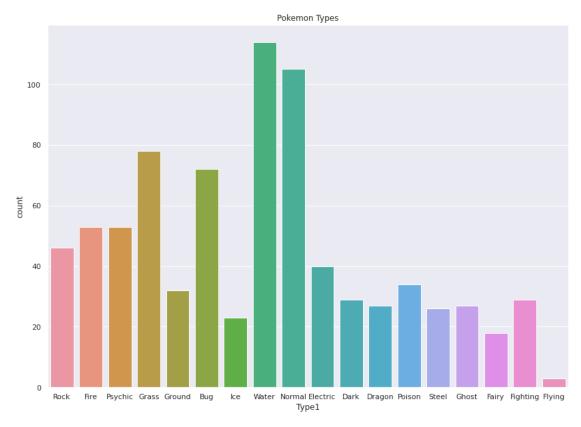
Pokemon Final Project.Ftrl

August 14, 2022

```
[164]: from google.colab import drive
       drive.mount('/content/drive')
      Drive already mounted at /content/drive; to attempt to forcibly remount, call
      drive.mount("/content/drive", force_remount=True).
[165]: import numpy as np
       import pandas as pd
       import seaborn as sns
       from matplotlib import pyplot as plt
       import matplotlib.image as mpimg
       import zipfile
       import os
       from pathlib import Path
       import re
       import tensorflow as tf
       from tensorflow import keras
       from keras import layers
       from PIL import Image
       import plotly.express as px
       import glob
[166]: data = pd.read_csv('/content/pokemon.csv')
[167]: data = data.sort_values(by=['Name'], ascending=True).reset_index(drop=True)
[168]: directory = r'/content/drive/MyDrive/images/images/'
       for filename in os.listdir(directory):
           if filename.endswith('.jpg'):
               prefix = filename.split('.jpg')[0]
               os.rename(os.path.join(directory, filename), os.path.join(directory,
        →prefix+'.png'))
           else:
               continue
[169]: train_dir = r'/content/drive/MyDrive/images/images/'
       train_path = Path(train_dir)
```

```
[170]: files = list(train_path.glob('*.png'))
       name = [os.path.split(x)[1] for x in list(train_path.glob('*.png'))]
       pokemon_image_df = pd.concat([pd.Series(name, name='Name'), pd.Series(files,__
        pokemon_image_df['Name'] = pokemon_image_df['Name'].apply(lambda x: re.sub(r'\.
        \hookrightarrow \backslash w+\$', '', x)
       pokemon_image_df
[170]:
                                                                        Filepath
                       Name
       0
                              /content/drive/MyDrive/images/images/aurorus.png
                    aurorus
       1
                   arcanine
                             /content/drive/MyDrive/images/images/arcanine.png
       2
                       abra
                                  /content/drive/MyDrive/images/images/abra.png
       3
                 barbaracle
                             /content/drive/MyDrive/images/images/barbaracl...
       4
                             /content/drive/MyDrive/images/images/amoonguss...
                  amoonguss
                             /content/drive/MyDrive/images/images/vikavolt.png
       804
                   vikavolt
                             /content/drive/MyDrive/images/images/wishiwash...
       805
            wishiwashi-solo
                              /content/drive/MyDrive/images/images/yungoos.png
       806
                    yungoos
       807
                             /content/drive/MyDrive/images/images/type-null...
                  type-null
                              /content/drive/MyDrive/images/images/zeraora.png
       808
                    zeraora
       [809 rows x 2 columns]
[171]: #list(train_path.qlob('*.pnq'))
       comb_df = pokemon_image_df.merge(data, on='Name')
       comb_df = comb_df.drop(['Name','Type2'], axis=1)
       comb_df
[171]:
                                                      Filepath
                                                                    Type1
       0
             /content/drive/MyDrive/images/images/aurorus.png
                                                                    Rock
       1
            /content/drive/MyDrive/images/images/arcanine.png
                                                                    Fire
       2
                /content/drive/MyDrive/images/images/abra.png
                                                                 Psychic
       3
            /content/drive/MyDrive/images/images/barbaracl...
                                                                  Rock
            /content/drive/MyDrive/images/images/amoonguss...
       4
                                                                 Grass
       804
            /content/drive/MyDrive/images/images/vikavolt.png
                                                                     Bug
            /content/drive/MyDrive/images/images/wishiwash...
       805
                                                                 Water
             /content/drive/MyDrive/images/images/yungoos.png
       806
                                                                  Normal
       807
            /content/drive/MyDrive/images/images/type-null...
                                                                Normal
       808
             /content/drive/MyDrive/images/images/zeraora.png
                                                                Electric
       [809 rows x 2 columns]
[172]: data['Type1'].unique()
```



```
path = r'/content/drive/MyDrive/images/images/'
fig,(ax1, ax2, ax3, ax4, ax5, ax6, ax7, ax8) = plt.subplots(1, 8, figsize=(15, u + 10))
ax = [ax1, ax2, ax3, ax4, ax5, ax6, ax7, ax8]
for i in range(8):
    img = mpimg.imread(path+data['Name'][i**3]+'.png', 0)
    ax[i].imshow(img)
    ax[i].set_title(data['Name'][i**3])
    ax[i].axis('off')
plt.tight_layout()
plt.show()
```

abomasnow abra alakazam audino braixen













cosmog





klefki

```
[175]: type_colors = [
       "#8ED752",
       "#F95643",
       "#53AFFE",
       "#C3D221",
       "#BBBDAF",
       "#AD5CA2",
       "#F8E64E",
       "#FOCA42",
       "#F9AEFE",
       "#A35449",
       "#FB61B4",
       "#CDBD72",
       "#7673DA",
       "#66EBFF",
       "#8B76FF",
       "#8E6856",
       "#C3C1D7",
       "#75A4F9"]
[176]: pokemon_types=data['Type1'].unique();
         pokemon_colors=dict(zip(pokemon_types,type_colors))
[177]: data['Type1'].value_counts()
[177]: Water
                    114
       Normal
                    105
       Grass
                    78
       Bug
                    72
       Fire
                    53
       Psychic
                    53
       Rock
                    46
       Electric
                    40
                    34
       Poison
       Ground
                    32
       Dark
                    29
       Fighting
                    29
       Dragon
                    27
       Ghost
                    27
```

```
Steel
                    26
       Ice
                    23
       Fairy
                    18
                     3
       Flying
       Name: Type1, dtype: int64
[178]: | # px.density_heatmap(comb_df, x="Type1", marginal_x="histogram", title='Pokemonu
        → Types')
[179]: | firewater = comb_df.query("Type1 == 'Fire' | Type1 == 'Water'")
       firewater
       print("Number of Fire Types:", len(firewater[firewater['Type1'] == 'Fire']))
       print("Number of Water Types:", len(firewater[firewater['Type1'] == 'Water']))
      Number of Fire Types: 53
      Number of Water Types: 114
[180]: | # grasswater = comb_df.query("Type1 == 'Grass' | Type1 == 'Water'")
       # grasswater
       # print("Number of Fire Types:", len(grasswater[grasswater['Type1'] ==__
       # print("Number of Water Types:", len(grasswater[grasswater['Type1'] ==_

    'Water'7))

[181]: # import os
       # from shutil import copyfile
       # os.mkdir('train/')
       # os.mkdir('test/')
       # os.mkdir('val/')
       # for class_ in firewater['Type1'].unique():
             os.mkdir('train/'+str(class_)+'/')
             os.mkdir('test/'+str(class )+'/')
             os.mkdir('val/'+str(class_)+'/')
[182]: # from sklearn.model_selection import train_test_split
       # X_train, X_test, y_train, y_test = train_test_split(
             firewater, firewater['Type1'], test_size=0.33, stratify=firewater['Type1'])
       # X_test, X_val, y_test, y_val = train_test_split(
             X_test, y_test, test_size=0.33,stratify=y_test)
[183]: # from shutil import copyfile, copy2
       # for image, type_ in zip(X_train['Filepath'], y_train):
             copy2(image, 'train/'+type_)
```

```
# for image, type_ in zip(X_test['Filepath'], y_test):
             copy2(image, 'test/'+type_)
       # for image, type_ in zip(X_val['Filepath'], y_val):
             copy2(image, 'val/'+type_)
[184]: # from keras.preprocessing.image import ImageDataGenerator
       # datagen = ImageDataGenerator(rescale=1./255)
       # train = datagen.flow_from_directory('train/',
             target size=(120, 120),
             color mode='rqba',
             class_mode='sparse',
             shuffle=True,
             seed=33)
       # test = datagen.flow_from_directory('test/',
             target_size=(120, 120),
       #
             color_mode='rqba',
             class_mode='sparse',
             shuffle=True,
             seed=33)
       # val = datagen.flow_from_directory('val/',
             target_size=(120, 120),
       #
             color mode='rqba',
             class_mode='sparse',
       #
             shuffle=True,
             seed=33)
[185]: # from keras.models import Sequential
       # from keras.layers import Conv2D, MaxPooling2D, Dropout, Flatten, Dense, u
       →Activation, BatchNormalization, Lambda
       # from keras.preprocessing.image import ImageDataGenerator
       # def build():
             model = Sequential()
       #
             IMAGE_WIDTH = 120
       #
             IMAGE\_HEIGHT = 120
             IMAGE_CHANNELS = 4
             model.add(Lambda(lambda x: x, input_shape=(IMAGE_WIDTH, IMAGE_HEIGHT, __
        → IMAGE_CHANNELS)))
             model.add(Conv2D(64, (4, 4), activation='relu'))
       #
             model.add(BatchNormalization())
       #
            model.add(MaxPooling2D(pool size=(3, 3)))
```

#

model.add(Dropout(0.25))

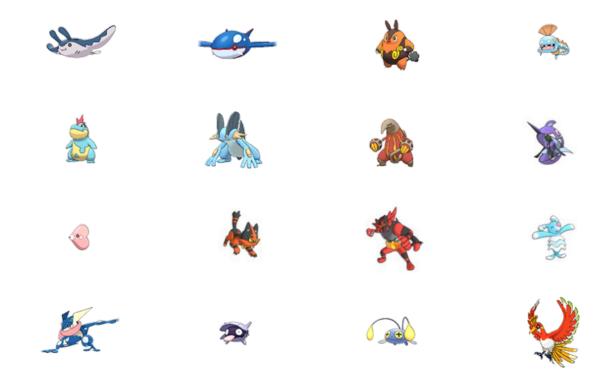
```
model.add(Conv2D(128, (4, 4), activation='relu'))
       #
             model.add(BatchNormalization())
       #
             model.add(MaxPooling2D(pool_size=(3, 3)))
             model.add(Dropout(0.25))
       #
             model.add(Conv2D(256, (4, 4), activation='relu'))
             model.add(BatchNormalization())
             model.add(MaxPooling2D(pool_size=(3, 3)))
       #
             model.add(Dropout(0.25))
             model.add(Flatten())
             model.add(Dense(512, activation='relu'))
       #
             model.add(BatchNormalization())
             model.add(Dropout(0.5))
             model.add(Dense(1, activation='softmax'))
       # # If you have a multi-class classification problem = there is only one "right" _{\sqcup}
        →answer" = the outputs are mutually exclusive, then use a softmax function. ⊔
        → The softmax will enforce that the sum of the probabilities of your output
        ⇔classes are equal to one, so in order to increase the probability of a⊔
        \hookrightarrowparticular class, your model must correspondingly decrease the probability\sqcup
        →of at least one of the other classes. Example: classifying images from the
        →MNIST data set of handwritten digits. A single picture of a digit has only
        one true identity - the picture cannot be a 7 and an 8 at the same time.
       # # "Will there be any pictures that have BOTH cat and dog in the same picture?
        →" If the answer is yes (or unknown), use sigmoid. If the answer is no, use
        \hookrightarrowsoftmax.
             model.compile(loss='binary_crossentropy', optimizer='adam',_
        →metrics=['acc', keras.metrics.AUC()])
       # # https://keras.io/api/callbacks/early_stopping/
             model.summary()
             return model
       # model = build()
       # history = model.fit(train, epochs=30, validation data=val,
           callbacks=[keras.callbacks.EarlyStopping(monitor='val_loss',patience=3,
               restore_best_weights=True), keras.callbacks.ReduceLROnPlateau()])
[186]: #shuffle the data
       firewater = firewater.sample(frac=1).reset_index(drop=True)
       train_gen = keras.preprocessing.image.ImageDataGenerator(
           validation_split=0.2,
                                  # split the dataset into a training set and a_{\sqcup}
        ⇔validation set in an 8:2 ratio
```

```
[187]: train_data = train_gen.flow_from_dataframe(
           firewater,
           x_col='Filepath',
           y_col='Type1',
           target_size=(120, 120),
           color_mode='rgba',
           class_mode='sparse',
           batch size=32,
           subset='training'
       val_data = train_gen.flow_from_dataframe(
           firewater,
           x_col='Filepath',
           y_col='Type1',
           target_size=(120, 120),
           color_mode='rgba',
           class_mode='sparse',
           batch_size=32,
           subset='validation'
       )
```

Found 134 validated image filenames belonging to 2 classes. Found 33 validated image filenames belonging to 2 classes.

```
[188]: sample_image = train_data.next()[0]

plt.figure(figsize=(16,10))
for i in range(16):
    plt.subplot(4, 4, i + 1)
    plt.imshow(sample_image[i, :, :, :])
    plt.axis('off')
plt.show()
```



[190]: # feature_extractor.summary()

```
x = layers.MaxPool2D()(x)
x = layers.Conv2D(filters=128, kernel_size=(9, 9), activation='relu')(x)
x = layers.MaxPool2D()(x)
x = layers.Conv2D(filters=256, kernel_size=(9, 9), activation='relu')(x)
x = layers.MaxPool2D()(x)
x = layers.Flatten()(x)
x = layers.Dense(512, activation='relu')(x)
x = layers.Dropout(0.5)(x)
output = layers.Dense(units=1, activation='sigmoid')(x)
model = keras.Model(inputs=inputs, outputs=output)
model.compile(
    optimizer='Ftrl',
    loss='binary_crossentropy',
   metrics=['acc', keras.metrics.AUC()]
# print model layers
model.summary()
```

Model: "model_6"

Layer (type)	Output Shape	 Param #
input_7 (InputLayer)	[(None, 120, 120, 4)]	0
conv2d_18 (Conv2D)	(None, 112, 112, 64)	20800
<pre>max_pooling2d_18 (MaxPoolin g2D)</pre>	(None, 56, 56, 64)	0
conv2d_19 (Conv2D)	(None, 48, 48, 128)	663680
<pre>max_pooling2d_19 (MaxPoolin g2D)</pre>	(None, 24, 24, 128)	0
conv2d_20 (Conv2D)	(None, 16, 16, 256)	2654464
<pre>max_pooling2d_20 (MaxPoolin g2D)</pre>	(None, 8, 8, 256)	0
flatten_6 (Flatten)	(None, 16384)	0

```
dense_12 (Dense)
                      (None, 512)
    dropout_6 (Dropout)
                       (None, 512)
                       (None, 1)
    dense 13 (Dense)
                                        513
    Total params: 11,728,577
    Trainable params: 11,728,577
    Non-trainable params: 0
[192]: history = model.fit(
       train_data,
       validation_data=val_data,
       epochs=30,
       callbacks=[
         keras.callbacks.EarlyStopping(
            monitor='val_loss',
            patience=3,
            restore_best_weights=True
         ),
         keras.callbacks.ReduceLROnPlateau()
       ]
    )
    Epoch 1/30
    - auc_6: 0.4894 - val_loss: 0.6927 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
    0.0010
    Epoch 2/30
    - auc_6: 0.5000 - val_loss: 0.6924 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
    0.0010
    Epoch 3/30
    - auc_6: 0.5000 - val_loss: 0.6922 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
    0.0010
    Epoch 4/30
    - auc_6: 0.5000 - val_loss: 0.6919 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
    0.0010
    Epoch 5/30
    - auc_6: 0.5000 - val_loss: 0.6917 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
    0.0010
    Epoch 6/30
```

8389120

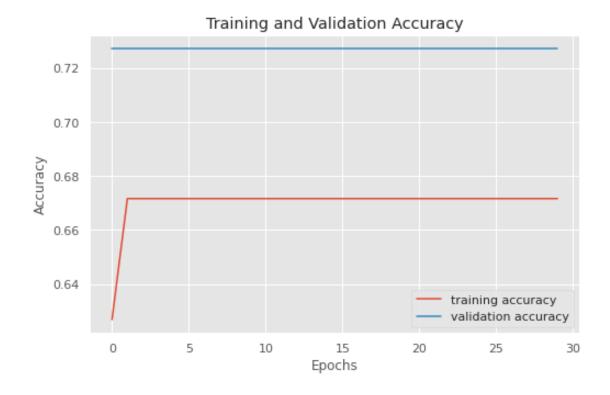
```
- auc_6: 0.5000 - val_loss: 0.6915 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 7/30
- auc_6: 0.5000 - val_loss: 0.6913 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 8/30
- auc_6: 0.5000 - val_loss: 0.6912 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 9/30
- auc_6: 0.5000 - val_loss: 0.6910 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 10/30
- auc_6: 0.4995 - val_loss: 0.6908 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 11/30
- auc_6: 0.5000 - val_loss: 0.6907 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 12/30
- auc_6: 0.5000 - val_loss: 0.6906 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 13/30
- auc_6: 0.5000 - val_loss: 0.6905 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 14/30
- auc_6: 0.5000 - val_loss: 0.6904 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 15/30
- auc_6: 0.5000 - val_loss: 0.6903 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 16/30
- auc_6: 0.5000 - val_loss: 0.6902 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 17/30
- auc_6: 0.5000 - val_loss: 0.6900 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 18/30
```

```
- auc_6: 0.5000 - val_loss: 0.6899 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 19/30
- auc_6: 0.5000 - val_loss: 0.6898 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 20/30
- auc_6: 0.5000 - val_loss: 0.6897 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 21/30
- auc_6: 0.5000 - val_loss: 0.6896 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 22/30
- auc_6: 0.5000 - val_loss: 0.6895 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 23/30
- auc_6: 0.5000 - val_loss: 0.6894 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 24/30
- auc_6: 0.5000 - val_loss: 0.6893 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 25/30
- auc_6: 0.5000 - val_loss: 0.6892 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 26/30
- auc_6: 0.5000 - val_loss: 0.6891 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 27/30
- auc_6: 0.5000 - val_loss: 0.6891 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 28/30
- auc_6: 0.5000 - val_loss: 0.6890 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 29/30
5/5 [============= ] - 46s 9s/step - loss: 0.6900 - acc: 0.6716
- auc_6: 0.5000 - val_loss: 0.6889 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr:
0.0010
Epoch 30/30
```

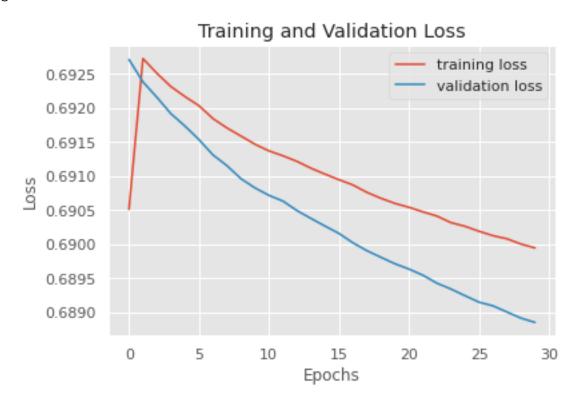
```
- auc_6: 0.5000 - val_loss: 0.6889 - val_acc: 0.7273 - val_auc_6: 0.5000 - lr: 0.0010
```

```
[193]: plt.style.use('ggplot')
       # retrieve accuracy history on training and validation data
       acc = history.history['acc']
       val_acc = history.history['val_acc']
       # retrieve loss history on training and validation data
       loss = history.history['loss']
       val_loss = history.history['val_loss']
       # get number of epochs
       epochs = range(len(acc))
       plt.figure(figsize=(8, 5))
       # plot training and validation accuracy per epoch
       plt.plot(epochs, acc, label='training accuracy')
       plt.plot(epochs, val_acc, label='validation accuracy')
       plt.title('Training and Validation Accuracy')
       plt.xlabel('Epochs')
       plt.ylabel('Accuracy')
      plt.legend()
       plt.figure(figsize=(8, 5))
       # plot training and validation loss per epoch
       plt.figure()
       plt.plot(epochs, loss, label='training loss')
       plt.plot(epochs, val_loss, label='validation loss')
       plt.title('Training and Validation Loss')
       plt.xlabel('Epochs')
       plt.ylabel('Loss')
       plt.legend()
```

[193]: <matplotlib.legend.Legend at 0x7fc293827050>



<Figure size 576x360 with 0 Axes>



```
[194]: # get true labels
      true_labels = val_data.labels
      # get predictions in the form of probablities
      predictions = model.predict(val_data)
      # convert probablities into binary values
      predictions = [1 \text{ if } n \ge 0.5 \text{ else } 0 \text{ for } n \text{ in predictions}]
      print("Model predictions: "+str(predictions))
      print("Actual labels:
                                "+str(true labels))
      # determine filepaths of misclassified pokemon
      num_misclasssified = 0
      misclassified_filepaths = []
      correctness = []
      for pred, label, i in zip(predictions, true_labels, range(len(predictions))):
        misclassified_filepaths.append(val_data.filepaths[i])
        if pred != label:
          correctness.append('incorrect')
          num misclasssified += 1
        else:
          correctness.append('correct')
      print("# of misclassified pokemon: "+str(num_misclasssified))
      1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1]
                     [0, 1, 1, 1, 1, 1, 1, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 1,
      Actual labels:
      1, 1, 1, 1, 1, 0, 1, 0, 0, 1, 0, 0]
      # of misclassified pokemon: 9
[195]: | # obtain the images from the filepath at the determined indices
      misclassified_imgs = []
      for filepath in misclassified_filepaths:
        misclassified_imgs.append(mpimg.imread(filepath, 0))
      # plot results
      f, axarr = plt.subplots(6,5, figsize=(20,10))
      count = 0
      for r in range(6):
        for c in range(5):
          axarr[r,c].imshow(misclassified_imgs[count])
          if correctness[count] == 'correct':
            axarr[r,c].set_title(correctness[count])
          else:
```

```
axarr[r,c].set_title(correctness[count], color='red')
axarr[r,c].set_axis_off()
count += 1
plt.show()
```

