Contact: BBAK Technologies – Ph: +91-9353205447
email: bbaktech@gmail.com for clarifications

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

Recently, the identification of human text and ChatGPT-generated text has become a hot research topic. The current study presents a Memory Recurrent Neural Network (LSTMRNN) model to detect both human as well as ChatGPT-generated text. The purpose of the proposed LSTMRNN method is to investigate the model's decision and detect the presence of any particular pattern. In addition to this, the LSTMRNN technique focuses on designing Term Frequency–Inverse Document Frequency (TF-IDF), word embedding, and count vectorizers for the feature extraction process. For the detection and classification processes, the LSTMRNN model is used. The simulation performance of the proposed LSTMRNN technique was investigated on benchmark databases, and the outcome demonstrated the advantage of the LSTMRNN system over other recent methods (SVM, Basic Deep learning, Convolution Model (CNN) LSTM Model (RNN)).

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. both are merged to get 10,000 records]. DataSet 10000 records - **human_chatgpt_genarated_dataset.csv**

2. LSTMRNN-based model was developed and trained using 9900 randomly selected records from the above dataset and the remaining 100 records were 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**

   a. **Confusion metrics for LSTMRNN approach on validating records (100).**

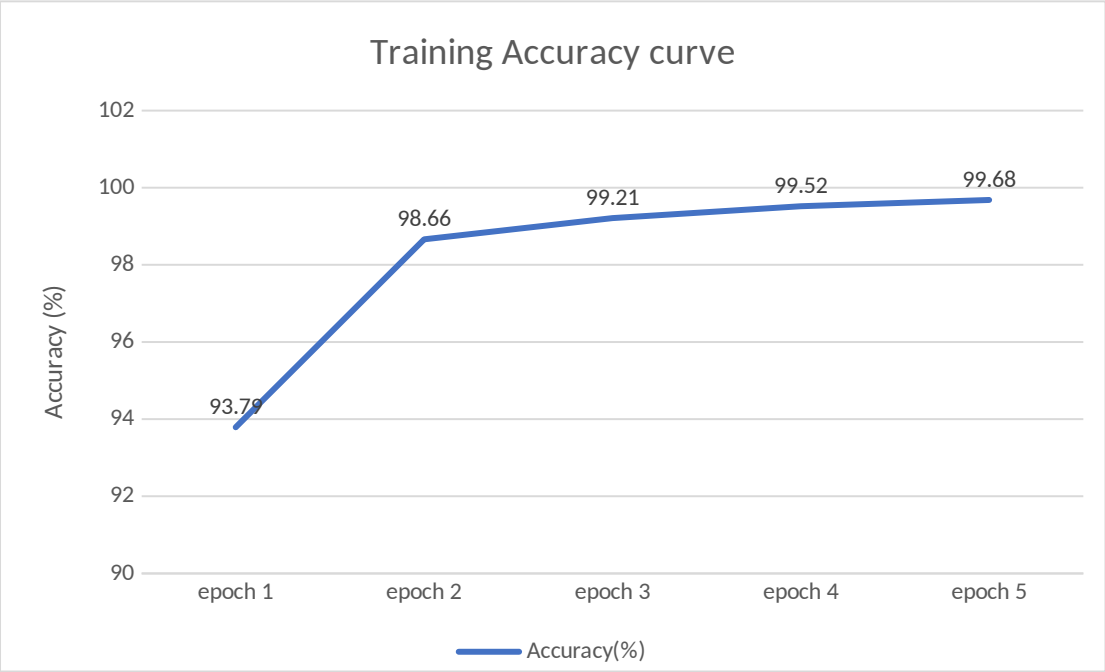
ChatGPT-Text	**45**	**0**
Human-Text	**3**	**52**
	ChatGPT-Text	**Human-Text**

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

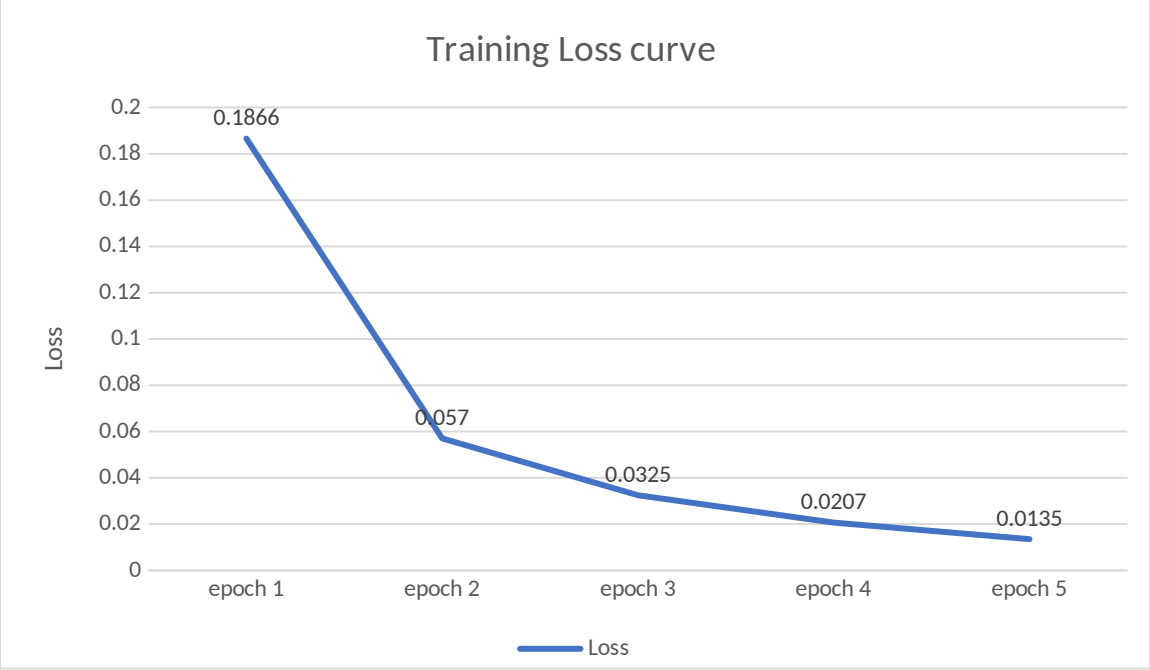
No of epochs	Accuracy (%)	Loss
1	93.79	0.1866
2	98.66	0.0570
3	99.21	0.0325
4	99.52	0.0207

5	99.68	0.0135

B1. Training Accuracy curve of the LSTMRNN approach.



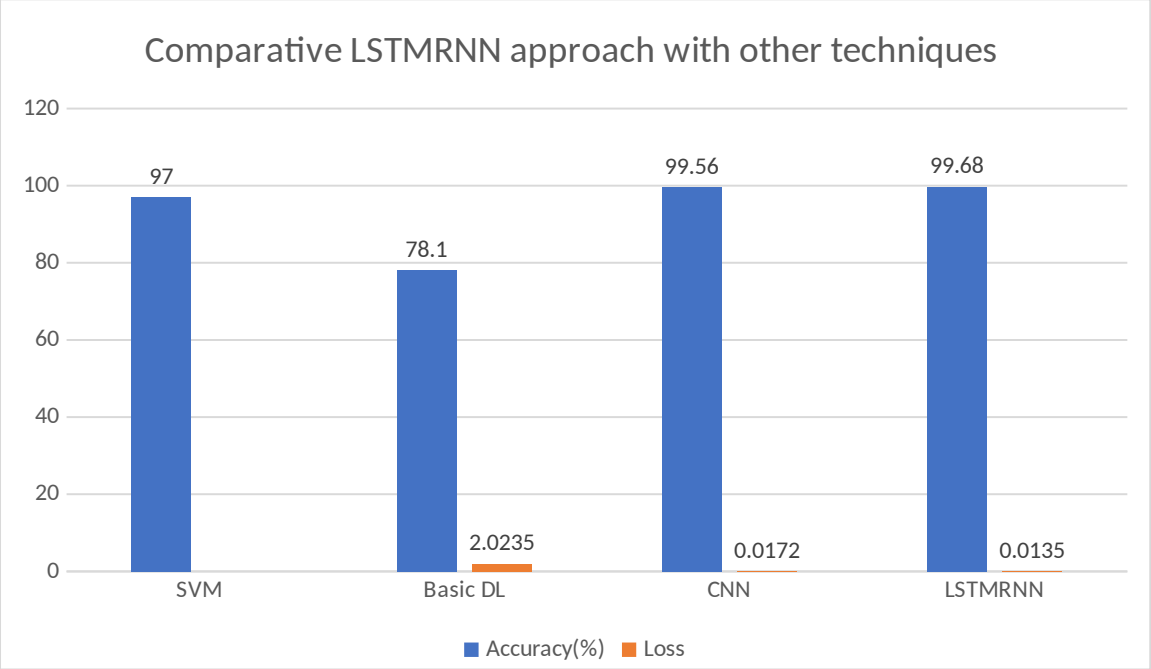
B2. Training Loss curve of the LSTMRNN approach.



3. 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). The 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**
5. **All these files are available at** https://github.com/bbaktech/Differentiating_ChatGPT_AND_Humans.git