CS5542 Big Data Apps and Analytics  
LAB ASSIGNMENT-9  
REPORT and SCREEN SHOTS

*By Rakesh Reddy Bandi(2) & Mark J Schultz(26)*

**Spark ML Lib (with Instagram streaming and data from smart devices) and Smart Application**

**1) Image collection and sentimental analysis based on the image tags using Instagram streaming (related to your project)**

**a. Training Datasets: Instagram Streaming/Categorized Image (e.g., Static UEC Food Dataset) and metadata**

**b. Testing Datasets e.g., Image, UserGroup, Category, Rating (Instagram streaming)**

**2) Image Classification based on the categories related to your project**

**3) Image-based Recommendation system (related to your own project)**

**a. The rating based on sentiment analysis of Instagram metadata**

**b. Expected outcome is to make a recommendation based on user image input or profile (e.g., preferences, location, gender, age)**

**4) Instagram trend notification to smartphone/smartwatch**

**5) Mobile Recommendation through smartphone/smartwatch using your ML application**

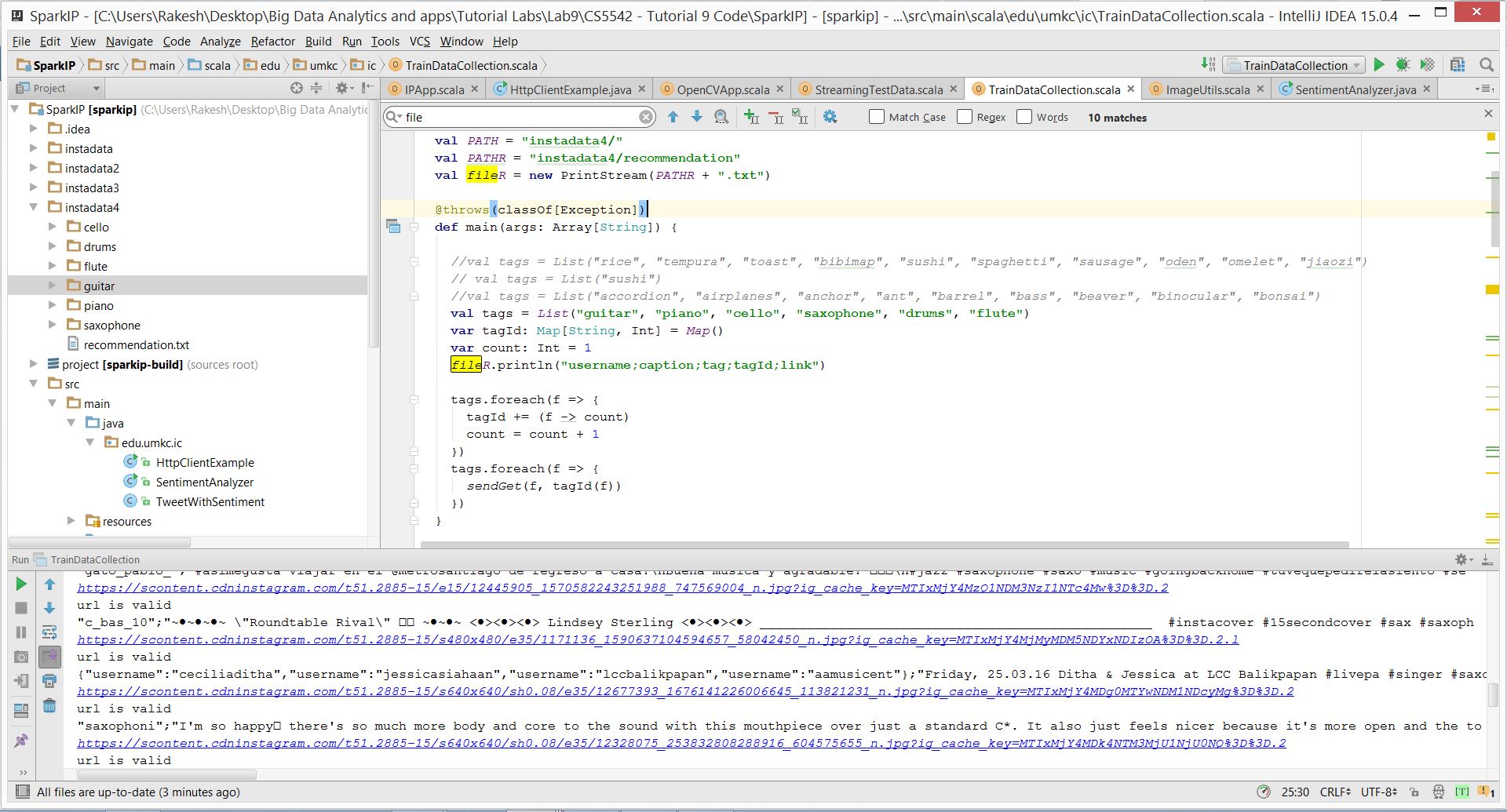
**1) Image collection and sentimental analysis based on the image tags using Instagram streaming (related to your project)**

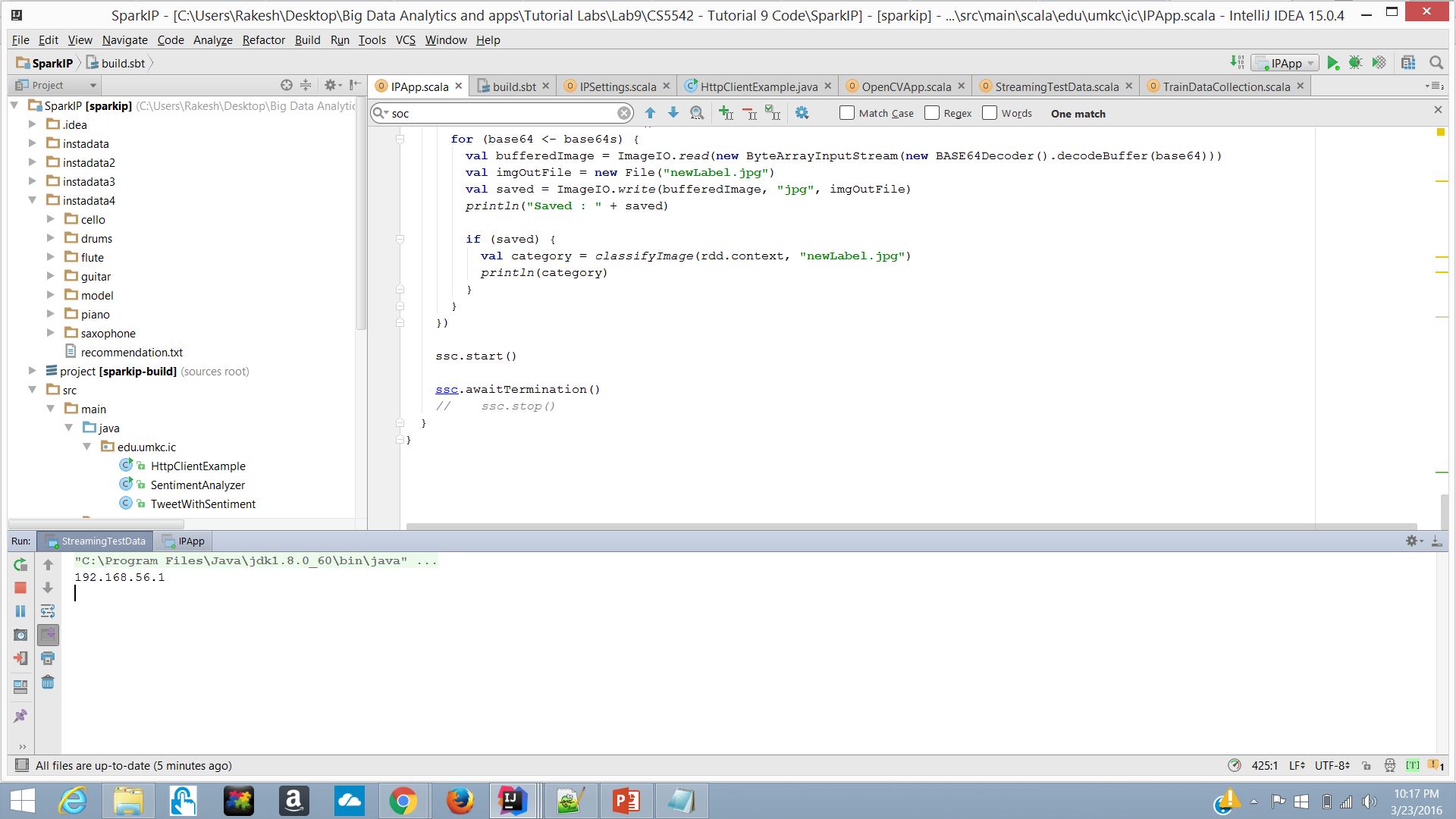
**a. Training Datasets: Instagram Streaming/Categorized Image (e.g., Static UEC Food Dataset) and metadata**

**b. Testing Datasets e.g., Image, UserGroup, Category, Rating (Instagram streaming)**

* We have collected the image datasets from the Instagram. We have collected the Image data for the hashtags musical instruments.
* The list of tag categories for musical instruments are List("guitar", "piano", "cello", "saxophone", "drums", "flute")
* Along with the Training data collection we have also collected the image tags based on user group and their image links and tags in Recommendation.txt file which is later used for the Recommendation. Format “username;caption;tag;tagId;link”
* Testing Data is live streaming Data from the Instagram.

1. We have collected the image datasets from the Instagram. We have collected the Image data for the hashtags musical instruments. **Data Folder: Instadata4 -** **List("guitar", "piano", "cello", "saxophone", "drums", "flute")**

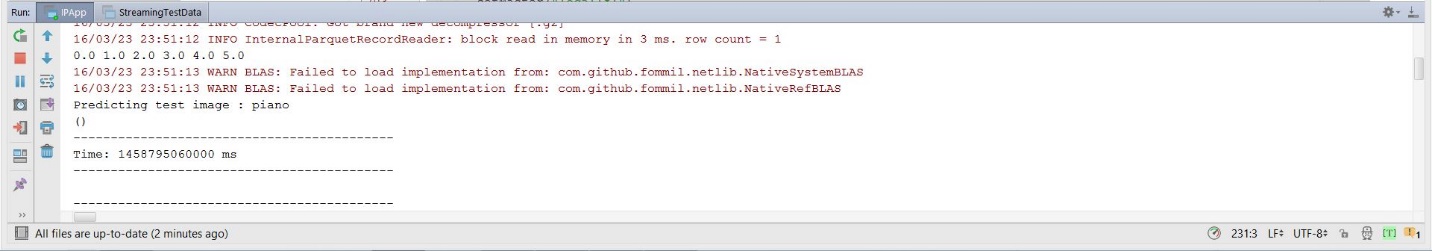


2. For Testing and Validation of the model we create a live streaming data to predict the current live image for the particular tags. We open a socket to listen the request from the client i.e. our classification model.

**2) Image Classification based on the categories related to your project**

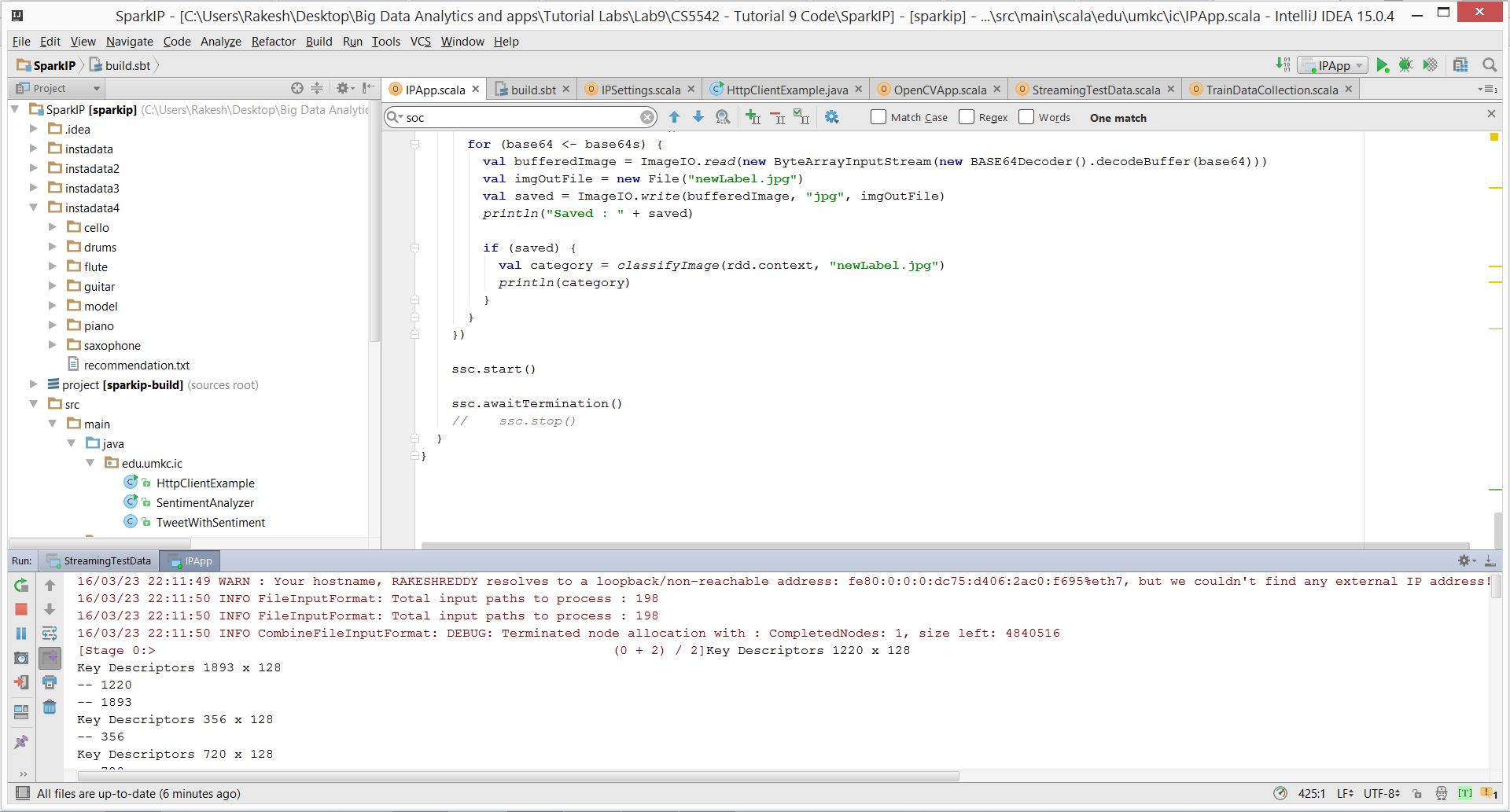
Image classification:

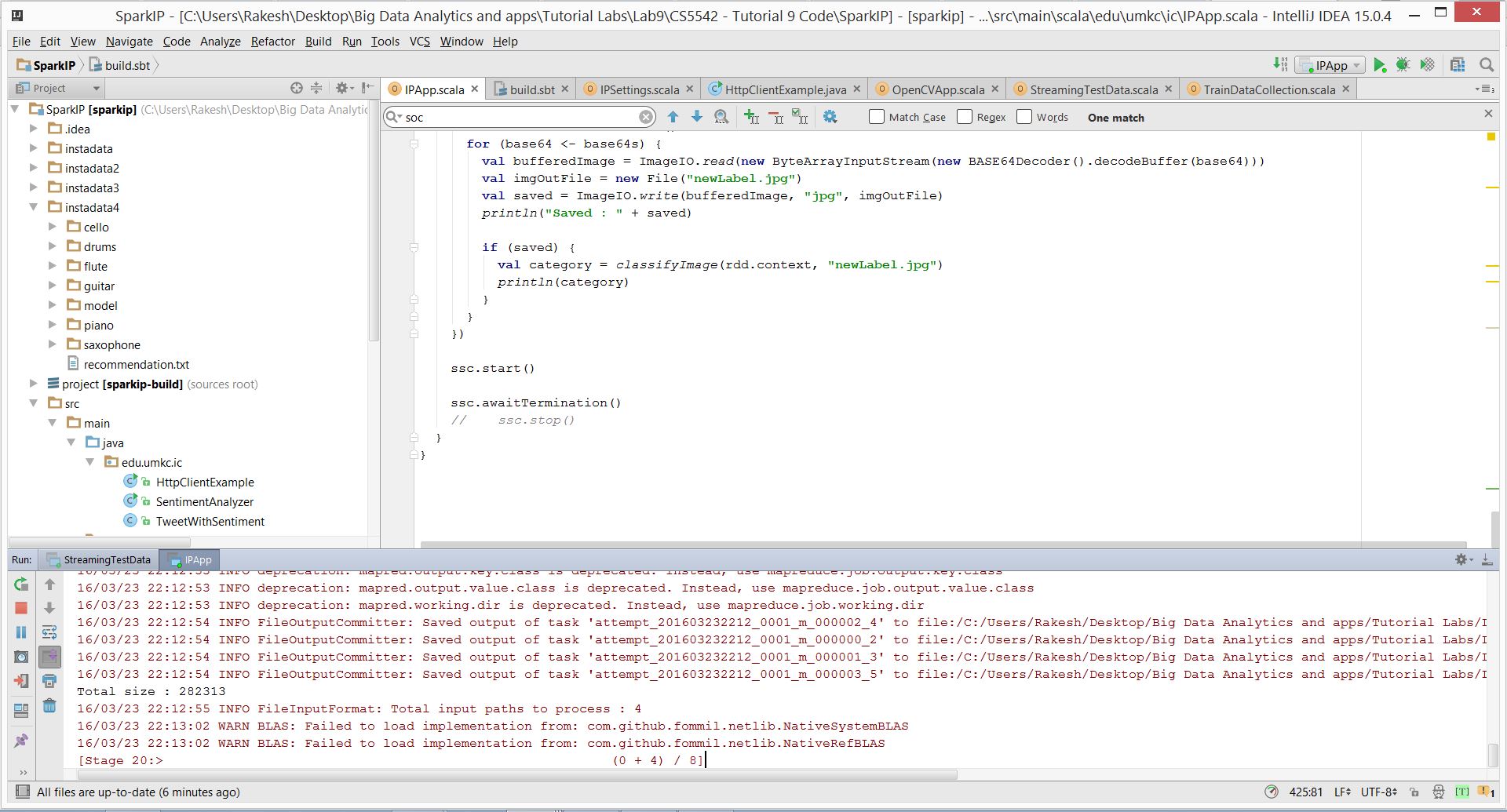
Predicting the tag for Live Image Data



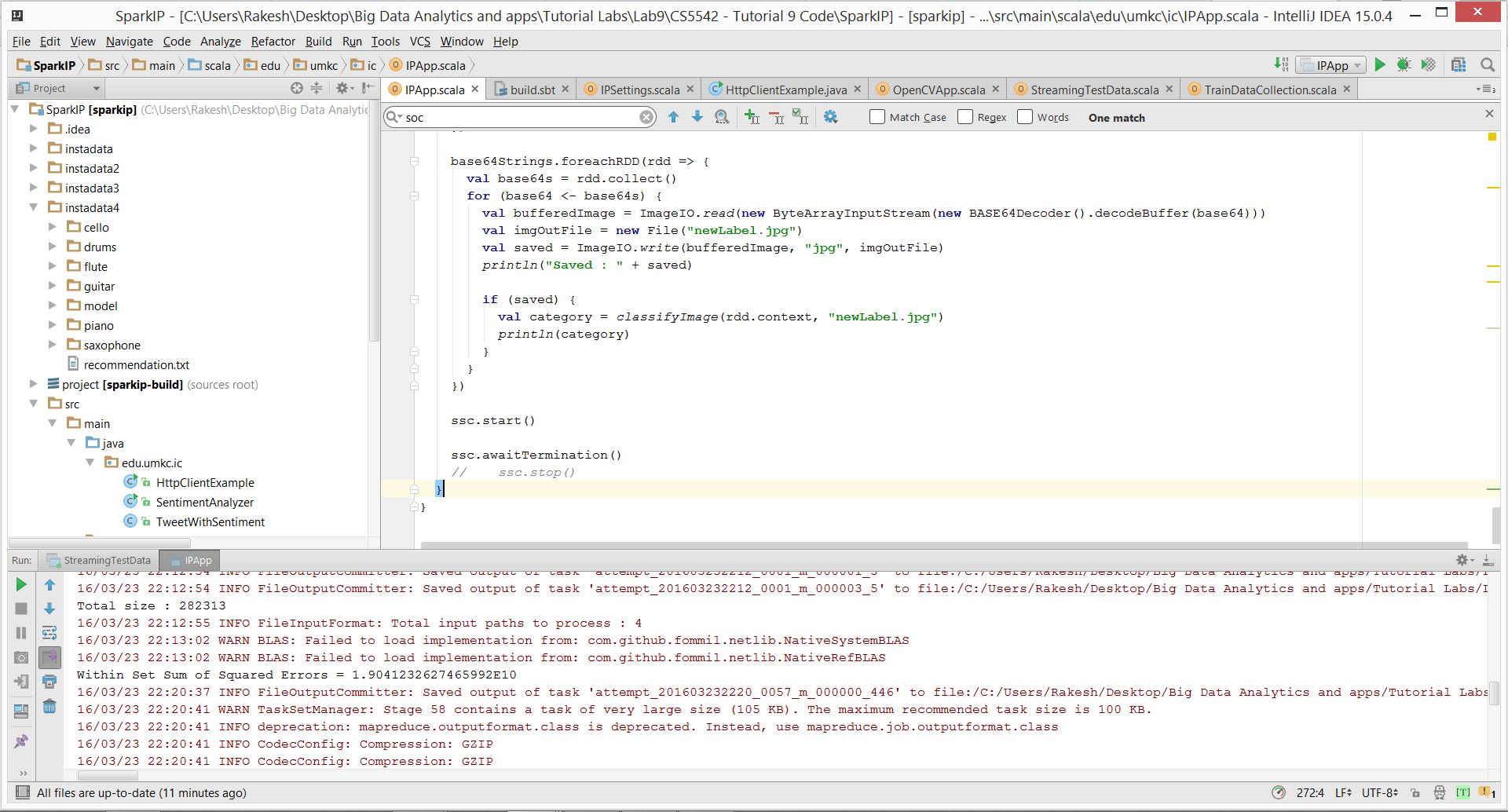
Steps:

1. First we get the key Descriptors for the all images

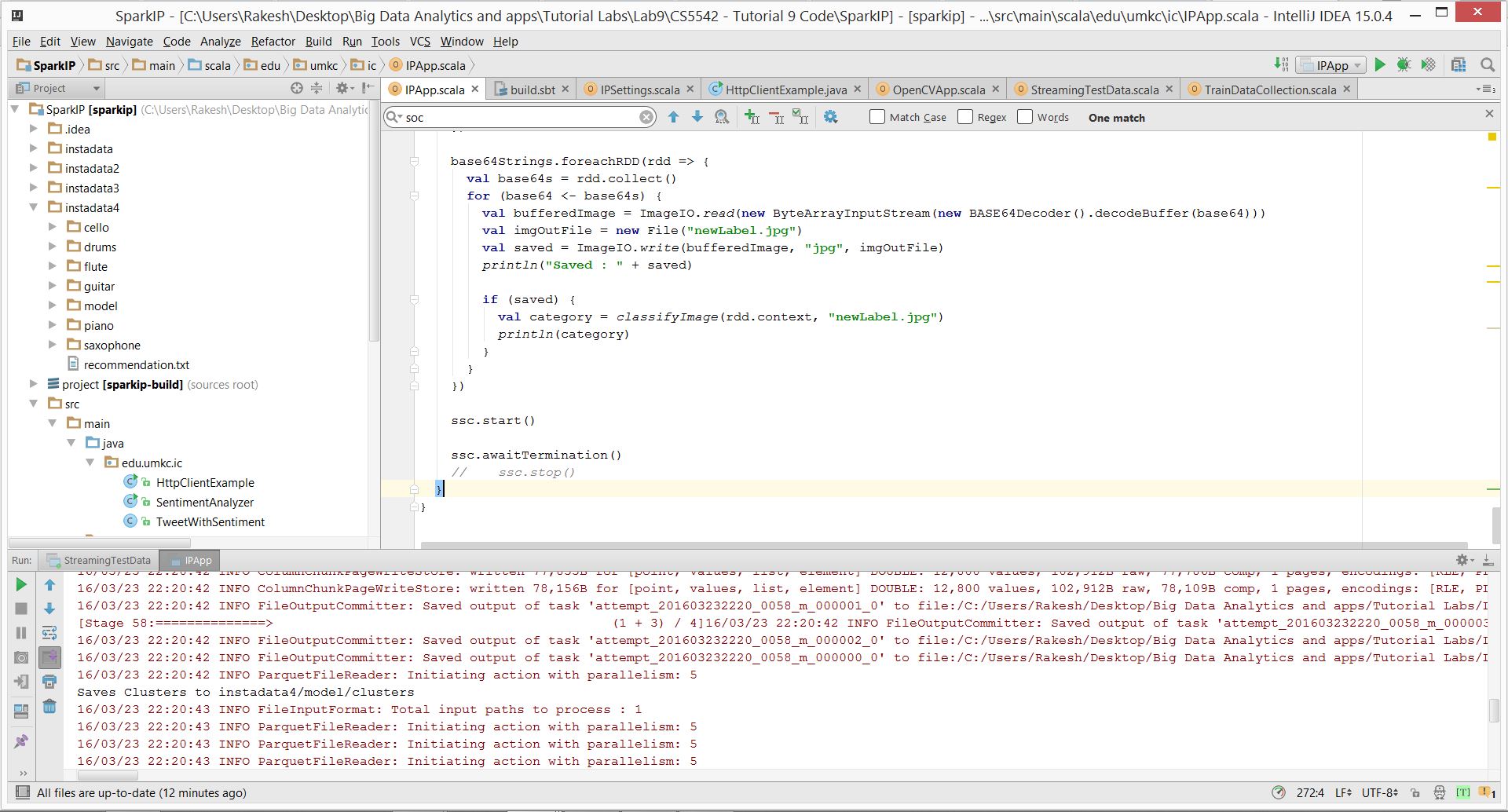




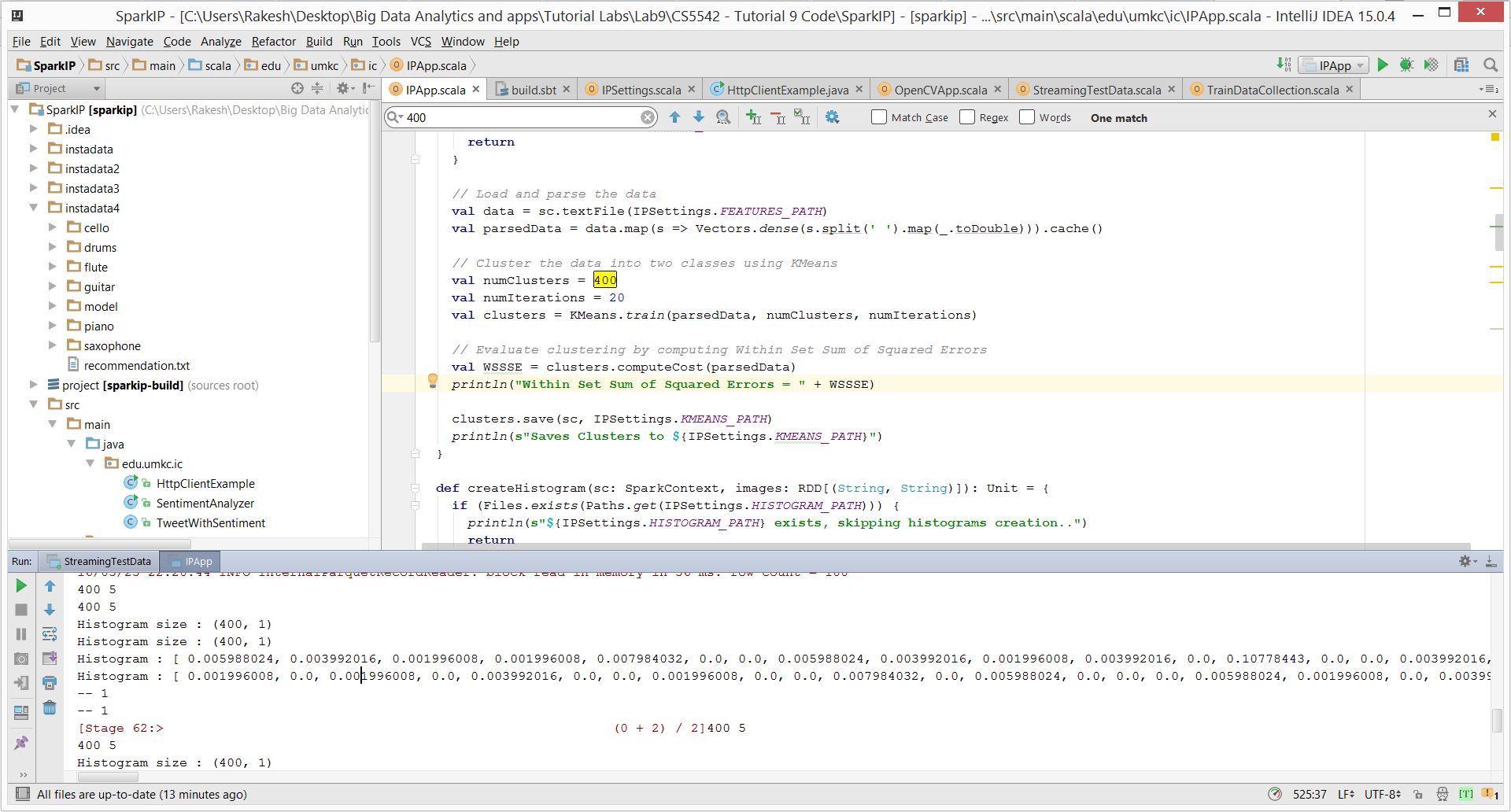
5. Displaying for squared errors



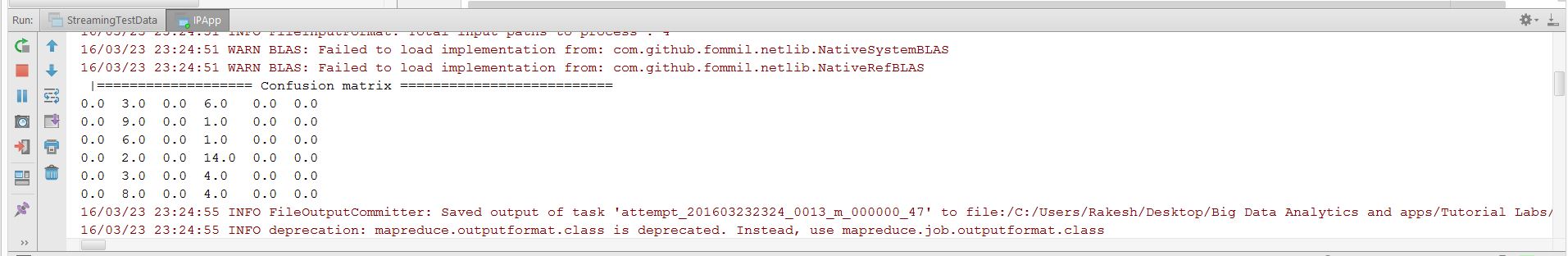
6. Clustering the image Bag of Visual words



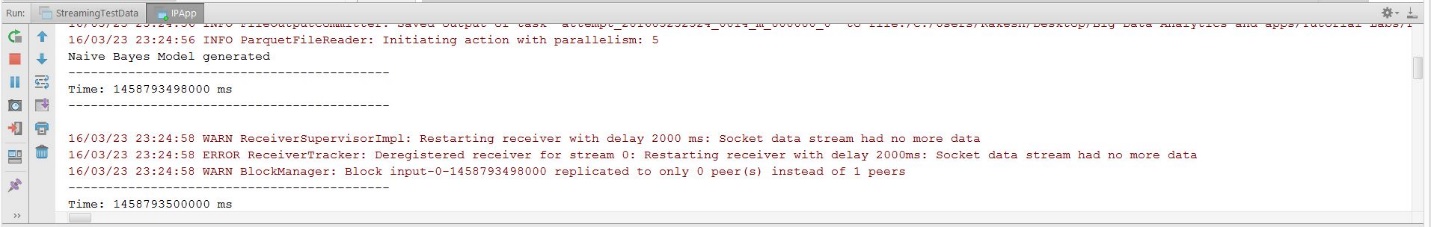
7. Histograms for features of images



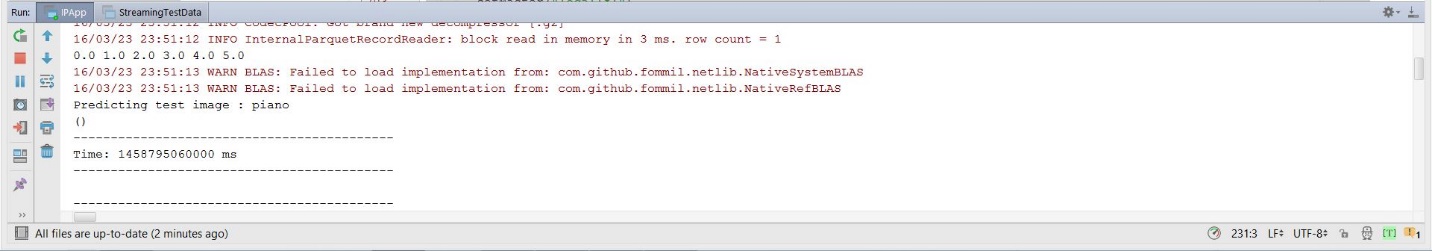
8. Confusion Matrix Almost we got accuracy of 45%



9. Built the Naïve Bayes Model



10. **Predicting the tag for Live Image Data – Piano tag**



11. Trained model stored in Data folder

