Assignment – 2 Report

Random Forest Regression

1.1. Using One-Hot Encoder:

- Mean Squared Error (MSE): 7.3914
- **R-squared:** 0.9998

```
Pipeline
Pipeline(steps=[('num_preprocess',
                 Pipeline(steps=[('imputer', SimpleImputer()),
                                 ('scaler', MinMaxScaler())])),
                ('cat preprocess',
                 Pipeline(steps=[('imputer',
                                  SimpleImputer(strategy='most frequent')),
                                 ('onehot',
                                  OneHotEncoder(drop='first',
                                                 sparse output=False))])),
                ('model', RandomForestRegressor(random_state=42))])
                           num_preprocess: Pipeline
                                 SimpleImputer
                                 ▶ MinMaxScaler
                            cat preprocess: Pipeline
Pipeline(steps=[('imputer', SimpleImputer(strategy='most_frequent')),
                ('onehot', OneHotEncoder(drop='first', sparse output=False))])
                                ▶ SimpleImputer
                                  OneHotEncoder
               OneHotEncoder(drop='first', sparse output=False)
                              RandomForestRegressor
                    RandomForestRegressor(random state=42)
```

1.2. Using Target Encoder:

• Mean Squared Error (MSE): 7.5275

```
Pipeline
Pipeline(steps=[('num_preprocess',
                 Pipeline(steps=[('imputer', SimpleImputer()),
                                 ('scaler', MinMaxScaler())])),
                ('cat preprocess',
                 Pipeline(steps=[('imputer',
                                  SimpleImputer(strategy='most_frequent')),
                                 ('targetencoder',
                                  TargetEncoder(cols=[0, 1, 2]))])),
                ('model', RandomForestRegressor(random_state=42))])
                         num_preprocess: Pipeline
                               ▶ SimpleImputer
                               ▶ MinMaxScaler
                         cat preprocess: Pipeline
                               SimpleImputer
                               ▶ TargetEncoder
                            RandomForestRegressor
                  RandomForestRegressor(random state=42)
```

2. Linear Regression (Using Package)

2.1. Using Target Encoder:

• Mean Squared Error (MSE): 15563.8545

```
Pipeline
Pipeline(steps=[('num_preprocess',
                 Pipeline(steps=[('imputer', SimpleImputer()),
                                  ('scaler', MinMaxScaler())])),
                ('cat preprocess',
                 Pipeline(steps=[('imputer',
                                  SimpleImputer(strategy='most_frequent')),
                                  ('target',
                                  TargetEncoder(target_type='continuous'))])),
                ('model', LinearRegression())])
                           num_preprocess: Pipeline
                                ▶ SimpleImputer
                                 ▶ MinMaxScaler
                            cat_preprocess: Pipeline
                                SimpleImputer
                                  TargetEncoder
                    TargetEncoder(target_type='continuous')

▼ LinearRegression

                              LinearRegression()
```

2.2. Using One-Hot Encoder:

• Mean Squared Error (MSE): 15563.8545

```
Pipeline
Pipeline(steps=[('num_preprocess',
                 Pipeline(steps=[('imputer', SimpleImputer()),
                                 ('scaler', MinMaxScaler())])),
                ('cat_preprocess',
                 Pipeline(steps=[('imputer',
                                  SimpleImputer(strategy='most_frequent')),
                                 ('onehot',
                                  OneHotEncoder(sparse_output=False))])),
                ('model', LinearRegression())])
                         num_preprocess: Pipeline
                              SimpleImputer
                               ▶ MinMaxScaler
                         reprocess: Pipeline
                              SimpleImputer
                              ▶ OneHotEncoder

→ LinearRegression

                            LinearRegression()
```

3. Linear Regression (From Scratch)

3.1. Using One-Hot Encoder:

• Mean Squared Error (MSE): 15563.8545

```
Pipeline
Pipeline(steps=[('num preprocess',
                 Pipeline(steps=[('imputer', SimpleImputer()),
                                 ('scaler', MinMaxScaler())])),
                ('cat preprocess'
                 Pipeline(steps=[('imputer',
                                  SimpleImputer(strategy='most frequent')),
                                 ('onehot',
                                  OneHotEncoder(sparse output=False))])),
                ('model',
                 < main .Linear Regression object at 0x7a6957a11d20>)])
                         num_preprocess: Pipeline
                              SimpleImputer
                               ▶ MinMaxScaler
                          cat_preprocess: Pipeline
                              SimpleImputer
                                OneHotEncoder
                    OneHotEncoder(sparse output=False)
                            ► Linear Regression
```

3.2. Using Target Encoder:

Mean Squared Error (MSE): 16359.2324

• **R-squared:** 0.4834

```
Pipeline
Pipeline(steps=[('num_preprocess',
                 Pipeline(steps=[('imputer', SimpleImputer()),
                                 ('scaler', MinMaxScaler())])),
                ('cat preprocess',
                 Pipeline(steps=[('imputer',
                                  SimpleImputer(strategy='most frequent')),
                                 ('target',
                                  TargetEncoder(target type='continuous'))])),
                ('model',
                   main .Linear Regression object at 0x7a695dbe9e70>)])
                          num_preprocess: Pipeline
                                SimpleImputer
                                MinMaxScaler
                            cat_preprocess: Pipeline
                                SimpleImputer
                                 TargetEncoder
                   TargetEncoder(target type='continuous')
                              ▶ Linear Regression
```

For the Linear Regression, two numerical features have been added in-order to improve the performance of the model.

1. hum x temp

2. hum x windspeed

Humidity, temperature and windspeed are essential features in-order to predict the target values.

From the analysis, Random Forest Regressor has complete superiority over Linear Regressor.

For the given data, Random Forest is well suited for this.

On the other hand, Linear Regressor suffers due to complex data.