1.Include a 2-3 sentence explanation for why the classifiers perform as they do on your data set.
The key is to include as many features as you possibly can because KNN does not perform well in high dimension. I set 50 attributes to be numerical and 48 of which are random samples from 1 to 50 which have nothing to do with the outputs and use the rest of two attributes to build a decision tree and determine the outputs. |J48 accuracy - Ibk accuracy| = |100% - 58.4%| = 41.6%

2. Include a 2-3 sentence explanation for why the classifiers perform as they do on your data set.
The key is to make attributes dependent because Naive Bayes assume attributes are independent. I set 17 attributes to be randomly sampled and the outputs are determined by the sum of samples. |Multilayer Perceptron - Naive Bayes| = |94.4% - 55.6%| = 38.8%

3. Consider the following four random variables, which can take values in the given sets:
Hours of Sleep: {0, 1, 2, 3, 4, 5, 6, 7, 8}
Studied: {None, Some, Lots}
LikesMaterial: {None, Some, Lots}
ExamScore: {A, B, C}
(a) How many independent parameters are needed to specify the joint distribution over these four variables? A joint distribution defined over k disjoint events requires k-1 independent parameters. (1/3pt)
There are 9 * 3 * 3 * 3 = 243 disjoint events and therefore 243 - 1 = 242 independent parameters are needed to specify the joint distribution over these four variables.

(b) Now assume you want to model the conditional distribution, P(ExamScore | HoursOfSleep, Studied, LikesMaterial). How many independent parameters does that conditional distribution contain? (1/3pt)
There are 9 * 3 * 3 * 2 = 162 independent parameters that conditional distribution contains.

(c) Now imagine modeling P(ExamScore | HoursOfSleep, Studied, LikesMaterial) with the Naïve Bayes assumption, that each of the three variables on the right-hand-side of the pipe is conditionally independent given ExamScore. How many independent parameters are needed to specify the conditional distribution under the Naïve Bayes assumption? (1/3pt)
# of I.P needed for P(HoursOfSleep|ExamScore) is 3 * 8 = 24
# of I.P needed for P(Studied|ExamScore) is 3 * 2 = 6
# of I.P needed for P(LikesMaterial|ExamScore) is 3 * 2 = 6
# of I.P needed for P(ExamScore) is 2
Total # of I.P needed under Naive Bayes assumption is 24 + 6 + 6 + 2 = 38

5. Report the similarity between each pair using VGG representation and pixel representation respectively (so you should report 6 numbers in total). Which pair is the

most similar in VGG representation and which pair is the most similar in pixel representation?

	vgg_rep	pixel_rep
mj1-mj2	**0.960011026178**	0.370861946454
mj1-cat	0.152535112667	0.473088523579
mj2-cat	0.141834212136	**0.619743990598**

The pair that has highest similarity in a representation is bolded.

6. Based on the result in #5 above, what problem does pixel representation have and why does the convolutional neural network fix it?

When comparing two vectors, we count any mismatched entry as a difference, but pixel representation does not tell us the relative position of each part of an image. So if two images that have the same figure but one is slightly shifted from the position of the other, pixel representation will tell you the two images are way different by generating two way different vectors. CNNs does well on comparing images piece by piece or feature by feature. CNNs do way better on looking for similarity than whole-image matching techniques such as pixel_rep by finding rough feature matched in roughly the same positions in two images.

7. Compare the two results from VGG representation and pixel representation – is one better than the other? Justify your answer in 1-2 sentences.

VGG representation is better than Pixel representation because VGG rep sees an image composed by different components whereas Pixel rep sees an image composed by different pixels. A matched pixel is way less likely to happen than a matched component for the comparison of two images.

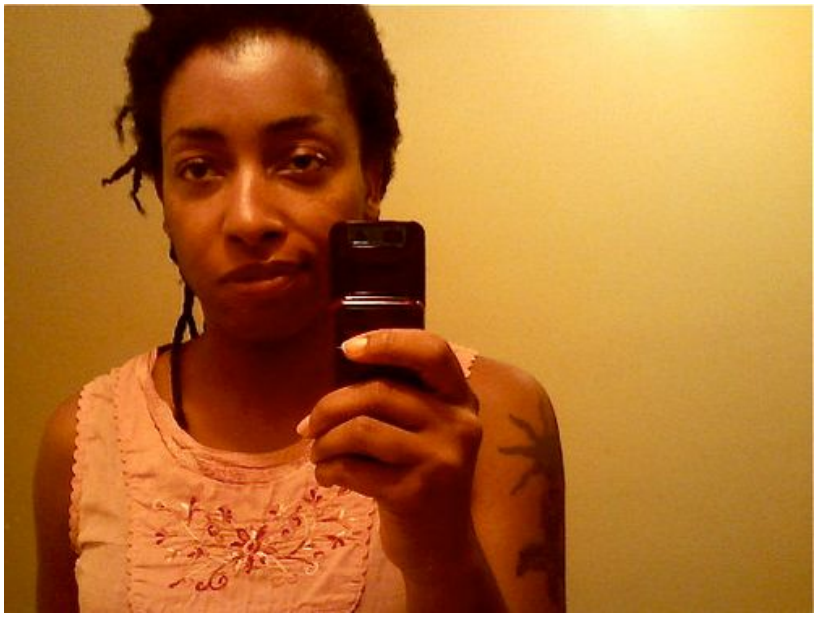
8. A machine learning model is usually not perfect, and analyzing its errors can usually tell us more about how the model works. From the captions that doesn't describe the image well, pick one example that uses VGG representation and one that uses pixel representation. Include the image and the generated caption, and describe why you think it generates that wrong caption. (1 pt)

VGG:

Given this picture

What VGG tells us this following picture is the most similar one

What I think the following picture is the most similar one instead



The reason is VGG cannot interpret what people are holding cellphones for. VGG only captures the action “holding a cellphone” whereas it fails to capture the action “taking pictures with a cellphone”

Pixel:
Given this picture:

What Pixel rep thinks the following picture is the most similar one

Needless to say, this picture is not the most similar one given the picture. I am surprised by Pixel captures the action “holding” which is just a coincident. The reason comes from there are tons of pixels matched on the top-left corner for the two pictures.