Differentiating ChatGPT from Humans: Detecting ChatGPT-Generated Text and Human Text Using Machine Learning Arabic Language.

1. The Data Set has 10,000 records. [5,000 records are taken from ChatGPT generated Arabic Dataset CIDAR and other 5000 records taken from human-written articles MNAD (Moroccan News Articles) Dataset and are both are merged]
2. LSTMRNN-based model developed and trained using 9900 randomly selected records from above dataset and other 100 records used for validating. Results and metrics are captured and presented. Note: LSTMRNN model source code could be found in **NLP\_LSTM2.py** and results in **lstm-rnn-mdl-result.txt**
   1. **Confusion metrics for LSTMRNN approach on validating records (100).**

|  |  |  |
| --- | --- | --- |
| **ChatGPT-Text** | **45** | **0** |
| **Human-Text** | **3** | **52** |
|  | **ChatGPT-Text** | **Human-Text** |

* 1. **Training Loss and Accuracy details for LSTMRNN approach.**

|  |  |  |
| --- | --- | --- |
| **No of epochs** | **Accuracy (%)** | **Loss** |
| 1 | 93.79 | 0.1866 |
| 2 | 98.66 | 0.0570 |
| 3 | 99.21 | 0.0325 |
| 4 | 99.52 | 0.0207 |
| 5 | 99.68 | 0.0135 |

B1. Training Accuracy curve of the LSTMRNN approach.

**B2. Training Loss curve of the LSTMRNN approach.**

1. Comparative outcomes of the LSTMRNN approach with other techniques on the said datasets (using 9900 randomly selected records from above and 100 other records for validating). Comparative table and graph are given below.

* SVM
* Basic Deep learning
* Convolution Model (CNN)
* LSTM Model (RNN)

|  |  |  |
| --- | --- | --- |
| **Approach** | **Accuracy (%)** | **Loss** |
| SVM | 97.00 |  |
| Basic Deep Learning | 78.10 | 2.0235 |
| CNN | 99.56 | 0.0172 |
| LSTMRNN | 99.68 | 0.0135 |

Note:

1. SVM model source code could be found in **NLP\_SVM.py** and results in **lstm-rnn-mdl-result.txt**
2. Basic DeepLearning model source code could be found in **NLP\_DL2.py** and results in **basic\_dl\_mdl\_result.txt**
3. CNN model source code could be found in **NLP\_CNN2.py** and results in **cnn\_mdl\_result.txt**
4. DataSet with 10000 records could be found in **human\_chatgpt\_genarated\_dataset.csv**