

```
1  ## Author: Wong Yee En
2  from pyspark.ml.feature import VectorAssembler
3  from pyspark.ml.classification import LogisticRegression
4  from pyspark.ml.evaluation import MulticlassClassificationEvaluator
5  from pyspark.sql import DataFrame
6
7  class ModelTrainer:
8
9      @staticmethod
10     def prepare_features(df: DataFrame, feature_cols: list) -> DataFrame:
11         """Combine multiple feature columns into a single feature vector."""
12         assembler = VectorAssembler(inputCols=feature_cols, outputCol="combinedFeatures")
13         return assembler.transform(df)
14
15     @staticmethod
16     def train_model(df: DataFrame, label_col: str):
17         """Train a Logistic Regression model."""
18         lr = LogisticRegression(featuresCol="combinedFeatures", labelCol=label_col)
19         model = lr.fit(df)
20         return model
```

```
21
22 @staticmethod
23 def evaluate_model(model, df: DataFrame, label_col: str):
24     """Evaluate the model using accuracy, precision, recall, and F1 score."""
25     predictions = model.transform(df)
26
27     evaluator = MulticlassClassificationEvaluator(
28         labelCol=label_col,
29         predictionCol="prediction",
30         metricName="accuracy"
31     )
32
33     accuracy = evaluator.evaluate(predictions)
34     evaluator.setMetricName("weightedPrecision")
35     precision = evaluator.evaluate(predictions)
36     evaluator.setMetricName("weightedRecall")
37     recall = evaluator.evaluate(predictions)
38     evaluator.setMetricName("f1")
39     f1_score = evaluator.evaluate(predictions)
40
41     print(f"Accuracy: {accuracy}")
42     print(f"Precision: {precision}")
43     print(f"Recall: {recall}")
44     print(f"F1 Score: {f1_score}")
45
46     return predictions, accuracy, precision, recall, f1_score
47
```