In [3]:
```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_absolute_error, mean_squared_error
import matplotlib.pyplot as plt
import joblib
import numpy as np
```

In [6]:
```
df = pd.read_csv("/Users/hassankhalid/Downloads/appliance_energy_da
```

In [7]:
```
X = df.drop(columns=["House_ID", "Monthly_kWh"])
y = df["Monthly_kWh"]
```

In [8]:
```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size
```

In [9]:
```
rf = RandomForestRegressor(
    n_estimators=200,
    max_depth=None,
    min_samples_split=2,
    min_samples_leaf=1,
    random_state=42      #
)
```

In [11]:
```
rf.fit(X_train, y_train)
```

Out[11]:

▾ RandomForestRegressor

▸ Parameters

In [12]:
```
y_pred = rf.predict(X_test)
```

In [13]:
```
mae = mean_absolute_error(y_test, y_pred)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
r2 = r2_score(y_test, y_pred)
```

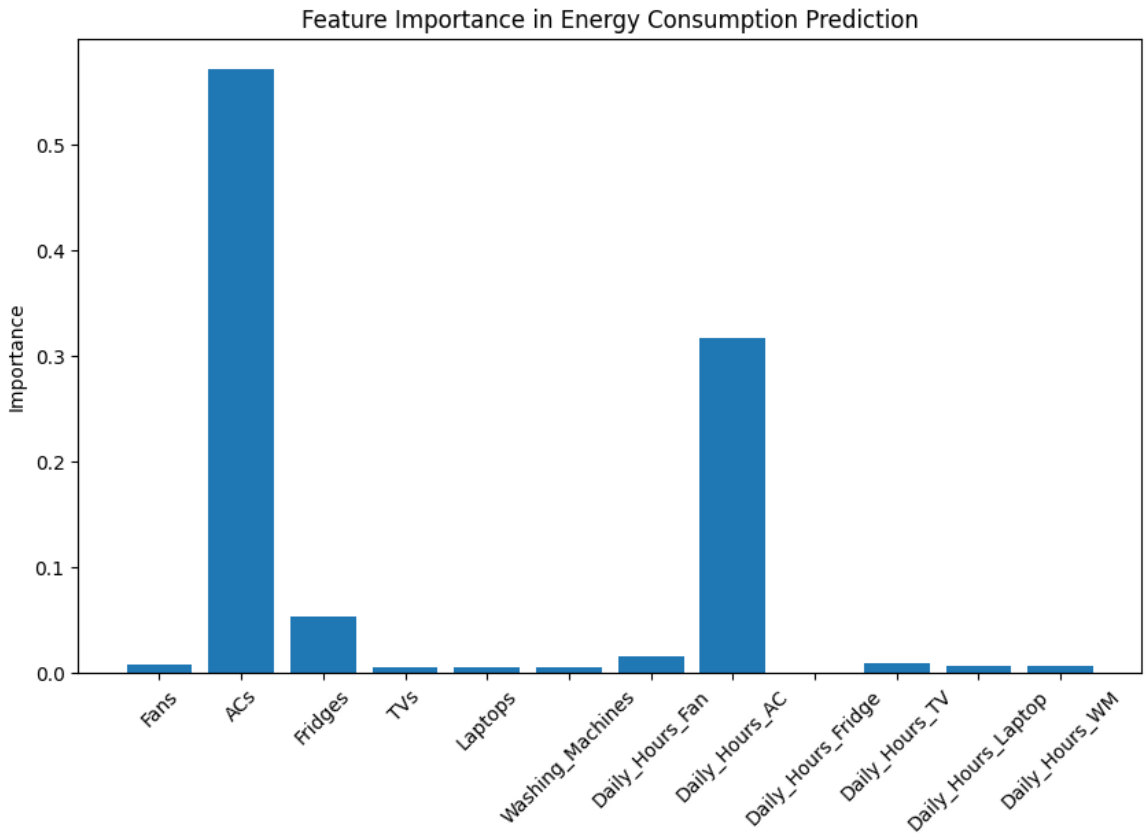
In [14]:
```
print(f"Mean Absolute Error (MAE): {mae:.2f} kWh")
print(f"Root Mean Squared Error (RMSE): {rmse:.2f} kWh")
print(f"R² Score: {r2:.3f}")
```

```
Mean Absolute Error (MAE): 89.83 kWh
Root Mean Squared Error (RMSE): 119.36 kWh
R² Score: 0.865
```

In [16]:
```
importances = rf.feature_importances_
features = X.columns
```

In [17]:
```
plt.figure(figsize=(10, 6))
plt.bar(range(len(importances)), importances)
plt.xticks(range(len(importances)), features, rotation=45)
plt.ylabel("Importance")
plt.title("Feature Importance in Energy Consumption Prediction")
```

```
plt.show()
```



In [19]:
```
joblib.dump(rf, "random_forest_model.pkl")
```

Out[19]:
```
['random_forest_model.pkl']
```

In [ ]:
```

```