```
C:\Users\Ys1ong\anaconda3\envs\ELEC0141\python.exe C:\Users\Ys1ong\DLNLP_23_SN22082567\main.py

**********Running Task A: Decision Tree**********

The current path of Task A is C:\Users\Ys1ong\DLNLP_23_SN22082567

Plot distribution of training, validation and test data

Training data:
joy: 5362 (33.51%)
sadness: 4666 (29.16%)
anger: 2159 (13.49%)
fear: 1937 (12.11%)
love: 1304 (8.15%)
surprise: 572 (3.57%)

Validation data:
joy: 704 (35.20%)
sadness: 550 (27.50%)
anger: 275 (13.75%)
fear: 212 (10.60%)
love: 178 (8.90%)
surprise: 81 (4.05%)

Test data:
joy: 695 (34.75%)
sadness: 581 (29.05%)
anger: 275 (13.75%)
fear: 224 (11.20%)
love: 159 (7.95%)
surprise: 66 (3.30%)

Process and cleaning training and validation dataset
Remove all URL links https?:\/\/\S+
Removing all punctuation !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
Convert all letters to lowercase

The shape of combined dataset is (18000, 16170)
The shape of training dataset is (16000, 16170)
The shape of validation dataset is (2000, 16170)

```

```
The shape of combined label is (18000,)
The shape of training label is (16000,)
The shape of validation label is (2000,)

Hyper-parameter tunning of minimum samples split
100%|██████████| 50/50 [04:55<00:00,  5.91s/it]


When Samples Split is 170, Validation Accuracy has the
  Highest Value

Choose the Minimum Sample: 170 and plot learning curve...
The Training Accuracy of decision tree is 0.8930
The Validation Accuracy of decision tree is 0.8597

Prepare test data
Remove all URL links https?:\/\/\S+
Removing all punctuation !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
Convert all letters to lowercase
The shape of test dataset is (2000, 16170)
The shape of test label is (2000,)

Time for training decision tree model is: 5.94s
The Test Accuracy of Decision Tree is 0.8420

Confusion Matrix of Decision Tree:
           sadness  joy  love  anger  fear  surprise
sadness        464   64     9     23    16         5
joy             28  613    32     11     5         6
love             0   36   121      0     0         2
anger            6    7     3    246    12         1
fear             4    2     2      6   198        12
surprise         1    9     0      0    14        42

Classification Report of Decision Tree:
              precision    recall  f1-score   support

           0       0.92      0.80      0.86       581
           1       0.84      0.88      0.86       695
           2       0.72      0.76      0.74       159
           3       0.86      0.89      0.88       275
           4       0.81      0.88      0.84       224
           5       0.62      0.64      0.63        66

```

```
 85     accuracy                           0.84      2000
 86    macro avg       0.80      0.81      0.80      2000
 87 weighted avg       0.85      0.84      0.84      2000
 88
 89
 90
 91 ******************************Running Task B: Multi-Layer
     Perceptron*******************************
 92
 93 The current path of Task B is C:\Users\Ys1ong\
    DLNLP_23_SN22082567
 94
 95 Plot distribution of training, validation and test data
 96
 97 Training data:
 98 joy: 5362 (33.51%)
 99 sadness: 4666 (29.16%)
100 anger: 2159 (13.49%)
101 fear: 1937 (12.11%)
102 love: 1304 (8.15%)
103 surprise: 572 (3.57%)
104
105 Validation data:
106 joy: 704 (35.20%)
107 sadness: 550 (27.50%)
108 anger: 275 (13.75%)
109 fear: 212 (10.60%)
110 love: 178 (8.90%)
111 surprise: 81 (4.05%)
112
113 Test data:
114 joy: 695 (34.75%)
115 sadness: 581 (29.05%)
116 anger: 275 (13.75%)
117 fear: 224 (11.20%)
118 love: 159 (7.95%)
119 surprise: 66 (3.30%)
120
121 Start data cleaning ...
122 Remove all URL links https?:\/\/\S+
123 Removing all punctuation !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
124 Convert all letters to lowercase
125
126 Start data tokenization of test data...
```

```
The shape of training dataset is (16000, 66)
The shape of validation dataset is (2000, 66)
The shape of training label is (16000, 6)
The shape of validation label is (2000, 6)

MLP model:
2023-05-01 04:38:03.890297: I tensorflow/core/platform/cpu_feature_guard.cc:193] This TensorFlow binary is optimized with oneAPI Deep Neural Network Library (oneDNN) to use the following CPU instructions in performance-critical operations:  AVX AVX2
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
2023-05-01 04:38:04.197630: I tensorflow/core/common_runtime/gpu/gpu_device.cc:1532] Created device /job:localhost/replica:0/task:0/device:GPU:0 with 5449 MB memory:  -> device: 0, name: NVIDIA GeForce RTX 3070 Ti, pci bus id: 0000:01:00.0, compute capability: 8.6
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 66, 100)           1618900

 global_average_pooling1d (G  (None, 100)              0
 lobalAveragePooling1D)

 dense (Dense)               (None, 128)               12928

 dropout (Dropout)           (None, 128)               0

 dense_1 (Dense)             (None, 6)
```

```
149 774
150
151 =================================================================
    ========
152 Total params: 1,632,602
153 Trainable params: 1,632,602
154 Non-trainable params: 0
155 _________________________________________________________________
    ________
156
157 Start training
158 Epoch 1/50
159 2023-05-01 04:38:05.150745: I tensorflow/stream_executor/
    cuda/cuda_blas.cc:1786] TensorFloat-32 will be used for
    the matrix multiplication. This will only be logged once.
160 500/500 [==============================] - 2s 3ms/step -
    loss: 1.5595 - accuracy: 0.3576 - val_loss: 1.4634 -
    val_accuracy: 0.4525
161 Epoch 2/50
162 500/500 [==============================] - 1s 3ms/step -
    loss: 1.0187 - accuracy: 0.6508 - val_loss: 0.7222 -
    val_accuracy: 0.7555
163 Epoch 3/50
164 500/500 [==============================] - 1s 3ms/step -
    loss: 0.4497 - accuracy: 0.8736 - val_loss: 0.4791 -
    val_accuracy: 0.8455
165 Epoch 4/50
166 500/500 [==============================] - 1s 3ms/step -
    loss: 0.2366 - accuracy: 0.9297 - val_loss: 0.4142 -
    val_accuracy: 0.8730
167 Epoch 5/50
168 500/500 [==============================] - 1s 3ms/step -
    loss: 0.1547 - accuracy: 0.9551 - val_loss: 0.4084 -
    val_accuracy: 0.8780
169 Epoch 6/50
170 500/500 [==============================] - 1s 3ms/step -
    loss: 0.1154 - accuracy: 0.9672 - val_loss: 0.4413 -
    val_accuracy: 0.8715
171 Epoch 7/50
172 500/500 [==============================] - 2s 3ms/step -
    loss: 0.0941 - accuracy: 0.9720 - val_loss: 0.4669 -
    val_accuracy: 0.8725
173 Epoch 8/50
```

```
500/500 [==============================] - 2s 3ms/step -
    loss: 0.0738 - accuracy: 0.9780 - val_loss: 0.4700 -
    val_accuracy: 0.8725
Epoch 9/50
500/500 [==============================] - 1s 3ms/step -
    loss: 0.0602 - accuracy: 0.9826 - val_loss: 0.4889 -
    val_accuracy: 0.8755
Epoch 10/50
500/500 [==============================] - 1s 3ms/step -
    loss: 0.0527 - accuracy: 0.9844 - val_loss: 0.5260 -
    val_accuracy: 0.8705

Time for training MLP model is: 15.72s
The Training Accuracy of MLP is 0.9844
The Validation Accuracy of MLP is 0.8705

Start data tokenization ...
The shape of validation dataset is (2000, 66)
The length of training label is 2000

Start evaluation
63/63 [==============================] - 0s 613us/step
The Test accuracy of MLP is 0.8760

Confusion Matrix of MLP:
           sadness  joy  love  anger  fear  surprise
sadness       521   30     1     16    13         0
joy             1  647    37      5     5         0
love            0   30   121      4     2         2
anger          15   10     0    233    16         1
fear           15    0     1      3   202         3
surprise        3    3    12      3    17        28

Classification Report of MLP:
               precision    recall  f1-score   support

            0       0.94      0.90      0.92       581
            1       0.90      0.93      0.91       695
            2       0.70      0.76      0.73       159
            3       0.88      0.85      0.86       275
            4       0.79      0.90      0.84       224
            5       0.82      0.42      0.56        66

    accuracy                           0.88      2000
```

```
212    macro avg       0.84      0.79      0.81      2000
213 weighted avg       0.88      0.88      0.87      2000
214
215
216
217 ******************************Running Task C: Long Short-
    Term Memory******************************
218
219 The current path of Task C is C:\Users\Ys1ong\
    DLNLP_23_SN22082567
220
221 Plot distribution of training, validation and test data
222
223 Training data:
224 joy: 5362 (33.51%)
225 sadness: 4666 (29.16%)
226 anger: 2159 (13.49%)
227 fear: 1937 (12.11%)
228 love: 1304 (8.15%)
229 surprise: 572 (3.57%)
230
231 Validation data:
232 joy: 704 (35.20%)
233 sadness: 550 (27.50%)
234 anger: 275 (13.75%)
235 fear: 212 (10.60%)
236 love: 178 (8.90%)
237 surprise: 81 (4.05%)
238
239 Test data:
240 joy: 695 (34.75%)
241 sadness: 581 (29.05%)
242 anger: 275 (13.75%)
243 fear: 224 (11.20%)
244 love: 159 (7.95%)
245 surprise: 66 (3.30%)
246
247 Start data cleaning ...
248 Remove all URL links https?:\/\/\S+
249 Removing all punctuation !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
250 Convert all letters to lowercase
251
252 Start data tokenization of test data...
253 The shape of training dataset is (16000, 66)
```

```
254 The shape of validation dataset is (2000, 66)
255
256 The shape of training label is (16000, 6)
257 The shape of validation label is (2000, 6)
258
259 LSTM model:
260 Model: "sequential_1"
261 _________________________________________________________
    ________
262  Layer (type)                Output Shape
    Param #
263 =========================================================
    ========
264  embedding_1 (Embedding)     (None, 66, 100)
    1618900
265
266  lstm (LSTM)                 (None, 66, 64)
    42240
267
268  dropout_1 (Dropout)         (None, 66, 64)             0
269
270  lstm_1 (LSTM)               (None, 32)
    12416
271
272  dense_2 (Dense)             (None, 6)
    198
273
274 =========================================================
    ========
275 Total params: 1,673,754
276 Trainable params: 1,673,754
277 Non-trainable params: 0
278 _________________________________________________________
    ________
279
280 Start training
281 Epoch 1/50
282 2023-05-01 04:38:25.785093: I tensorflow/stream_executor/
```

```
282 cuda/cuda_dnn.cc:384] Loaded cuDNN version 8201
283 500/500 [==============================] - 7s 11ms/step
     - loss: 1.4086 - accuracy: 0.3756 - val_loss: 1.1589 -
    val_accuracy: 0.4605
284 Epoch 2/50
285 500/500 [==============================] - 5s 9ms/step -
    loss: 1.0530 - accuracy: 0.4652 - val_loss: 1.0356 -
    val_accuracy: 0.5245
286 Epoch 3/50
287 500/500 [==============================] - 5s 9ms/step -
    loss: 0.9487 - accuracy: 0.5580 - val_loss: 1.5195 -
    val_accuracy: 0.4500
288 Epoch 4/50
289 500/500 [==============================] - 4s 9ms/step -
    loss: 0.9885 - accuracy: 0.5861 - val_loss: 1.0856 -
    val_accuracy: 0.6225
290 Epoch 5/50
291 500/500 [==============================] - 4s 9ms/step -
    loss: 1.0607 - accuracy: 0.5899 - val_loss: 1.3227 -
    val_accuracy: 0.4070
292 Epoch 6/50
293 500/500 [==============================] - 5s 9ms/step -
    loss: 1.0527 - accuracy: 0.5873 - val_loss: 1.2815 -
    val_accuracy: 0.5445
294 Epoch 7/50
295 500/500 [==============================] - 5s 9ms/step -
    loss: 0.8903 - accuracy: 0.6869 - val_loss: 0.9421 -
    val_accuracy: 0.7165
296 Epoch 8/50
297 500/500 [==============================] - 5s 9ms/step -
    loss: 0.8287 - accuracy: 0.6770 - val_loss: 1.0116 -
    val_accuracy: 0.5820
298 Epoch 9/50
299 500/500 [==============================] - 5s 9ms/step -
    loss: 0.7895 - accuracy: 0.6283 - val_loss: 0.9690 -
    val_accuracy: 0.5880
300 Epoch 10/50
301 500/500 [==============================] - 5s 9ms/step -
    loss: 0.6691 - accuracy: 0.7218 - val_loss: 0.8361 -
    val_accuracy: 0.7715
302 Epoch 11/50
303 500/500 [==============================] - 5s 9ms/step -
    loss: 0.5376 - accuracy: 0.8526 - val_loss: 0.8003 -
    val_accuracy: 0.8055
```

```
Epoch 12/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.4128 - accuracy: 0.9077 - val_loss: 0.6696 -
    val_accuracy: 0.8345
Epoch 13/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.4092 - accuracy: 0.9007 - val_loss: 0.7605 -
    val_accuracy: 0.8275
Epoch 14/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.3293 - accuracy: 0.9233 - val_loss: 0.5660 -
    val_accuracy: 0.8675
Epoch 15/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.2417 - accuracy: 0.9498 - val_loss: 0.4548 -
    val_accuracy: 0.8875
Epoch 16/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.1884 - accuracy: 0.9674 - val_loss: 0.4905 -
    val_accuracy: 0.8975
Epoch 17/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.1816 - accuracy: 0.9699 - val_loss: 0.4908 -
    val_accuracy: 0.8960
Epoch 18/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.1560 - accuracy: 0.9729 - val_loss: 0.5214 -
    val_accuracy: 0.8805
Epoch 19/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.1309 - accuracy: 0.9789 - val_loss: 0.4707 -
    val_accuracy: 0.8930
Epoch 20/50
500/500 [==============================] - 4s 9ms/step -
    loss: 0.1208 - accuracy: 0.9801 - val_loss: 0.4446 -
    val_accuracy: 0.8920
Epoch 21/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.1160 - accuracy: 0.9799 - val_loss: 0.4860 -
    val_accuracy: 0.8925
Epoch 22/50
500/500 [==============================] - 5s 9ms/step -
    loss: 0.1052 - accuracy: 0.9831 - val_loss: 0.4713 -
    val_accuracy: 0.8815
```

```
Epoch 23/50
500/500 [==============================] - 5s 9ms/step - loss: 0.0981 - accuracy: 0.9845 - val_loss: 0.4905 - val_accuracy: 0.8920
Epoch 24/50
500/500 [==============================] - 5s 9ms/step - loss: 0.0935 - accuracy: 0.9847 - val_loss: 0.4589 - val_accuracy: 0.8990
Epoch 25/50
500/500 [==============================] - 5s 9ms/step - loss: 0.1143 - accuracy: 0.9796 - val_loss: 0.4875 - val_accuracy: 0.8885

Time for training LSTM model is: 117.54s
The Training Accuracy of LSTM is 0.9796
The Validation Accuracy of LSTM is 0.8885

Start data tokenization ...
The shape of validation dataset is (2000, 66)
The length of training label is 2000

Start evaluation
63/63 [==============================] - 1s 3ms/step
The Test accuracy of BILSTM is 0.8870

Confusion Matrix of LSTM:
           sadness  joy  love  anger  fear  surprise
sadness       533   35     2      7     4         0
joy             9  639    36      0     5         6
love            0   23   123      1     3         9
anger          15    8     0    243     9         0
fear           10    3     0     11   199         1
surprise        2    3     0      0    24        37

Classification Report of LSTM:
              precision    recall  f1-score   support

           0       0.94      0.92      0.93       581
           1       0.90      0.92      0.91       695
           2       0.76      0.77      0.77       159
           3       0.93      0.88      0.91       275
           4       0.82      0.89      0.85       224
           5       0.70      0.56      0.62        66

```

```
364     accuracy                           0.89      2000
365    macro avg       0.84      0.82      0.83      2000
366 weighted avg       0.89      0.89      0.89      2000
367
368
369
370 ******************************Running Task D:
    Bidirectional Long Short-Term Memory
    ******************************
371
372 The current path of Task D is C:\Users\Ys1ong\
    DLNLP_23_SN22082567
373
374 Plot distribution of training, validation and test data
375
376 Training data:
377 joy: 5362 (33.51%)
378 sadness: 4666 (29.16%)
379 anger: 2159 (13.49%)
380 fear: 1937 (12.11%)
381 love: 1304 (8.15%)
382 surprise: 572 (3.57%)
383
384 Validation data:
385 joy: 704 (35.20%)
386 sadness: 550 (27.50%)
387 anger: 275 (13.75%)
388 fear: 212 (10.60%)
389 love: 178 (8.90%)
390 surprise: 81 (4.05%)
391
392 Test data:
393 joy: 695 (34.75%)
394 sadness: 581 (29.05%)
395 anger: 275 (13.75%)
396 fear: 224 (11.20%)
397 love: 159 (7.95%)
398 surprise: 66 (3.30%)
399
400 Start data cleaning ...
401 Remove all URL links https?:\/\/\S+
402 Removing all punctuation !"#$%&'()*+,-./:;<=>?@[\]^_`{|}~
403 Convert all letters to lowercase
404
```

```
Start data tokenization ...
The shape of training dataset is (16000, 66)
The shape of validation dataset is (2000, 66)

The shape of training label is (16000, 6)
The shape of validation label is (2000, 6)

BILSTM model
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding_2 (Embedding)     (None, 66, 100)           1618900

 dense_3 (Dense)             (None, 66, 128)           12928

 bidirectional (Bidirectiona  (None, 66, 128)          98816
 l)

 dropout_2 (Dropout)         (None, 66, 128)           0

 bidirectional_1 (Bidirectio  (None, 64)               41216
 nal)

 dense_4 (Dense)             (None, 6)                 390

```

```
430
431 =================================================================
    ========
432 Total params: 1,772,250
433 Trainable params: 1,772,250
434 Non-trainable params: 0
435 _________________________________________________________________
    ________
436
437 Start training
438 Epoch 1/50
439 500/500 [==============================] - 12s 19ms/step
     - loss: 1.0757 - accuracy: 0.6176 - val_loss: 0.6626 -
    val_accuracy: 0.7925
440 Epoch 2/50
441 500/500 [==============================] - 8s 15ms/step
     - loss: 0.4626 - accuracy: 0.8781 - val_loss: 0.3847 -
    val_accuracy: 0.9095
442 Epoch 3/50
443 500/500 [==============================] - 8s 15ms/step
     - loss: 0.2617 - accuracy: 0.9484 - val_loss: 0.3128 -
    val_accuracy: 0.9170
444 Epoch 4/50
445 500/500 [==============================] - 8s 15ms/step
     - loss: 0.1878 - accuracy: 0.9659 - val_loss: 0.3237 -
    val_accuracy: 0.9060
446 Epoch 5/50
447 500/500 [==============================] - 8s 15ms/step
     - loss: 0.1766 - accuracy: 0.9711 - val_loss: 0.3167 -
    val_accuracy: 0.9220
448 Epoch 6/50
449 500/500 [==============================] - 8s 15ms/step
     - loss: 0.1451 - accuracy: 0.9765 - val_loss: 0.3007 -
    val_accuracy: 0.9180
450 Epoch 7/50
451 500/500 [==============================] - 8s 15ms/step
     - loss: 0.1236 - accuracy: 0.9812 - val_loss: 0.3266 -
    val_accuracy: 0.9195
452 Epoch 8/50
453 500/500 [==============================] - 8s 15ms/step
     - loss: 0.1155 - accuracy: 0.9824 - val_loss: 0.4115 -
    val_accuracy: 0.8940
454 Epoch 9/50
455 500/500 [==============================] - 8s 15ms/step
```

```
455  - loss: 0.1315 - accuracy: 0.9789 - val_loss: 0.4509 -
    val_accuracy: 0.8875
456 Epoch 10/50
457 500/500 [==============================] - 8s 15ms/step
     - loss: 0.1225 - accuracy: 0.9825 - val_loss: 0.3357 -
    val_accuracy: 0.9150
458 Epoch 11/50
459 500/500 [==============================] - 8s 15ms/step
     - loss: 0.0893 - accuracy: 0.9891 - val_loss: 0.4334 -
    val_accuracy: 0.8865
460
461 Time for training BILSTM model is: 88.38s
462 The Training Accuracy of BILSTM is 0.9891
463 The Validation Accuracy of BILSTM is 0.8865
464
465 Start data tokenization of test data...
466 The shape of validation dataset is (2000, 66)
467 The length of training label is 2000
468
469 Start evaluation
470 63/63 [==============================] - 1s 6ms/step
471 The Test accuracy of BILSTM is 0.9125
472
473 Confusion Matrix of BILSTM:
474           sadness  joy  love  anger  fear  surprise
475 sadness       568    3     2      4     4         0
476 joy             7  634    42      3     5         4
477 love            3   18   135      1     1         1
478 anger          18    2     1    250     4         0
479 fear            9    0     0     10   196         9
480 surprise        5    2     0      0    17        42
481
482 Classification Report of BILSTM:
483               precision    recall  f1-score   support
484
485            0       0.93      0.98      0.95       581
486            1       0.96      0.91      0.94       695
487            2       0.75      0.85      0.80       159
488            3       0.93      0.91      0.92       275
489            4       0.86      0.88      0.87       224
490            5       0.75      0.64      0.69        66
491
492     accuracy                           0.91      2000
493    macro avg       0.86      0.86      0.86      2000
```

```
494 weighted avg       0.91      0.91      0.91      2000
495
496
497
498 ******************************Summary
    ******************************
499                  DT     MLP    LSTM  BILSTM
500 training    0.8930  0.9844  0.9796  0.9891
501 validation  0.8597  0.8705  0.8885  0.8865
502 accuracy    0.8420  0.8760  0.8870  0.9125
503
504 Time for running all is: 576.64s
505
506 Process finished with exit code 0
507
```