```
from pyspark.sql import SparkSession
from pyspark.ml import Pipeline
from pyspark.sql.functions import mean,col,split, col, regexp_extract, when, lit
from pyspark.ml.feature import StringIndexer
from pyspark.ml.feature import VectorAssembler
from pyspark.ml.evaluation import MulticlassClassificationEvaluator
from pyspark.ml.feature import QuantileDiscretizer
```

```
spark = SparkSession \
    .builder \
    .appName("Spark ML example on Social Media") \
    .getOrCreate()
```

```
22/12/22 19:52:39 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
```

```
df = spark.read.csv("Social Media.csv",header = 'True',inferSchema='True')
```

```
df.printSchema()
```

```
root
 |-- Timestamp: string (nullable = true)
 |-- Gender: string (nullable = true)
 |-- Age_Range: string (nullable = true)
 |-- Social_Media_Platforms_use: string (nullable = true)
 |-- Time_spend_in_Social_media_during_the_Lockdown: string (nullable = true)
 |-- Mental_health_by_Social_media: string (nullable = true)
 |-- Social_Media_Platform_most_helpful_while_talking_about_Mental_health: string (nullable = true)
 |-- Social_Media_Platform_least_helpful_while_talking_about_Mental_health: string (nullable = true)
 |-- faced_Mental_health_Issue: string (nullable = true)
 |-- Talk_about_your_mental_health: string (nullable = true)
 |-- Have_you_received_support_from_friends_because_of_something_you_posted_Social_Media: string (nullable = true)
 |-- With_whom_do_you_talk_on_social_media_regarding_your_mental_health: string (nullable = true)
 |-- Positive_or_negative_effect_on_mental_health: string (nullable = true)
 |-- Social_media_status_that_your_friend_has_posted_relating_to_his_her_mental_health: string (nullable = true)
 |-- If_yes_then_did_you_lend_any_support_to_such_person14: string (nullable = true)
 |-- Suicidal_status_or_any_such_status_that_any_of_your_friend's_have_posted: string (nullable = true)
 |-- If_yes_then_did_you_lend_any_support_to_such_person16: string (nullable = true)
```

```
df.show()
```

```
22/12/22 19:55:09 WARN CSVHeaderChecker: CSV header does not conform to the schema.
 Header: Timestamp, Gender, Age_Range, Social_Media_Platforms_use, Time_spend_in_Social_media_during_the_Lockdown, Mental_health_by
 Schema: Timestamp, Gender, Age_Range, Social_Media_Platforms_use, Time_spend_in_Social_media_during_the_Lockdown, Mental_health_by
Expected: If_yes_then_did_you_lend_any_support_to_such_person14 but found: If_yes_then_did_you_lend_any_support_to_such_person
CSV file: file:///home/taniya/Untitled%20Folder/Social%20Media.csv
+----------+------+---------+--------------------------+----------------------------------------------+----------------------------
| Timestamp|Gender|Age_Range|Social_Media_Platforms_use|Time_spend_in_Social_media_during_the_Lockdown|Mental_health_by_Social_medi
+----------+------+---------+--------------------------+----------------------------------------------+----------------------------
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     2-3 hours|                           N
|04-07-2021|  Male|    18-21|      Whatsapp, Snapcha...|                                     3-6 hours|                          Ye
|04-07-2021|Female|    22-30|      Facebook, Whatsap...|                                     2-3 hours|                           N
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     2-3 hours|                           N
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     3-6 hours|                           N
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     3-6 hours|                           N
|04-07-2021|Female|    22-30|      Facebook, Whatsap...|                                     3-6 hours|                           N
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     2-3 hours|                           N
|04-07-2021|  Male|    18-21|      Whatsapp, Instagr...|                                     3-6 hours|                          Ye
|04-07-2021|Female|    18-21|      Facebook, Whatsap...|                                     3-6 hours|                           N
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     6-8 hours|                           N
|04-07-2021|Female|    22-30|      Facebook, Whatsap...|                             Less than 2 hours|                          Ye
|04-07-2021|Female|    41-60|      Facebook, Whatsap...|                             Less than 2 hours|                           N
|04-07-2021|  Male|    18-21|      Facebook, Whatsap...|                                     3-6 hours|                           N
|04-07-2021|Female|    41-60|      Facebook, Whatsap...|                             Less than 2 hours|                           N
|04-07-2021|Female|    41-60|      Facebook, Whatsap...|                                     3-6 hours|                           N
|04-07-2021|Female|    18-21|      Facebook, Whatsap...|                             Less than 2 hours|                           N
|04-07-2021|Female|    22-30|      Facebook, Whatsap...|                                     6-8 hours|                           N
|04-07-2021|Female|    22-30|      Facebook, Whatsap...|                             Less than 2 hours|                          Ye
|04-07-2021|  Male|    22-30|      Facebook, Whatsap...|                                     3-6 hours|                           N
+----------+------+---------+--------------------------+----------------------------------------------+----------------------------
only showing top 20 rows
```

```
df.select("Age_Range","Gender","Positive_or_negative_effect_on_mental_health").show()
```

```
+---------+------+--------------------------------------------+
|Age_Range|Gender|Positive_or_negative_effect_on_mental_health|
+---------+------+--------------------------------------------+
|    18-21|  Male|                                    Negative|
|    18-21|  Male|                                     Negative|
|    22-30|Female|                                     Negative|
|    18-21|  Male|                                     Negative|
|    18-21|  Male|                                     Positive|
|    18-21|  Male|                                     Negative|
|    22-30|Female|                                     Negative|
|    18-21|  Male|                                     Negative|
|    18-21|  Male|                                     Negative|
|    18-21|Female|                                     Positive|
|    18-21|  Male|                                     Negative|
|    22-30|Female|                                     Positive|
|    41-60|Female|                                     Positive|
|    18-21|  Male|                                     Negative|
|    41-60|Female|                                     Positive|
|    41-60|Female|                                     Positive|
|    18-21|Female|                                     Negative|
|    22-30|Female|                                     Negative|
|    22-30|Female|                                     Negative|
|    22-30|  Male|                                     Negative|
+---------+------+---------------------------------------------+
only showing top 20 rows
```

```
df=df.drop("Timestamp","Mental_health_by_Social_media","Social_Media_Platform_most_helpful_while_talking_about_Mental_health")
```

```
df=df.drop("Social_Media_Platform_least_helpful_while_talking_about_Mental_health","faced_Mental_health_Issue","Talk_about_your_mental_he
```

```
df=df.drop("With_whom_do_you_talk_on_social_media_regarding_your_mental_health")
```

```
df=df.drop("Social_media_status_that_your_friend_has_posted_relating_to_his_her_mental_health","If_yes_then_did_you_lend_any_support_to_s
```

```
df=df.drop("Suicidal_status_or_any_such_status_that_any_of_your_friend's_have_posted")
```

```
df=df.drop("If_yes_then_did_you_lend_any_support_to_such_person")
```

```
df.show()
```

```
22/12/22 20:16:32 WARN CSVHeaderChecker: CSV header does not conform to the schema.
 Header: Gender, Age_Range, Social_Media_Platforms_use, Time_spend_in_Social_media_during_the_Lockdown, Have_you_received_support_f
 Schema: Gender, Age_Range, Social_Media_Platforms_use, Time_spend_in_Social_media_during_the_Lockdown, Have_you_received_support_f
Expected: If_yes_then_did_you_lend_any_support_to_such_person14 but found: If_yes_then_did_you_lend_any_support_to_such_person
CSV file: file:///home/taniya/Untitled%20Folder/Social%20Media.csv
+------+---------+--------------------------+----------------------------------------------+---------------------------------------
|Gender|Age_Range|Social_Media_Platforms_use|Time_spend_in_Social_media_during_the_Lockdown|Have_you_received_support_from_friends_
+------+---------+--------------------------+----------------------------------------------+---------------------------------------
|  Male|    18-21|      Facebook, Whatsap...|                                     2-3 hours|
|  Male|    18-21|      Whatsapp, Snapcha...|                                     3-6 hours|
|Female|    22-30|      Facebook, Whatsap...|                                     2-3 hours|
|  Male|    18-21|      Facebook, Whatsap...|                                     2-3 hours|
|  Male|    18-21|      Facebook, Whatsap...|                                     3-6 hours|
|  Male|    18-21|      Facebook, Whatsap...|                                     3-6 hours|
|Female|    22-30|      Facebook, Whatsap...|                                     3-6 hours|
|  Male|    18-21|      Facebook, Whatsap...|                                     2-3 hours|
|  Male|    18-21|      Whatsapp, Instagr...|                                     3-6 hours|
|Female|    18-21|      Facebook, Whatsap...|                                     3-6 hours|
|  Male|    18-21|      Facebook, Whatsap...|                                     6-8 hours|
|Female|    22-30|      Facebook, Whatsap...|                             Less than 2 hours|
|Female|    41-60|      Facebook, Whatsap...|                             Less than 2 hours|
|  Male|    18-21|      Facebook, Whatsap...|                                     3-6 hours|
|Female|    41-60|      Facebook, Whatsap...|                             Less than 2 hours|
|Female|    41-60|      Facebook, Whatsap...|                                     3-6 hours|
|Female|    18-21|      Facebook, Whatsap...|                             Less than 2 hours|
|Female|    22-30|      Facebook, Whatsap...|                                     6-8 hours|
|Female|    22-30|      Facebook, Whatsap...|                             Less than 2 hours|
|  Male|    22-30|      Facebook, Whatsap...|                                     3-6 hours|
+------+---------+--------------------------+----------------------------------------------+---------------------------------------
only showing top 20 rows
```

```
df=df.drop("Social_Media_Platform_use","Time_spend_in_Social_media_during_the_lockdown")
```

```
df.show()
```

```
22/12/22 20:18:57 WARN CSVHeaderChecker: CSV header does not conform to the schema.
 Header: Gender, Age_Range, Social_Media_Platforms_use, Have_you_received_support_from_friends_because_of_something_you_posted_Soci
 Schema: Gender, Age_Range, Social_Media_Platforms_use, Have_you_received_support_from_friends_because_of_something_you_posted_Soci
Expected: If_yes_then_did_you_lend_any_support_to_such_person14 but found: If_yes_then_did_you_lend_any_support_to_such_person
CSV file: file:///home/taniya/Untitled%20Folder/Social%20Media.csv
+------+---------+--------------------------+-----------------------------------------------------------------------------------+--
|Gender|Age_Range|Social_Media_Platforms_use|Have_you_received_support_from_friends_because_of_something_you_posted_Social_Media|Po
```

```
+------+---------+-------------------------+------------------------------------------------------------------------------------+--
|  Male|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Whatsapp, Snapcha...|                                                                I haven't posted ...|
|Female|    22-30|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Facebook, Whatsap...|                                                                                 Yes|
|Female|    22-30|     Facebook, Whatsap...|                                                                                 Yes|
|  Male|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Whatsapp, Instagr...|                                                                                 Yes|
|Female|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|Female|    22-30|     Facebook, Whatsap...|                                                                                 Yes|
|Female|    41-60|     Facebook, Whatsap...|                                                                I haven't posted ...|
|  Male|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|Female|    41-60|     Facebook, Whatsap...|                                                                I haven't posted ...|
|Female|    41-60|     Facebook, Whatsap...|                                                                I haven't posted ...|
|Female|    18-21|     Facebook, Whatsap...|                                                                I haven't posted ...|
|Female|    22-30|     Facebook, Whatsap...|                                                                                 Yes|
|Female|    22-30|     Facebook, Whatsap...|                                                                                 Yes|
|  Male|    22-30|     Facebook, Whatsap...|                                                                I haven't posted ...|
+------+---------+-------------------------+------------------------------------------------------------------------------------+--
only showing top 20 rows
```

```
df=df.drop("Social_Media_Platforms_use","Have_you_received_support_from_friends_because_of_something_you_posted_Social_Media","If_yes_the
```

```
df.show()
```

```
+------+---------+--------------------------------------------+
|Gender|Age_Range|Positive_or_negative_effect_on_mental_health|
+------+---------+--------------------------------------------+
|  Male|    18-21|                                    Negative|
|  Male|    18-21|                                    Negative|
|Female|    22-30|                                    Negative|
|  Male|    18-21|                                    Negative|
|  Male|    18-21|                                    Positive|
|  Male|    18-21|                                    Negative|
|Female|    22-30|                                    Negative|
|  Male|    18-21|                                    Negative|
|  Male|    18-21|                                    Negative|
|Female|    18-21|                                    Positive|
|  Male|    18-21|                                    Negative|
|Female|    22-30|                                    Positive|
|Female|    41-60|                                    Positive|
|  Male|    18-21|                                    Negative|
|Female|    41-60|                                    Positive|
|Female|    41-60|                                    Positive|
|Female|    18-21|                                    Negative|
|Female|    22-30|                                    Negative|
|Female|    22-30|                                    Negative|
|  Male|    22-30|                                    Negative|
+------+---------+--------------------------------------------+
only showing top 20 rows
```

```
count=df.count()
print(count)
```

```
136
```

```
df.groupBy("Positive_or_negative_effect_on_mental_health").count().show()
```

```
+--------------------------------------------+-----+
|Positive_or_negative_effect_on_mental_health|count|
+--------------------------------------------+-----+
|                                    Positive|   54|
|                                    Negative|   82|
+--------------------------------------------+-----+
```

```
df.groupBy("Age_Range","Positive_or_negative_effect_on_mental_health").count().show()
```

```
+---------------+--------------------------------------------+-----+
|      Age_Range|Positive_or_negative_effect_on_mental_health|count|
+---------------+--------------------------------------------+-----+
|22-30 years old|                                    Negative|    3|
|          18-21|                                    Positive|   25|
|          41-60|                                    Negative|   26|
|          18-21|                                    Negative|   36|
|          22-30|                                    Negative|   15|
|31-40 years old|                                    Positive|    1|
|          31-40|                                    Negative|    2|
|          41-60|                                    Positive|   18|
|          31-40|                                    Positive|    2|
|          22-30|                                    Positive|    8|
+---------------+--------------------------------------------+-----+
```

```
df = df.replace(['22-30 years old','31-40 years old'],['22-30','31-40'])
```

```
df.select("Age_Range").show()
```

```
+---------+
|Age_Range|
+---------+
|    18-21|
|    18-21|
|    22-30|
|    18-21|
|    18-21|
|    18-21|
|    22-30|
|    18-21|
|    18-21|
|    18-21|
|    18-21|
|    22-30|
|    41-60|
|    18-21|
|    41-60|
|    41-60|
|    18-21|
|    22-30|
|    22-30|
|    22-30|
+---------+
only showing top 20 rows
```

```
df.groupBy("Age_Range").count().show()
```

```
+---------+-----+
|Age_Range|count|
+---------+-----+
|    18-21|   61|
|    22-30|   26|
|    31-40|    5|
|    41-60|   44|
+---------+-----+
```

```
df.groupBy("Age_Range","Positive_or_negative_effect_on_mental_health").count().show()
```

```
+---------+--------------------------------------------+-----+
|Age_Range|Positive_or_negative_effect_on_mental_health|count|
+---------+--------------------------------------------+-----+
|    18-21|                                    Positive|   25|
|    41-60|                                    Negative|   26|
|    18-21|                                    Negative|   36|
|    22-30|                                    Negative|   18|
|    31-40|                                    Negative|    2|
|    41-60|                                    Positive|   18|
|    31-40|                                    Positive|    3|
|    22-30|                                    Positive|    8|
+---------+--------------------------------------------+-----+
```

```
df.groupBy("Gender","Positive_or_negative_effect_on_mental_health").count().show()
```

```
+------+--------------------------------------------+-----+
|Gender|Positive_or_negative_effect_on_mental_health|count|
+------+--------------------------------------------+-----+
|  Male|                                    Positive|   29|
|Female|                                    Positive|   25|
|Female|                                    Negative|   36|
|  Male|                                    Negative|   46|
+------+--------------------------------------------+-----+
```

```
df.show()
```

```
+------+---------+--------------------------------------------+
|Gender|Age_Range|Positive_or_negative_effect_on_mental_health|
+------+---------+--------------------------------------------+
|  Male|    18-21|                                    Negative|
|  Male|    18-21|                                    Negative|
|Female|    22-30|                                    Negative|
|  Male|    18-21|                                    Negative|
|  Male|    18-21|                                    Positive|
|  Male|    18-21|                                    Negative|
|Female|    22-30|                                    Negative|
|  Male|    18-21|                                    Negative|
|  Male|    18-21|                                    Negative|
|Female|    18-21|                                    Positive|
|  Male|    18-21|                                    Negative|
|Female|    22-30|                                    Positive|
|Female|    41-60|                                    Positive|
|  Male|    18-21|                                    Negative|
|Female|    41-60|                                    Positive|
|Female|    41-60|                                    Positive|
|Female|    18-21|                                    Negative|
|Female|    22-30|                                    Negative|
|Female|    22-30|                                    Negative|
|  Male|    22-30|                                    Negative|
+------+---------+--------------------------------------------+
only showing top 20 rows
```

```
indexers = [StringIndexer(inputCol=column, outputCol=column+"_index").fit(df) for column in ["Age_Range","Gender","Positive_or_negative_e
pipeline = Pipeline(stages=indexers)
df = pipeline.fit(df).transform(df)
```

```
df.show()
```

```
+------+---------+--------------------------------------------+---------------+------------+---------------------------------------
|Gender|Age_Range|Positive_or_negative_effect_on_mental_health|Age_Range_index|Gender_index|Positive_or_negative_effect_on_mental_h
+------+---------+--------------------------------------------+---------------+------------+---------------------------------------
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|Female|    22-30|                                    Negative|            2.0|         1.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|  Male|    18-21|                                    Positive|            0.0|         0.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|Female|    22-30|                                    Negative|            2.0|         1.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|Female|    18-21|                                    Positive|            0.0|         1.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|Female|    22-30|                                    Positive|            2.0|         1.0|
|Female|    41-60|                                    Positive|            1.0|         1.0|
|  Male|    18-21|                                    Negative|            0.0|         0.0|
|Female|    41-60|                                    Positive|            1.0|         1.0|
|Female|    41-60|                                    Positive|            1.0|         1.0|
|Female|    18-21|                                    Negative|            0.0|         1.0|
|Female|    22-30|                                    Negative|            2.0|         1.0|
|Female|    22-30|                                    Negative|            2.0|         1.0|
|  Male|    22-30|                                    Negative|            2.0|         0.0|
+------+---------+--------------------------------------------+---------------+------------+---------------------------------------
only showing top 20 rows
```

```
df=df.drop("Gender","Age_Range","Positive_or_negative_effect_on_mental_health")
```

```
df.show()
```

```
+---------------+------------+--------------------------------------------------+
|Age_Range_index|Gender_index|Positive_or_negative_effect_on_mental_health_index|
+---------------+------------+--------------------------------------------------+
|            0.0|         0.0|                                               0.0|
|            0.0|         0.0|                                               0.0|
|            2.0|         1.0|                                               0.0|
|            0.0|         0.0|                                               0.0|
|            0.0|         0.0|                                               1.0|
|            0.0|         0.0|                                               0.0|
|            2.0|         1.0|                                               0.0|
|            0.0|         0.0|                                               0.0|
|            0.0|         0.0|                                               0.0|
|            0.0|         1.0|                                               1.0|
|            0.0|         0.0|                                               0.0|
|            2.0|         1.0|                                               1.0|
|            1.0|         1.0|                                               1.0|
|            0.0|         0.0|                                               0.0|
|            1.0|         1.0|                                               1.0|
|            1.0|         1.0|                                               1.0|
|            0.0|         1.0|                                               0.0|
|            2.0|         1.0|                                               0.0|
|            2.0|         1.0|                                               0.0|
|            2.0|         0.0|                                               0.0|
+---------------+------------+--------------------------------------------------+
only showing top 20 rows
```

```
feature = VectorAssembler(inputCols=df.columns[1:],outputCol="features")
feature_vector= feature.transform(df)
```

```
feature_vector.show()
```

```
+---------------+------------+--------------------------------------------------+---------+
|Age_Range_index|Gender_index|Positive_or_negative_effect_on_mental_health_index| features|
+---------------+------------+--------------------------------------------------+---------+
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            2.0|         1.0|                                               0.0|[1.0,0.0]|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            0.0|         0.0|                                               1.0|[0.0,1.0]|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            2.0|         1.0|                                               0.0|[1.0,0.0]|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            0.0|         1.0|                                               1.0|[1.0,1.0]|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            2.0|         1.0|                                               1.0|[1.0,1.0]|
|            1.0|         1.0|                                               1.0|[1.0,1.0]|
|            0.0|         0.0|                                               0.0|(2,[],[])|
|            1.0|         1.0|                                               1.0|[1.0,1.0]|
|            1.0|         1.0|                                               1.0|[1.0,1.0]|
|            0.0|         1.0|                                               0.0|[1.0,0.0]|
|            2.0|         1.0|                                               0.0|[1.0,0.0]|
|            2.0|         1.0|                                               0.0|[1.0,0.0]|
|            2.0|         0.0|                                               0.0|(2,[],[])|
+---------------+------------+--------------------------------------------------+---------+
only showing top 20 rows
```

```
feature_vector.select("Positive_or_negative_effect_on_mental_health_index","features").show()
```

```
+--------------------------------------------------+---------+
|Positive_or_negative_effect_on_mental_health_index| features|
+--------------------------------------------------+---------+
|                                               0.0|(2,[],[])|
|                                               0.0|(2,[],[])|
|                                               0.0|[1.0,0.0]|
|                                               0.0|(2,[],[])|
|                                               1.0|[0.0,1.0]|
|                                               0.0|(2,[],[])|
|                                               0.0|[1.0,0.0]|
|                                               0.0|(2,[],[])|
|                                               0.0|(2,[],[])|
|                                               1.0|[1.0,1.0]|
|                                               0.0|(2,[],[])|
|                                               1.0|[1.0,1.0]|
|                                               1.0|[1.0,1.0]|
|                                               0.0|(2,[],[])|
|                                               1.0|[1.0,1.0]|
|                                               1.0|[1.0,1.0]|
|                                               0.0|[1.0,0.0]|
|                                               0.0|[1.0,0.0]|
|                                               0.0|[1.0,0.0]|
|                                               0.0|(2,[],[])|
+--------------------------------------------------+---------+
only showing top 20 rows
```

```
(trainingData, testData) = feature_vector.randomSplit([0.8, 0.2],seed = 11)
```

```
from pyspark.ml.classification import LinearSVC
svm = LinearSVC(labelCol="Positive_or_negative_effect_on_mental_health_index", featuresCol="features")
svm_model = svm.fit(trainingData)
svm_prediction = svm_model.transform(testData)
svm_prediction.select("prediction", "Positive_or_negative_effect_on_mental_health_index", "features").show()
```

```
+----------+--------------------------------------------------+---------+
|prediction|Positive_or_negative_effect_on_mental_health_index| features|
+----------+--------------------------------------------------+---------+
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       1.0|                                               1.0|[0.0,1.0]|
|       1.0|                                               1.0|[0.0,1.0]|
|       1.0|                                               1.0|[0.0,1.0]|
|       1.0|                                               1.0|[0.0,1.0]|
|       0.0|                                               0.0|[1.0,0.0]|
|       0.0|                                               0.0|[1.0,0.0]|
|       1.0|                                               1.0|[1.0,1.0]|
|       0.0|                                               0.0|(2,[],[])|
|       0.0|                                               0.0|(2,[],[])|
|       1.0|                                               1.0|[0.0,1.0]|
|       1.0|                                               1.0|[0.0,1.0]|
|       0.0|                                               0.0|[1.0,0.0]|
+----------+--------------------------------------------------+---------+
only showing top 20 rows
```

```
evaluator = MulticlassClassificationEvaluator(labelCol="Positive_or_negative_effect_on_mental_health_index", predictionCol="prediction",
```

```
svm_accuracy = evaluator.evaluate(svm_prediction)
print("Accuracy of Support Vector Machine is = %g"% (svm_accuracy))
print("Test Error of Support Vector Machine = %g " % (1.0 - svm_accuracy))
```

```
Accuracy of Support Vector Machine is = 1
Test Error of Support Vector Machine = 0
```