(CS6630 - Fall 2016)

Data Visualization Final Project
MILESTONE 1

Stefan Hamilton
u0812842
stefhamilton@gmail.com

Anil Dhungel
U1000489
anildgl@gmail.com

TABLE OF CONTENT

Overview and Motivation:

Currently Behavior Analysts at the University of Utah observe classrooms in person or through video to evaluate behaviors of clients. Specific items are looked for such as out of seat, hitting, non-compliance, etc and the information is used to assess or change treatment plans.

Currently Behavior analysts analyze the client's behavior by going through the videos manually .Going through video is time-consuming and error prone (~70% of actual events are detected by any one observer). In an ideal situation an algorithm would process video and provide clips as well as statistics to the behavior analyst.

The goal of this project is to create the gui portion of a dashboard that allows a behavior analyst to build a model that detects an event of interest in an image while minimizing the number of images to go through.

An assumption exists that a model will be generated using the following process.

1. Tag a broad sample of images as having a hand raise or not
2. Train a model using clarifai's API on 80% of the tagged images, evaluate on 20% of the images
3. If the learning curve shows that the model is good, stop
4. Tag all images where the probabilities are between 40 and 60% chance of having a hand raise and include in train/test data
5. When the model misclassifies an image, tag images before and after the misclassification to provide more data on classification boundaries.
6. Re-build and re-evaluate model and go to step 3

By following steps 4 and 5 it is expected that more data will be collected at model boundaries, thus speeding up and lowering the cost of the model building process.

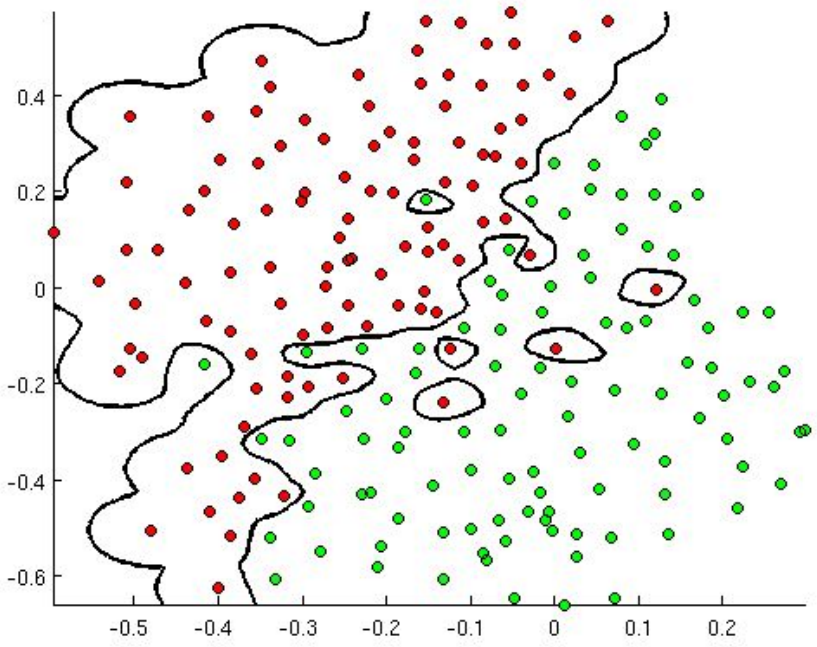


Image 1 - Scatterplot showing the classification model's result

We envision the visualization something like this on the picture above where each dot represents an activity occurrence. Color can be used to identify misclassification or correct classification.

Related Work:

The project was inspired by the need to improve the process in a real world project.Since the project has such a custom requirements there isn't much work available similar to what we are looking for.

On the backend we will use Clarifai's machine learning API (https://developer.clarifai.com/).The data will processed with Clarifai's API which will create a model for our data based on the events detected. The model's finding will be displayed as a scatter plot (See Image 1 above) .

Also, another element of visualization was inspired from our meeting discussion with our advisor Sean. We also found Metamind.io a predictive analytics software that outputs its result in the similar format what we discussed in our meeting (See Image 2 below).

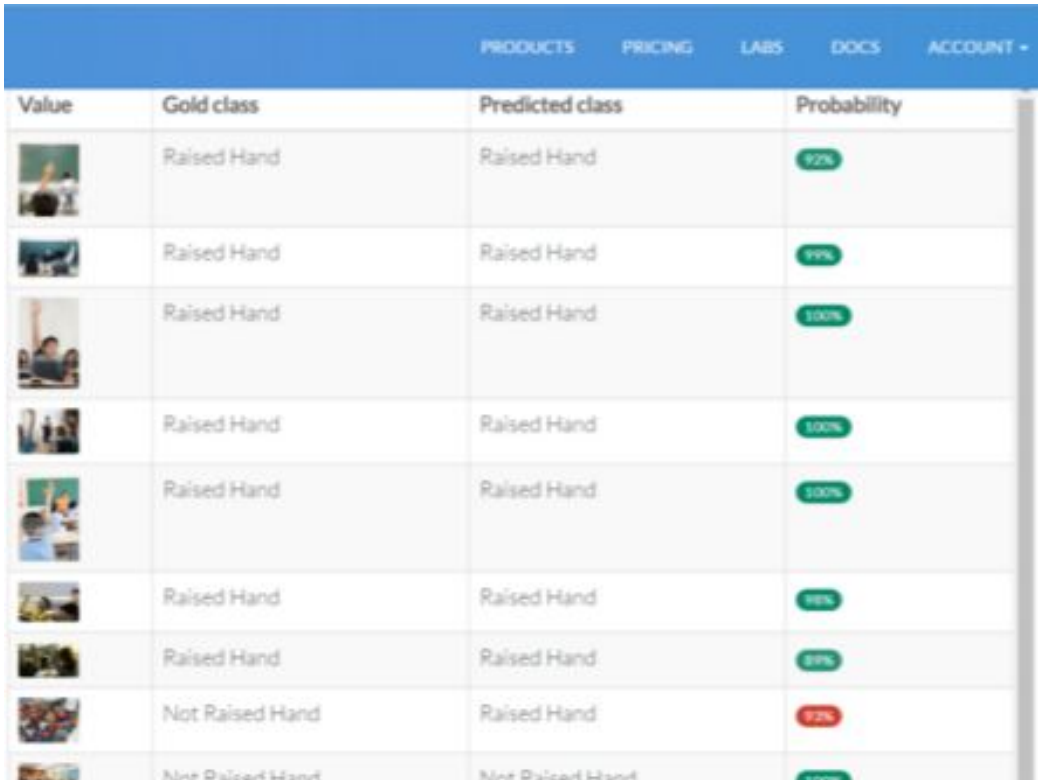


Image 2- Metamind.io result display format.

Data

The main source of data for the project is going to be video taken by the behavior analysts. These video will be processed through video editing software to get snap shot of every 10 seconds. These images will be uploaded to the machine learning application which will classify each images based on the activities detected. For this project we will be only focusing the images that consists of hand raises event. However, the machine learning algorithms can be trained to detect other events of interest from the video as well. The result obtained from the model will be stored in a csv file. The csv file with classification results along with the images created by processing the videos will be used to create our visualization.

Since the actual data is HIPPA protected we are not able to use the actual images. Therefore, images of hand raising were pulled off the internet, resized. Images were obtained using the google search. In our first attempt we ran all the images through a pre-trained neural net and a logistic regression model build from the 4096 features.

Implementation:

After running the visualization through the pre-trained neural net and a logistic regression model we realized it might not be better fit for our project due to complexity of implementation.Implementation of just the machine learning process would take large amount of our time and we would rather use that time in building front end of the project. Also, the performance of the model was not on par with other options we had seen on the internet.

In order to improve the performance and reduce the complexity of implementation we will be using Clarifai's api. Another option we might look into is the google cloud machine learning.

Another aspect of our project is provide behavior analysts ability to tag any images that are misclassified by the machine learning algorithm. The image tagging functionality will be similar to the FastPhotoTagger (http://fastphototagger.sourceforge.net/fastphototagger.html). Basically, every time a behavior analyst is tagging an image they are adding metadata to the image. This metadata information will help to improve the performance of machine learning model.

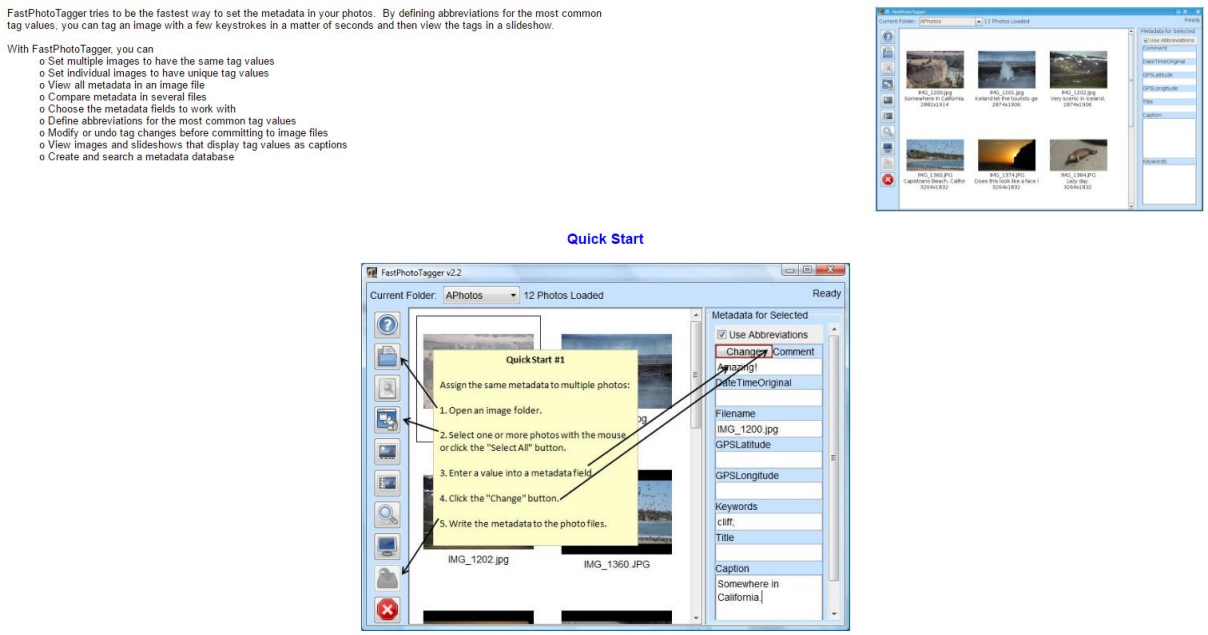


Image 3 - FastPhotoTagger tagging fuctionality

Design Evolution - Iteration 1

Based on the discussion and requirements of Behavior Analysts we started sketching few ideas.

Design 1 -

Image 4 - Design Sketch #1

Design 2 -

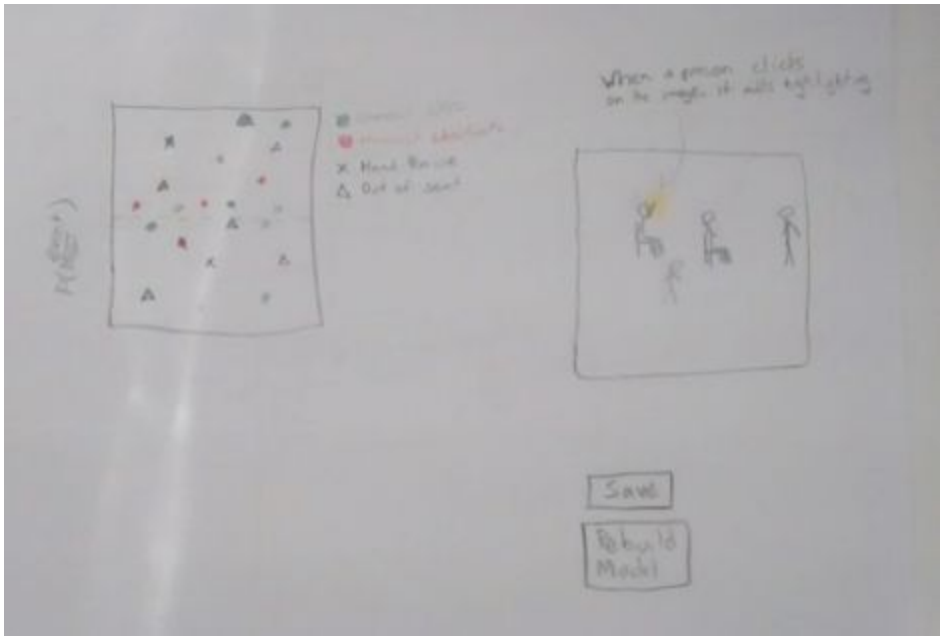


Image 5 - Design Sketch #2

Design 3 -

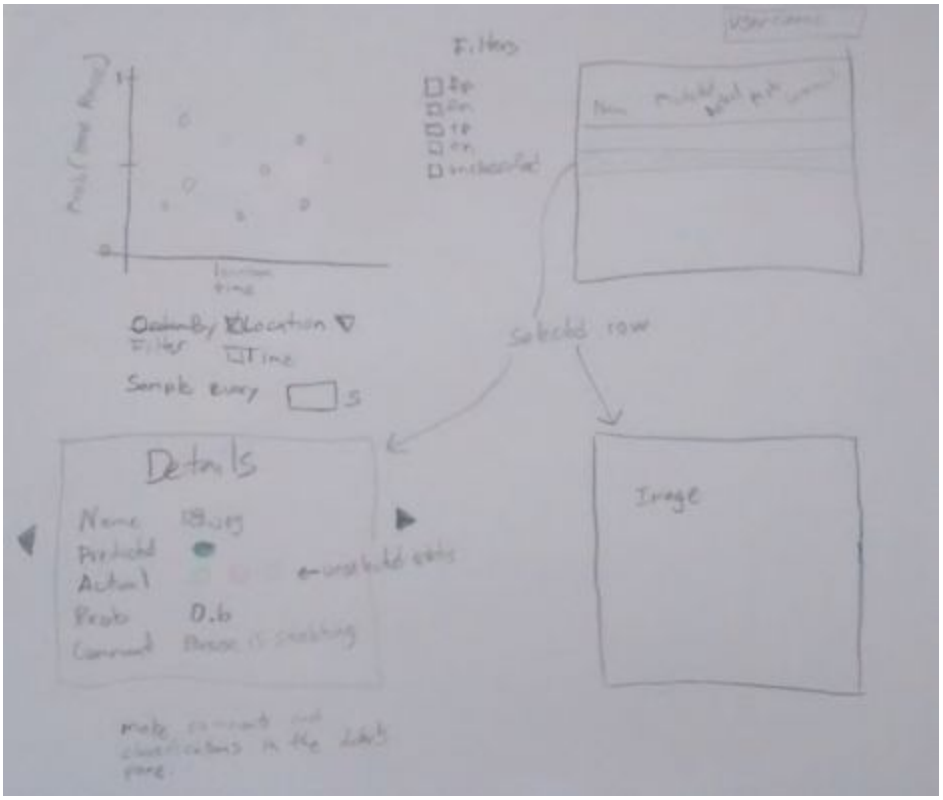


Image 6 - Design Sketch # 3

Final Design 1st iteration -

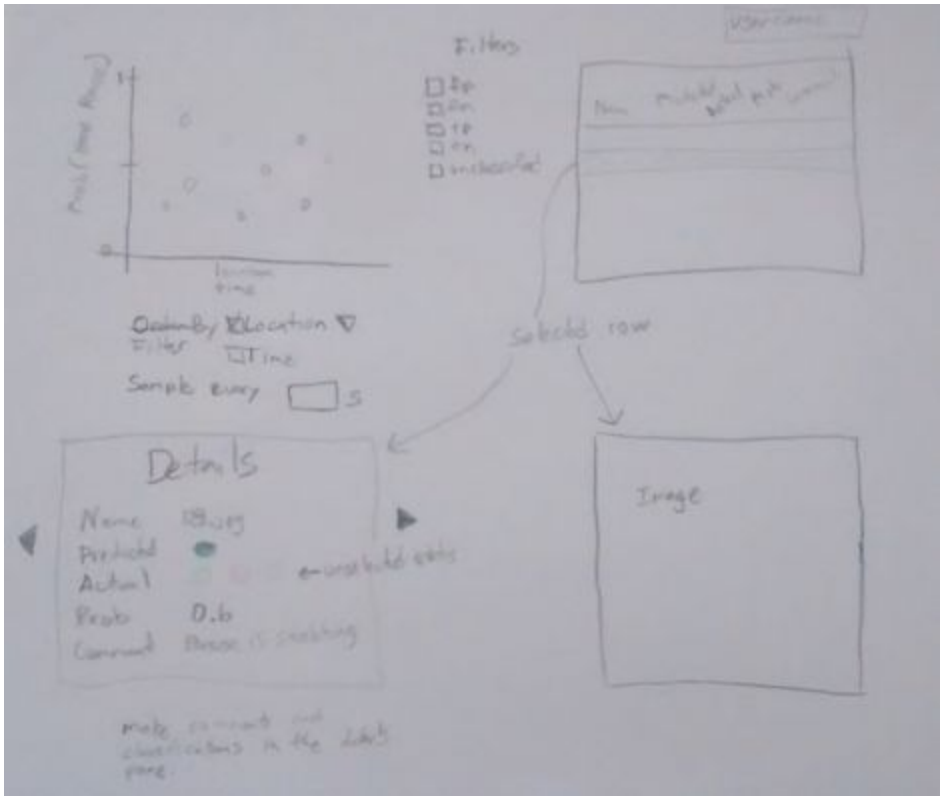


Image 7 - Final Design 1st Iteration

Based on our first iteration we plan on having four elements on our visualization.

Session view on the top left (scatter) showing the overall classifcation result of machine learning. Scatter plot with the probability on the y-axis and time of the event on the x-axis. Each dot on the plot will represent an event or an image.

Brushed view on the top right is a table with more detail info on each image which is generated by brushing the dots on the scatter plot described above. This element will have name of the image for each event along with probability , prediction and actual. Clicking on the row will generate a picture of the event on the bottom left corner.

Also clicking on the scatter plot's dot or row of the Brushed view will generate a detail view on the bottom left corner. This will contain more detail of the event.

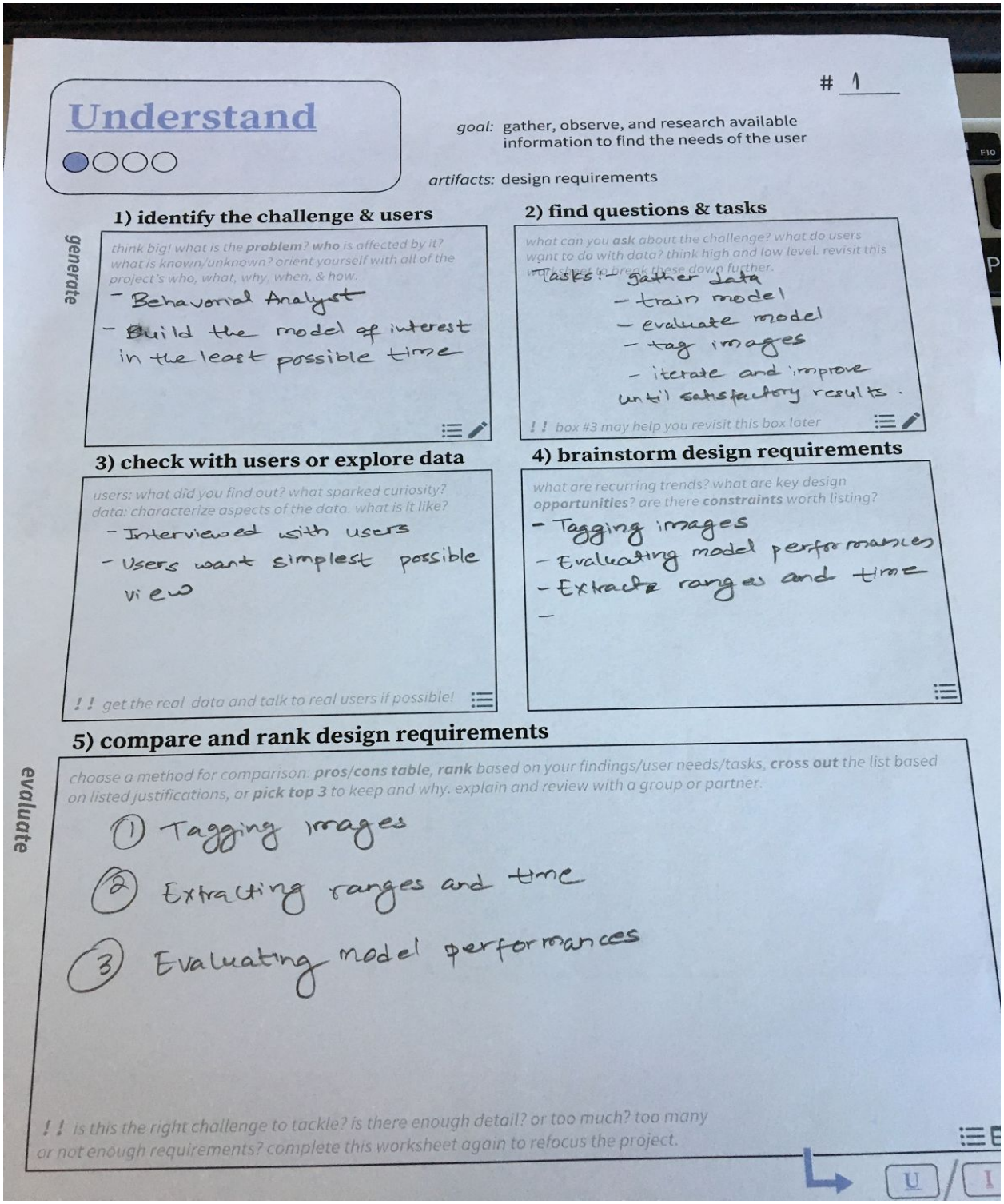
Feedback from the Behavior Analysts (Users)

At this point we decided to meet with Behavior Analysts (user) to get the feedback on the designs we had sketched out. The raised a concern that they want to be able to tag the images that were misclassified. Our current design allowed the users to explore the result and machine learning classify the event but it didn't allow the user to reclassify the images that were misclassified.

So, we started focusing on how we could redesign our visualization so the users can tag the misclassified images.

Design Evolution - Iteration 2

This time we decided to use the Design worksheet.



1

Understand

goal: gather, observe, and research available information to find the needs of the user

artifacts: design requirements

generate

1) identify the challenge & users

*think big! what is the **problem**? **who** is affected by it? what is known/unknown? orient yourself with all of the project's who, what, why, when, & how.*

- Behavorial Analyst
- Build the model of interest in the least possible time

2) find questions & tasks

*what can you **ask** about the challenge? what do users want to do with data? think high and low level. revisit this with how to break these down further.*

Tasks:
- Gather data
- train model
- evaluate model
- tag images
- iterate and improve until satisfactory results.

!! box #3 may help you revisit this box later

3) check with users or explore data

users: what did you find out? what sparked curiosity? data: characterize aspects of the data. what is it like?

- Interviewed with users
- Users want simplest possible view

!! get the real data and talk to real users if possible!

4) brainstorm design requirements

*what are recurring trends? what are key design **opportunities**? are there **constraints** worth listing?*

- Tagging images
- Evaluating model performances
- Extracte ranges and time
-

evaluate

5) compare and rank design requirements

*choose a method for comparison: **pros/cons table**, **rank** based on your findings/user needs/tasks, **cross out** the list based on listed justifications, or **pick top 3** to keep and why. explain and review with a group or partner.*

1. Tagging images
2. Extracting ranges and time
3. Evaluating model performances

!! is this the right challenge to tackle? is there enough detail? or too much? too many or not enough requirements? complete this worksheet again to refocus the project.

Image 8 - Understand Design Worksheet

Ideate

#2

goal: generate good concepts and ideas for supporting some of the project's design requirements

artifacts: ideas & sketches

generate

1) select a design requirement

how might we *address the challenge using the requirement? which questions would a user ask? revisit this worksheet for each important design requirement.*

Tagging images

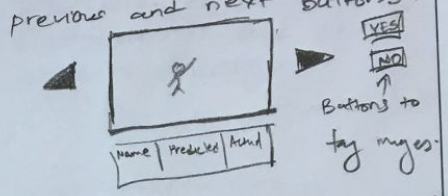
- Generated when session view is brushed

!! revisit this worksheet for all important design requirements for your project

2) sketch first idea

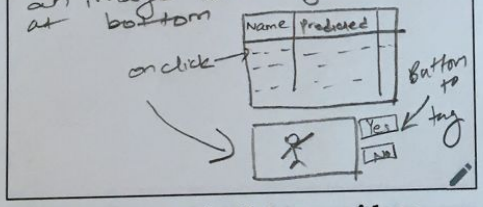
show how to address this requirement using an ***informal sketch*** *- focus on the big idea, not the fine details*

one picture at a time with previous and next buttons.



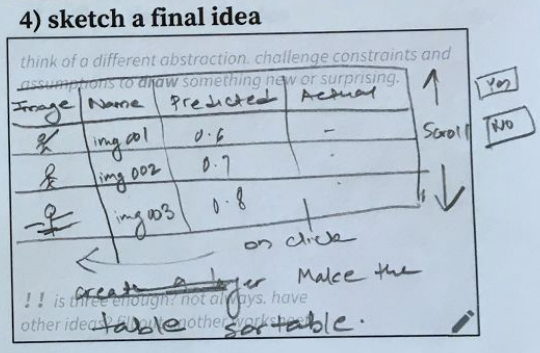
3) sketch another idea

no images, just the name of image with description. On click of the name an image will be generated at bottom



4) sketch a final idea

think of a different abstraction. challenge constraints and assumptions to draw something new or surprising.



!! is this enough? not always. have other ideas? explore other sketches

evaluate

5) compare and relate your ideas

for each sketch, break apart ***what works well (+)*** *and* ***what doesn't (-)*** *in the* ***table*** *below. make connections. reflect on best parts. can you* ***combine ideas****? review the table with a partner or group.*

sketch #1	sketch #2	sketch #3
+ Easy to identify what's going on in the image easily - only one image at a time	+ Can fit more images in table - Not able to see images without clicking	+ It has both images and name. So can see many images at once + fast and easy to read - may be technically difficult to create them o. - Might to be too compact

!! combining ideas and sketches is not easy. sometimes it may open up new possibilities and ideas - guess what, ideate again!

U / I / M

Image 9 - Ideate Design Worksheet

Make

#_____

goal: concretize ideas into tangible prototypes which are approximations of a product in some aspects

artifacts: prototypes

generate

1) set an achievable goal

*what should the prototype **achieve**? what are the specific **criteria for success**? break a larger goal into parts with clearer feature sets.*

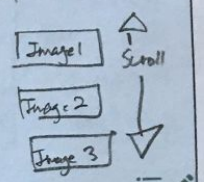
1- Create layout for the table
2- implement image table.
3. develop a working prototype

!! break a goal apart into multiple and create a worksheet for each sub-goal

2) plan encodings & layouts

*what are good visualization **encodings** or **layouts** for what you've ideated? map with and remember to justify for users and their tasks.*

Table with scrollable view
* Similar to final idea sketch in Ideate



3) plan support for interactions

***what can the user do?** what is required given the chosen encodings? **justify** your design decisions.*

* Scroll to view images
* Buttons to tag the images

4) sketching additional views

what other parts of the data must be seen? brainstorm how to show this data in the tool.

* Buttons for tagging image

!! if you are thinking up new ideas to visualize, go back to the Ideate activity!

evaluate

5) build the prototype and check-in

*are your **goals met** by the prototype? test with users if possible. are design decisions properly justified? do any need to be revisited? were any new constraints or limitations discovered? write down your progress and additional justifications below. review this progress and the prototype with a partner or your group.*

Implement table with image :- can be tricky to scale the large amount of image.

Buttons to tag the image :- where do you place the buttons.

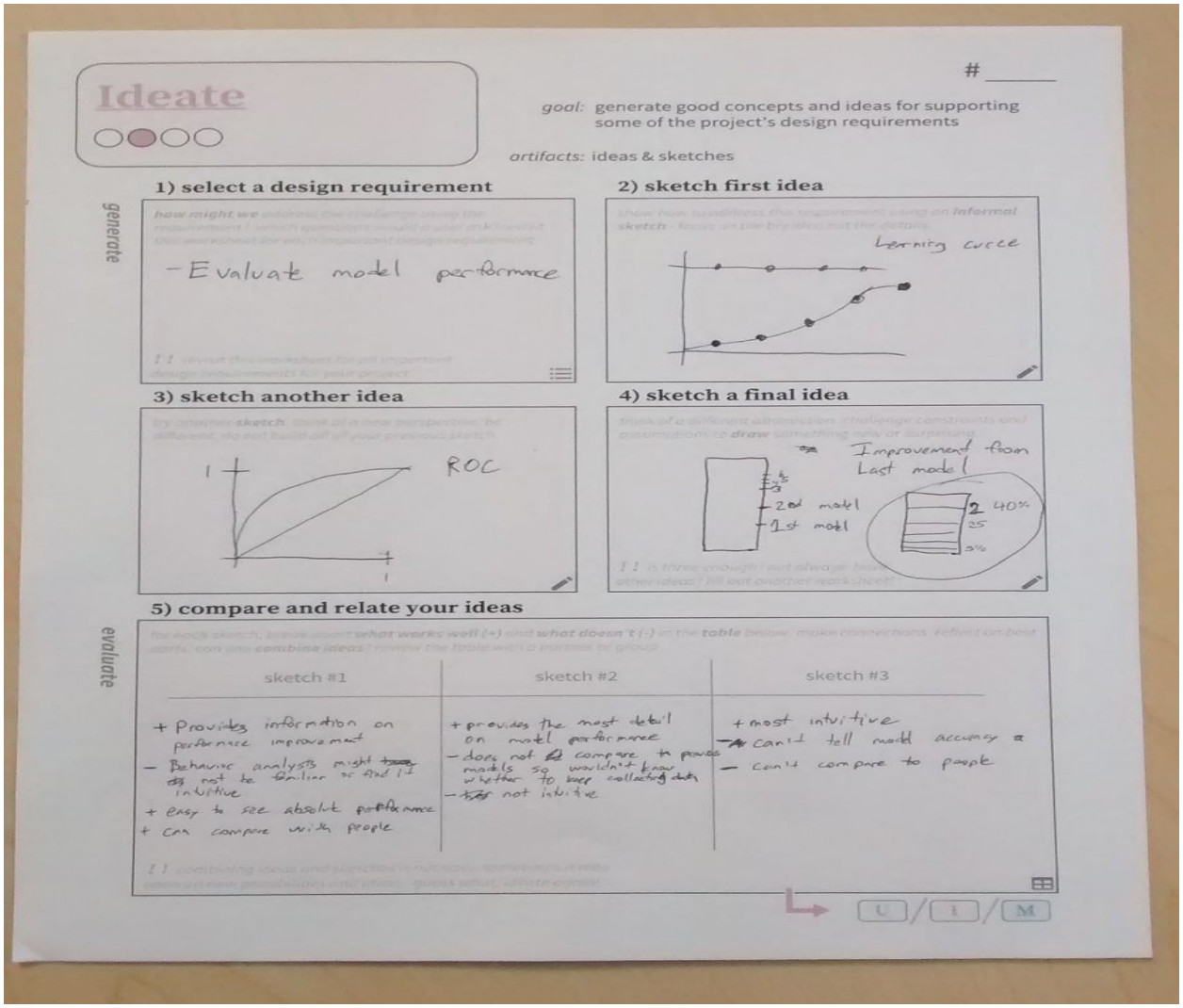
Tagg

!! did the prototype meet its goal/s? measure its success. make sure you have addressed the design requirement. does the prototype try to do too much?

U / I / M / D

Image 10 - Make Design Worksheet

Evaluate model Performance worksheets



Ideate

_____

goal: generate good concepts and ideas for supporting some of the project's design requirements

artifacts: ideas & sketches

generate

1) select a design requirement

- Evaluate model performance

2) sketch first idea

Learning curve

3) sketch another idea

ROC

4) sketch a final idea

Improvement from Last model

2nd model

1st model

2 40%

25

3%

evaluate

5) compare and relate your ideas

sketch #1	sketch #2	sketch #3
+ Provides information on performance improvement – Behavior analysts might not be familiar or find it intuitive + Easy to see absolut performance + Can compare with people	+ provides the most detail on model performance – does not compare to previous models so wouldn't know whether to keep collecting data – not intuitive	+ most intuitive – can't tell model accuracy – can't compare to people

U / I / M

Make

#_____

goal: concretize ideas into tangible prototypes which are approximations of a product in some aspects

artifacts: prototypes

generate

1) set an achievable goal

- creat ~~mock~~ UP
- have chart react to a button press (real data is not needed)

2) plan encodings & layouts

Learning curve is similar to what behavior analysts use and was performed. See sketch 2

3) plan support for interactions

Interaction with Charts or google machine learning is ideal but is complicated optional

4) sketching additional views

evaluate

5) build the prototype and check-in

U / I / M / D

Visualization Implementation

After going through the design worksheet we plan on having three main elements and one optional elements in our visualization.

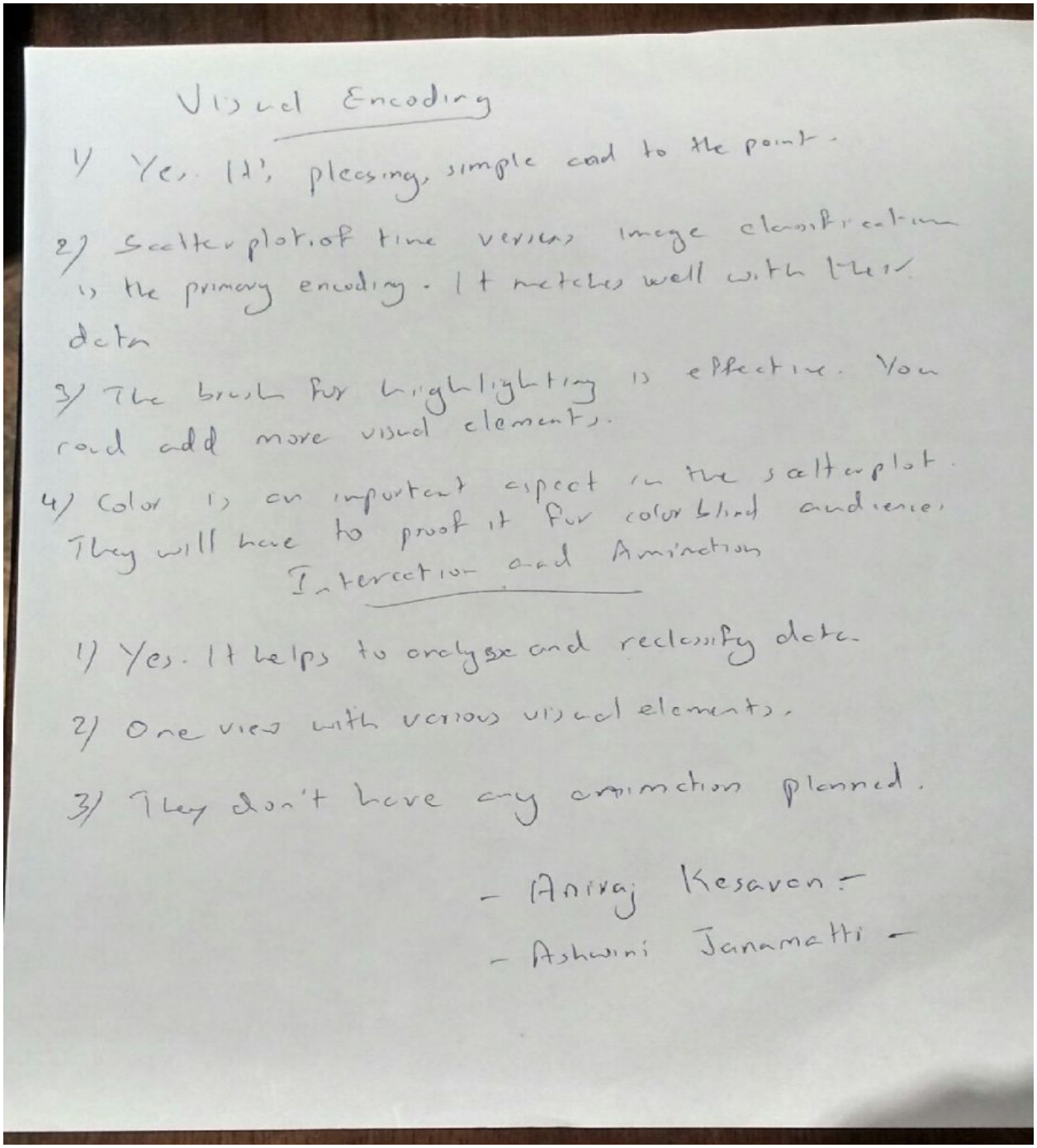
Element 1 Session View - This view will show the overall result of classification model. Still in process of what final

Element 2 Scrollable Table with images - generated based on the brush tool on the session view.

Element 3 Detail view - when an image is clicked on the Element 2

Optional Element 4 Chart to evaluate the performance of the model (ROC curve etc)

Feedback from Peer Review:



Visual Encoding

1/ Yes. It's pleasing, simple and to the point.

2/ Scatterplot of time versus image classification is the primary encoding. It matches well with the data

3/ The brush for highlighting is effective. You could add more visual elements.

4/ Color is an important aspect in the scatterplot. They will have to proof it for color blind audience.

Interaction and Animation

1/ Yes. It helps to analyse and reclassify data.

2/ One view with various visual elements.

3/ They don't have any animation planned.

- Aniruj Kesavan -
- Ashwini Janamatti -

Screenshot Current Prototype:

Automated Behavior Metrics

Session View

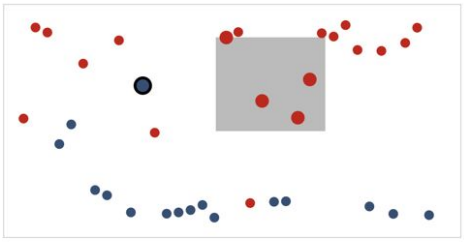


Image Detail

Prob(Hand Raise) 68.3%
Actual ToDo put in a radio button selections
ToDo support comments

Name	Predicted	Actual	Prob	Comment
images (2)_raised.jpg			0.931449	
images (18)_raised.jpg			0.603393	
images (15)_raised.jpg			0.516786	
images (22)_raised.jpg			0.714503	