

2020



Made by : **doha essam ahmed zaki**
B.Sc. Computer engineer

[TASKS DOCUMENT]

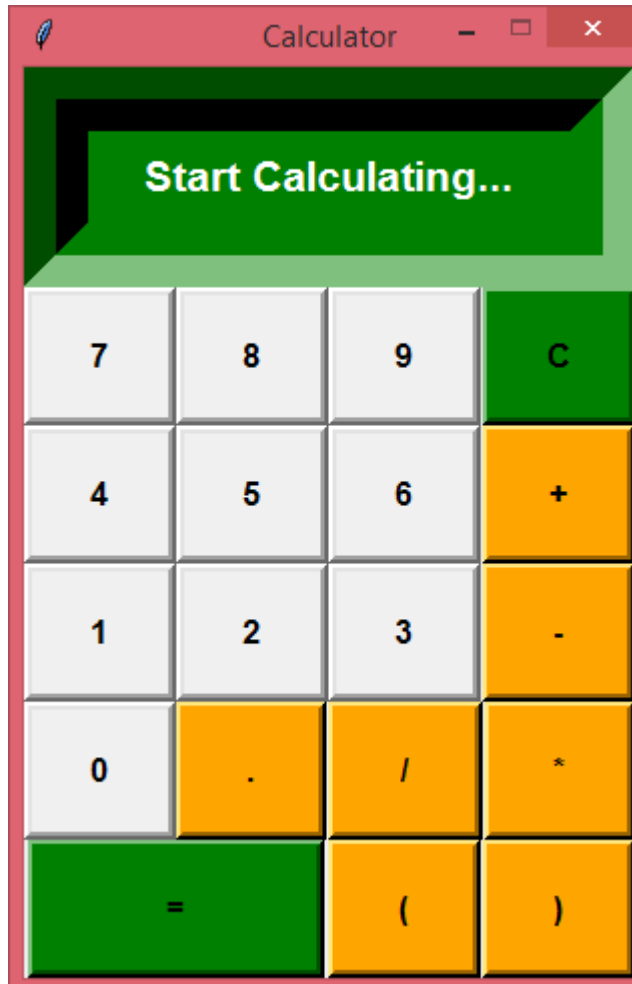
U Intern Task , digit-recognizer , fake-and-real-news

TASKS

- Calculator -----2
- Definitions -----4
- Recognition digit ----- 6
- Fake & real news -----9

Calculator:

- Using python package named “tkinter”



- Then to create window use
`x = tkinter.tk()`
where
`Tk(screenName=None, baseName=None, className='Tk', useTk=1)`
- When click on button

```
def button_click(item):  
    global expression  
  
    expression = expression + str(item)  
  
    input_text.set(expression)
```

```
expression = ""
```

```
# 'StringVar()' is used to get the instance of input field
```

```
input_text = StringVar()
```

Then to show value of input on screen

```
# set the screen of displaying text
```

```
input_field = Entry(root, font=('arial', 16, 'bold'), textvariable = input_text, foreground="white", bd = 32, insertwidth = 4, bg="green", justify
```

Then using styling for input field

```
input_field.grid(row=0, column=0)
```

```
input_field.pack(ipady=10) # 'ipady' is internal padding to increase the height of input field
```

```
btns_frame = Frame(root, width=312, height=272.5, bg="#eee")
```

```
btns_frame.pack()
```

Define button

```
#Define buttons
```

```
button_1 = Button(btns_frame, relief=RAISED, borderwidth=4, text="1", fg="black", font=('arial', 12, 'bold'), padx=25, pady=17, command=lambda
```

- At the end use `mainloop()`
where is used when your application is ready to run.

Definitions:

- **Data Cleansing means**
 - Check for there is no null data .
 - Check for redundancy
 - Remove noisy data by :
 - Removing stop words (is like removing unnecessary data).
- **Stemming:** (is the way to treat all words with same way , It removes suffices, like “ing”, “ly”, “s”,....
- **Lemmatization :** It is better than stemming as it uses a dictionary-based approach
- **Vectorizing Data: Bag-Of-Words :**
 - **Vectorizing** means convert text into vector.
 - **Bag of Words (BoW)** or CountVectorizer describes the presence of words within the text data.
 - Create document-matrix in each text document , so if result is 1 : means present of word in sentence and 0 for not present .
- **Image data Augmentation :**
 - Used to improve classifier
 - That by enlarge our training dataset (thus reducing overfitting)
 - where from a single image we were able to generate multiple distinct samples using rotation, shifting, zooming etc.
- **To_categorical :**
used to return binary matrix representation of input

```
>>> y = [0, 1, 2, 3]
>>> tf.keras.utils.to_categorical(y, num_classes=4)
array([[1., 0., 0., 0.],
       [0., 1., 0., 0.],
       [0., 0., 1., 0.],
       [0., 0., 0., 1.]], dtype=float32)
```
- **Prepare training & validation sets**
 - Separate data into training set and validation set.
 - Make model which
 - Learn from training set
 - Validate from validation set .

- **Sequential Model :**
 - Is Model where each layer has one input tensor and one output tensor
- **Compile Model :** has parameters to define
 - **Loss Function:** This is the function that evaluated how well your algorithm models your data set.
 - **Optimizer:** This is a method that finds the weights that minimize your loss function.
 - **Metrics:** This allows us to keep track of the loss as the model is being trained.
- **Evaluation :**
 - Using `.predict_generator`
 - Used to evaluate model after training
- **Fitting Model**
 - That when the model is well-fitted produces more accurate outcomes.
 - The by passing value to epochs(number of iterations over the data) to try to improve accuracy

Digit Recognizer :

Description:

Classification of handwritten digits, 10 classes (0–9).

Given a dataset of labeled handwritten images, (kaggle dataset)

Output : build a classifier that would assign correct labels to the new images.

Process :

- Load data sets
- Pre-processing :
 - Labeling data.
 - Reshape data_
 - Encoding data

```
# One-Hot encoding
from keras.utils.np_utils import to_categorical
y = to_categorical(y, num_classes=CLASSES)

print(y)
```

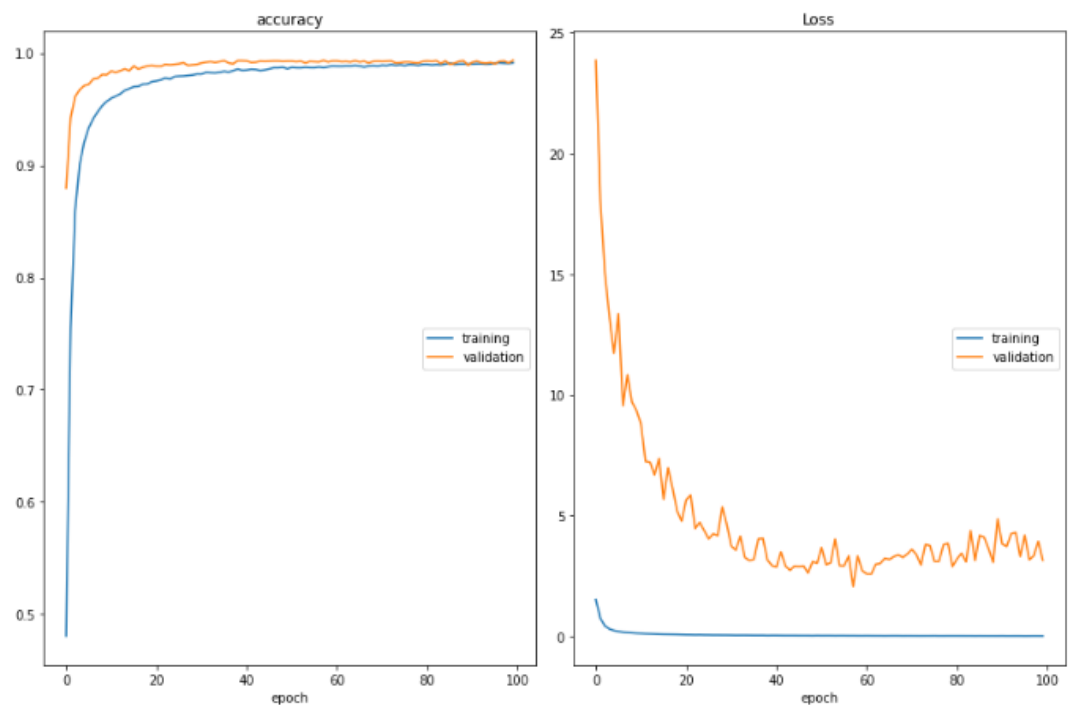
```
[[0. 1. 0. ... 0. 0. 0.]
 [1. 0. 0. ... 0. 0. 0.]
 [0. 1. 0. ... 0. 0. 0.]
 ...
 [0. 0. 0. ... 1. 0. 0.]
 [0. 0. 0. ... 0. 0. 0.]
 [0. 0. 0. ... 0. 0. 1.]]
```

- Prepare training & validation sets
- Build sequential Model
- Compile model
- Image data Augmentation:

```
# Data augmentation
data_generator = ImageDataGenerator(rescale=1./255,
                                    rotation_range=10,
                                    zoom_range=0.15,
                                    width_shift_range=0.1,
                                    height_shift_range=0.1)

data_generator.fit(x_training)
```

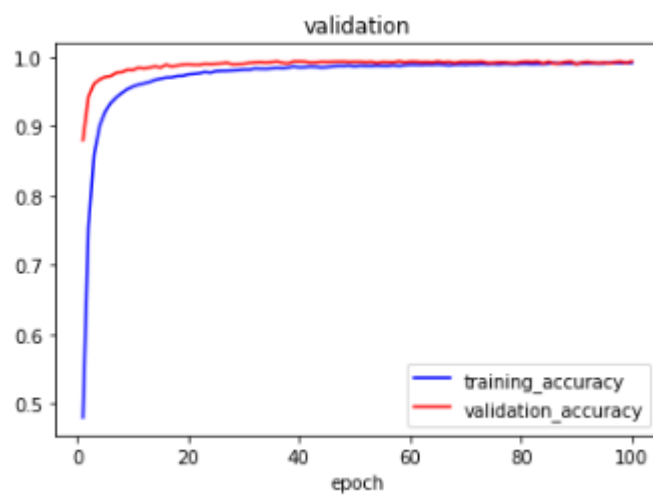
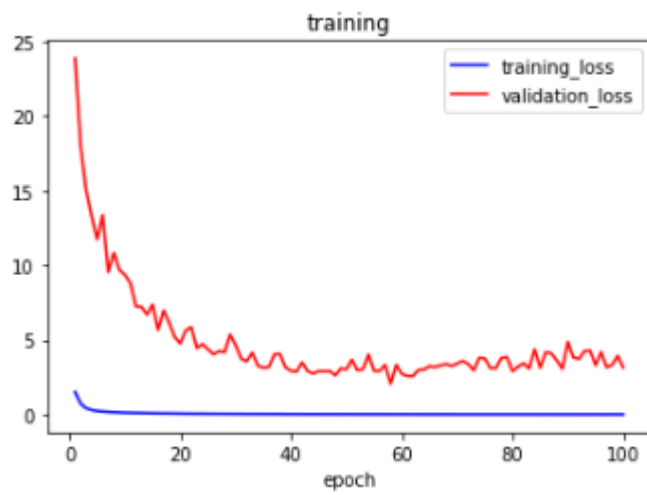
- **Training Model using .fit_generator() because using data augmentation**



The accuracy :

```
accuracy
      training      (min:  0.481, max:  0.991, cur:  0.991)
      validation    (min:  0.880, max:  0.994, cur:  0.994)
Loss
      training      (min:  0.028, max:  1.525, cur:  0.028)
      validation    (min:  2.077, max: 23.849, cur:  3.162)
73/73 [=====] - 11s 157ms/step - loss: 0.0284 - accuracy: 0.
9913 - val_loss: 3.1620 - val_accuracy: 0.9936
```

- **Evaluation :**
here predicts classes for images from the testing set and outputs the results to the .csv file.



News-real-fake detection:

Description :

Use data set to make an algorithm able to determine if an article is fake news or not

Given dataset of fake data and real data

Output : build classifier to detect sentence is real or fake

- **Process :**

- **Loading real dataset and fake dataset**
- **Define new column in each of datasets named “category” so**
 - **If category value = 0 means fake data**
 - **Else means real data**

	title	text	subject	date	category
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1
2	Senior U.S. Republican senator: 'Let Mr. Mue...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1

- **Merge datasets**
- **Pre-Processing Data , where using Nlp techniques that by :**
 - **Data Cleansing .**
 - **using Stemming .**
 - **Lemmatization .**
 - **Vectorizing Data: Bag-Of-Words .**
- **Splitting data into training set and test set.**
- **Building Model :**
 - **Using Sequential .**
- **Compile Model**
- **Fitting Model (Training Model)**

```
model.fit(cv_train_reviews,y_train , epochs = 5)
```

```
Epoch 1/5
33673/33673 [=====] - 551s 16ms/step - loss: 0.0397 - accuracy: 0.9897
Epoch 2/5
33673/33673 [=====] - 551s 16ms/step - loss: 0.0399 - accuracy: 0.9897
Epoch 3/5
33673/33673 [=====] - 552s 16ms/step - loss: 0.0395 - accuracy: 0.9897
Epoch 4/5
33673/33673 [=====] - 554s 16ms/step - loss: 0.0393 - accuracy: 0.9897
```

- **Summary of Model**

```
# Summary of neural network
model.summary()
```

```
Model: "sequential_2"
-----
Layer (type)                 Output Shape              Param #
-----
dense_6 (Dense)              (None, 100)              195826900
-----
dense_7 (Dense)              (None, 50)               5050
-----
dense_8 (Dense)              (None, 25)               1275
-----
dense_9 (Dense)              (None, 10)               260
-----
dense_10 (Dense)             (None, 1)                11
-----
Total params: 195,833,496
Trainable params: 195,833,496
Non-trainable params: 0
-----
```

- **Predict Model**
- **Calculate Accuracy.**

```
#accuracy
accuracy_score(pred,y_test)
```

```
0.9020044543429844
```

- **Evaluation**

- **Using classification report**

```

      precision    recall  f1-score   support

     0       0.90      0.92      0.91       5845
     1       0.91      0.88      0.90       5380

 accuracy          0.90       11225
 macro avg          0.90      0.90      0.90       11225
 weighted avg          0.90      0.90      0.90       11225
```

- **Plotting using confusion matrix :**

