9584848771999018, 'colsample_bytree':
0.8449657932186474, 'gamma': 0.4114356082578148, 'reg_alpha':
0.3205594754646863, 'reg_lambda': 0.8653289841632166}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:47,361] Trial 98 finished with value: 1.4175026818176115 and
parameters: {'n_estimators': 461, 'max_depth': 3, 'learning_rate':
0.04990322039172632, 'subsample': 0.978705952430842, 'colsample_bytree':
0.8000807233714402, 'gamma': 0.26109235691204297, 'reg_alpha':
0.29225830642356326, 'reg_lambda': 0.8309672149613861}. Best is trial 71 with
value: 1.4152957952269598.
[I 2025-03-10 19:02:51,089] Trial 99 finished with value: 1.4158817975664406 and
parameters: {'n_estimators': 449, 'max_depth': 4, 'learning_rate':
0.014310197057801247, 'subsample': 0.9663001852045823, 'colsample_bytree':
0.8750538451726101, 'gamma': 0.20540729492603269, 'reg_alpha':
0.05549225029693461, 'reg_lambda': 0.90364885240155}. Best is trial 71 with
value: 1.4152957952269598.
```

```
{'n_estimators': 432, 'max_depth': 3, 'learning_rate': 0.03821265131885728,
'subsample': 0.9510624507419833, 'colsample_bytree': 0.9068339501068202,
'gamma': 0.38683110367329177, 'reg_alpha': 0.34480120951271076, 'reg_lambda':
0.8074297227631062, 'n_jobs': -1}
                   Model  Overall MSE  Lat MSE  Lon MSE     MAE     R^2
0      Linear Regression       1.4630   0.4469   2.4791  0.4539  0.7449
1  Polynomial Regression       1.4488   0.4310   2.4666  0.4472  0.7497
2          Random Forest       1.8168   0.5227   3.1109  0.5486  0.6898
```

3	XGBoost	1.4958	0.4444	2.5473	0.4505	0.7417
4	XGBoost (Optimized)	1.4153	0.4207	2.4098	0.4381	0.7556



[42]:
```python
# Save the model
import joblib
joblib.dump(xgb_model, 'vessel_prediction_model.pkl')
```

[42]: ['vessel_prediction_model.pkl']