

Figure 3: Last fifty entries of *Iris.csv* read line by line and written into a particular topic in Kafka.

Problem 2

Write a subscriber.py file that uses spark streaming (can be receiver-based, dstream or structured) for producing real-time predictions on these rows by utilizing the model trained in Lab 5 and calculates the accuracy (the real-time predictions, true labels and accuracy all should get printed on console).

Solution:

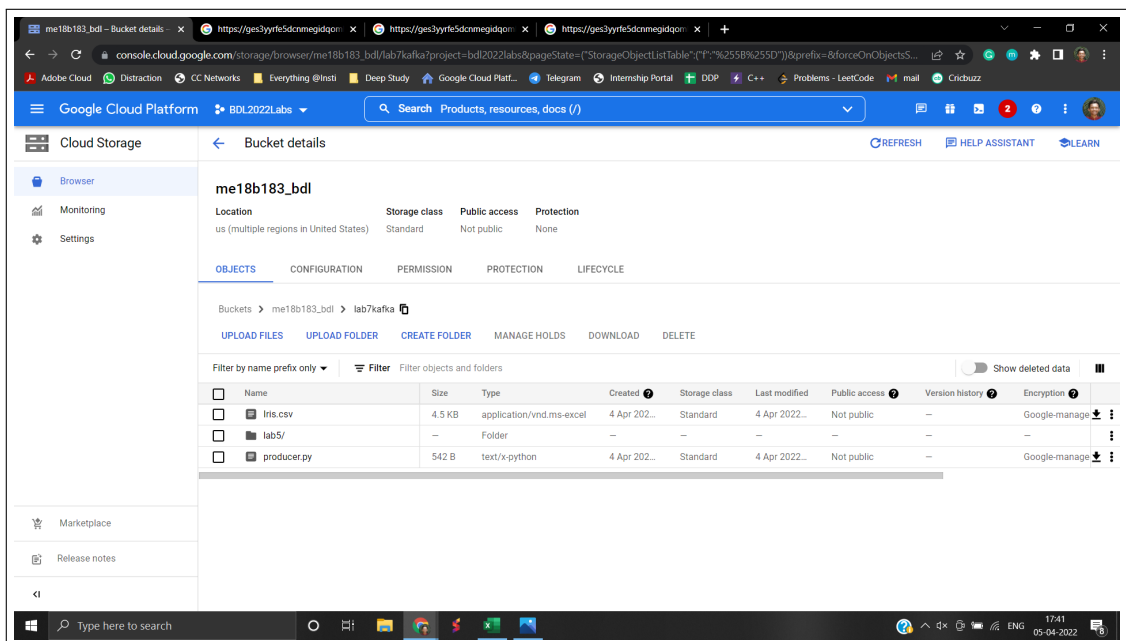


Figure 4: Cloud storage screenshot for lab 7 tasks.

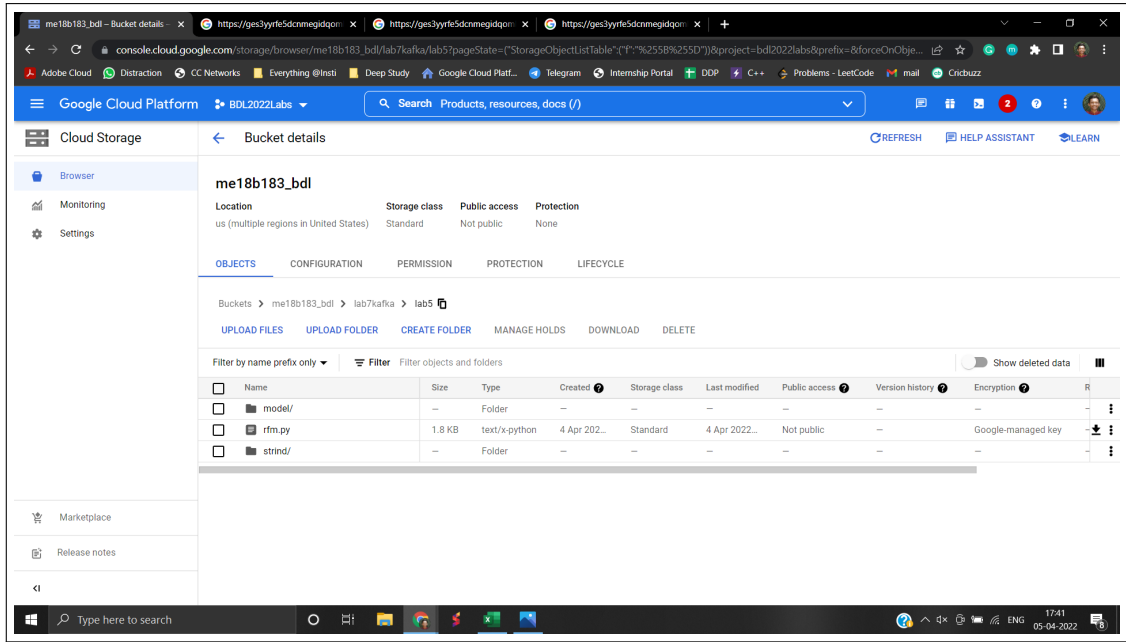


Figure 5: Random Forest Classifier model as saved from lab 5.

Batch: 8		

Actual_Species	Predicted_Species	Accuracy

Iris-setosa	Iris-setosa	100

Batch: 9		

Actual_Species	Predicted_Species	Accuracy

Iris-setosa	Iris-setosa	100

Figure 6: Correct prediction of *Iris.csv* data in real-time.

Batch: 91		

Actual_Species	Predicted_Species	Accuracy

Iris-versicolor	Iris-versicolor	100

Batch: 92		

Actual_Species	Predicted_Species	Accuracy

Iris-versicolor	Iris-versicolor	100

Figure 7: Correct prediction of *Iris.csv* data in real-time.

Batch: 81		
Actual_Species	Predicted_Species	Accuracy
Iris-versicolor	Iris-virginica	0

Figure 8: Incorrect prediction of *Iris.csv* data in real-time.

* * **End of Assignment** * *