NLP-based System for Resume Screening with DeepLearning Models

Kevin – Jason – Justin – Rayes

Problem Backgrounds

One of the difficulties an HRD experienced when hiring new employees was reading resumes. To hire new employees, each organization has its own requirements, and those requirements can vary greatly. There will be a lot of candidates when a company posts a job opening since they are interested in working there. HRD will struggle because they must examine each applicant's résumé to determine whether or not they meet the requirements of the firm.

Purpose

We aim to relieve HRD of the burden of selecting candidates based solely on their resumes. Their resumes will then be evaluated into categories to be later classified into suitable job roles.

Related Works

Related Works #1

Resume Screening Using LSTM

- **Using One of deep learning method (LSTM) to do resume screening**
- **The application of LSTM can reduce the time spent on screening resume**

[1] Divya Mule, Samiksha Doke, Sakshi Navale, and Prof. S.K.Said, "RESUME SCREENING USING LSTM," International Research Journal of Engineering and Technology (IRJET), vol. 09, no. 04, pp. 3129-3132, Apr. 2022. [Online]. Available: www.irjet.net.

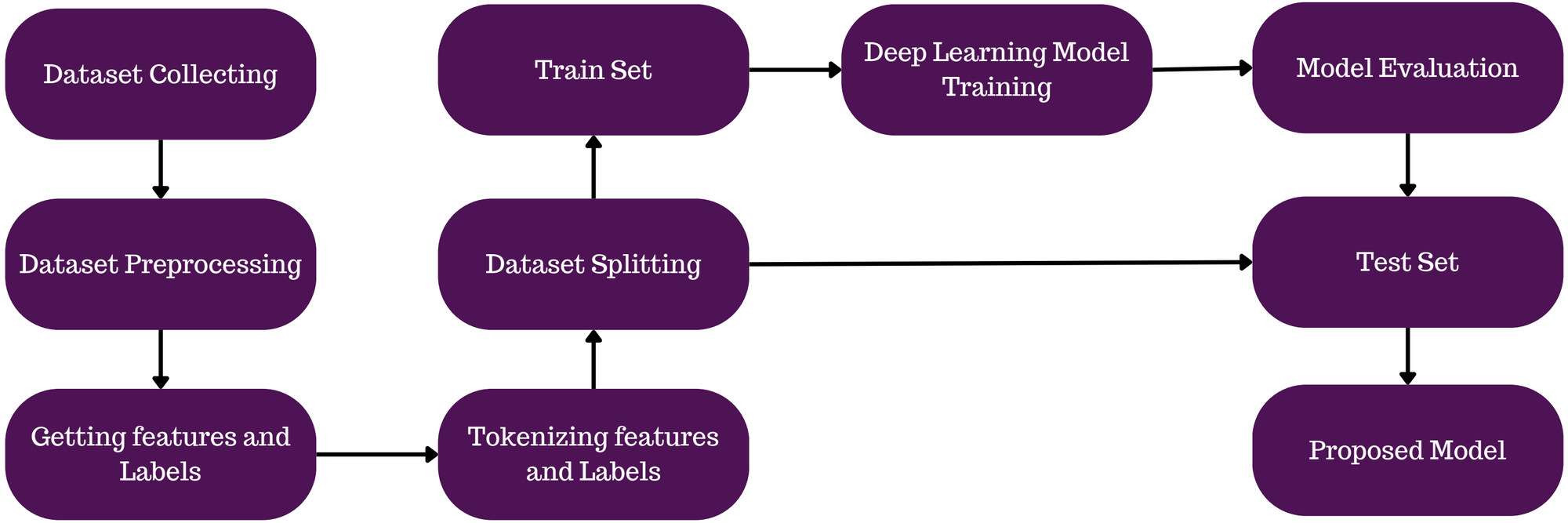
Related Works #2

Resume Screening Using Machine Learning and NLP: A Proposed System

- **Using machine learning methods and neural network SVM, KNN, Word2VEC, Cosine Similarity**
- **SVM shows the best performance with an accuracy of 78.98%**

[2]Bhushan Kinge, Shrinivas Mandhare, Pranali Chavan, and S. M. Chaware, "Resume Screening using Machine Learning and NLP: A proposed system," International Journal of Scientific Research in Computer Science, Engineering and Information Technology, pp. 253–258, Apr. 2022, doi: 10.32628/cseit228240.

Model development Flowchart



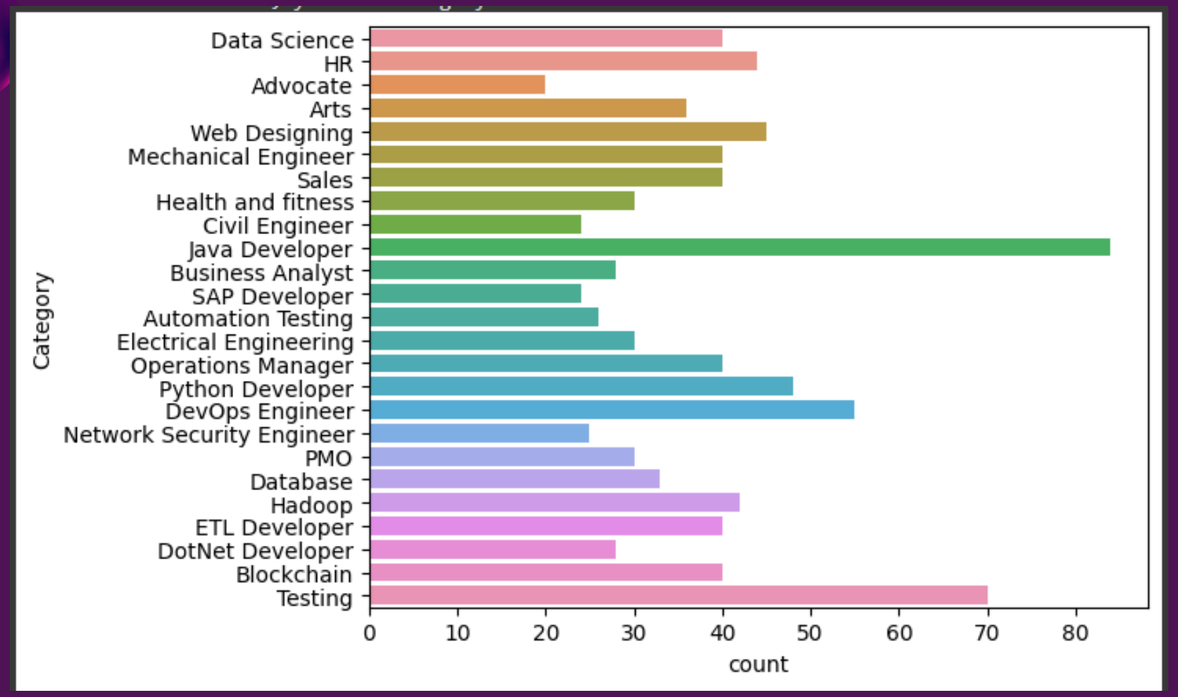
Experiments

DATASET

Resume Dataset

by GAURAV DUTTA

Source : KAGGLE



	Category	Resume
0	Data Science	Skills * Programming Languages: Python (pandas...
1	Data Science	Education Details \r\nMay 2013 to May 2017 B.E...
2	Data Science	Areas of Interest Deep Learning, Control Syste...
3	Data Science	Skills â¢ R â¢ Python â¢ SAP HANA â¢ Table...
4	Data Science	Education Details \r\n MCA YMCAUST, Faridab...

	Category	Resume
957	Testing	Computer Skills: â¢ Proficient in MS office (...
958	Testing	â Willingness to accept the challenges. â ...
959	Testing	PERSONAL SKILLS â¢ Quick learner, â¢ Eagerne...
960	Testing	COMPUTER SKILLS & SOFTWARE KNOWLEDGE MS-Power ...
961	Testing	Skill Set OS Windows XP/7/8/8.1/10 Database MY...

Past experiment result

by : PG student, Department of Computer Science and System Engineering Andhra University

KNN

	precision	recall	f1-score	support
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	12
3	0.62	1.00	0.76	13
4	1.00	1.00	1.00	8
5	0.67	1.00	0.80	8
6	0.64	1.00	0.78	9
7	1.00	1.00	1.00	6
8	0.62	1.00	0.76	8
9	0.71	0.50	0.59	10
10	1.00	1.00	1.00	9
11	1.00	1.00	1.00	9
12	1.00	1.00	1.00	11
13	1.00	1.00	1.00	9
14	1.00	1.00	1.00	7
15	0.00	0.00	0.00	10
16	1.00	1.00	1.00	7
17	1.00	1.00	1.00	3
18	1.00	0.43	0.60	7
19	0.25	0.38	0.30	8
20	1.00	0.80	0.89	5
21	1.00	0.38	0.55	8
22	1.00	1.00	1.00	3
23	0.33	0.22	0.27	9
24	1.00	1.00	1.00	1
25	0.33	0.33	0.33	3
accuracy			0.80	193
macro avg	0.81	0.80	0.79	193
weighted avg	0.79	0.80	0.77	193

Logistic Regression

	precision	recall	f1-score	support
1	0.83	1.00	0.91	10
2	1.00	1.00	1.00	12
3	1.00	0.92	0.96	13
4	1.00	1.00	1.00	8
5	1.00	1.00	1.00	8
6	1.00	1.00	1.00	9
7	1.00	1.00	1.00	6
8	1.00	1.00	1.00	8
9	0.91	1.00	0.95	10
10	1.00	1.00	1.00	9
11	1.00	1.00	1.00	9
12	1.00	1.00	1.00	11
13	1.00	1.00	1.00	9
14	0.78	1.00	0.88	7
15	1.00	1.00	1.00	10
16	1.00	1.00	1.00	7
17	1.00	1.00	1.00	3
18	1.00	1.00	1.00	7
19	1.00	1.00	1.00	8
20	1.00	1.00	1.00	5
21	1.00	0.75	0.86	8
22	1.00	1.00	1.00	3
23	1.00	1.00	1.00	9
24	1.00	1.00	1.00	1
25	1.00	0.33	0.50	3
accuracy			0.97	193
macro avg	0.98	0.96	0.96	193
weighted avg	0.98	0.97	0.97	193

SVM

	precision	recall	f1-score	support
1	0.83	1.00	0.91	10
2	1.00	1.00	1.00	12
3	0.86	0.92	0.89	13
4	1.00	1.00	1.00	8
5	1.00	1.00	1.00	8
6	1.00	1.00	1.00	9
7	1.00	1.00	1.00	6
8	1.00	1.00	1.00	8
9	1.00	1.00	1.00	10
10	1.00	1.00	1.00	9
11	1.00	1.00	1.00	9
12	1.00	1.00	1.00	11
13	1.00	1.00	1.00	9
14	1.00	1.00	1.00	7
15	0.91	1.00	0.95	10
16	1.00	1.00	1.00	7
17	1.00	1.00	1.00	3
18	1.00	1.00	1.00	7
19	1.00	1.00	1.00	8
20	1.00	1.00	1.00	5
21	1.00	0.75	0.86	8
22	1.00	1.00	1.00	3
23	1.00	1.00	1.00	9
24	1.00	1.00	1.00	1
25	1.00	0.33	0.50	3
accuracy			0.97	193
macro avg	0.98	0.96	0.96	193
weighted avg	0.98	0.97	0.97	193

Random Forest

	precision	recall	f1-score	support
1	1.00	1.00	1.00	10
2	1.00	1.00	1.00	12
3	0.86	0.92	0.89	13
4	1.00	1.00	1.00	8
5	1.00	1.00	1.00	8
6	1.00	1.00	1.00	9
7	1.00	1.00	1.00	6
8	1.00	1.00	1.00	8
9	1.00	1.00	1.00	10
10	1.00	1.00	1.00	9
11	1.00	1.00	1.00	9
12	1.00	1.00	1.00	11
13	1.00	1.00	1.00	9
14	1.00	1.00	1.00	7
15	0.91	1.00	0.95	10
16	1.00	1.00	1.00	7
17	1.00	1.00	1.00	3
18	1.00	1.00	1.00	7
19	1.00	1.00	1.00	8
20	1.00	1.00	1.00	5
21	1.00	0.75	0.86	8
22	1.00	1.00	1.00	3
23	0.82	1.00	0.90	9
24	1.00	1.00	1.00	1
25	1.00	0.33	0.50	3
accuracy			0.97	193
macro avg	0.98	0.96	0.96	193
weighted avg	0.98	0.97	0.97	193

OVERFITTING

Source : https://jespublication.com/upload/2022-V13I9053.pdf

DEEP LEARNING

1D-CNN

1 Dimensional Convolutional Neural Network

LSTM

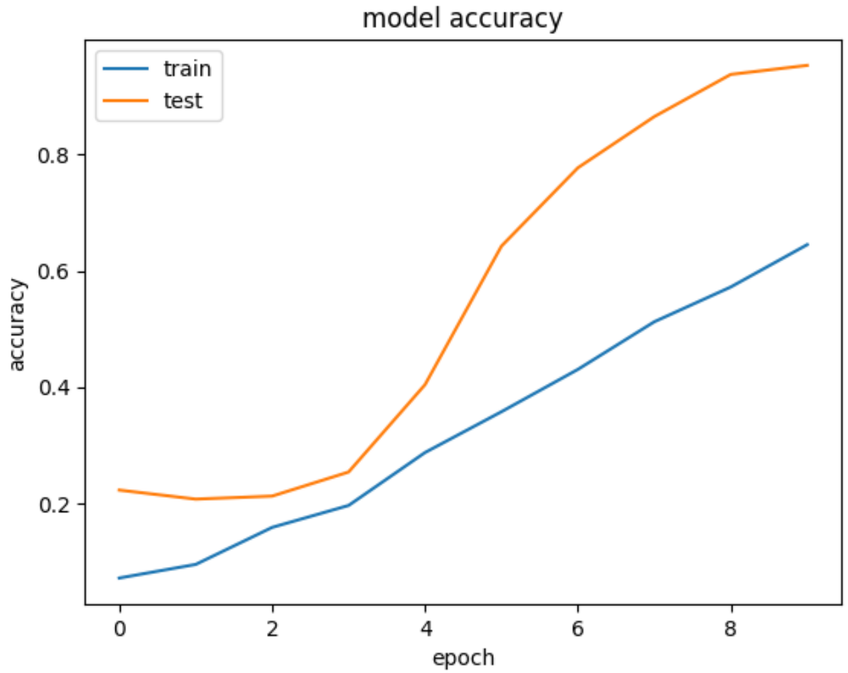
Long-Short Term Memory

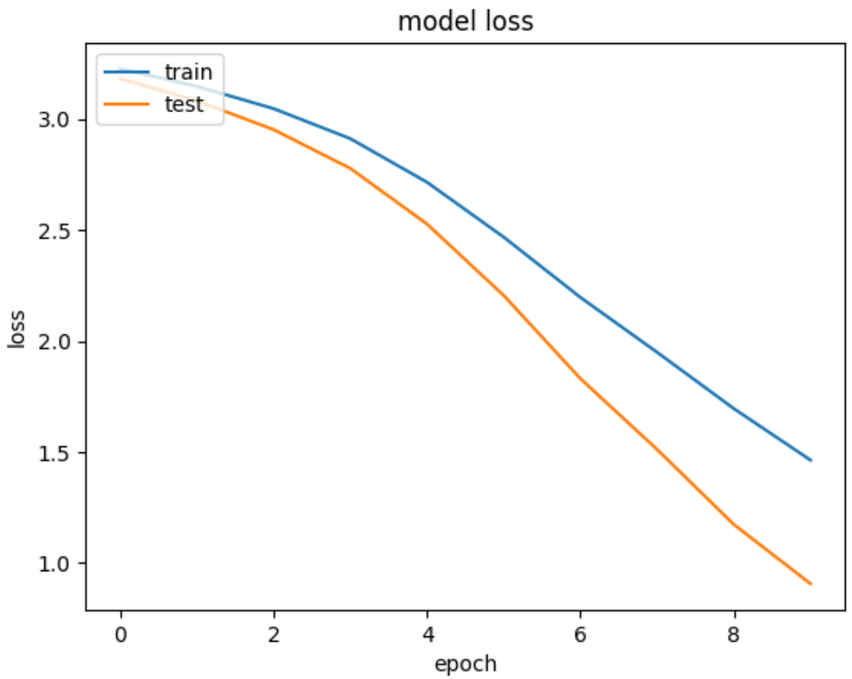
GRU

Gated Recurrent Unit

Results

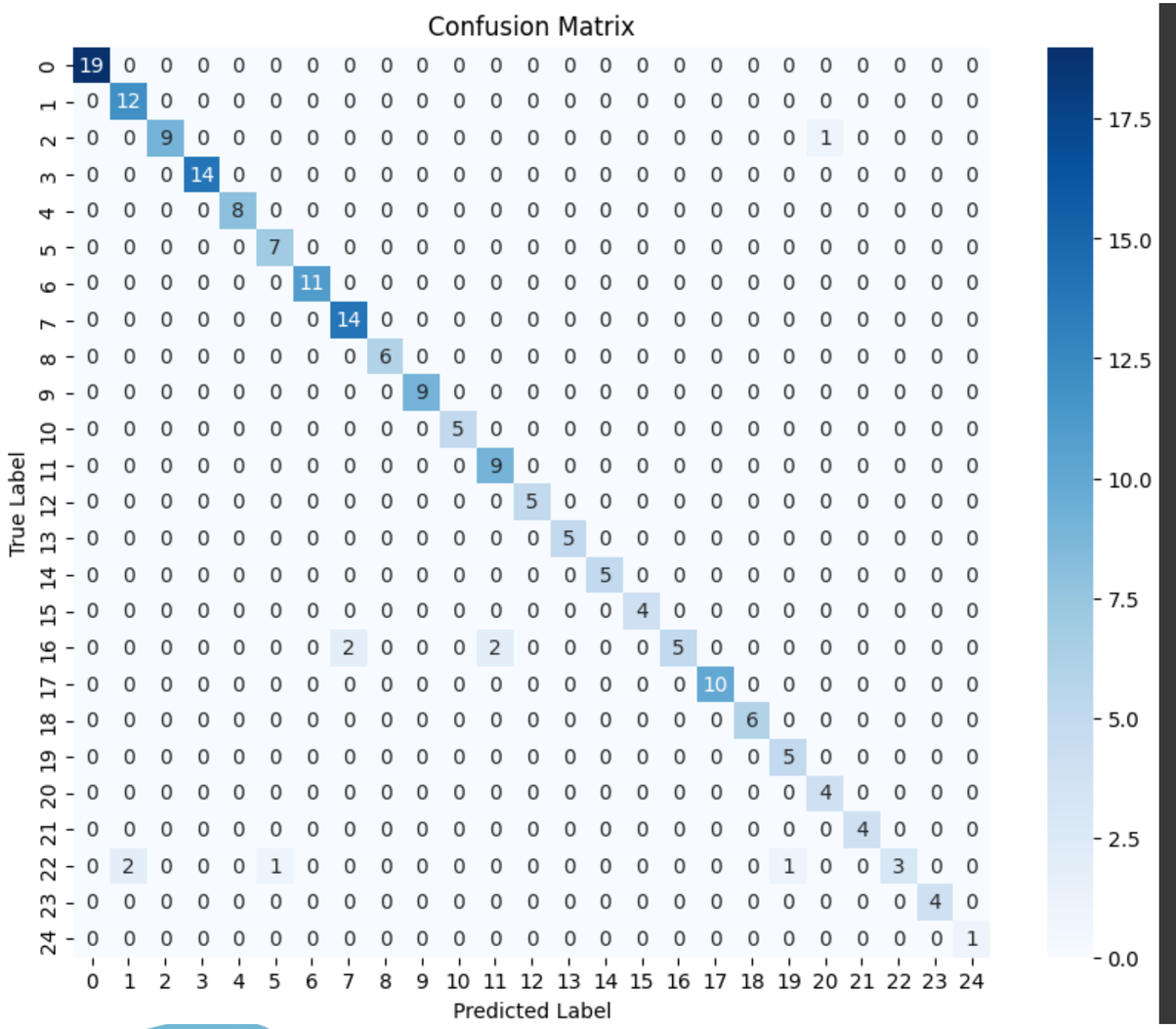
1D - CNN





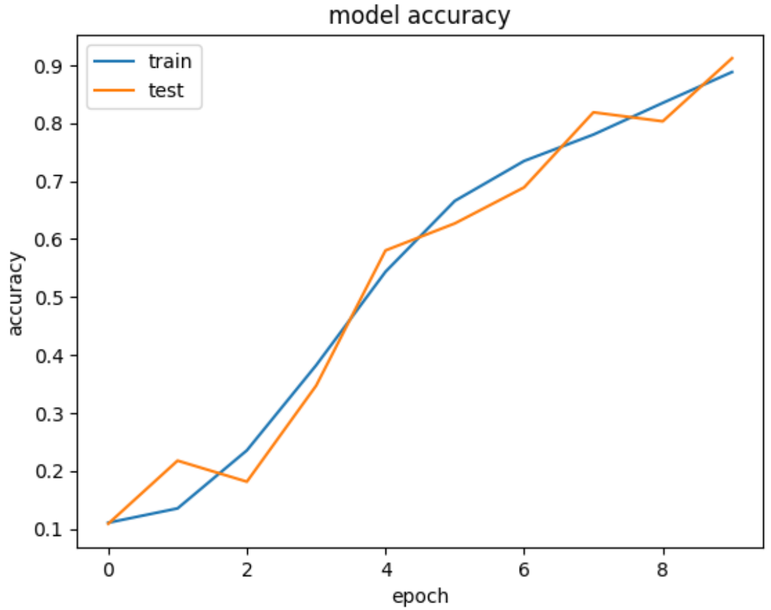

```
7/7 [==============================] - 0s 6ms/step - loss: 0.9062 - accuracy: 0.9534
Test Score: 0.9061946868896484
Test Accuracy: 0.9533678889274597
```

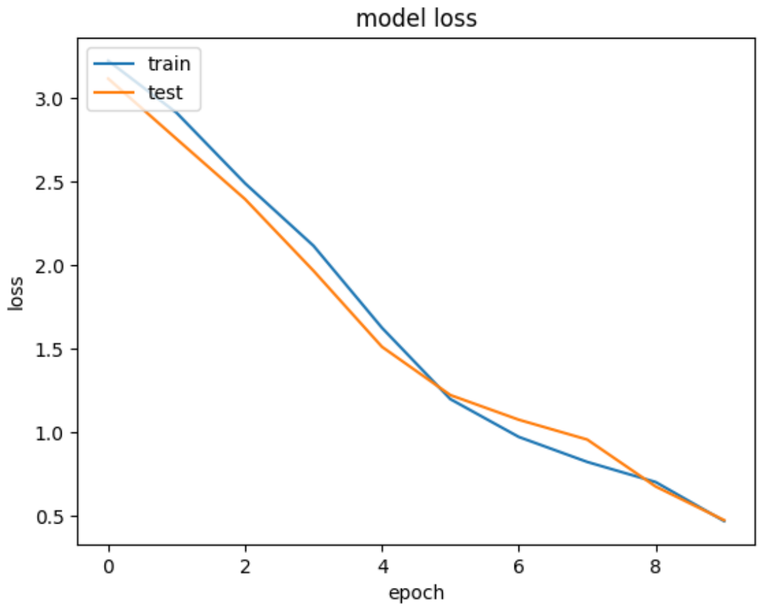
1D - CNN




```
1D-CNN Model Evaluation
------------------------------------------------------------
7/7 [==============================] - 0s 7ms/step
              precision    recall  f1-score   support

           1       1.00      1.00      1.00        19
           2       0.86      1.00      0.92        12
           3       1.00      0.90      0.95        10
           4       1.00      1.00      1.00        14
           5       1.00      1.00      1.00         8
           6       0.88      1.00      0.93         7
           7       1.00      1.00      1.00        11
           8       0.88      1.00      0.93        14
           9       1.00      1.00      1.00         6
          10       1.00      1.00      1.00         9
          11       1.00      1.00      1.00         5
          12       0.82      1.00      0.90         9
          13       1.00      1.00      1.00         5
          14       1.00      1.00      1.00         5
          15       1.00      1.00      1.00         5
          16       1.00      1.00      1.00         4
          17       1.00      0.56      0.71         9
          18       1.00      1.00      1.00        10
          19       1.00      1.00      1.00         6
          20       0.83      1.00      0.91         5
          21       0.80      1.00      0.89         4
          22       1.00      1.00      1.00         4
          23       1.00      0.43      0.60         7
          24       1.00      1.00      1.00         4
          25       1.00      1.00      1.00         1

    accuracy                           0.95       193
   macro avg       0.96      0.96      0.95       193
weighted avg       0.96      0.95      0.95       193

************************************************************
```

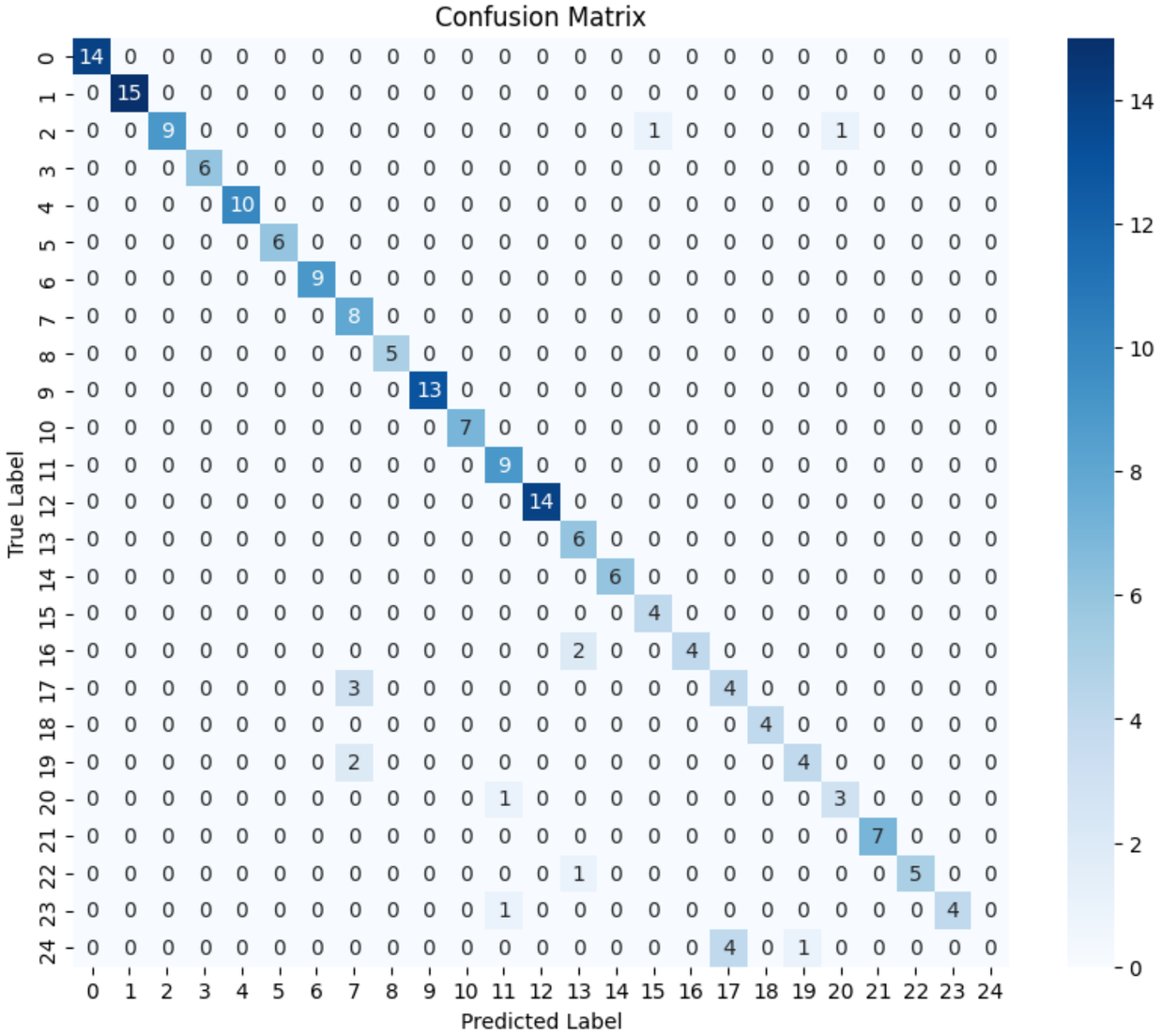
LSTM






```
7/7 [==============================] - 1s 108ms/step - loss: 0.4768 - accuracy: 0.9119
Test Score: 0.47675821185112
Test Accuracy: 0.9119170904159546
```

LSTM



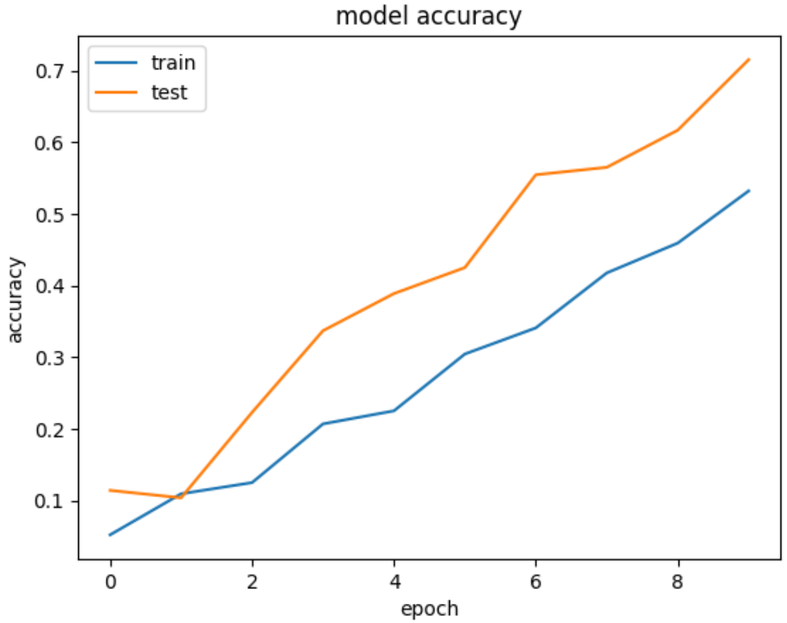
LSTM Model Evaluation

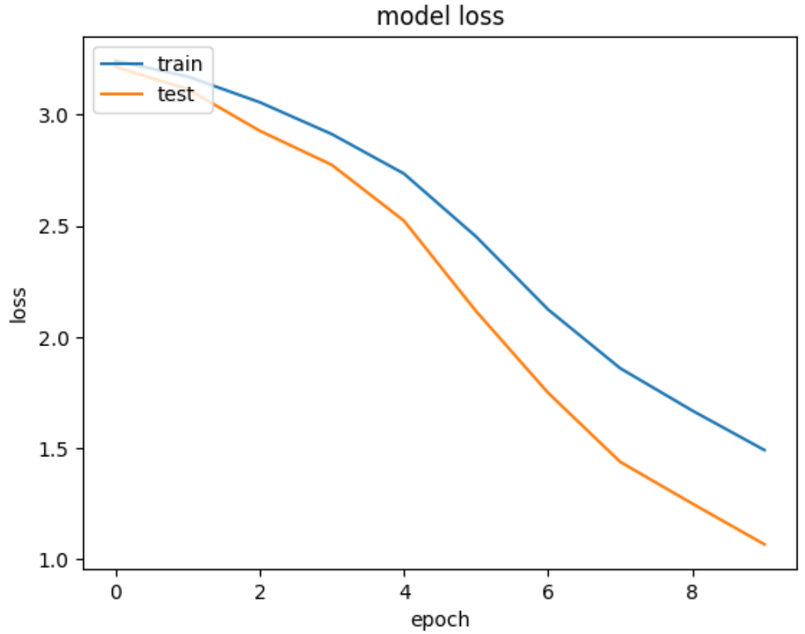
7/7 [==============================] - 1s 142ms/step

	precision	recall	f1-score	support
1	1.00	1.00	1.00	14
2	1.00	1.00	1.00	15
3	1.00	0.82	0.90	11
4	1.00	1.00	1.00	6
5	1.00	1.00	1.00	10
6	1.00	1.00	1.00	6
7	1.00	1.00	1.00	9
8	0.62	1.00	0.76	8
9	1.00	1.00	1.00	5
10	1.00	1.00	1.00	13
11	1.00	1.00	1.00	7
12	0.82	1.00	0.90	9
13	1.00	1.00	1.00	14
14	0.67	1.00	0.80	6
15	1.00	1.00	1.00	6
16	0.80	1.00	0.89	4
17	1.00	0.67	0.80	6
18	0.50	0.57	0.53	7
19	1.00	1.00	1.00	4
20	0.80	0.67	0.73	6
21	0.75	0.75	0.75	4
22	1.00	1.00	1.00	7
23	1.00	0.83	0.91	6
24	1.00	0.80	0.89	5
25	0.00	0.00	0.00	5
accuracy			0.91	193
macro avg	0.88	0.88	0.87	193
weighted avg	0.91	0.91	0.90	193

****

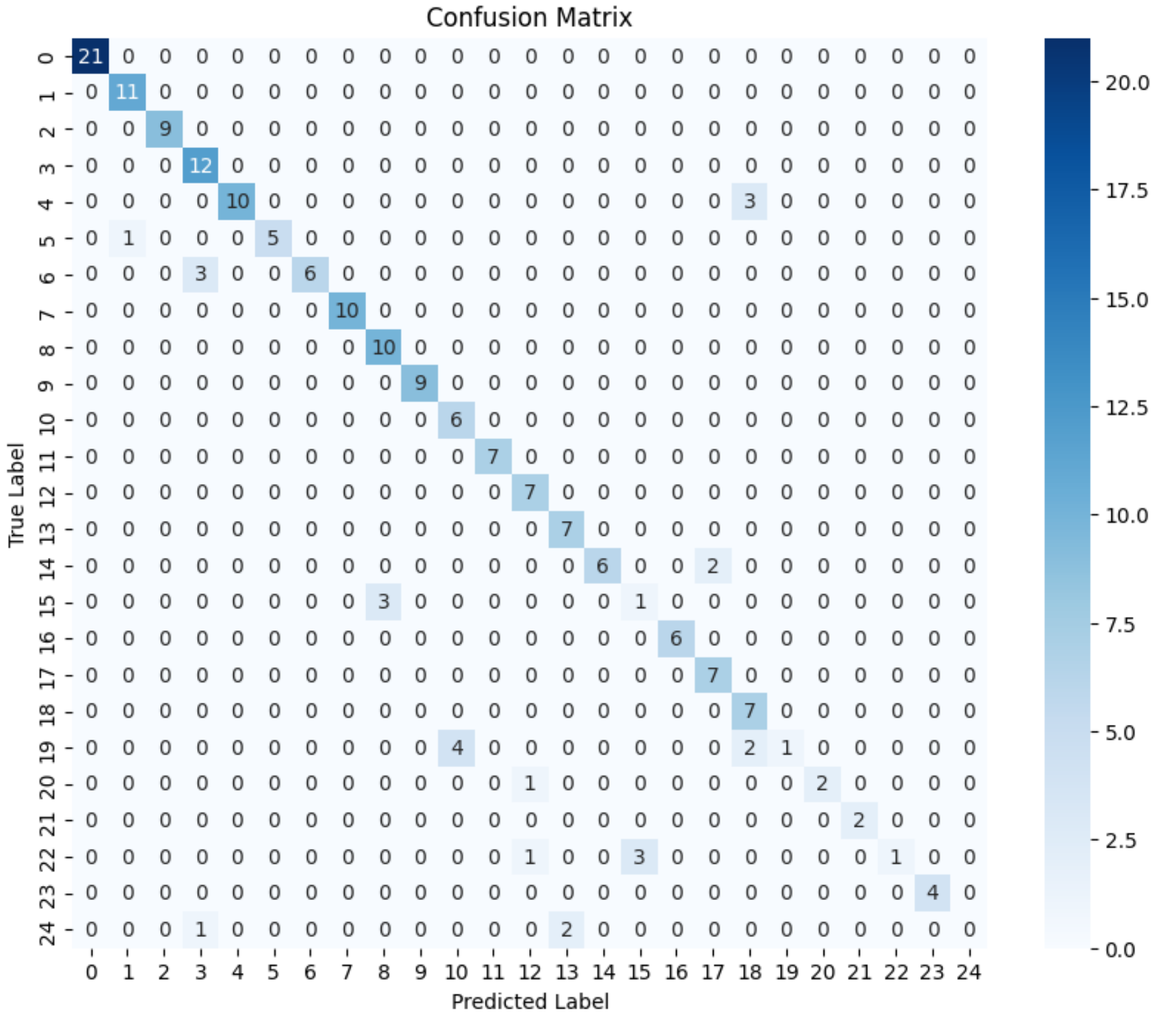
GRU






```
7/7 [==============================] - 0s 60ms/step - loss: 1.0667 - accuracy: 0.7150
Test Score: 1.0667073726654053
Test Accuracy: 0.7150259017944336
```

GRU



GRU Model Evaluation

7/7 [==============================] - 1s 186ms/step

	precision	recall	f1-score	support
1	0.94	1.00	0.97	16
2	0.74	1.00	0.85	17
3	0.90	1.00	0.95	9
4	0.38	1.00	0.55	12
5	0.90	0.90	0.90	10
6	0.60	0.30	0.40	10
7	1.00	0.50	0.67	10
8	0.83	1.00	0.91	10
9	1.00	1.00	1.00	9
10	0.00	0.00	0.00	6
11	0.83	1.00	0.91	5
12	1.00	1.00	1.00	11
13	1.00	0.56	0.71	9
14	0.12	1.00	0.21	2
15	0.50	1.00	0.67	3
16	1.00	0.43	0.60	7
17	0.00	0.00	0.00	3
18	1.00	1.00	1.00	8
19	1.00	0.40	0.57	5
20	0.00	0.00	0.00	5
21	1.00	0.43	0.60	7
22	1.00	1.00	1.00	4
23	1.00	0.17	0.29	6
24	1.00	0.17	0.29	6
25	0.00	0.00	0.00	3
accuracy			0.72	193
macro avg	0.71	0.63	0.60	193
weighted avg	0.78	0.72	0.69	193

****

Conclusion

Conclusion

Proposed model :

After some tinkering in the model's parameters, out of the three models presented, LSTM still produces the best result due to its low overfitting trait compared to the rest of the model, followed by GRU and lastly 1D-CNN.

Compared to the previous experiment :

DeepLearning model suffers less overfitting, compared to the other Machine Learning method in this experiment.

Thank You

Source codes :

LSTM :
https://colab.research.google.com/drive/1ePTMKVbHlmbIa7fqJaoUd9EbHJOEjgEZ?usp=sharing#scrollTo=BKfu1mHPxorl

GRU :
https://colab.research.google.com/drive/1TkEptk-QHJTsDglnX6tXE4F7km47vk2a#scrollTo=FTDVMiqeu8IE

1D-CNN :
https://colab.research.google.com/drive/1sT-XC-yowDw4r7wkJ4bpc8btX29U_4Dl#scrollTo=7mgaGp-oEqON