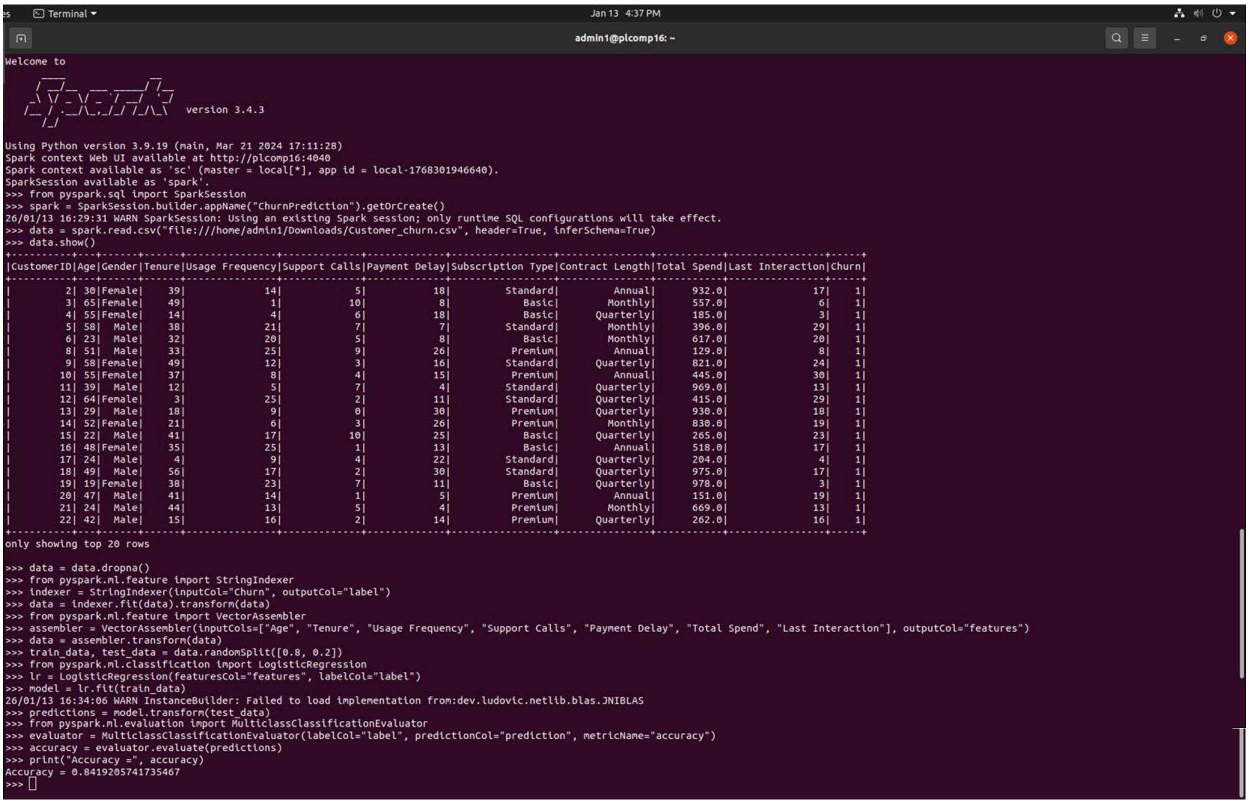
# ASSIGNMENT – II

**TITLE :**
Develop a machine learning model to predict customer churn based on historical data.

**NAME :** Shinde Shubham Dnyandev. **ROLL NO :** 23107121. **BATCH :** B



```
Welcome to
      ____              __
     / __/__  ___ _____/ /__
    _\ \/ _ \/ _ `/ __/  '_/
   /__ / .__/\_,_/_/ /_/\_\   version 3.4.3
      /_/

Using Python version 3.9.19 (main, Mar 21 2024 17:11:28)
Spark context Web UI available at http://plcomp16:4040
Spark context available as 'sc' (master = local[*], app id = local-1768301946640).
SparkSession available as 'spark'.
>>> from pyspark.sql import SparkSession
>>> spark = SparkSession.builder.appName("ChurnPrediction").getOrCreate()
26/01/13 16:29:31 WARN SparkSession: Using an existing Spark session; only runtime SQL configurations will take effect.
>>> data = spark.read.csv("file:///home/admin1/Downloads/Customer_churn.csv", header=True, inferSchema=True)
>>> data.show()
+----------+---+------+------+---------------+-------------+-------------+-----------------+---------------+-----------+----------------+-----+
|CustomerID|Age|Gender|Tenure|Usage Frequency|Support Calls|Payment Delay|Subscription Type|Contract Length|Total Spend|Last Interaction|Churn|
+----------+---+------+------+---------------+-------------+-------------+-----------------+---------------+-----------+----------------+-----+
|         2| 30|Female|    39|             14|            5|           18|         Standard|         Annual|      932.0|              17|    1|
|         3| 65|Female|    49|              1|           10|            8|            Basic|        Monthly|      557.0|               6|    1|
|         4| 55|Female|    14|              4|            6|           18|            Basic|      Quarterly|      185.0|               3|    1|
|         5| 58|  Male|    38|             21|            7|            7|         Standard|        Monthly|      396.0|              29|    1|
|         6| 23|  Male|    32|             20|            5|            8|            Basic|        Monthly|      617.0|              20|    1|
|         8| 51|  Male|    33|             25|            9|           26|          Premium|         Annual|      129.0|               8|    1|
|         9| 58|Female|    49|             12|            3|           16|         Standard|      Quarterly|      821.0|              24|    1|
|        10| 55|Female|    37|              8|            4|           15|          Premium|         Annual|      445.0|              30|    1|
|        11| 39|  Male|    12|              5|            7|            4|         Standard|      Quarterly|      969.0|              13|    1|
|        12| 64|Female|     3|             25|            2|           11|         Standard|      Quarterly|      415.0|              29|    1|
|        13| 29|  Male|    18|              9|            0|           30|          Premium|      Quarterly|      930.0|              18|    1|
|        14| 52|Female|    21|              6|            3|           26|          Premium|        Monthly|      830.0|              19|    1|
|        15| 22|  Male|    41|             17|           10|           25|            Basic|      Quarterly|      265.0|              23|    1|
|        16| 48|Female|    35|             25|            1|           13|            Basic|         Annual|      518.0|              17|    1|
|        17| 24|  Male|     4|              9|            4|           22|         Standard|      Quarterly|      204.0|               4|    1|
|        18| 49|  Male|    56|             17|            2|           30|         Standard|      Quarterly|      975.0|              17|    1|
|        19| 19|Female|    38|             23|            7|           11|            Basic|      Quarterly|      978.0|               3|    1|
|        20| 47|  Male|    41|             14|            1|            5|          Premium|         Annual|      151.0|              19|    1|
|        21| 24|  Male|    44|             13|            5|            4|          Premium|        Monthly|      669.0|              13|    1|
|        22| 42|  Male|    15|             16|            2|           14|          Premium|      Quarterly|      262.0|              16|    1|
+----------+---+------+------+---------------+-------------+-------------+-----------------+---------------+-----------+----------------+-----+
only showing top 20 rows

>>> data = data.dropna()
>>> from pyspark.ml.feature import StringIndexer
>>> indexer = StringIndexer(inputCol="Churn", outputCol="label")
>>> data = indexer.fit(data).transform(data)
>>> from pyspark.ml.feature import VectorAssembler
>>> assembler = VectorAssembler(inputCols=["Age", "Tenure", "Usage Frequency", "Support Calls", "Payment Delay", "Total Spend", "Last Interaction"], outputCol="features")
>>> data = assembler.transform(data)
>>> train_data, test_data = data.randomSplit([0.8, 0.2])
>>> from pyspark.ml.classification import LogisticRegression
>>> lr = LogisticRegression(featuresCol="features", labelCol="label")
>>> model = lr.fit(train_data)
26/01/13 16:34:06 WARN InstanceBuilder: Failed to load implementation from:dev.ludovic.netlib.blas.JNIBLAS
>>> predictions = model.transform(test_data)
>>> from pyspark.ml.evaluation import MulticlassClassificationEvaluator
>>> evaluator = MulticlassClassificationEvaluator(labelCol="label", predictionCol="prediction", metricName="accuracy")
>>> accuracy = evaluator.evaluate(predictions)
>>> print("Accuracy =", accuracy)
Accuracy = 0.8419205741735467
>>> 
```