**Voice-Interactive, Immersive Sports Experience**

A Question and Answering system for sports domain using video analysis.

**Team 6 - Affinity**

Gulnoza Khakimova

Ganesh Taduri

Raghava Kundavajjala

Steven Evans



**1.PROJECT OBJECTIVES:**

**1.1 Background**

Since the popularization of virtual reality, the focus has predominantly centered around gaming. This can be attributed to a variety of factors such as the gaming community’s tendency to embrace new technologies and the familiarity of computer-generated graphics to gamers. Now that virtual reality has become more familiar to mainstream consumers there is a push to expand its use into other areas of entertainment, including professional sports. Over the past year or so major sports broadcasting companies, such as FOX Sports and NBC, have partnered with tech startups to release 360°, immersive VR coverage of select sporting events like the 2016 Summer Olympic Games and the 2015 NBA season opener.

While VR for sports content has proven to be a promising market, there continues to be a number of aspects that can be enhanced to provide a more enjoyable user experience. One problem is where and how to display relevant information to the user. In a 360° environment care must be taken not to obstruct the user’s view while displaying information. Another opportunity for improvement is the lack of control for users to determine what information they want and when they want it. Finally, a third concern is the limited methods of interaction between the user and the environment. The aim of this project is to address these issues by allowing the user to interact with the VR environment via voice commands. Voice interaction would improve the first issue by allowing the system to respond to voice commands with audio feedback, rather than visual, thus reducing the amount of screen clutter. User’s would also gain control of the information they receive by using voice queries to request the information they need. By utilizing voice commands users would have another, more natural, method of interaction with system in addition to headset buttons and remote controls that are included with many current VR headsets.

**1.2 Significance**

Voice controlled computer systems have been in existence for some time and used in a variety of academic and commercial fields. However, the use of voice commands within virtual reality environments has appeared only recently. For virtual reality professional sports coverage, the ability for viewers to use voice-interaction is non-existent. This can most likely be attributed to relatively recent emergence of the VR sports market. The majority of effort in early development was placed on capturing high-quality footage in a format that could be viewed in 360°. Now that the concept of VR sports has been established and shown promise in marketability, efforts should shift to enhancing the viewer’s experience by providing new features.

**1.3 Objectives**

Although this project will focus on improving interactions with VR sports content, the ideas and results would apply to a variety of other media. Our primary purpose is to take an experience that has traditionally been static and enable it to be user controlled in a natural way. In doing so, we expect the outcome to improve the user’s visual experience by reducing the amount of data displayed over visual content, enhance the overall experience by providing relevant contextual information when the user wants it, and promote increased adoption of VR for entertainment purposes outside of gaming through intuitive user interactions.

**1.4 System Features**

The main feature of the system is to answer a specific question asked by a user related to a sport video. A few of the examples that user could ask the system are:

* What sport is it?
* *Game info - Who’s winning? What quarter is it? How many hard until first down?*
* *Player statistics - What’s the pitcher’s ERA? How many assists has Kobe made so far?*
* *Switch view perspectives - Take me to the fifty-yard line.*

**2. APPROACH**

**2.1 Data Sources**

We collected large amount of data sets - images and videos for training and testing the model. The major data sources that we used for collecting the data are google images and videos from youtube. For image classification it was a little difficult to find better images on all different categories of sports so instead we extracted main frames from the video we collected from youtube and used these as images for training the model.

**2.2 Analytic tools**

**2.3 Analytical tasks**

**2.4 expected Inputs/Outputs**

This application is a typical Question and Answering system where a user can ask a question about a sport Image/video and system answers the question. A few of the expected questions and answers from a user and the system respectively are

1. Q: What game is it?

A: This is Basketball

1. What is the current score?

A: The current score is Nadal:3; Federer : 5

1. Who would win the game?

A: According to previous Statistics seems like Federer would win the Game.

2.

**2.5 Algorithms**

The major algorithms we used for this project are related to Machine Learning algorithms. For phase 1 to implement the image classification we Decision Tree and Random Forest algorithms. The Main intention of this work is to study how different machine learning algorithms are performing on different sets. In the future phases of the project we are also planning to implement Naive Bayes Algorithm. Also, we are planning to implement Neural Networks for Deep learning in the future.

**3. RELATED WORK**

Currently, there are several tech startups which partner with broadcasting companies and major league sports. A few of the projects that are related to this project are:

● *Fox Sports VR[7]*

● *NextVR[11]*

● *EON Sports[12]*

● *LiveLikeVR[13]*

● Virtually Live[14] - virtual reconstruction, social aspect

Research and products incorporating voice control and VR:

● *VoiceBot[8] - voice powered game control*

●  *Modbox[9] - VR multiplayer sandbox game*

●  *VRIO[10] - speech processing unit for VR and real-world scenarios*

● *VR Speech Synthesis[6]*

**4. APPLICATION SPECIFICATION**

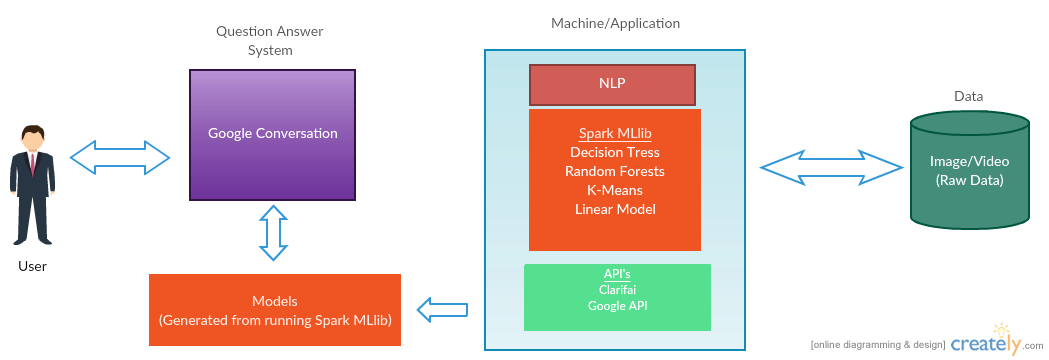
**4.1 System Specification**

Operating System: Windows 10/ IOS

Memory: 8GB

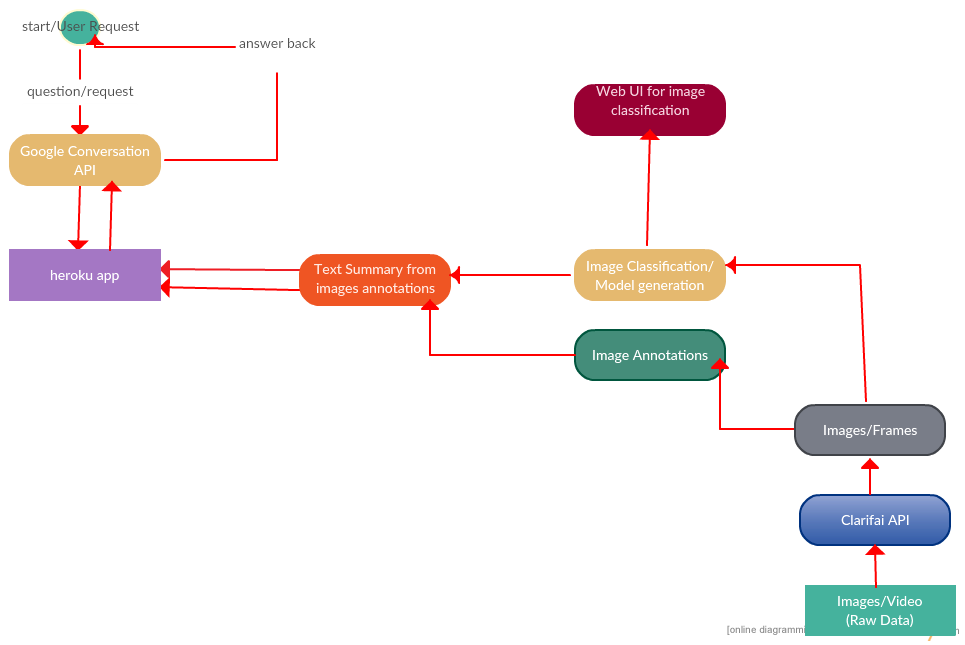
Processor: intel i5 6th gen/ AMD

**4.1.1 Software Architecture**

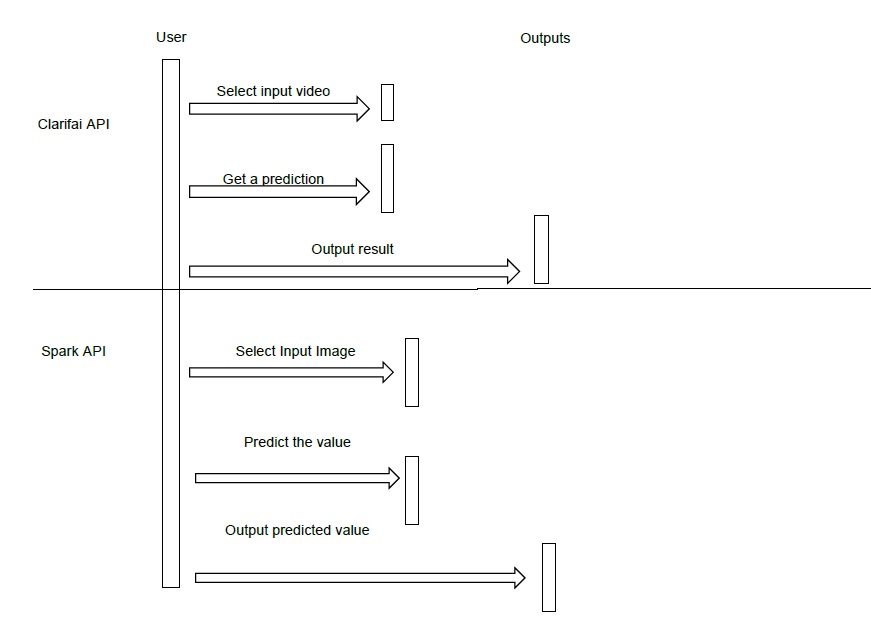


**4.1.2 Features, workflow, technologies**

* **Activity Diagram**



* **Sequence Diagram**



* **Feature specification**

1. Identify what kind of game it is by analyzing input video.
2. Identify what kind of game it is by analyzing input image using Decision tree and Random Forest Models.

* **Operation Specification**

1. Videos no more than 2MB in size.
2. Do not use huge dataset for training and testing.

**4.1.3 Existing Applications/Services Used: Name, Description, UR**

* Clarifai API
* Spark API

**5. IMPLEMENTATION**

**5.1 Implementation of the application using Clarifai API**

Clarifai APi offers video and image recognition. Clarifai API is built around a simple idea – you feed input file and Clarifai API service returns prediction. There are several models available in Clarifai API - for example food, travel, wedding, etc. In our project we are using the General Model.

This part of the project predicts what type of sport game is played on the provided video file. Below are steps used to predict video game using external API:

1. Key Frames detection- take video file as input and iterate over video frames to detect key frames:

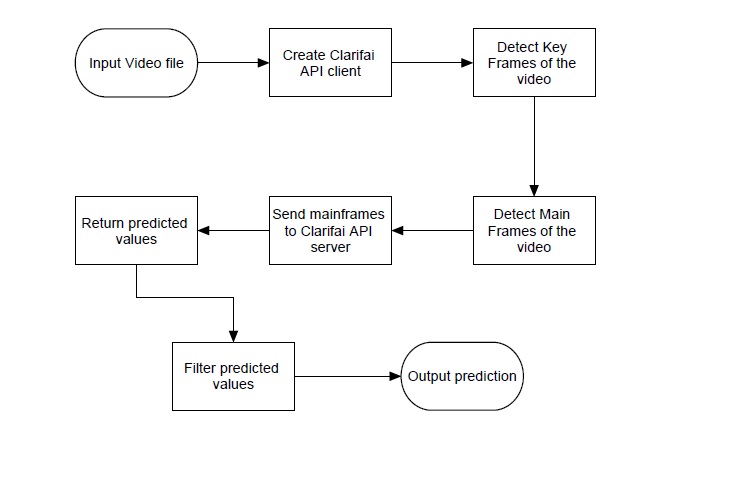


2. Detect main frames – after detecting key frames, main frames need to be pointed out so they could be uploaded to Clarifai API server to get predictions.



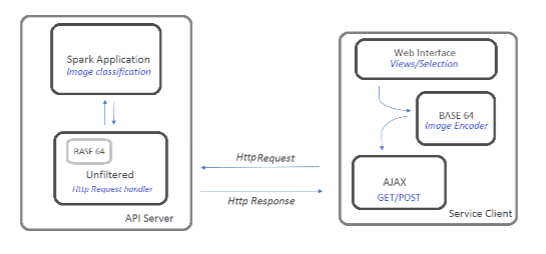
3. Annotation – after detection main frames and uploading them to Clarifai server for processing we get back predictions on what sport game is shown on the main frame. At this level we need to filter some of the predicted values since we are interested only in what sport is shown on the video. We created the program which can predict the following sport games: soccer/football, tennis, basketball, volleyball, swimming, boxing, badminton and cricket. Other games can be added as needed.

Diagram:



**5.2 Implementation of the application using Spark Client API**

In the phase 1 of the project, we implemented a simple application using a spark Client API. This application , given an image will identify what kind of sport it is. We considered five different categories of sports as five different classes while training the model. Cricket, BasketBall, Swimming, Boxing and Tennis are the five sport categories used for this application. We collected large amount of data set( Sports images) and categorized them into five different classes with respect to the sport the belong and used decision tree model to generate the model for image classification. SIFT features of the images are used for implementing the classification algorithm. Then, we used this model from server end to test an image from client side. Basically, We implemented the following architecture shown in the figure.



*Fig. Architecture of Spark Client API.*

A client can request a service from the server using GET and POST calls for sending an image to the server and to get a response of the prediction of the Image. Once the model gets an image it predicts the image sports category using SIFT features and displays the result at the client end.

We, also considered Random Forest algorithm for generating the model for image classification expecting it would fetch more accuracy than Decision tree. On Contrary Decision tree model has more Accuracy of 84% than Random forest which has around 72%. Screenshots of the results are included in the documentation.

**5.3 Implementation of the application using Google Conversation API**



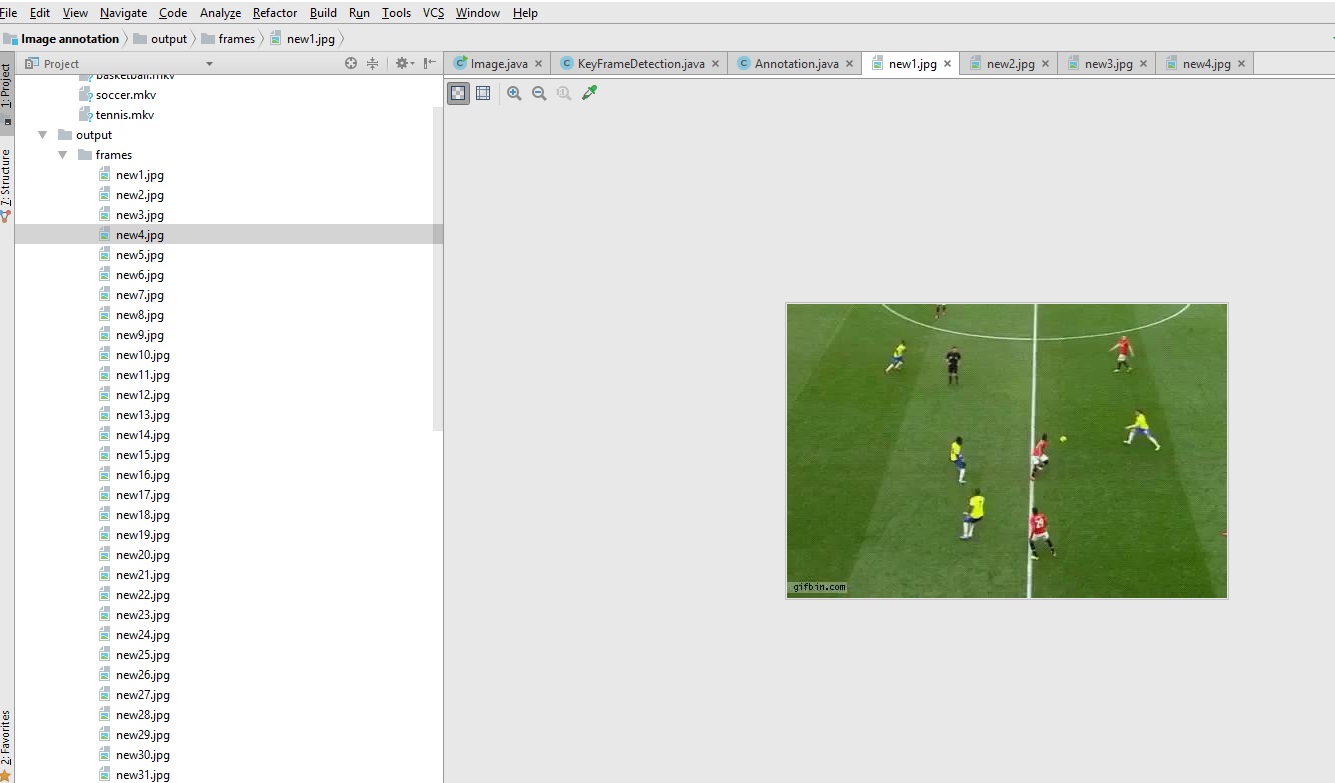
[15]API.AI is a natural language understanding platform that makes it easy for developers (and non-developers) to design and integrate intelligent and sophisticated conversational user interfaces into mobile apps, web applications, devices, and bots.

We used api.api to create an agent VoiceInteractiveISE. We have created intent and entities as part of the conversation. We have given the summary of the video and several other data as inputs and trained the agent.

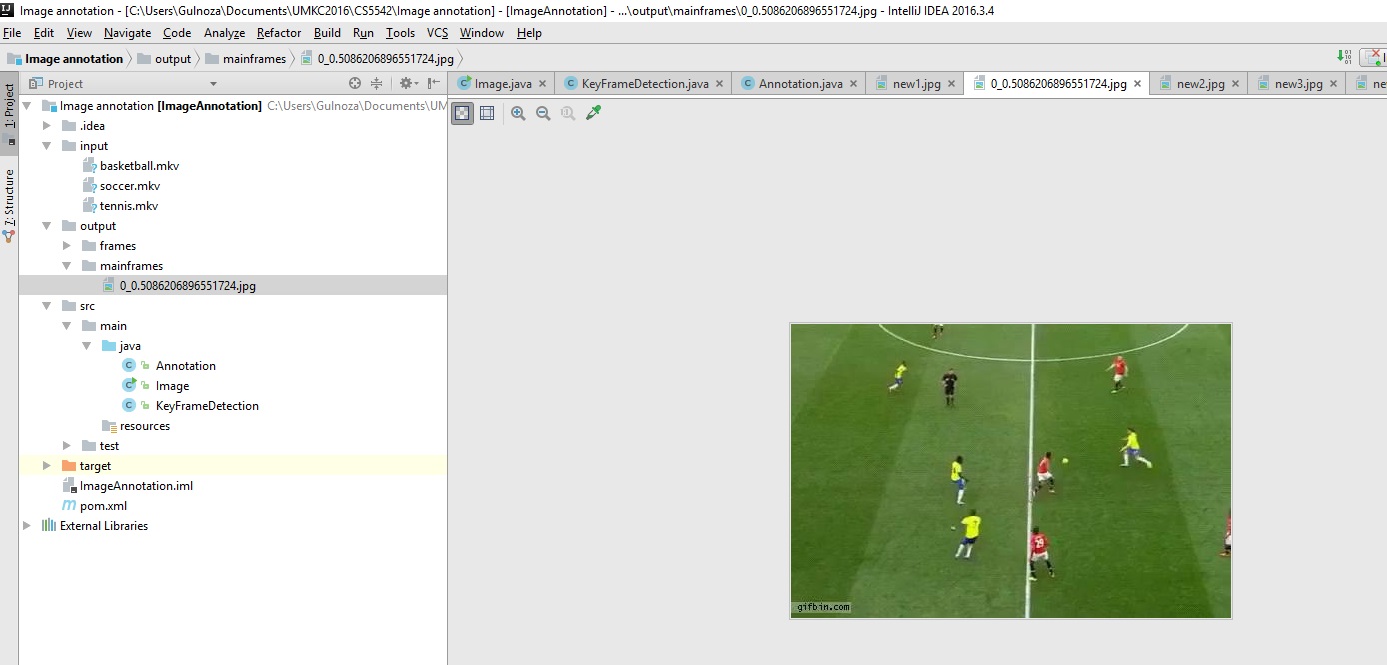
**6. DOCUMENTATION**

* **Screenshots of application using clarifai API**

Detected key Frames:

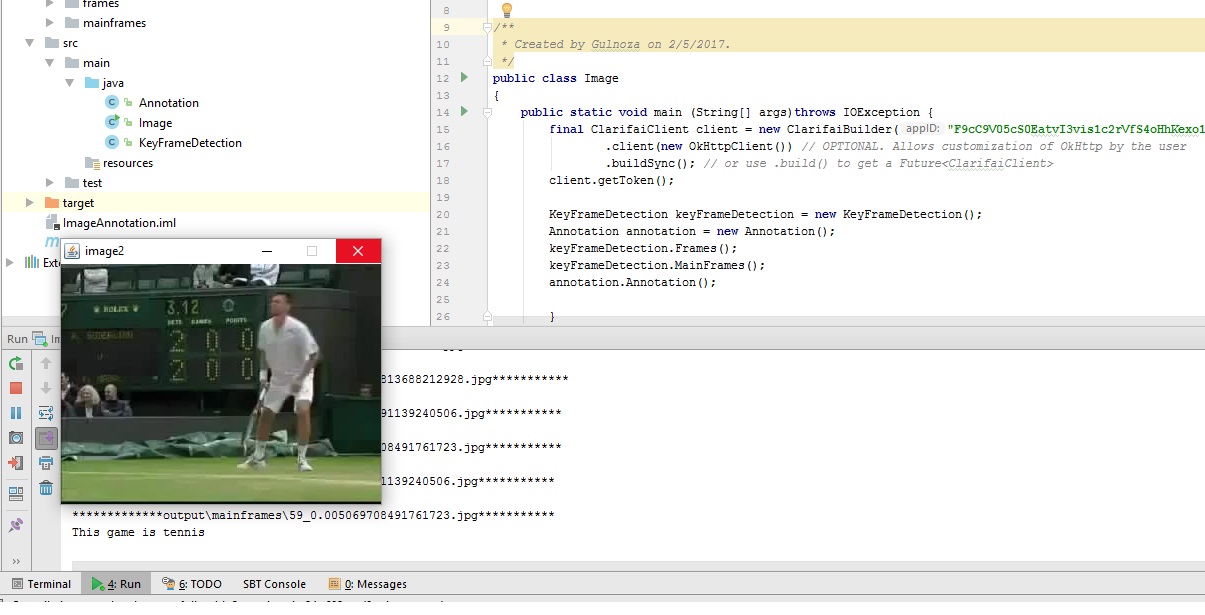


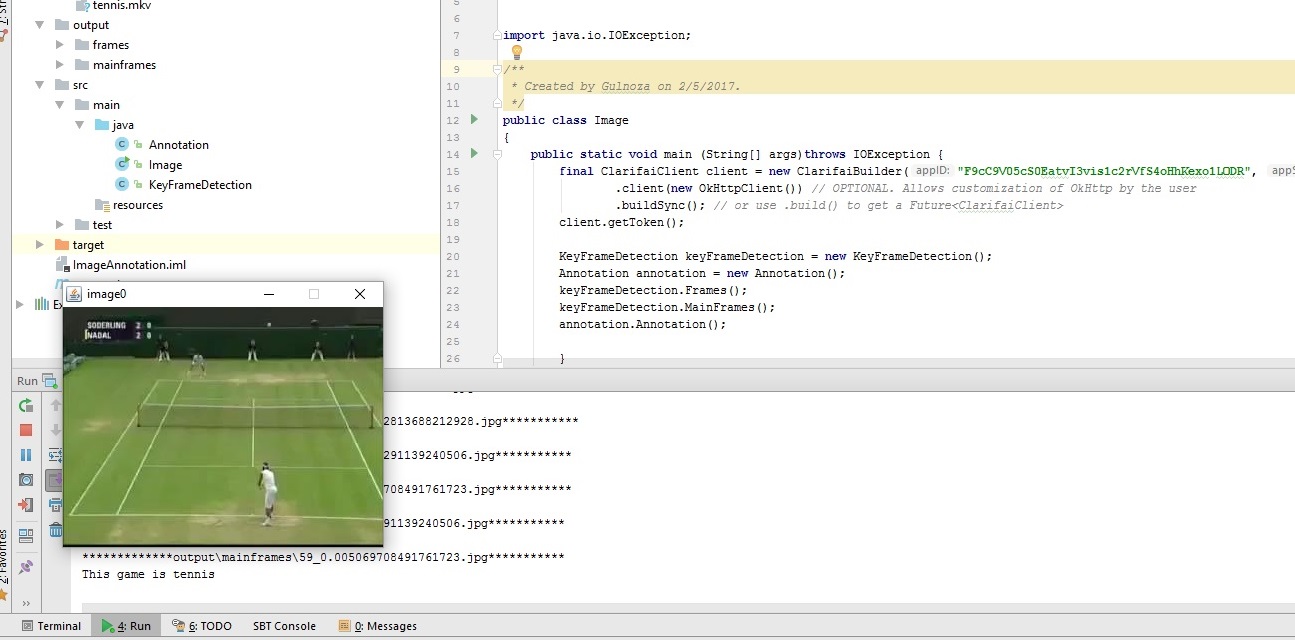
Detected Main Frames:



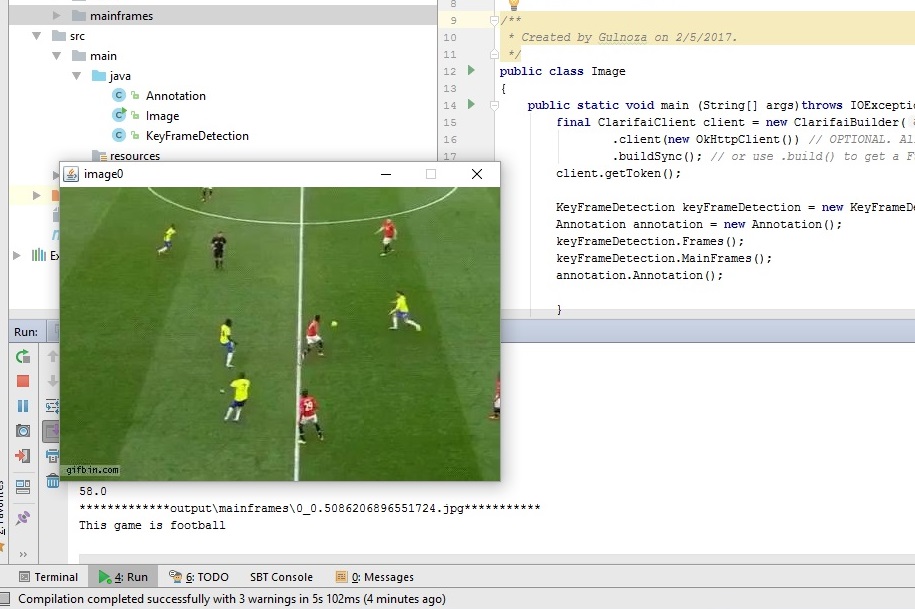
Predicting types of games:

Tennis:

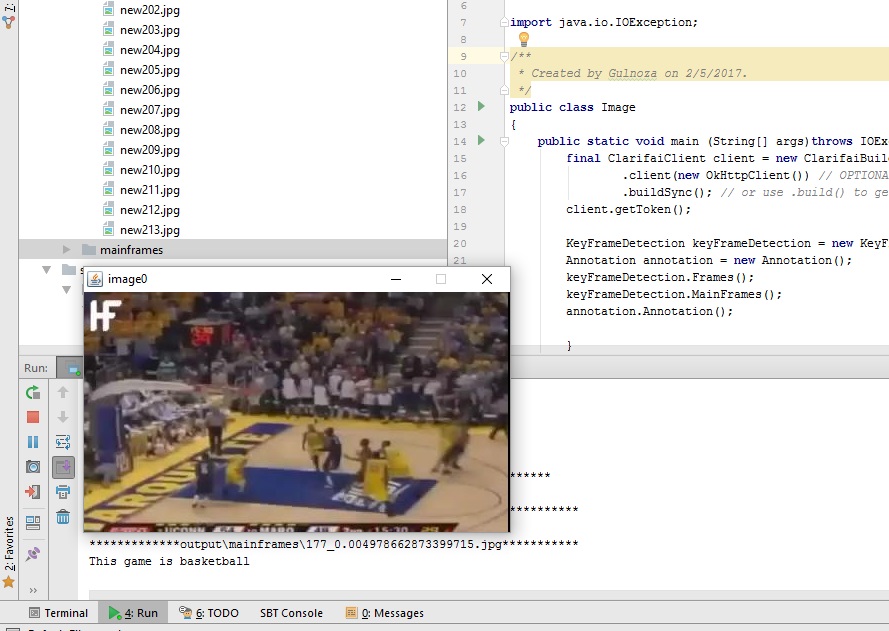




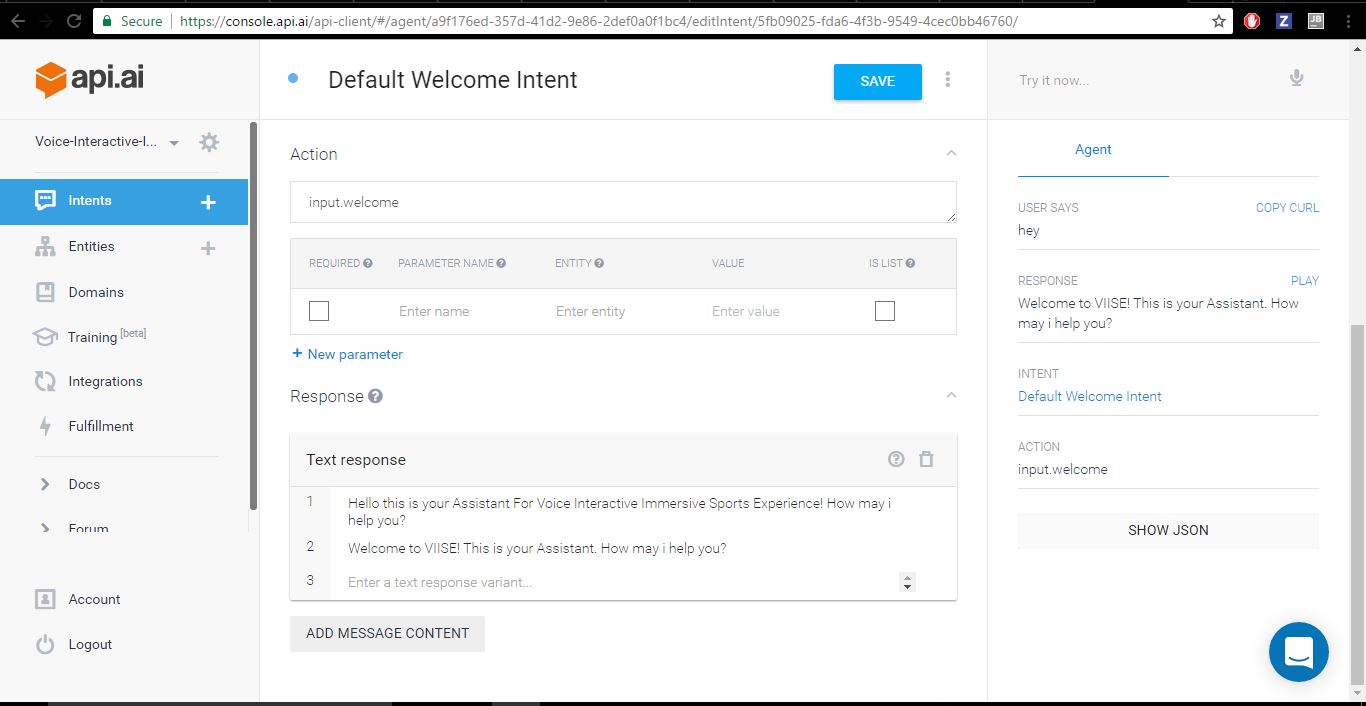
Soccer/football:

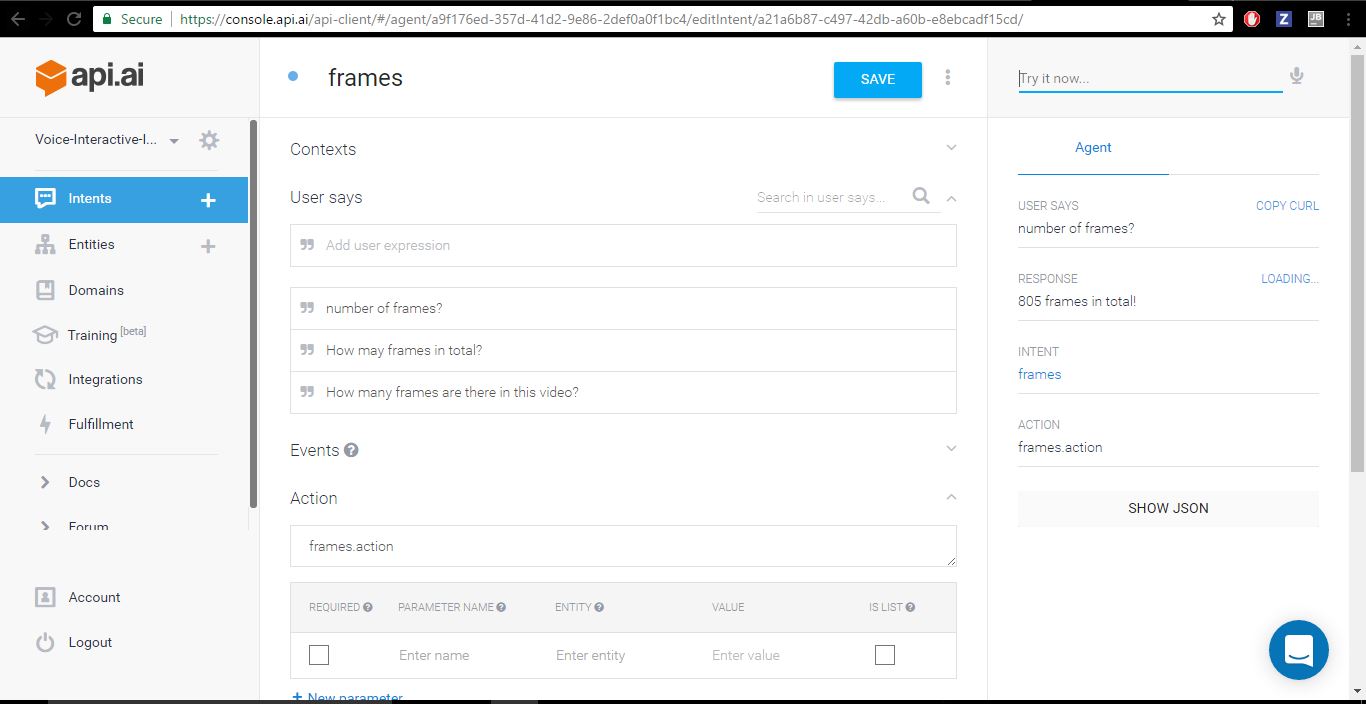


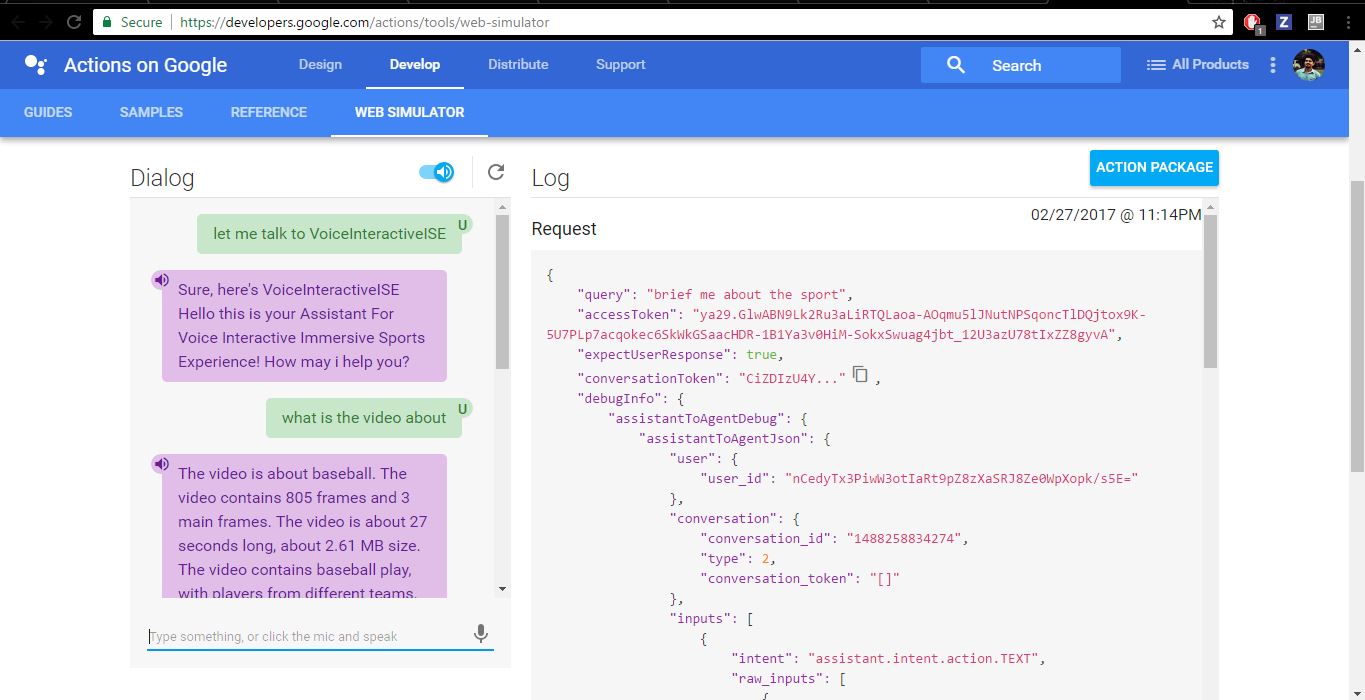
Basketball:

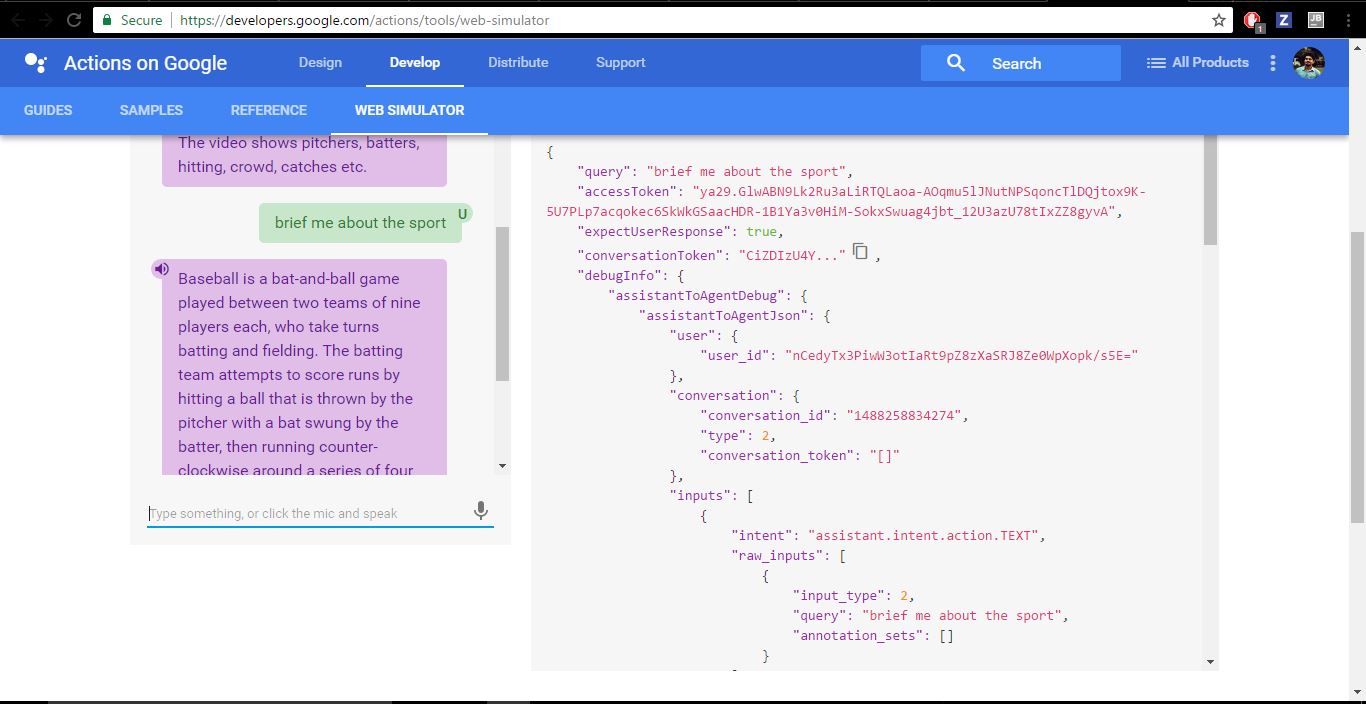


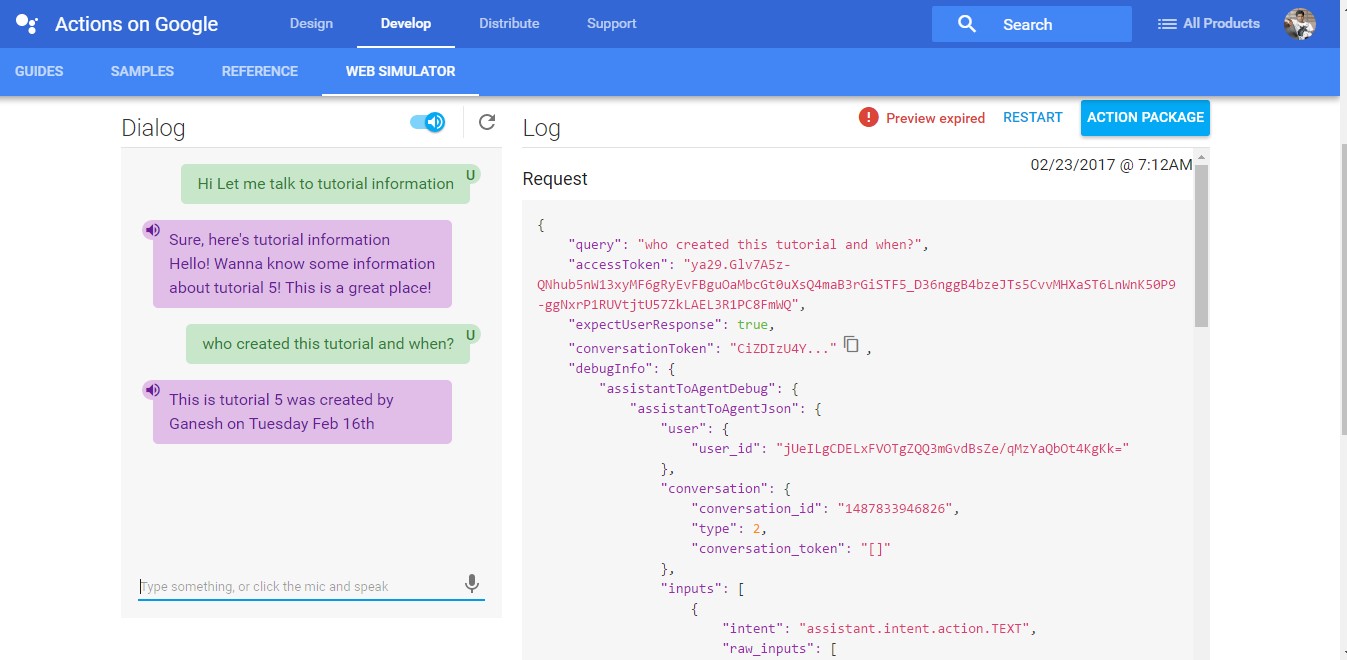
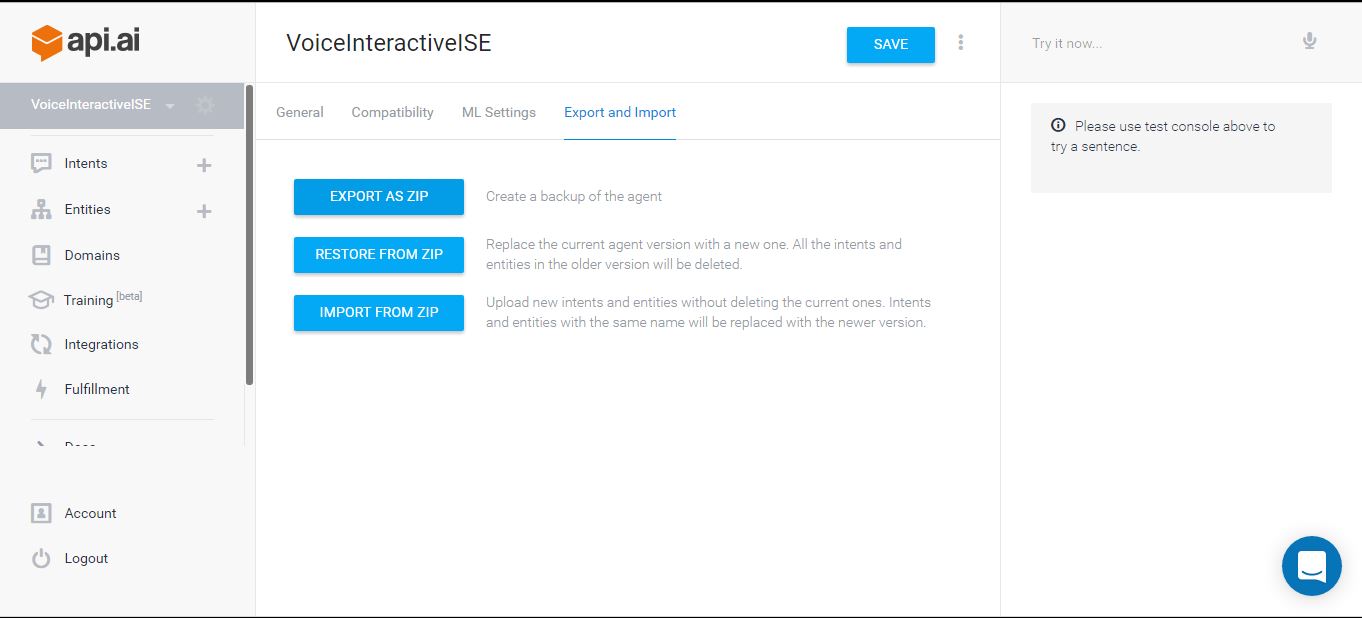
* **Screenshots of application using Spark Client API**

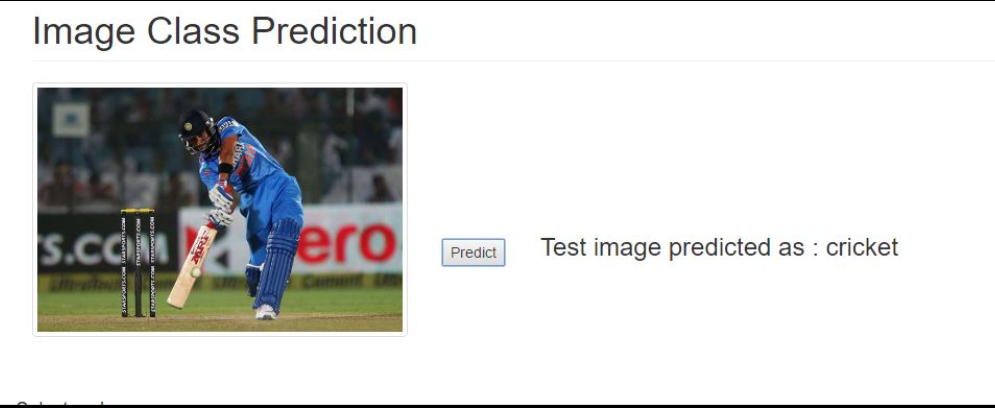




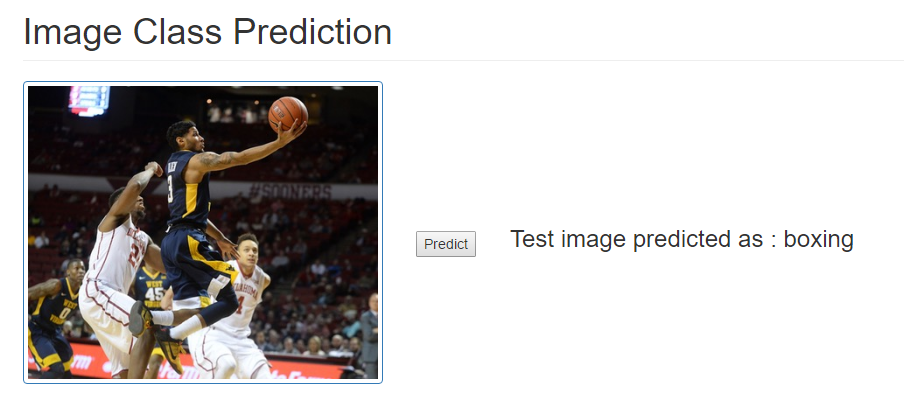




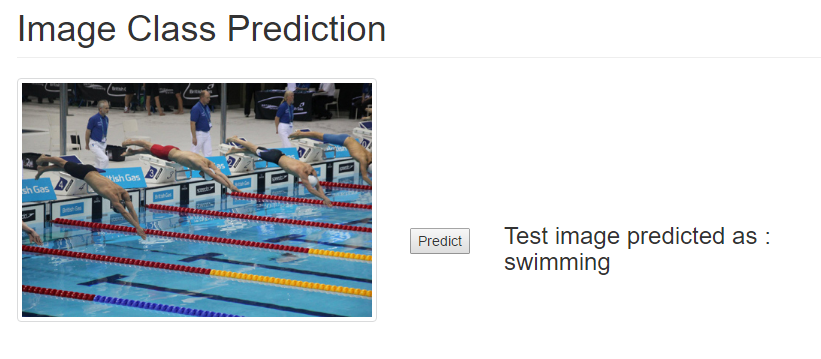




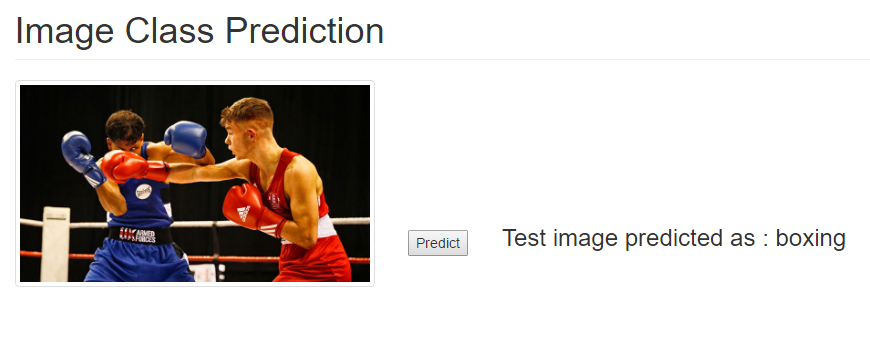
*Fig: Model predicted the cricket image as cricket*



*Fig: model predicted the basketball as Boxing*



*Fig: model predicting swimming as swimming*



* **Screenshots of application using Google conversation API**

**7. PROJECT MANAGEMENT**

**7.1 Plan & Project Timelines , Members, Task Responsibility**

**Implementation status report**

* **Work Completed:**

1. **Description:**

We finished Implementing the Clarifai API, Spark Client API and Google conversation API for phase 1. We also generated two different models for image classification using decision tree and random forest.

1. **Responsibility**

Gulnoza Khakimova: Implemented application using Clarifai API, worked on the diagrams and project report.

Ganesh Taduri: Worked on model generation using Decision tree, Random Forest and Spark Client API

Raghava Kundavajjala: Worked on Data collection/Clarifai API and Image Classification and Model Generation

Steven Evans: N/A

1. **Time taken**

Clarifai API: Around 12 hours

Spark Client API: Around 15 hours

Google Conversation API:

1. **Contributions**

Gulnoza Khakimova: 33%

Ganesh Taduri: 33%

Raghava Kundavajjala: 33%

Steven Evans: N/A

* **Work to be completed**

1. **Description**

We need to implement neural nets for Deep learning and Naive Bayes algorithms for predicting the game stats in next phases.

1. **Responsibility**

Future Phases work yet to be shared depending on the Work load.

1. **Time to be taken**

Expected time for the implementation os arou

**8. References**

[1]<http://dl.acm.org.proxy.library.umkc.edu/citation.cfm?id=2983530&CFID=894804323&CFTOKEN=31026051>

[2]<http://dl.acm.org.proxy.library.umkc.edu/citation.cfm?id=2535589&CFID=894804323&CFTOKEN=31026051>

[3]<http://dl.acm.org.proxy.library.umkc.edu/citation.cfm?id=2830309&CFID=894804323&CFTOKEN=31026051>

[4]<http://dl.acm.org.proxy.library.umkc.edu/citation.cfm?id=1466673&CFID=894804323&CFTOKEN=31026051>

[5]<https://techcrunch.com/2016/09/15/how-virtual-reality-is-transforming-the-sports-industry/>

[6]<http://www.idiap.ch/~mcernak/public/MCernak_RTT02_eng.pdf>

[7]<http://www.foxsports.com/presspass/latestnews/2016/09/13/fox-sports-and-livelikevr-team-to-present-ohio-state-at-oklahoma>

[8]<https://www.voicebot.net/>

[9]<http://store.steampowered.com/app/414120/>

[10]<https://www.researchgate.net/publication/265917403_VRIO_A_Speech_Processing_Unit_for_Virtual_Reality_and_Real-World_Scenarios_-_An_Experience_Report>

[11]<http://www.nextvr.com/news>

[12]<http://fortune.com/2016/04/29/mlb-eon-sports-vr/>

[13]<http://www.livelikevr.com/#sectionHome>

[14]<http://virtuallylive.com/>

[15] <https://docs.api.ai/docs/welcome>